This thesis has been submitted in fulfilment of the requirements for a
postgraduate degree (e.g. PhD, MPhil, DClinPsychol) at the University of
Edinburgh. Please note the following terms and conditions of use:

- This work is protected by copyright and other intellectual property rights,
which are retained by the thesis author, unless otherwise stated.
- A copy can be downloaded for personal non-commercial research or
study, without prior permission or charge.
- This thesis cannot be reproduced or quoted extensively from without
first obtaining permission in writing from the author.
- The content must not be changed in any way or sold commercially in
any format or medium without the formal permission of the author.
- When referring to this work, full bibliographic details including the
author, title, awarding institution and date of the thesis must be given.
Epigenetic signatures of chronic inflammation and their relation to brain structure and function across the lifecourse

Eleanor Lucy Shepherd Conole

THE UNIVERSITY of EDINBURGH

Doctor of Philosophy
The University of Edinburgh
2023
Abstract

Chronic inflammation is considered a key contributor to individual differences in brain ageing. However, there remains conflicting evidence about the exact brain structural and functional consequences of chronic inflammation. Part of this ambiguity comes from the lack of robust biomarkers used to characterise chronic inflammation, with most studies conflating acute vs chronic inflammation and relying on single time point sampling of highly phasic inflammatory markers from blood. Recent research indicates the potential of epigenetics to circumvent this issue, in particular the use of DNA methylation (DNAm) to provide a biological archive of inflammatory burden. DNAm is an epigenetic mechanism that regulates gene expression and acts as an interface by which lifestyle and environment can influence phenotype; alterations in DNAm are increasingly investigated as proxies for certain exposures, traits, and conditions.

This thesis explores chronic inflammation and its relation to brain health using a multi-omic approach, examining associations of proteomic and DNA methylation signatures of inflammation with brain measures from structural and diffusion MRI and cognitive ability at different stages of the lifecourse. Following introductory chapters presenting overviews of epigenetics, inflammation and neuroimaging and cognitive metrics, the empirical work of this thesis examines these interrelationships in three population cohorts covering the human lifespan: from infancy in Theirworld Edinburgh Birth cohort (TEBC; age < 1 year), to mid to late adulthood in Stratifying Longitudinal Resilience & Depression Longitudinally cohort (STRADL; age ~ 60 years, range 28 – 81 years), to older-age in the Lothian Birth Cohort 1936 (LBC1936; age ~ 73 years).

First, neuroimaging and cognitive associations of chronic inflammation in older age are examined. In 521 individuals from the LBC 1936 cohort, associations between a DNA methylation predictor of C-Reactive Protein (DNAm CRP) and brain structure and cognition are shown to be consistently stronger (6.4-fold greater on average) than those with traditional serum CRP measures: with higher DNAm CRP levels significantly associated with global and regional brain atrophy (β range |0.200| to |0.150|), differences in white matter microstructure and white matter hyperintensity burden (β range |0.099| to |0.162|), and poorer global and domain- specific cognitive functioning (β range |0.095| to |0.158|). This paper also demonstrates that the association between inflammation and cognitive ability is partially mediated by brain structure (up to 29.7%), dependent on lifestyle and health factors. The second empirical chapter replicates this association of an epigenetic inflammatory signature with brain and postnatal health...
outcomes in early life. In a neonatal cohort of 258 infants, the relationship between DNA methylation (DNAm) CRP with perinatal health and neuroimaging outcomes is investigated. The results support the theory that DNAm may be leveraged to capture a more cumulative impact of inflammatory burden, with DNAm CRP being higher in preterms compared to term infants and higher DNAm CRP levels associating with perinatal inflammatory-related morbidities (such as late onset sepsis and histologic chorioamnionitis) both individually and in aggregate (OR range [2.00 | to |4.71]). The main finding – that elevated DNAm CRP associates with poorer measures of white matter, both globally and regionally (β range [0.206] to [0.371]) – is considered in the context of how early birth associates with an increase in immune-related risks, which coincide with windows of neurodevelopmental plasticity, highlighting the vulnerability of developing white matter to inflammatory insults.

Finally, the brain health associations of signatures of inflammation in midlife are considered. In 709 participants from the STRADL cohort, DNAm signatures for CRP alongside a range of other inflammatory-immune mediators are constructed and compared against protein levels in relation to neuroimaging and cognitive metrics. 73 unique DNAm signatures associated with numerous aspects of global brain structure and cognitive ability (β range [0.097] to [0.200]), alongside regional atrophy across the brain's cortex (β range [0.087] to [0.260]) and focal vulnerability of specific white matter tract microstructure (β range [0.103] to [0.185]). Many of these DNAm-brain associations were larger than the analogous proteome-brain associations, broadly independent of immune-cell proportions, clinical risk factors, and had previously been linked to various age-related diseases, reinforcing the central role that inflammation plays in health trajectories. The results of this chapter suggest that associations with chronic inflammation and brain structure are apparent in mid to late adulthood and may precipitate and underlie changes in cognitive functioning seen in older-age.

This PhD characterises the global and regional associations of chronic inflammation on brain structure at different stages of life, elucidating new perspectives on how these mechanisms may be contributing to individual differences in cognitive ability. I highlight the potential of the developing field of epigenetics to offer a solution to the traditional limitations of assessing inflammatory burden in human cohorts, and how this may be used to gain a better understanding of how inflammation relates to aberrant neurodevelopment and cognitive decline. These findings provide new insights about the extent of inflammation’s impact on brain health, as well as highlighting the utility of DNA methylation for risk prediction and stratification in relation to brain health outcomes.
Lay Summary

Inflammation is the body’s first-line response to damage or infection. We recognise it as heat, redness, swelling, pain, and sometimes loss of function in the affected area. We all encounter it – from the irritation of a scabbed knee, the swelling of a stubbed toe, to the malaise accompanying a common cold. When we refer to inflammation, we tend to think of its acute presentation – considering it a short-lived reaction to some initiating insult. But inflammation can linger, relatively undetected, causing subtle but significant damage to cells, tissues, and organs. We characterise this type of inflammation as ‘chronic’, where the balance between immune resolution and regulation has gone awry. This type of inflammation has been linked to increased risk of mortality and age-related diseases, and there is particular interest in how it contributes to differences in brain ageing.

At the molecular level, inflammation is characterised by the coordination of a vast number of different cell types and mechanisms by which they communicate with each other. Different signalling cascades are set off by various compounds that patrol the blood, primed to detect signs of infection, damage or decay. Once alerted, specialised immune cells migrate to the sites of injury, set on their paths to ingesting dying or foreign cells. Simultaneously, blood vessels widen and blood flow increases – an effect that we experience as redness and heat. The walls of these vessels become more permeable, permitting cells, proteins and fluids to leak out – a process that causes swelling when it occurs at the skin’s surface and more hidden damage at deeper sites, such as at the interface between brain and blood (the blood brain barrier). Thousands of different inflammatory mediators – cytokines, chemokines, growth factors and acute phase proteins – are key cogs in the molecular machinery of these interactions, their expression regulated by an equally large number of genes.

This intricate array of responses highlights how chronic inflammation is a complex phenomenon, with no one protein apt to define it. Despite this, researchers typically rely on singular inflammatory proteins in the blood, such as C-reactive protein (CRP), to assess inflammation levels. Because inflammatory protein levels go up and down a lot within narrow time frames, measuring just one protein, at a single time-point, may give a false account of general inflammation levels over a longer period of time. This could be one of the reasons why the evidence for inflammation’s impact on brain structure and function is so mixed, with conflicting reports of the exact structural and functional correlates of chronic inflammation. To overcome the limitations of this sampling bias,
this thesis explores the use of DNA methylation (DNAm) as an epigenetic mechanism to characterise chronic inflammation. DNAm refers to chemical modifications of DNA that have the ability to activate or deactivate genes, representing a distinctive interface between environmental factors and genetic processes. Preliminary work has shown the potential of DNAm to capture variability in blood protein levels and its value in examining associations with incident health outcomes. This PhD aims to investigate different measures of chronic inflammation in relation to both brain structure and cognitive ability, utilising data from three cohort studies spanning the human lifespan. By integrating DNAm with these assessments, we can gain valuable insights into the role of chronic inflammation in brain ageing.

The introductory chapters provide a critical overview of the literature on chronic inflammation, brain and cognitive ageing, and DNA methylation. Chapter 1 focuses on chronic inflammation, our approaches to characterising it, and how it associates with age, before evaluating the evidence for how peripheral inflammation relates to neuroinflammation and related brain health outcomes. Chapter 2 introduces the lifespan perspective of brain ageing, and how we study cognitive and brain ageing with neuroimaging and cognitive testing methods, before a literature overview of the determinants of poor cognitive and brain structural outcomes. Chapter 3 introduces the epigenetic mechanism DNA methylation and its use-case for signposting both the risk of factors for disease and disease phenotypes, concluding with a focused literature review of DNA methylation signatures of exposures, including inflammation. Chapter 4 summarises the main aspects of the introductory chapters that have informed the research objectives of this thesis, outlining the gaps and weaknesses in the literature to date and how the aims of this thesis attempt to address these. The final introductory chapter (5) describes the population cohorts used throughout this PhD and methodologies used to conduct the studies. The empirical work is then presented in Chapters 6-8 which investigate the relationship between chronic inflammation and brain health across the lifecourse. These analyses aim to assess the validity of DNA methylation biomarkers to explore the downstream effects of chronic low-grade inflammation, map the associated brain structural alterations linking inflammation to cognitive ability, and clarify the influence of lifestyle factors on these relationships. Chapter 6 is set in a cohort of older-age adults in the Lothian Birth Cohort 1936, Chapter 7 in a mixed cohort of term and preterm infants in Theirworld Edinburgh Birth Cohort and Chapter 8 in a cohort spanning midlife to old age in the Stratifying Resilience and Depression Longitudinally Cohort. A summary and discussion on the main findings of this thesis, and points for future work, are then presented in Chapter 9.
Declaration of Originality

I declare that this thesis is my own composition and that it has not been submitted for any other degree or professional qualification at this university or any other institution. Parts of the work comprising this thesis have been previously published and include contributions from coauthors, the details of which are included in the Statement of Authorship section below. The included publications comprising the main empirical chapters are my own work, except where indicated otherwise.

Signed: [REDACTED signature]  Date: 27/05/2023
Statement of Authorship

Most of the data used in the analyses presented in this thesis were pre-processed and made available by the individual population cohort groups, the details of which are included in Chapter 5 of this thesis. This includes all proteomic, DNA methylation, lifestyle- and demographic variables, data on cognitive tests, and MRI imaging. All analyses for the following chapters can be found at the relevant repositories at https://github.com/EleanorSC. Figures throughout the thesis, unless otherwise stated, are drawn by E.L.S.C using Adobe Cloud Creative Suite platform and BioRender. As this thesis contains work which led to jointly authored publications, my contributions and those of other authors to this work are indicated below.

The work presented in Chapter 6 has been published in Neurology. https://doi.org/10.1212/WNL.0000000000012997. Author contributions are as follows: E.L.S.C, R.E.M and S.R.C conceived and designed the research; E.L.S.C conducted the statistical analyses; A.J.S, S.M, M.A.H, M.E.B, J.M.W, I.J.D contributed to data collection and preparation; E.L.S.C drafted the article, with assistance from supervisors (R.E.M, H.C.W, V.E.M, S.J.C); all authors reviewed the manuscript.

The work presented in Chapter 7 has been published in Brain, Behaviour and Immunity. https://doi.org/10.1016/j.bbi.2023.03.011. Author contributions are as follows: E.L.S.C, R.E.M, J.P.B and S.R.C. conceived and designed the research; M.B.C, G.S, L.M, M.J.T, M.E.B, A.J.Q., J.P.B., contributed to data collection and preparation; E.L.S.C conducted the analysis; E.L.S.C, R.E.M, J.P.B and S.R.C drafted the article; all authors reviewed the manuscript.

The work presented in Chapter 8 contains a discussion of analyses examining DNAm CRP in the STRADL cohort that was published in Brain, Behaviour and Immunity. https://doi.org/10.1016/j.bbi.2020.11.024 where C.G, H.C.W and S.R.C conceived and designed the research; C.G analysed the data; E.L.S.C assisted in statistical analyses and all authors reviewed the manuscript. The main analyses reported in Chapter 8 is in preparation for submission. Author contributions are as follows: E.L.S.C, R.E.M and S.R.C. conceived and designed the research; D.A.G generated the weights for the DNAm signatures; E.L.S.C conducted the analysis; X.S, A.S, M.E.B, contributed to data collection and preparation; E.L.S.C, R.E.M and S.R.C drafted the article.
Other publications or preprints arising during the course of the PhD that were not directly related to the PhD thesis are listed below:

- Jaggi, A; **Conole, E.L.S**; Raisi-Estabragh, Z; Gkontra, P; McCracken, C; Neubauer, S; Petersen, S; Cox, S. R; Lekadir, K. A Structural Heart-Brain Axis Mediates the Association Between Cardiovascular Risk and Cognitive Function. (2023) medRxiv. [https://doi.org/10.1101/2022.09.15.22279275](https://doi.org/10.1101/2022.09.15.22279275) [under review at Imaging Neuroscience]


Dedicated to my parents Mike Shepherd and Grainne Conole
Acknowledgements

I had always planned to dedicate this thesis to my dad, the reason I’m a scientist, who died suddenly just before I started my undergraduate degree in 2014. In a weird twist of circumstance, and frankly terrible timing, my mum fell critically ill during the latter half of this PhD. As a result, I’ve written sections of this thesis in hospital cafés, beside her bedside, and in hospital corridors. With what felt like heavy irony, I found myself writing about C-Reactive Protein (CRP) readings when mum was in ICU with a CRP level of 680 mg/L. In the wake of this, it is difficult to get across the level of support and understanding I have received at every turn and from every direction. Not least of which from mum herself, who this thesis is now jointly dedicated to.

First and foremost, I am immensely grateful to my supervisors, Simon Cox, Riccardo Marioni, Heather Whalley, and Veronique Miron – for pushing me when I needed challenging, and advocating for me when I needed supporting. Their input, critique and guidance has been invaluable in helping me to learn how to do research and enjoy it in the process. Thanks also to my thesis committee chair Mark Bastin for his forensic thesis draft edits, constant encouragement and practical advice re matters of care and power of attorney. The fact that I have loved my PhD and want to stay in academia is a testament to my supervisory team and the environment they fostered. I am especially grateful to my primary supervisor, Simon, for endless generosity with his time, knowledge, and kindness.

My sincere thanks go to my collaborators and co-authors for their expertise and contributions to various projects, and to the numerous experts who have provided me with guidance, feedback, and suggestions along the way. Thanks to Rob Hillary, James Boardman, Ian Deary in particular for their subject-specific expertise. I would also like to express my gratitude to the members of the Lothian Birth Cohort 1936, Generation Scotland, and Theirworld Edinburgh Birth Cohort, as well as the research team members and funders who contributed to the studies. Thanks too to my F1 colleagues Joanne Moodie, Colin Buchannan, Sarah Harris, Gail Davies and everyone else who made 7GS such a great place to work. I am also grateful for the continued academic and administrative support of the Wellcome Trust Translational Neuroscience programme directors and coordinators.

There are of course many people in my personal life who have kept me afloat. To my sister Tabby, for endless phone calls and for making lockdown so memorable (meep); to Elodie Cox for pushing me to ‘just submit’ when I most needed it; and to my flatmates Rosalyn, Fred, Jamie and Ollie for being my favourite people to ever live with. Particular thanks to my best and oldest friend, Dr. Rosalyn Pearson - Gnomey – thanks for teaching me everything I know about coding and for so much more. I could not have done this PhD without you.

1 for context on just how high this is, see page 19
Finally, to Ollie. We have talked about how when I first met you at 19, you set this unfair precedent of what my time at university would be like. That I thought there would be all these people I’d connect with such instant intensity… only to realise that, actually, that only happens a handful of times. Thank you for cooking for me for the entire time I wrote this up, for driving me to be with mum, and for never complaining that I work too late. Thanks for fixing things, explaining things, making things. I’m forever glad I met you, re-met you, and tricked you into being so involved in my life. I love you!

Funding acknowledgement

This 4-year PhD in Translational Neuroscience was funded entirely by the Wellcome Trust, grant number 108890/Z/15/Z.
# Table of Contents

1 **Chronic inflammation** ................................................................. 2
  1.1 Introduction ................................................................................. 2
  1.2 Inflammaging ............................................................................. 3
  1.3 Acute vs chronic inflammation .................................................. 4
  1.4 The Innate and Adaptive Immune System .................................... 4
    1.4.1 Early-life inflammation and immune dynamics ...................... 7
    1.4.2 Dynamics and changes of the immune system with age .......... 9
  1.5 Measuring chronic inflammation: approaches and biomarkers .... 11
    1.5.1 Acute phase proteins .......................................................... 12
    1.5.2 C-Reactive protein ............................................................. 14
    1.5.3 Why is CRP the most popular biomarker of inflammation? ...... 29
    1.5.4 Other individual inflammatory biomarkers ........................... 30
    1.5.5 Composite and multi-omic biomarkers of inflammation ......... 34
  1.6 Neuroinflammation .................................................................... 36
    1.6.1 Systemic inflammation vs neuroinflammation ...................... 37
    1.6.2 Blood brain barrier ............................................................. 40
    1.6.3 Mechanisms of immune-crosstalk between periphery and brain .. 42
  1.7 Summary ...................................................................................... 45

2 **Brain and cognitive ageing** ......................................................... 47
  2.1 Introduction ................................................................................. 47
  2.2 Lifespan perspective of brain ageing ........................................... 48
  2.3 Cognitive ability ......................................................................... 50
    2.3.1 Assessing cognitive ability ................................................... 50
  2.4 Cognitive ageing ......................................................................... 54
    2.4.1 Pathological cognitive ageing (dementia) ............................... 54
    2.4.2 Non-pathological cognitive ageing ........................................ 56
    2.4.3 Inflammation and cognitive ageing ...................................... 57
  2.5 Neuroimaging ............................................................................. 60
    2.5.1 Structural Magnetic Resonance Imaging (MRI) ....................... 60
    2.5.2 Diffusion MRI .................................................................... 61
    2.5.3 Diffusion Tensor Imaging ..................................................... 62
  2.6 Assessing brain ageing with MRI ................................................ 66
# Global and regional brain changes with age and chronic inflammation

2.6.1 Global and regional brain changes with age and chronic inflammation .................................................67

2.6.2 ‘Brain-age’ methods .................................................................................................................................71

2.7 Early life experiences influence the brain throughout life .................................................................73

2.7.1 Preterm birth...........................................................................................................................................74

2.7.2 Encephalopathy of Prematurity: neuroimaging substrates of PTB.........................................................75

2.7.3 Inflammation, white matter tract development, and contribution to EoP .............................................77

2.8 Summary.....................................................................................................................................................80

## 3 DNA methylation

3.1 Introduction ..................................................................................................................................................81

3.2 Discovery ....................................................................................................................................................81

3.3 Molecular mechanisms ............................................................................................................................82

3.4 Measurement ..............................................................................................................................................84

3.5 Analysis: EWAS vs Candidate Gene approaches .......................................................................................85

3.6 Epigenetic epidemiology ..........................................................................................................................86

3.6.1 Association of DNAm with health and disease......................................................................................88

3.6.2 Association of DNAm with lifestyle and environmental exposure.......................................................90

3.7 The case for using DNAm to characterise chronic inflammation .............................................................92

3.7.1 DNAm and CRP: Candidate Gene and EWAS studies ...........................................................................93

3.8 DNAm signatures (methylation risk scores).............................................................................................99

3.8.1 DNAm signatures of age (epigenetic clocks) .........................................................................................102

3.8.2 DNAm signatures of exposures and risk factors .....................................................................................103

3.8.3 DNAm signatures of inflammation ........................................................................................................109

3.8.4 Application of DNAm signatures to examine brain health outcomes ..................................................111

3.9 Summary.....................................................................................................................................................113

## 4 Research gaps and aims of thesis

4.1 Introduction ..................................................................................................................................................114

4.2 Gaps in the literature ..................................................................................................................................115

4.3 Aims of thesis ..............................................................................................................................................119

4.3.1 Chapter 6 Aims and Hypotheses: ............................................................................................................120

4.3.2 Chapter 7 Aims and Hypotheses: ............................................................................................................121

4.3.3 Chapter 8 Aims and Hypotheses: ............................................................................................................122

4.4 Summary.....................................................................................................................................................123

## 5 Study cohorts and methods

5 Study cohorts and methods ...........................................................................................................................124
5.1 The Lothian Birth Cohort 1936.................................................................124
  5.1.1 Ethics and funding............................................................................125
  5.1.2 Clinical assessment........................................................................125
  5.1.3 Inflammatory protein data in the Lothian Birth Cohort 1936...........125
  5.1.4 Cognitive data in LBC1936..............................................................126
  5.1.5 Neuroimaging data in LBC1936.......................................................128
  5.1.6 DNA methylation data in the Lothian Birth Cohort 1936.................132
  5.1.7 Inflammatory-DNAm signature.......................................................132
5.2 Theirworld Edinburgh Birth Cohort (TEBC):.........................................133
  5.2.1 Ethics and funding............................................................................133
  5.2.2 Clinical assessment (health and lifestyle data).................................133
  5.2.3 Neuroimaging data in Theirworld Edinburgh Birth Cohort..............134
  5.2.4 DNA methylation data in Theirworld Edinburgh Birth Cohort..........137
  5.2.5 DNAm signature pipeline...............................................................138
5.3 Stratifying Resilience and Depression Longitudinally cohort (STRADL) ...139
  5.3.1 Ethics and funding............................................................................139
  5.3.2 Proteomics data in STRADL..............................................................139
  5.3.3 Cognitive data in STRADL...............................................................140
  5.3.4 Neuroimaging data in STRADL.......................................................142
  5.3.5 DNA methylation data in STRADL..................................................146
  5.3.6 Machine learning approaches for selecting relevant CpG sites..........146
  5.3.7 Inflammatory-related DNAm signature pipeline.............................149
5.4 Summary.............................................................................................151
6 Chronic inflammation and brain health in old age.................................152
  6.1 Introduction.......................................................................................152
  6.2 DNA methylation and protein markers of chronic inflammation and their
       associations with brain and cognitive ageing......................................154
  6.3 Conclusion........................................................................................168
7 Chronic inflammation and brain health in early life.............................169
  7.1 Introduction.......................................................................................169
  7.2 Immuno-epigenetic signature derived in saliva associates with the
       encephalopathy of prematurity and perinatal inflammatory disorders ....170
  7.3 Conclusion........................................................................................188
8 Chronic inflammation and brain health in midlife................................191
  8.1 Introduction.......................................................................................191
8.2 Examining multi-omic signatures of inflammation in relation to brain structure, cognitive ability, and brain ageing.................................................................192
8.3 Conclusion ........................................................................................................232

9 Discussion ...........................................................................................................235
9.1 Overview ............................................................................................................235
9.2 The case for DNAm signatures as indices of chronic inflammation...............240
9.3 Multi-omic signatures reveal inflammation-brain ageing relationships.........243
9.4 Links between chronic inflammation, brain structure and cognition ............246
9.5 Inflammation-related DNAm and white matter microstructure .................249
9.6 The influence of health and lifestyle on inflammation-brain health relationships257
9.7 Limitations ........................................................................................................260
  9.7.1 Cohorts and generalisability ......................................................................261
  9.7.2 Neuroimaging metrics ..............................................................................262
  9.7.3 Proteomic assays .......................................................................................266
  9.7.4 DNAm arrays ............................................................................................267
  9.7.5 DNAm tissue and cell specificity ..............................................................268
  9.7.6 Sources of confounding ...........................................................................274
  9.7.7 Temporal stability of DNAm ....................................................................276
  9.7.8 Inconsistencies across the chapters in terms of study design .................277
  9.7.9 Causality ..................................................................................................281
9.8 Recommendations for future research .............................................................287
9.9 Final summary ...................................................................................................292

10 References .........................................................................................................293

11 Appendix – Supplementary data .....................................................................336
11.1 Chapter 6 Supplementary data .....................................................................336
11.2 Chapter 7 Supplementary data .....................................................................360
11.3 Chapter 8 Supplementary data .....................................................................389
## List of Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>AD</td>
<td>Alzheimer's disease</td>
</tr>
<tr>
<td>AIM2</td>
<td>Absent in Melanoma 2</td>
</tr>
<tr>
<td>AHRR</td>
<td>Aryl-hydrocarbon receptor repressor</td>
</tr>
<tr>
<td>APOE</td>
<td>Apolipoprotein E</td>
</tr>
<tr>
<td>BBB</td>
<td>Blood brain barrier</td>
</tr>
<tr>
<td>BPD</td>
<td>Bronchopulmonary dysplasia</td>
</tr>
<tr>
<td>CI</td>
<td>Confidence interval</td>
</tr>
<tr>
<td>CNS</td>
<td>Central nervous system</td>
</tr>
<tr>
<td>COPD</td>
<td>Chronic obstructive pulmonary disease</td>
</tr>
<tr>
<td>CpG</td>
<td>Cytosine-Guanine dinucleotide</td>
</tr>
<tr>
<td>CRP</td>
<td>C-reactive protein</td>
</tr>
<tr>
<td>CP</td>
<td>Cerebral palsy</td>
</tr>
<tr>
<td>CSF</td>
<td>Cerebrospinal fluid</td>
</tr>
<tr>
<td>CVD</td>
<td>Cardiovascular disease</td>
</tr>
<tr>
<td>DAMPs</td>
<td>Damaged associated molecular patterns</td>
</tr>
<tr>
<td>DALY</td>
<td>Disability-adjusted life years</td>
</tr>
<tr>
<td>dMRI</td>
<td>Diffusion Magnetic Resonance Imaging</td>
</tr>
<tr>
<td>DNA</td>
<td>Deoxyribonucleic acid</td>
</tr>
<tr>
<td>DNAm</td>
<td>DNA methylation</td>
</tr>
<tr>
<td>DNMT</td>
<td>DNA methyltransferase</td>
</tr>
<tr>
<td>DOHaD</td>
<td>Developmental Origins of Health and Disease</td>
</tr>
<tr>
<td>DSM</td>
<td>Diagnostic and statistical manual of mental disorders</td>
</tr>
<tr>
<td>eGFR</td>
<td>Estimated glomerular filtration rate</td>
</tr>
<tr>
<td>EHR</td>
<td>Electronic health record</td>
</tr>
<tr>
<td>eQTL</td>
<td>Expression quantitative trait locus</td>
</tr>
<tr>
<td>EoP</td>
<td>Encephalopathy of Prematurity</td>
</tr>
<tr>
<td>ESR</td>
<td>Erythrocyte sedimentation rate</td>
</tr>
<tr>
<td>EWAS</td>
<td>Epigenome-wide association studies</td>
</tr>
<tr>
<td>F2RL3</td>
<td>F2R-like thrombin or trypsin receptor 3</td>
</tr>
<tr>
<td>FA</td>
<td>Fractional anisotropy</td>
</tr>
<tr>
<td>FDR</td>
<td>False discovery rate</td>
</tr>
<tr>
<td>FTD</td>
<td>Frontotemporal dementia</td>
</tr>
<tr>
<td>FUMA</td>
<td>Functional annotation and mapping</td>
</tr>
<tr>
<td>Term</td>
<td>Definition</td>
</tr>
<tr>
<td>---------------</td>
<td>-------------------------------------------------</td>
</tr>
<tr>
<td>GO</td>
<td>Gene ontology</td>
</tr>
<tr>
<td>GWAS</td>
<td>Genome-wide association studies</td>
</tr>
<tr>
<td>HD</td>
<td>Huntington’s disease</td>
</tr>
<tr>
<td>HPA</td>
<td>Hypothalamic-Pituitary-Adrenal</td>
</tr>
<tr>
<td>HCA</td>
<td>Histologic chorioamnionitis</td>
</tr>
<tr>
<td>HR</td>
<td>Hazard ratio</td>
</tr>
<tr>
<td>IBD</td>
<td>Inflammatory bowel disease</td>
</tr>
<tr>
<td>ICV</td>
<td>Intracranial volume</td>
</tr>
<tr>
<td>IHDI</td>
<td>Ischemic heart disease</td>
</tr>
<tr>
<td>IL</td>
<td>Interleukin</td>
</tr>
<tr>
<td>IDP</td>
<td>Imaging-derived phenotype</td>
</tr>
<tr>
<td>IQ</td>
<td>Intelligence quotient</td>
</tr>
<tr>
<td>JAK-STAT</td>
<td>Janus kinase/signal transduction and activator of transcription</td>
</tr>
<tr>
<td>LASSO</td>
<td>Least absolute shrinkage and selection operator</td>
</tr>
<tr>
<td>LBC1936</td>
<td>Lothian Birth Cohort 1936</td>
</tr>
<tr>
<td>LBD</td>
<td>Lewy body dementia</td>
</tr>
<tr>
<td>MAPK</td>
<td>Mitogen-activated protein kinase</td>
</tr>
<tr>
<td>MD</td>
<td>Mean diffusivity</td>
</tr>
<tr>
<td>MHC</td>
<td>Major histocompatibility complex</td>
</tr>
<tr>
<td>MMP</td>
<td>Matrix metalloproteinase</td>
</tr>
<tr>
<td>MMSE</td>
<td>Mini-mental state examination</td>
</tr>
<tr>
<td>MND</td>
<td>Motor neuron disease</td>
</tr>
<tr>
<td>ML</td>
<td>Machine learning</td>
</tr>
<tr>
<td>mQTL</td>
<td>Methylation quantitative trait locus</td>
</tr>
<tr>
<td>MR</td>
<td>Mendelian randomisation</td>
</tr>
<tr>
<td>MRI</td>
<td>Magnetic resonance imaging</td>
</tr>
<tr>
<td>mRNA</td>
<td>Messenger RNA</td>
</tr>
<tr>
<td>MS</td>
<td>Mass spectrometry</td>
</tr>
<tr>
<td>NDI</td>
<td>Neurite density index</td>
</tr>
<tr>
<td>NEC</td>
<td>Necrotising Enterocolitis</td>
</tr>
<tr>
<td>NF-κB</td>
<td>Nuclear factor kappa beta</td>
</tr>
<tr>
<td>NK</td>
<td>Natural Killer [T-cells]</td>
</tr>
<tr>
<td>NODDI</td>
<td>Neurite orientation dispersion and density imaging</td>
</tr>
<tr>
<td>ns</td>
<td>Non-significant</td>
</tr>
<tr>
<td>ODI</td>
<td>Orientation dispersion index</td>
</tr>
<tr>
<td>OLS</td>
<td>Ordinary least squares</td>
</tr>
<tr>
<td>OR</td>
<td>Odds ratio</td>
</tr>
<tr>
<td>PAMPs</td>
<td>Pathogen-associated molecular patterns</td>
</tr>
<tr>
<td>PCA</td>
<td>Principal component analysis</td>
</tr>
<tr>
<td>PCR</td>
<td>Polymerase chain reaction</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Full Form</td>
</tr>
<tr>
<td>--------------</td>
<td>-----------</td>
</tr>
<tr>
<td>PD</td>
<td>Parkinson’s disease</td>
</tr>
<tr>
<td>PNT</td>
<td>Probabilistic neighbourhood tractography</td>
</tr>
<tr>
<td>pPROM</td>
<td>Preterm Premature Rupture of the Membranes</td>
</tr>
<tr>
<td>pQTL</td>
<td>Protein quantitative trait locus</td>
</tr>
<tr>
<td>PRS</td>
<td>Polygenic risk score</td>
</tr>
<tr>
<td>ROI</td>
<td>Region of interest</td>
</tr>
<tr>
<td>ROP</td>
<td>Retinopathy of Prematurity</td>
</tr>
<tr>
<td>ROS</td>
<td>Reactive Oxygen Species</td>
</tr>
<tr>
<td>SAA</td>
<td>Serum Amyloid A</td>
</tr>
<tr>
<td>SASP</td>
<td>Senescence Associated Secretory Phenotype</td>
</tr>
<tr>
<td>SCID</td>
<td>Structured Clinical Interview for DSM-IV</td>
</tr>
<tr>
<td>SD</td>
<td>Standard deviation</td>
</tr>
<tr>
<td>SE</td>
<td>Standard error</td>
</tr>
<tr>
<td>SIMD</td>
<td>Scottish index of multiple deprivation</td>
</tr>
<tr>
<td>SLE</td>
<td>Systemic lupus erythematosus</td>
</tr>
<tr>
<td>SMR</td>
<td>Scottish morbidity records</td>
</tr>
<tr>
<td>SNP</td>
<td>Single-nucleotide polymorphism</td>
</tr>
<tr>
<td>SOMAmers</td>
<td>Slow Off-rate Modified Aptamers</td>
</tr>
<tr>
<td>STRADL</td>
<td>Stratifying Resilience and Depression Longitudinally</td>
</tr>
<tr>
<td>TBCA</td>
<td>Tubulin-specific chaperone A</td>
</tr>
<tr>
<td>TBSS</td>
<td>Tract-based spatial statistics</td>
</tr>
<tr>
<td>TLR</td>
<td>Toll-Like Receptor</td>
</tr>
<tr>
<td>TREM2</td>
<td>Triggering receptor expressed on myeloid cells 2</td>
</tr>
<tr>
<td>T2D</td>
<td>Type 2 diabetes</td>
</tr>
<tr>
<td>VEGFA</td>
<td>Vascular endothelial growth factor A</td>
</tr>
<tr>
<td>WMH</td>
<td>White matter hyperintensity</td>
</tr>
</tbody>
</table>
List of Figures

Figure 1.1 Roles of the innate and adaptative immune system ................................................. 6
Figure 1.2 Susceptibility to chronic inflammation in early life and old age ....................... 10
Figure 1.3 Changes in plasma concentrations of acute phase proteins .............................. 13
Figure 1.4 C-Reactive Protein as a molecular alarm system .................................................. 16
Figure 1.5 CRP Epidemiology: associations of CRP with 1,673 diseases / phenotypes assimilated from literature text-mining......................................................... 24
Figure 1.6: Drivers and consequences of chronic inflammation ........................................... 28
Figure 1.7 Classification of inflammatory biomarkers and presence of inflammatory mediators in different inflammatory processes......................................................... 31
Figure 1.8 Differences in cell populations between the peripheral immune system and brain ................................................................................................................. 38
Figure 1.9 Mechanisms of neuroinflammation: microglial priming ..................................... 39
Figure 1.10 Blood brain barrier ............................................................................................... 41
Figure 1.11 Blood brain barrier disruption ............................................................................... 43
Figure 2.1 Lifespan perspective on brain and cognitive ageing .............................................. 49
Figure 2.2 MRI principles: distinguishing tissue types.......................................................... 60
Figure 2.3. Diffusion Tensor Model. .......................................................................................... 63
Figure 2.4 FA map, using anisotropy and isotropy in brain imaging ..................................... 64
Figure 2.5. Trajectories of change in MRI phenotypes across the lifespan. ......................... 68
Figure 2.6 Brain-age vs chronological age distributions in STRADL ................................. 72
Figure 3.1 The process of DNA methylation (cytosine to 5-methylcytosine) .................... 83
Figure 3.2. Epigenetic epidemiological frameworks: how epigenetics relates to genotype, exposure and disease ................................................................................. 88
Figure 3.3 Flow chart for appropriate methods for creating DNAm signatures .......... 101
Figure 5.1 LBC1936 cortical thickness data ........................................................................... 129
Figure 5.2 PNT derived white matter tract ROIs in LBC1936 participants ....................... 131
Figure 5.3 Representative neonatal brain maps ................................................................... 135
Figure 5.4 Scree plot demonstrating a strong single factor capturing common variance across white matter tracts ................................................................. 136
Figure 5.5 DNAm CRP signature pipeline ............................................................................. 138
Figure 5.6 Distributions of DNAm CRP in TEBC participants ............................................ 139
Figure 5.7 Distributions of cognitive test metrics in STRADL.................................. 142
Figure 5.8 Distributions of neuroimaging metrics used in STRADL ..................... 144
Figure 5.9 Ridge, LASSO and elastic net .......................................................... 148
Figure 5.10 Elastic-net DNAm signatures of circulating protein levels............. 150
Figure 9.1 Overview of thesis aims ................................................................. 237
Figure 9.2 DNAm biomarkers of exposure ....................................................... 242
Figure 9.3 Grey and white matter exhibit distinct trajectories starting in the perinatal period ........................................................................................................... 252
Figure 9.4 Chronic inflammation vs acute inflammation in preterm infants ........ 278
Figure 9.5 Future of DNAm biomarkers: potential, hurdles, solutions ............ 290
List of Tables

Table 1.1 Positive and negative acute phase proteins ................................. 12
Table 1.2 Interpretation of plasma CRP levels in humans ............................ 19
Table 3.1 Epigenome Wide Association Studies of C-Reactive Protein .......... 96
Table 3.2 Epigenetic clocks ........................................................................ 103
Table 3.3 DNAm signatures of exposure (smoking, alcohol consumption and inflammation) ................................................................. 107
Table 5.1 Clinical data of inflammatory exposures .................................... 134
Table 5.2 STRADL Cognitive test score summary .................................... 141
Table 5.3 STRADL white matter tract ROIs (DTI data) ............................. 145
1 Chronic inflammation

1.1 Introduction

Inflammation is a natural response of the body to injury or infection with recognisable characteristics, such as redness, swelling, irritation, pain and loss of function. The purpose of inflammation is to fight off infection, protect the body from further damage and restore homeostasis. Inflammatory responses that fail to subside are considered chronic; this type of inflammation seems to increase with age and has been linked to a broad spectrum of health consequences, including those concerning the brain.

This Chapter begins by addressing the main theory of ‘inflammaging’, before going on to discuss the differences between acute vs chronic inflammation, and the function of the adaptive and innate immune system. The ways in which researchers attempt to characterise inflammation are then discussed, starting with the acute phase proteins. An in-depth assessment of the most widely researched marker of inflammation, C-Reactive Protein (CRP), follows, with commentary on its popularity, genetics, relation to health and disease, and a critique of its suitability for profiling chronic inflammation in population cohorts. Other approaches to indexing inflammation are then discussed, including using composite metrics or multi-omics to measure multiple inflammatory mediators and their associations with age related diseases. The link between systemic inflammation and neuroinflammation is then addressed: with an overview of the differences between the peripheral immune system and neural immune system and the key mechanistic pathways that enable crosstalk between them. The chapter concludes with an overview of systemic inflammation, neuroinflammation and how these perpetuating cycles can lead to alterations in brain structure, setting us up for a discussion of brain and cognitive ageing in Chapter 2.
1.2 Inflammaging

Our populations are ageing. In 2020, the number of people aged 85 years and over in the UK was estimated to be 1.7 million (totalling around 2.5% of the UK population), a number which is projected to increase almost 2-fold by 2045 to 3.1 million ~ 4.3% of the UK population (Robards et al., 2022).

However, data indicates that this increase in how long we’re living does not translate to how well we’re living (Welsh, Matthews and Jagger, 2021), with cases of cognitive decline in the UK almost doubling in the last ten years (Hallam et al., 2022). In general, there is a widening gap between the two profiles of lifespan and healthspan (how long we live in the absence of chronic illness and disability), with few countries demonstrating increases of disability-free life expectancy in line with overall life expectancy (Jagger et al., 2020). Because of this, understanding the risk factors underlying age-related diseases has become a public health priority.

Inflammation has been suggested as the common link between numerous diseases, and a key driver of age-related phenotypes. In the 2000s, Franceschi and colleagues coined this process ‘inflammm-aging’ (Franceschi et al., 2000) – a phrase to describe how chronic inflammation is not only a dominant feature of advancing age but may be an active driving force for both age-related morbidity and risk of mortality. Inflammaging is also believed to be a contributing factor to age-related cognitive decline and individual differences in cognitive ageing trajectories. Understanding the underlying mechanisms driving inflammaging and its impacts on different aspects of brain structure and function is crucial for identifying factors that can alleviate its effects and promote healthy cognitive aging.

In order to discuss the implications of inflammaging on brain health outcomes, we first need to examine how chronic inflammation relates to the broader context of the immune system.
1.3 Acute vs chronic inflammation

The inflammatory response is a coordinated system of events: the inflammatory stimuli, the receptors that detect this, the inflammatory mediators created by the receptors, and the target tissues that inflammatory mediators affect. Each of these parts can take multiple forms, and their distinct combinations are governed by different inflammatory pathways (Medzhitov, 2010). The regulation and coordination of these components is what initiates, sustains and resolves inflammation. Similarly, the difference between an affective immune response and tissue damage, cell death and illness, is dependent on the balanced regulation of the network of inflammatory mediators and the immune cells responding to them.

Acute inflammation is a short-term reaction to insult – characterised by both quick onset and resolution. It is usually triggered by an infection (caused by a bacteria or virus) or injury (such as frostbite, burn, or tissue trauma). Chronic inflammation is a long-term – often low-grade – response that fails to resolve, even after the initiating stimulus has been dealt with.

To understand how acute inflammation and chronic inflammation relate to each other, and the transition from efficient self-resolving inflammatory responses to chronic ones, the two main arms of the immune system need to be discussed.

1.4 The Innate and Adaptive Immune System

The immune system has two main lines of defence: the innate immune response and the adaptive immune response. These can be thought of as generalised vs specific lines of attack that operate on different time scales.

The innate immune response is the body’s first line of defence against infection and injury and the related pathogens or damage that result from this. It is a non-specific, acute response that is activated immediately upon
the recognition of either pathogens or dying cells, and is composed of both physical barriers, such as skin and mucous membranes, as well as chemical and cellular responses. Innate immunity is what we typically think of when we talk about acute inflammation. **Figure 1.1** illustrates the key cells of the innate immune response, which include monocytes (which can develop into macrophages and dendritic cells), mast cells and the granulocytes (eosinophils, basophils, and neutrophils).

The macrophages, neutrophils and dendritic cells are the primary sentinels of the systemic circulation: these have Toll-like receptors (TLRs) on their cell surface, which can detect pathogens via pathogen-associated molecular patterns (PAMPs) and signs of dead or dying cells via damaged associated molecular patterns (DAMPs) (Cronkite and Strutt, 2018; Gong et al., 2020). When a foreign invader, sign of cell death, or injury is detected, these innate immune cells secrete a variety of inflammatory mediators (discussed in depth in section 1.5.4), such as cytokines, prostaglandins, and chemokines. These chemical mediators have various roles both in priming local vascular tissue, activating other innate immune cells into reactive phenotypes, and sending out signals to the adaptive immune cells to join in on the action.

The adaptive immune response is a more specific, slightly lagged response that follows (see bottom panel, **Figure 1.1**). Instead of generalised patrollers and non-specific recruits, it is composed of specialised immune cells (B cells and T cells) that recognise specific antigens and mount a targeted response in retaliation to the perceived threat. When signalled by the innate immune cells, the white blood cells (leucocytes) of the adaptive immune response migrate to the site of inflammation. Though there are these broad differences, some immune cells are considered both part of the innate and adaptive response, such as the Natural Killer (NK) T cells and Gamma-delta (γδ) T cells.
Figure 1.1 Roles of the innate and adaptative immune system

**Schematic of cells involved in innate and adaptive immune response and their timelines of action**. Adapted from (Dranoff, 2004; Sette and Crotty, 2021).

These two systems are reciprocal and synergistic, working alongside each other to protect the body and communicating between each other. The innate response is activated first and triggers the adaptive response; the adaptive response is then able to recognise specific antigens, allowing for a more targeted response. Because the innate immune system effectively acts as a gatekeeper for initiating an efficient and effective adaptive immune response, a failure in this exchange results in a weak adaptive immune response, which

---

2 Figure redrawn with top panel of figure informed by separate sources: top panel - Dranoff (2004) and bottom panel adapted from Sette & Crotty (2021); upper panel of figure adapted with permission from Macmillan Publishers Ltd: Nature Reviews Cancer. Dranoff G. Nat Rev Cancer. 2004;4:11-22, © 2004.
can result in prolonged non-specific inflammation. Equally, the B-cells and T-cells of the adaptive immune response can regulate innate immune cells via their secretion of antibodies and cytokines (Paulnock, 1992) – this reciprocal activation is just one example of how self-perpetuating inflammation can arise.

Something to consider when asking questions about how inflammation relates to various states of health and disease is that these two aspects of immunity are not static and exhibit age-related changes. Both early-life and later-life appear to be higher periods of risk for dysregulated immune dynamics and represent specific windows of vulnerability for various health outcomes linked to chronic inflammation. This is perhaps evidenced most plainly in how both the young and old are more susceptible to infection than other age groups (Dowling and Levy, 2014; Simon, Hollander and McMichael, 2015), with a recent example of this being COVID-19. How chronic inflammation varies at these two extremes of the lifespan, potentially predisposing individuals to adverse health effects, will be discussed in depth next.

1.4.1 Early-life inflammation and immune dynamics

Compared to the adult immune system, the neonatal immune system relies on different dynamics to afford protection against infection and injury. Infants display immune responses that consist of naïve cell-types that mature both in utero and postnatally, with step-changes in maturity occurring during the third trimester of pregnancy and birth (Schultz et al., 2004; Gibbons et al., 2009, 2014; Dowling and Levy, 2014). Disruption or dysregulation in these innate immune system dynamics renders neonates vulnerable to sustained inflammation and its associated comorbidities and consequences, such as infection, sepsis and impaired neurodevelopment (Tsafaras, Ntontsi and Xanthou, 2020).
In very early-life, these immune dynamics manifest as infants displaying heightened inflammatory responses, consisting of unbalanced ratios of pro-inflammatory and anti-inflammatory mediators (Schultz et al., 2004). This effect is pronounced when infants are born too early (preterm), where the prevalence of unremitting and sustained inflammation is much higher than in term infants (Skogstrand et al., 2008; Leviton, Fichorova, et al., 2013; Dammann and Leviton, 2014; Dammann et al., 2016), a topic further reviewed in Chapter 2.7. The presence of this type of inflammation is also linked to the development and progression of multiple adverse health outcomes, including those related to neurodevelopment (Hansen-Pupp et al., 2008; Bose et al., 2013; O’Shea et al., 2013; Dammann and Leviton, 2014; Van Steenwinckel et al., 2014; Yanni et al., 2017; Bennet et al., 2018; Cuestas et al., 2019; Humberg et al., 2020; Reiss et al., 2022).

On a parallel to the cyclical patterns that appear to initiate and sustain chronic inflammation in later-life, inflammation is simultaneously a cause and consequence of preterm birth, with maternal, fetal and postnatal origins. Multiple drivers of inflammation are considered triggers for early labour (from intrauterine infection, prolonged rupture of the membranes, and aspects of maternal health in pregnancy such as preeclampsia) and similarly, the mode of preterm delivery (caesarean vs vaginal birth), risk of co-morbid inflammatory conditions postnatally, and neonatal intensive care unit (NICU) related exposures, interventions and medications, can all confer increased risk of inflammation (Hagberg et al., 2015; Humberg et al., 2020). Teasing apart these various contributions to sustained inflammation in the preterm neonate and their individual and combined influence on neurodevelopmental outcomes is key to identifying the potential modifiers of preterm birth and its consequences (Malaeb and Dammann, 2009; Yanni et al., 2017; Volpe, 2019; Humberg et al., 2020). This topic is further introduced in Chapter 2.7.3 and examined empirically in Chapter 7 (using an epigenetic signature of inflammation to examine brain dysmaturation in preterm birth) of this thesis.
1.4.2 Dynamics and changes of the immune system with age

At the other end of the lifecourse, immune dysregulation in both innate and adaptive systems can occur with advancing age: the innate immune system can become chronically activated (Salminen et al., 2008), and the adaptive immune response can become less responsive (Chambers et al., 2021), with an increase in circulating inflammatory mediators considered to be both a result and driver of this effect (Kasler and Verdin, 2021). In some cases, this pattern of reduced responsiveness and increased inflammation is directly linked to dire consequences, as highlighted by the vulnerability of older adults to COVID-19 (Bartleson et al., 2021), and the reduced effectiveness of immunotherapy treatment in cancer patients of advanced age (Hiam-Galvez, Allen and Spitzer, 2021).

One key driver of the rise of inflammatory mediators in the systemic circulation with advancing age is the chemical and metabolic stimuli released from damaged and dying cells – the DAMPs discussed earlier (the endogenous signals released from senescent cells that immune cells pick up on) which are part of the ‘senescence-associated secretory phenotype’ (SASP) that cells adopt. These phenomena prime the local environment for continual, low-level inflammation until such cellular ‘debris’ is cleared, as immune cells mount an inflammatory response upon detecting these compounds. Studies illustrate that presence of this cellular detritus interferes with the homeostatic mechanisms of clearance and tissue rejuvenation (Muñoz-Espín and Serrano, 2014) and many now attribute products of cellular senescence (age-related death of cells) and apoptosis (programmed cell death) to be key contributors to inflammaging (Franceschi et al., 2000; Gruver, Hudson and Sempowski, 2007; Campisi, 2013; Salvioli et al., 2013; Bauer and De la Fuente, 2014; Schafer et al., 2020; Chambers et al., 2021).

Alongside the death of immune cells (‘immuno-senescence’) there are general patterns of age observed in immune cell populations. This includes dysregulated ratios of immature / mature immune cells, reduced function of lymphocytes and reduced diversity of B cells and T cells (Montecino-
Rodriguez, Berent-Maoz and Dorshkind, 2013). Combined, these contribute to reduced responsiveness of the immune system and its ability to resolve inflammation in later life. Figure 1.2 illustrates these changes, demonstrating that both early-life and old-age and are both periods of vulnerability for chronic inflammation.

**Figure 1.2 Susceptibility to chronic inflammation in early life and old age**

Schematic illustrating shared mechanisms between inflammation in early life and the elderly. To the left and right of the circle are factors that contribute to dysregulated inflammation in (Left) preterm infants and (Right) elderly individuals. How these factors relate to chronic inflammation is illustrated by the shared mechanisms in the centred-circle. Opposite sides of the lifespan represent windows of vulnerability to chronic inflammation: in the first year of life, when the immune system is still developing, and in old-age, where it enters functional decline.

These age-related changes that contribute to chronic inflammation have also been linked to epigenetic changes in immune-cell signalling, immune-cell effectiveness and in the regulation of programmed cell death, which will be discussed in greater depth in Chapter 3.6 of this thesis.

---

3 Figure drawn as summary of the literature discussed in sections 1.4.1-2. Abbreviations: bronchopulmonary dysplasia (BPD); necrotising enterocolitis (NEC); retinopathy of prematurity (ROP); interleukins (IL6, IL10, IL37); transforming growth factor beta (TGF-β); tumour-necrosis factor alpha (TNFα); reactive oxygen species (ROS); regulatory T-cells (T-regs); C-X-C motif chemokine ligand 10 (CXCL10); senescence-associated secretory phenotype (SASP); Damage-associated-molecular-patterns (DAMPs)
1.5 Measuring chronic inflammation: approaches and biomarkers

To test for inflammation, clinicians have a standard battery of tests available at their disposal: those that measure levels of certain inflammatory markers, white blood cell counts, and coagulation properties of blood. These are what are routinely used to signpost risk of infection, autoimmune conditions, and certain malignancies, with various cut-offs and thresholds used as indicators of potential ill-health. The oldest of these inflammatory tests are those that measure inflammatory-related processes that happen in the blood, such as erythrocyte sedimentation rate (ESR) – the distance (in mm) that red blood cells settle in anticoagulated whole blood within an hour, where a faster ESR indicates inflammation – or plasma viscosity (the force needed to send plasma down a thin tube in a given time). As time has gone on, this pool of inflammatory tests has expanded considerably, to include multiple potential biomarkers with varying degrees of applicability to clinical and research settings.

There are multiple individual components that govern and regulate chronic inflammation, arising from different systems which can act in either pro-inflammatory or anti-inflammatory capacities. These components can be classed as cellular (such as neutrophils, T-cells, B-cells, macrophages, mast-cells and monocytes); humoral (such as cytokines, chemokines, histamines, adipokines or the eicosanoids – prostaglandins, leukotrienes and thromboxanes); vascular (intracellular adhesion molecules, vasodilators); viral (interferons), or molecular, such as the reactive and nitrogen oxide species (ROS and NOS). There are also the proteins that regulate different levels of these mediators or go up and down in reaction to them, such as the acute-phase proteins discussed next.

Because of this, literature pertaining to chronic inflammation has used various measurement approaches: from examining immune cell populations, assessing individual inflammatory protein levels, to measuring combinations of inflammatory and immune markers that are associated with disease risk.
1.5.1 Acute phase proteins

Acute phase proteins are phasic and adaptively regulated in response to perturbations in the systemic circulation, with roles of detecting pathogens and cellular damage and initiating a molecular alarm. They are some of the fastest acting components of the inflammatory response, and one of the most frequently measured factors when assessing inflammation in humans.

Acute phase proteins are categorised by whether they go up (positive acute phase proteins) or down (negative acute phase proteins) by at least 25% during inflammation (Gabay and Kushner, 1999), see Table 1.1.

<table>
<thead>
<tr>
<th>Direction of effect</th>
<th>Category</th>
<th>Acute phase protein</th>
</tr>
</thead>
<tbody>
<tr>
<td>Positive</td>
<td>Complement system</td>
<td>C3, C4, C9, C4b-binding protein, C-Reactive Protein</td>
</tr>
<tr>
<td></td>
<td>Coagulation and fibrinolytic system</td>
<td>Fibrinogen, Plasminogen, Tissue plasminogen activator, Protein S, Plasminogen-activator inhibitor 1</td>
</tr>
<tr>
<td></td>
<td>Anti-proteases-Protease inhibitors</td>
<td>Anti-chymotrypsin, Pancreatic secretory trypsin inhibitor</td>
</tr>
<tr>
<td></td>
<td>Transport proteins</td>
<td>Ceruloplasmin, Haptoglobin, Hemopexin</td>
</tr>
<tr>
<td></td>
<td>Others</td>
<td>Serum amyloid A, a1-acid glycoprotein</td>
</tr>
<tr>
<td>Negative</td>
<td>Others</td>
<td>Albumin, Transferrin, Alpha-fetoprotein, Thyroxin-binding globulin, Insulin-like growth factor I, Factor XII</td>
</tr>
</tbody>
</table>

Table 1.1 Positive and negative acute phase proteins

The degree to which these acute phase proteins go up or down varies greatly – by two-fold in the case of ceruloplasmin and over 1000-fold in the case of CRP or serum amyloid A (SAA). The magnitude of increase is what
determines the sensitivity of that marker. Of all the acute phase proteins, CRP and SAA emerge as the most sensitive, demonstrating the greatest response speed, magnitude and dynamic range in blood plasma following an initial inflammatory insult, as illustrated in Figure 1.3, an adaptation of the work by (Gitlin et al., 1987; Gabay and Kushner, 1999).

![Figure 1.3 Changes in plasma concentrations of acute phase proteins](image)

Adapted from seminal review by Gabay & Kushner (1999).

When sustained elevation in acute phase proteins is observed there is accompanying undue havoc on cellular systems and tissues. For example, in rheumatoid arthritis, 5% of patients can go on to develop AA amyloidosis, a condition characterised by chronic elevations of the acute phase protein serum amyloid A (SAA). While SAA is initially useful in isolating and neutralising pathogens, chronic elevation of SAA causes an onslaught of nonspecific cellular activation and tissue damage (in the case of AA amyloidosis, ultimately resulting in kidney failure). This type of dysregulation:

---

of acute phase proteins can occur at lower levels with general ageing, as outlined in section 1.4.2.

Of all the acute phase proteins, CRP’s sensitivity and responsiveness heralds it the inflammatory biomarker of choice for most studies investigating inflammation. It has unique kinetics, biology and widespread associations with health, disease, and brain-ageing, which I will cover next.

1.5.2 C-Reactive protein

The discovery of C-Reactive Protein (CRP) came from two clinicians who were treating patients with pneumonia in 1930 (Tillett and Francis, 1930). In the blood samples of these patients, the two scientists found that serum precipitated with an extract they termed ‘Fraction C’, which was a C-polysaccharide (PnC) isolated from the cell wall of the pneumococcus bacterium. In patients who had died from this infection, Fraction C was high; in recovered patients, this compound was no longer present. 11 years later, in 1941, this compound was characterised as a protein, and was identified as an elevated marker of inflammation across a range of illnesses and injuries (Abernethy and Avery, 1941).

CRP is now classed as one of the most sensitive and dynamic markers of systemic inflammation and is widely used clinically because of these properties. In hospitals, clinicians use CRP to rapidly identify signs of illness or infection, and are able to monitor trajectories of illness, recovery, and responses to treatments.
1.5.2.1 CRP Biology: structure, function, and kinetics

CRP is a ring-shaped protein under the pentraxin family and is a key player in the innate immune system (Agrawal et al., 2009). CRP is predominately made in the liver by liver cells (hepatocytes), though there are secondary synthesisers of CRP, such as smooth muscle cells (Calabro et al., 2005), lymphocytes and macrophages (Haider et al., 2006), endothelial cells (Pasceri, Willerson and Yeh, 2000; Gould and Weiser, 2001; Alexandrov, Kruck and Lukiw, 2015) and adipocytes (Calabro et al., 2005) – which have all been shown to produce small amounts of CRP (refer to Figure 1.4). As the primary synthesisers, hepatocytes release CRP when they detect elevations in pro-inflammatory cytokines. While interleukin-6 (IL6) is considered the primary upstream inducer of CRP expression (Pepys and Hirschfield, 2003), hepatocyte-transcription of CRP is also regulated by other pro-inflammatory cytokines such as interleukins 17 and 1-beta (L-17, IL-1β) and tumour necrosis factor alpha (TNF-α) (Sproston and Ashworth, 2018). The CRP produced in the liver is then secreted into the blood stream.
Figure 1.4 C-Reactive Protein as a molecular alarm system

Schematic overviewing CRP transduction, synthesis, and molecular targets. CRP is induced by upstream inflammatory mediators (IL6, TNF-α, IL-1β) and is released by hepatocytes but also other cells in smaller quantities (adipocytes, endothelial cells, immune cells and smooth muscle cells). Following synthesis, CRP is released into the circulation where its primary role is to act as molecular alarm by binding to and opsonising dying cells and their remnants, which facilitates phagocytosis via immune cells. CRP also activates the complement cascade by binding to C1q. CRP exists as two isoforms, mCRP and pCRP; the biological properties of mCRP partly overlap the pCRP effects but mCRP is generally ascribed a more active and proinflammatory profile with greater ability to activate cells. Abbreviations: interleukins 6 and 1-beta (IL6, IL1β); tumour-necrosis factor alpha (TNF-α); monomeric and pentameric C-Reactive Protein (mCRP, pCRP); complement component 1q (C1q)
As CRP circulates in the blood, it binds to microbes and damaged cells, and in doing so, marks them as ready to be phagocytosed by white blood cells. As well as designating which cells need to be destroyed (a process called opsonisation), CRP activates the complement cascade by binding to C1q to set off a molecular alarm (Kaplan and Volanakis, 1974). The complement cascade is one of the first lines of defence by the innate immune system involved in the recognition and clearance of cellular debris and pathogenic particles. It is a cascade made up of 35 plasma proteins (Sproston and Ashworth, 2018), and CRP’s role is in amplifying and maintaining the inflammatory response until the threat of infection or injury is resolved (McFadyen et al., 2018). Overall, rises in CRP levels trigger a host of cellular, molecular and signalling effects which go on to affect tissues and systems (Luan and Yao, 2018); Figure 1.4.

Originally CRP was considered as a single entity, and it was only in 1993 that CRP was discovered to have two isoforms with distinct biological and inflammatory properties: pentametric CRP (pCRP) and monomeric CRP (mCRP) (Polevshikov, Nazarov and Berestovaia, 1993). Pentametric CRP is what freely circulates surveilling the systemic circulation. Upon binding to cells at sites of infection, pCRP can break apart into its 5 subunits of mCRP, which have a greater affinity to interact with a host of other cells – promoting platelets, microglia, leukocytes, and endothelial cells into pro-inflammatory phenotypes. The molecular biology of this process is beyond the scope of this thesis, but has been laid out in reviews (McFadyen et al., 2018; Sproston and Ashworth, 2018); it is worth noting that most literature does not refer to the two isoforms of CRP, and for most purposes, mCRP and pCRP can be considered bioequivalent, with mCRP considered the more ‘active’ isoform in terms of which cells it can affect (as illustrated in Figure 1.4).

Studies typically measure blood CRP with immunoassays (covered in section 1.5.2.2). Even the more modern of these, high sensitivity assays, only measure pCRP; there are very few available assays specifically designed to measure mCRP. At the time of writing, only 5 studies have published mCRP measurements, and these are not widely commercially available (Melnikov et al., 2022).
These two main functions of CRP (its localised surveillance and amplifier of action) designate CRP as an alarm to the rest of the immune system, which can kickstart a positive feedback loop of elevated inflammation. This process is what makes CRP an ‘acute-phase reactive’ protein; it is a swift, early reaction of the body to infection or injury, an attempt to restore balance and remove harm.

1.5.2.2 CRP measurement and relevance – what constitutes a ‘normal’ level?

CRP is typically measured by immunoassays. Traditional enzyme immunoassays are cheap, accurate and fast for detecting levels between 3-8 mg/L but limited in their sensitivity and precision at high and low CRP concentrations. To detect lower levels of CRP (< 3 mg/L), several high-sensitivity assays (hsCRP) have been developed (Roberts et al., 2001; Dominici, Luraschi and Franzini, 2004). These include immunoturbidimetric assays (a method that measures the absorbance of light from a sample) and immunonephelometric assays (a method that quantifies the reflection rather than absorbed light), and tend to report CRP concentrations in mg/dL over mg/L.

Over the years there has been debate over what constitutes a healthy baseline CRP level. The most recent evidence base suggests that healthy individuals typically have plasma levels of CRP of under 3mg/L. However, classifications of ‘raised’ levels are frequently debated, and normal levels are usually referenced as being between 1-8mg/L, with most research setting 10mg/L as a cut off for ‘high’ inflammation (see Table 1.2). The concentration of CRP is considered directly related to the severity of infection/insult, with 90% of what is considered severe inflammation (CRP level >500mg/L) arising from bacterial infections or acute organ failure (Landry et al., 2017; Nehring et al., 2022).
### Table 1.2 Interpretation of plasma CRP levels in humans

<table>
<thead>
<tr>
<th>CRP levels</th>
<th>Health interpretation</th>
</tr>
</thead>
<tbody>
<tr>
<td>&lt; 3 mg/L</td>
<td>Normal (seen in most healthy adults)</td>
</tr>
<tr>
<td>3 – 10 mg/L</td>
<td>Normal-to minor elevation (observed in pregnancy, depression, diabetes, smoking, gingivitis, obesity)</td>
</tr>
<tr>
<td>10 – 100 mg/L</td>
<td>Moderate to high elevation (considered systemic inflammation; common in patients with autoimmune diseases, as well as chronic illness)</td>
</tr>
<tr>
<td>100 – 500 mg/L</td>
<td>High to excessive inflammation (seen in acute bacterial and viral infection and major tissue trauma)</td>
</tr>
<tr>
<td>&gt; 500 mg/L</td>
<td>Severe inflammation (severe infection, liver-failure)</td>
</tr>
</tbody>
</table>

Though in healthy individuals CRP levels can rest below 3mg/L, in the event of acute inflammatory response this can increase up to 10,000-fold within just 48 hours and decrease just as swiftly when the source of inflammation is resolved, with an estimated half-life in plasma of ~ 20 hours (Pepys and Hirschfield, 2003; Sproston and Ashworth, 2018). CRP’s ability to rise from as little as from 5µg/L to 500mg/L over a short-time course is what enables clinicians to monitor the severity and course of inflammatory responses, and to ascertain the effectiveness of a given treatment (such as antibiotics).

Outside of clinical use (e.g. in hospital settings monitoring an acute case of inflammation or infection, such as sepsis – or in primary care settings, screening for inflammatory-related pathophysiology that could substantiate symptomatology), CRP levels are often used in epidemiological research for risk stratification of various morbidities. As an example, a recent study on 17,464 patients used hs-CRP levels to stratify risk of cardiovascular disease and associated mortality, finding a linear increase in risk of cardiac events and death between hsCRP levels of 1 – 5mg/L, which levelled out to a high and sustained risk of both cardiac events and death at concentrations > 5mg/L (Carrero et al., 2019). Raised CRP levels are also associated with

---

6 Adapted to display mg/L over mg/dL readings from (Eklund, 2009, p. 5)
increased risk of mortality generally, with a recent study demonstrating that elevated CRP levels (classified here as levels > 6.8mg/L) are also a strong predictor of mortality in the short term (Watson et al., 2020). In Watson et al. (2020)'s study, elevated CRP was related to one-year mortality in primary care settings (GP surgeries), with a better predictive accuracy than several common frailty indices. Equally, a study comprising 136,961 patients demonstrated that CRP has a higher C-statistic for mortality than for cancer, infections or autoimmune diseases (Watson et al., 2019). These studies demonstrate that inflammatory markers have both diagnostic and prognostic utility for a broad range of pathologies rather than just the typical inflammatory-related conditions (where it is routinely screened in) but also highlight that there is variation in set 'thresholds' indicating elevated or sustained inflammation.

An additional caveat to when reference CRP levels can be used as reliable inferences of inflammatory responses is in cases of preterm birth. As explained in section 1.5.2.1, CRP is primarily produced in the liver, which is still developing when infants are born before 37 weeks gestational age (this is also why some infants are prone to jaundice in the first few weeks of life owing to immature bilirubin production). In preterm infants, serum CRP responses have been shown to be dysregulated compared to term infants, with inconsistent responses in reaction to infection. Because of this, the use of CRP in profiling inflammation in preterm infants has been called into question (Chiesa et al., 2011; Macallister et al., 2019). This caveat informs the rationale for Chapter 7 of this thesis which examines an alternative biomarker of inflammation in preterm cohort.

---

7 This study collated data obtained from GP records, so a caveat to address is that the reason for why inflammatory marker tests were obtained in the first place is not available. This may skew the findings of this study, as CRP is part of a standard blood panel to investigate illness, i.e., there is likely an implicit selection bias for sicker individuals in these results. Regardless, this study is impressive in its sample size with inflammatory biomarker and death-linkage data available for over 100,000 individuals; of these, patients with a raised CRP (n = 47,797) had an overall 1-year mortality of 6.89%, compared with 1.41% in those with normal CRP levels (p<0.001). This association was seen across all age groups except the under 30s (the authors binned age ranges from < 30 to >80 into 7 distinct categories), and was stark in the oldest age category (80 years +), where raised CRP was associated with a 1-year mortality of 21.8% vs 8.6% with normal CRP levels (i.e. < 6.8 mg/L).
1.5.2.3 CRP heritability and intra-individual variability

There appears to be a genetic component to variation in blood CRP levels. Family studies are a common method used to estimate genetic contributions to blood protein levels. An example of this type of study in relation to CRP was performed by Pankow et al. (2001), where blood CRP levels were examined within related individuals (parents, siblings and children), with the hypothesis that if two family members have similar blood CRP levels, it is likely that there is a genetic component to their variation. With this method, the heritability of variation in blood CRP levels was initially estimated to be as much as 40% (Pankow et al., 2001). Twin studies also pointed towards a degree of heritability within baseline CRP levels that are independent of factors such as age, BMI and sex (Pepys and Hirschfield, 2003). Within the last decade of research, twin and family studies point towards a moderate variability in baseline CRP, with estimates ranging from 10-45% (Vickers et al., 2002; Sas et al., 2017).

The cross-sectional nature of these studies, which prevalingly only record one sample of CRP instead of repeated samples at different time points, may explain the large range of this estimation. Very few studies – genetic or otherwise – have adopted a longitudinal approach to profiling CRP, or take multiple measurements of CRP to obtain an average reading, which is a significant weakness of the literature given the reported intra-individual variability in CRP noted over time (Bower et al., 2012; DeGoma et al., 2012; Wu et al., 2012; Bogaty et al., 2013; Nash et al., 2013).

Genome-Wide Association studies (GWAS), with much larger sample sizes, have provided more conservative estimates of genetic contributions to blood CRP, though confirm that individual polymorphisms account for non-trivial variation (7-16%) in population differences of circulating CRP. The largest of these to date (Said et al., 2022) was conducted in the UK Biobank (UKB) cohort and Cohorts for Heart and Aging Research in Genomic Epidemiology
(CHARGE) consortia (N = 575,531), where 266 loci were found to associate with CRP, 57 of which were replications of associations found in previous GWASs (Dehghan et al., 2011; Ligthart et al., 2018). This latest study demonstrated that multiple causal loci (clustered in immune and liver biology systems) appear to contribute to chronic inflammation and variation in circulating CRP levels. This recent study also derived a polygenic risk score (PRS; an approach in genetic research explained in relation to recent application of this methodology to epigenetic data in Chapter 3.8) of CRP, finding $\text{PRS}_{\text{CRP}}$ to be associated with 27 disease outcome measures including chronic obstructive pulmonary disease (COPD), schizophrenia, atherosclerosis and rheumatoid arthritis (Said et al., 2022). The causality of these associations was assessed using two sample Mendelian randomisation (MR)\(^8\), with risk of COPD and schizophrenia linked to genetically raised CRP, a topic further discussed in Chapter 9.7.9.

While these findings corroborate findings of a causal role of CRP in contribution to disease states reported in other population samples (Dahl et al., 2011; Prins et al., 2016; Si et al., 2021), including those related to brain-health phenotypes (Larsson et al., 2017; Bottigliengo et al., 2022), it should be stressed that there remains mixed evidence for a direct causal relationship between blood CRP and disease phenotypes. This was recently highlighted by a review which examined 196 MR studies pertaining to CRP and its associations (Markozannes et al., 2021). The authors of this study stress that CRP alone, as a single inflammatory biomarker, is unlikely to fully capture the specific inflammatory mediating pathways leading to the development and progression of disease states, making it less of a useful target for drug discovery than once anticipated. Despite the mixed causal evidence between CRP and incident disease found so far, there remains overwhelming observational evidence linking raised inflammation levels to the development and severity of various disease states (discussed next),

\(^8\) A more thorough examination of inferring causality in relationships between chronic inflammation and brain health outcomes, such as via use of MR, is provided in the discussion chapter of this thesis (9.7.9).
lending support to the hypothesis that chronic low-grade inflammation contributes to the development of various deviations from health.

1.5.2.4 CRP epidemiology: associations with age, mortality and morbidity

Higher levels of CRP have consistently been observed in populations of participants with certain diseases and have been examined as a potential risk factor for numerous conditions. As several seminal reviews have covered this topic in great depth (Pepys and Hirschfield, 2003; Luan and Yao, 2018; Sproston and Ashworth, 2018; Furman et al., 2019) and the literature on this topic is extensive – with over 1,500 phenotypes linked to CRP, see Figure 1.5 – I will focus on discussing the most replicated CRP-disease relationships and robustly supported risk factors for chronic inflammatory states.

Any disease with the ‘-itis’ suffix is characterised by inflammation – e.g. endocarditis, arthritis, appendicitis, colitis, etc. Many of these conditions classify as conditions of acute inflammation, presenting with markedly raised CRP levels (Landry et al., 2017). However, many other conditions are accompanied by a lower-level ‘chronic’ profile of inflammation of <10mg/L CRP, such disorders related to metabolism – from type 2 diabetes and obesity (Festa et al., 2002; Ridker et al., 2003; Greenfield et al., 2004; Calabro et al., 2005; Bastard et al., 2006; Timpson et al., 2011; Rodríguez-Hernández et al., 2013; Svensson et al., 2014), cardiovascular and pulmonary diseases (Koenig et al., 1999; Raitakari et al., 2005; Carrero et al., 2019), cancer (Ben-Baruch, 2006; Berasain et al., 2009; Zitvogel, Pietrocola and Kroemer, 2017; Das, Karthik and Taneja, 2021), autoimmune conditions (Ellinghaus et al., 2016; Pope and Choy, 2021), to those pertaining to the brain, such as psychiatric conditions, stroke and neurodegenerative diseases (Emsley et al., 2003; Wardlaw et al., 2003; Brown et al., 2019).

---

9 *itis* translates as ‘inflammation of’, so colitis = inflammation of the colon
Figure 1.5 CRP Epidemiology: associations of CRP with 1,673 diseases / phenotypes assimilated from literature text-mining

Associations of CRP to disease-outcome measures, grouped according to broad categories of interest. Each circle represents a disease-phenotype associated with CRP in the literature, with the size of the circles relating to the number of studies supporting this association and the saturation indicating the strength of these relationships – though it should be noted that this metric is a heuristic score based on the availability of published associations to date, and as such should not be interpreted as a confidence score on the target - disease association. An interactive map of CRP-epidemiological associations from which figure was based on can be accessed at: https://platform.opentargets.org/target/ENSG00000132693/associations/bubbles using the Open Targets software platform (Ochoa et al., 2021).

---

Figure Source: Open Targets (Ochoa et al. 2021) release 22.11. Permission requested 11/01/23. This figure has been adapted from the generated interactive version for clarity, but the original source can be accessed at the URL above. https://platformdocs.opentargets.org/associations#association-scores should be reviewed for details on how the score that determines the saturation of bubbles (strength of CRP-disease/phenotype associations) is derived. This generates a score between 0 and 1 where the higher the score = the stronger the association. Because this ‘score’ is generated based on the number of publications available, there is bias in the strength of associations whereby under-studied diseases are unlikely to produce high scores due to the lack of available evidence.
Many of these diseases share common pathophysiology and risk factors, as is the case in cardiovascular health outcomes and brain health outcomes (Mun and Hinman, 2022). CRP’s association with cardiovascular outcomes, such as atherosclerosis, has historically been a keen focus of CRP epidemiology, with several lines of evidence now pointing towards a mechanistic basis for this association with CRP’s direct action on blood vessel walls (Ridker et al., 2003; Ridker, 2016) and its indirect contribution via the recruitment of leukocytes in early atheroma formation (Libby and Ridker, 1999; Libby et al., 2014; Shabani et al., 2021); Figure 1.4. Of clinical interest here was the (at the time – largely unexpected) success of the CANTOS trial\(^\text{11}\), which brought home the cardio-protective effects of anti-inflammatory treatment in CVD populations (Ridker et al., 2017).

The vascular-inflammation link is particularly relevant when considering brain health outcomes, given the vasculature supply of brain tissue, with many studies investigating inflammation’s association with comorbid presence of atherosclerosis and stroke and small vessel disease (Emsley et al., 2003; Schmidt et al., 2006; Gottesman et al., 2014; Raina et al., 2017) and the role CRP in contributing to blood brain barrier permeability (a topic developed on in section 1.6.2).

When it comes to demographic and population differences in circulating CRP levels, socioeconomic status (SES), particularly in early-life, emerges as a more reliable predictor than ethnicity or country of residence (Pollitt et al., 2007; Liu et al., 2017). However, population differences have been recorded:

\(^{11}\) The Canakinumab Anti-inflammatory Thrombosis Outcome Study (CANTOS) trial was the dénouement of much of Ridker and Libby’s focus on CRP’s role in atherosclerosis (which went against the prevailing view held by most clinicians of CRP’s relevance to CVD at the time). CANTOS was a randomised double-blind trial that aimed to evaluate whether an anti-inflammatory treatment that specifically impacted CRP levels could prevent recurrent vascular events. The drug, canakinumab (a monoclonal antibody that inhibits the production of interleukin-1β, an upstream activator of CRP), was found to significantly lower the recurrence of major cardiac events following previous heart attacks and has also had success in lowering the risk of incident disease outcomes like arthritis, gout, type 2 diabetes (T2D) and various cancers.
a systematic review of 32 studies found that, compared to European populations, higher levels of CRP were consistently observed in individuals with South Asian, African and Latin American heritage even after controlling for potential mediators and confounders of this association (Nazmi and Victora, 2007). These differences may still be a product of SES, however. Across these 32 studies, the 9 that adjusted for SES consistently found higher CRP levels in those of lower SES. This finding was corroborated by a recent meta-analysis of 42 studies (combined N = 111,156) where individuals with lower SES displayed higher levels of systemic inflammation as measured by circulating CRP and IL6 (Muscatell et al., 2020). This association between SES and inflammation may be underpinned by poorer lifestyle and health differences in those of lower income backgrounds, as in Muscatell’s overview of 42 studies, effect sizes were found to be smaller in the studies that controlled for factors such as BMI or smoking. Higher levels of CRP are also linked to higher prevalence of chronic infections, such as HIV, which may underscore the raised levels of CRP seen in African population cohorts (Escadafal et al., 2020), and parasitic risk encountered by the Amazonian populations, such as the indigenous Tsimane people (Blackwell et al., 2016). The lowest levels of CRP are consistently observed in Japanese populations, with indicators that diet and lifestyle factors are significant determinants of low circulating CRP levels (Saito et al., 2007; Coe et al., 2020).

With regards to differences in sex, though early epidemiological evidence pointed towards CRP levels being higher in females (Garcia-Moll et al., 2000; Wener, Daum and McQuillan, 2000; Rifai and Ridker, 2003), this finding has not been consistently upheld in more recent research across different ethnic populations, with some studies reporting the opposite effect of sex (Tang et al., 2018). What’s more, adjustment for oral contraceptive use or hormone replacement therapy appears to attenuate these effects of sex (Rifai and Ridker, 2003; Raitakari et al., 2005), and research controlling for other confounding variables suggests higher body-fat percentage is likely to be the primary reason for this observed difference (Bochud et al., 2009). Several
studies point out that most of the observed ethnic disparities in CRP levels also appear to be mediated by BMI (Kelley-Hedgepeth et al., 2008). BMI, obesity and disorders of metabolism are in general one of the more studied areas of inflammation research, with higher fat levels consistently linked to elevations in baseline CRP (Greenfield et al., 2004; Bastard et al., 2006; Timpson et al., 2011; Rodríguez-Hernández et al., 2013). Weight loss interventions have resulted in marked reduction of CRP levels, with an average reduction of 3.1 mg/dL per kg of weight (Selvin, Paynter and Erlinger, 2007). One of the suspected reasons for this association is due to fat cells (adipocytes) secreting inflammatory mediator IL6, the upstream inducer of CRP production. Between 15-35% of IL6 in the blood is estimated to originate from adipocytes (Mohamed-Ali et al., 1997), and subcutaneous fat tissue (present in higher amounts in obese individuals) accounts for a higher proportion of circulating IL6 (Jonas et al., 2015), with multiple studies demonstrating that higher fat mass associates with higher proportions of IL6 and CRP (Bochud et al., 2009; Sproston and Ashworth, 2018).

In many of these age-related diseases where elevated inflammation is noted, the initiating trigger is not well defined, and continual presence of this trigger (for example, an active infection or open wound) does not have to be noted for chronic inflammation to persist. Because the presence of sustained inflammation and inflammatory risk factors is shared across many diseases, perpetuating cycles interlinking inflammatory-risk factors, chronic inflammation and disease-related outcomes have been proposed (Furman et al., 2019) (see Figure 1.6). Positive feedback loops of self-promoting inflammation are seen in atherosclerosis, cancer, and psychiatric and neurodegenerative phenotypes – where elements of lifestyle can lead to the development of illness, and subsequent attributes of these illnesses can prolong inflammation and increase the risk of developing subsequent conditions – which may explain the comorbid presence of certain conditions in specific populations. These types of reciprocal relationships may be responsible for the chronic nature of these inflammatory conditions, distinguishing them from diseases characterised by acute-inflammatory
episodes which may resolve with treatment or intervention, like an appendicitis or bout of gastroenteritis.

Figure 1.6: Drivers and consequences of chronic inflammation

Finally, lifestyle factors such as smoking, alcohol consumption, physical inactivity, poor diet, poor sleep, pollution and stress are all associated with higher circulating CRP levels (Ohsawa et al., 2005; Furman et al., 2019; Elisia et al., 2020); Figure 1.6. Given these reported interrelationships, a key weakness of the literature to date that examines the association of CRP with brain-related outcome metrics is the lack of consistent accounting for these factors in analyses. There is also a lack of defined population characteristics that acknowledge how different health conditions can influence inflammation-cognition associations (and attempt to control for this: e.g. with robust exclusion criteria or sub-population sensitivity analyses). Though there are some evidence-based guidelines for what to control for in these research protocols (O’Connor et al., 2009). Further discussion of the link between chronic inflammation and brain health is expanded on in the coming sections (1.6, 2.4.3 and 2.7.3).
1.5.3 **Why is CRP the most popular biomarker of inflammation?**

So why is CRP such a dominant marker of inflammation in the literature to date? It’s partly historical, but it’s also because of its non-specific nature, its ease of measurement, and its general predictability. It’s a measure of inflammation that can be tested easily, at relatively low cost, with good reliability. Unlike other inflammatory biomarkers, CRP is uniquely reactive to other components of the inflammatory cascade, responding quickly and resolving quickly. CRP levels do not appear to vary diurnally or in response to food over short time spans (though chronic malnutrition, shift-work, and lifestyle factors appear to influence baseline levels) (Ridker, 2016; Pavanello et al., 2017). Other than specific drugs that aim to resolve acute inflammatory episodes (such as antibiotics to treat infection, or NSAIDs), CRP is not influenced by most drugs directly. There are also not many health conditions that chronically dysregulate\(^\text{12}\) CRP aside from direct damage to the liver, such as cirrhosis and other forms of liver-failure (Pepys and Hirschfield, 2003).

These features are what make CRP, as a measure of acute-inflammation, excellent – but are also what make it a less than ideal measure of chronic inflammation. The vacillating nature of CRP and variability within short time scales means that individuals may present with transiently elevated or lowered CRP levels upon sample collection in population research studies. Not all studies utilise the high sensitivity assays of CRP to detect levels < 3mg/L, despite evidence indicating that this sensitivity is required for more accurate risk stratification (Carrero et al., 2019) or screen for (and omit) cases of elevated inflammation suggestive of acute-inflammation >10 mg/L. More concerningly, a feature of most research to date examining CRP and its associations is that a one-off measure of CRP is used (rather than repeated sampling to obtain an average reading), typically at a single-time point, to

---

\(^{12}\) With the term ‘chronically dysregulate’ here, I refer to a fundamental alteration in CRP production/synthesis; elevated CRP levels, suggestive of chronic inflammation, can be seen in an array of conditions as discussed in [section 1.5.2.4](#), but most of these are not resulting from damage to the liver (where CRP is primarily produced).
index inflammation, despite the high variation in baseline CRP within individuals overtime (Bower et al., 2012; DeGoma et al., 2012; Wu et al., 2012; Bogaty et al., 2013; Nash et al., 2013). Finally, in preterm infant populations, CRP levels are considered a less than ideal capture of inflammatory processes because the hepatic enzymatic machinery for its generation is maturation-dependent over the 2nd and 3rd trimesters of pregnancy, with studies suggesting that reference levels of CRP suggestive of elevated inflammation should be dependent on gestational age (Chiesa et al., 2001; Matoba et al., 2009; Chiesa et al., 2011; Hofer and Resch, 2011; Hofer et al., 2012; Macallister et al., 2019; Borowski, Shchors and Bar-Meir, 2022). Because of these drawbacks, there is a precedent to consider other inflammatory biomarkers for characterising the full profile of chronic inflammation and its effects.

1.5.4 Other individual inflammatory biomarkers

The list of components that are considered to regulate inflammation is too exhaustive to discuss each mediator in turn, with well over 1,000 distinct inflammation-related genes identified by different genomic approaches (Heller et al., 1997; Loza et al., 2007).

Broadly classifying, other than the acute phase proteins, there are the cytokines, prostaglandins, chemokines, leukotrienes, proteases, adipokines, interferons, vaso-active compounds, transcription factors and growth factors. These mediators can be grouped according to different classification criteria, with overlapping functions and play a multitude of roles in both the innate and adaptive immune response. Some are secreted by immune cells in response to infection or injury; others are upstream or downstream components in various cellular signalling cascades (such as MAPK, PI3k-AKT, JAK-STAT, and complement cascade pathways); or as factors in dynamic extracellular matrix reorganisation, angiogenesis and apoptosis. Ultimately, these are all linked to inflammation and have been examined as inflammatory biomarkers in relation to health and disease outcomes. An example of these categories
and some (but by no means all!) of the inflammatory mediators within them are given in Figure 1.7, alongside an overview of different inflammatory processes that inflammatory mediators are involved in. Many mediators overlap in their targets, functions and regulation and have different roles in multiple aspects of inflammation depending on the type of initial stimulus for an inflammatory response and the area of the body this occurs in.

Figure 1.7 Classification of inflammatory biomarkers and presence of inflammatory mediators in different inflammatory processes

An overview of the range of mediators involved in inflammation. Top panel of figure outlines broad classification categories of different inflammatory mediators. Bottom panel of figure illustrates the range of inflammatory processes these mediators participate in. Broad inflammatory processes (circles) are accompanied by examples of inflammatory proteins and mediators noted by previous studies; note this is by no means an exhaustive list and is designed to demonstrate the array of different inflammatory biomarkers and overlapping roles they play in inflammation.

Original figure based on literature discussed in section 1.5.4.
With the SASP phenotype that senescent cells adopt, the types of inflammatory mediators released by cells is dependent on cell-type

Robustly classifying individual mediators’ role in chronic inflammation is difficult, as many of these mediators have overlapping classes and functions (for example, many cytokines are also chemokines) and are differentially expressed across tissues in the human body. Some of these proteins are very localised, and as such are good prognostic biomarkers for certain conditions or tumour types; others are widespread and can be awry across multiple disease states. While some inflammatory mediators are classically pro-inflammatory and others ‘pro-resolution’ another difficulty is that many inflammatory mediators have pleiotropic or even opposing functions, acting in either pro- or anti-inflammatory capacities depending on context or where they are expressed (Capucetti, Albano and Bonecchi, 2020).

Equally, there are many molecular mediators that are considered to be involved in the regulation of inflammation without having a direct pro- or anti-inflammatory action. For example, the semaphorins (such as SEMA3E, a protein examined in Chapter 8) are a family of proteins involved in the regulation of cell migration, adhesion, and proliferation, and can act as either pro-inflammatory or anti-inflammatory molecules depending on the context. In a pro-inflammatory capacity, semaphorins can act as chemoattractants, promoting the migration of immune cells to the site of inflammation; conversely, they can also act as inhibitors of inflammation, suppressing the production of pro-inflammatory cytokines and chemokines from innate immune cells. In their interaction with vascular and endothelial cells, semaphorins can act as regulators of angiogenesis, promoting the formation of new blood vessels in the area of inflammation (Suzuki, Kumanogoh and Kikutani, 2008; Takamatsu, Okuno and Kumanogoh, 2010; Kermarrec et al., 2019). Another example of inflammatory mediators lacking a clear pro or anti-inflammatory status are those in the matrix metalloproteinase (MMP) family,

14 Pro-inflammatory factors have been detailed: the cytokines IL6, IL1, IL8 and TNFα are some seminal examples. Some anti-inflammatory classical mediators include interleukin-10 (IL10) and transforming growth factor-β (TGF-β).
which consist of proteins with roles in extracellular matrix remodelling. These can have somewhat contradictory roles in the contribution to adverse outcomes: the protease MMP9, for example, has been shown to contribute to the pathogenesis of influenza-mediated intracerebral haemorrhage (Muhammad, Planz and Schwaninger, 2016) via its effects on blood-brain barrier permeability (discussed in greater depth in Chapter 1.6.2). But, conversely, has also been shown to have protective roles in the brain, enhancing neurovascular remodelling and associating with improvement in clinical presentation post stroke (Zhao et al., 2006; Abdelnaseer et al., 2017). These two examples reinforce the duality of effects inflammatory mediators can have, and the importance of context, localisation and duration of inflammatory responses in reaction to events. The discussion of these different proteins and their relevance to characterising chronic inflammation and brain health outcomes is expanded upon in Chapter 8 of this thesis, where 109 inflammatory-related markers are examined in relation to brain health outcomes. A final note is that there are also inflammasomes, which are distinct from the previous mediators outlined in that they are not molecules or individual proteins but multi-protein complexes that are formed within cells as part of the innate immune response. These sense PAMPs or DAMPs and trigger the cleavage of various cytokines into their more active forms, and are implicated across a large continuum of disease states (Guo, Callaway and Ting, 2015).

In clinical settings, it is accepted that no one individual measure of inflammation is perfect. High signal-to-noise ratios, limited sensitivity and specificity all render isolated measures of inflammation imperfect and make ‘baseline’ inflammation levels easy to mischaracterise – for both false positive and negative error rate. Equally, because so many inflammatory biomarkers have significant overlap in their origins, targets, and functions, assigning one biomarker as the ‘most exemplary’ or ‘most reflective’ of chronic inflammation is reductive and potentially misleading. Selecting ‘downstream’ mediators is one way to overcome this (attempting to select the most pertinent inflammatory mediator that reflects a chronic inflammatory
response), more general broad-spectrum associations of inflammation with less variable half-lives, or, as discussed next, creating ‘composite’ inflammatory scores that accrue information about inflammation across multiple data sources.

1.5.5 Composite and multi-omic biomarkers of inflammation

Though overwhelmingly in epidemiological research there is a dominance of singular measures of individual inflammatory mediators such as CRP, IL6 or TNF-α to index inflammation, the use of an aggregate or composite measure of inflammation is not uncommon in certain scientific disciplines. For example, in cancer research, the derivation of ‘immunoscores’, ‘Immune Cell Scores’, ‘Immune activation scores’, ‘immune signatures’, and ‘systemic inflammation indices’ is a common practice for characterising immune and inflammatory responses. These are composite metrics built by measuring the relative abundance of different immune cell types, gene expression or cytokine/chemokine levels from tumour biopsies or peripheral blood samples, and have been shown to be strong prognostic markers for various cancer types and stages (Pagès et al., 2018). There are also examples of population-based epidemiological studies that adopt this multivariate approach to examine the association of inflammation with age and morbidity (Hsu et al., 2009; Morrisette-Thomas et al., 2014), typically leveraging dimensionality reduction methods like principle components analysis (PCA) to uncover axes of variation in inflammation in relation to health outcomes.

More recently, other areas of research have turned to alternative methods (both technically and statistically) to get a more comprehensive biological capture of inflammation. Multi-omics is a term used to describe the integration of multiple sources of biological (or ‘omic’) data layers, such as genomics, transcriptomics, proteomics, metabolomics, and epigenomics. It is an emerging field of research that has the potential to provide a more comprehensive understanding of the molecular architecture of inflammation, its endogenous and exogenous drivers, its association with various health
outcomes, and chronic inflammation’s specific role in disease pathogenesis. By combining multiple data types, multi-omics holds the potential to offer a more holistic view of inflammation and inflammatory-related events than individual isolated markers of inflammation.

Multi-omic characterisation of both chronic inflammation and immune dysregulation have become increasingly explored in relation to age and age-related diseases (Alpert et al., 2019; Sayed et al., 2021; Al-Nesf et al., 2022; Walker et al., 2022). This step-change in characterisation is a result of the rise in high-throughput technology and advances in computational methods to assess such data (Chu et al., 2021), including platforms which are able to measure hundreds of immune and inflammatory analytes from human tissue samples. More studies are moving away from the historical reliance of single protein measurements to characterise inflammation, bolstered by the availability of such panels (such as Olink® Explore 384 Inflammation I and Olink® Explore 384 Inflammation II panel, which profile ~384 inflammatory proteins apiece, and SomaLogic® Inflammation and Immune Response panel, which can assess 938 inflammatory analytes). More details on these platforms and the strengths and limitations of their approaches can be found in reviews (Joshi and Mayr, 2018) and are briefly discussed in Chapter 5 of this thesis.

Integrating omics data with clinical variables has shown enhanced performance in terms of disease prediction, classification, and prognostic outcomes. In a study that investigated 14 different cancer types from biological samples from 3,382 patients, researchers evaluated the prognostic prediction of genomic, epigenomic, and transcriptomic profiles individually, iteratively, and in combination with clinical variables – finding epigenomic data to perform the best (Zhu et al., 2017). This is a field that goes hand in hand with the development of machine learning (ML) approaches to anticipate health outcomes owing the vast amount of high dimensional data of clinical and biological significance (MacEachern and Forkert, 2021; Feldner-Busztin et al., 2023).
For inflammatory diseases, the integration of proteomic and DNA methylation data has shown potential in enabling expedited diagnosis of, and treatment response to, ulcerative colitis (Mäki-Nevala et al., 2021; Janker et al., 2023). Another study examined how multiple peripheral inflammatory biomarkers associated with age-associated morbidity (Alpert et al., 2019), creating an overall composite measure of immune-related dysregulation (which included traditional inflammatory proteins levels as well as cellular immune markers, from T-cells, monocytes and B-cells) they termed ‘IMM-AGE’, which demonstrated that higher inflammatory burden was associated with increased risk of various diseases. Another utilised blood-based immune biomarkers in 1,001 individuals from the Stanford 1000 Immunones Project to generate a single metric for age-related chronic inflammatory burden ‘iAge’ (Sayed et al., 2021). They adopted a ML approach\(^{15}\) to capture the non-linear associations of various immune protein levels in the systemic circulation and derive a compact representation of age-related inflammation.

Overall, these composite metrics of inflammation may reduce the noise that accompanies single sampling approaches of individual protein levels that are highly phasic in the systemic circulation. The specific use of epigenetic composite metrics, which leverage epigenetic information pertinent to inflammatory-related proteins, is expanded upon in Chapter 3 of this thesis.

1.6 Neuroinflammation

From cross-sectional studies linking elevated inflammatory mediators to brain health outcomes (Osimo et al., 2019; Osimo et al., 2020), to investigations of comorbid inflammatory conditions alongside depression and Alzheimer’s disease (AD) (Kozora et al., 2001; Franceschi and Campisi, 2014; Postal et al., 2017; Schrepf et al., 2018; Pope and Choy, 2021; Craig et al., 2022), to

\(^{15}\) The ML approach used in the study by Sayed et al (2021) was a guided auto-encoder (GAE) a type of neural net that uses nonlinear equations to reduce noise, which they used to highlight relevant biological relationships
longitudinal profiling of inflammation alongside brain health phenotypes (Khandaker et al., 2014; Osimo et al., 2020) there is a clear case for peripheral inflammation having an impact on brain functioning, with increasing evidence towards there being a causal link between observed associations between peripheral inflammation and brain health (Larsson et al., 2017; Bottigliengo et al., 2022).

Only recently have these relationships been examined in greater depth. Partly, this is because the immune system and central nervous system (CNS) have long been studied in isolation from each other due to the compartmentalisation ‘cottage-culture’ of disciplines that shaped the academic landscape in the past. However, recent research has demonstrated that these two systems are closely interconnected, with the CNS regulating the peripheral immune response, and the peripheral immune response influencing the neuro-immune landscape. There are two main factors which distinguish peripheral inflammation from neuroinflammation (1) the relative proportion, and type, of resident immune cells present in the systemic circulation vs brain (2) the separation of the blood brain barrier (BBB) between them. I will outline these differences before discussing the crosstalk between the two systems.

1.6.1 Systemic inflammation vs neuroinflammation

As outlined in section 1.3-1.4, chronic inflammation can occur in the systemic circulation due to a dysregulation of innate and adaptive immunity. In the brain, there are similar innate and adaptive immune responses, but some key differences in the primary immune-cell types (illustrated below in Figure 1.8).
Figure 1.8 Differences in cell populations between the peripheral immune system and brain

Schematic demonstrating the differences in immune regulation in periphery vs brain. In the peripheral circulation, immune activity is conducted by interaction between B-cells, T-cells, lymphocytes, and granulocytes (eosinophils, basophils, and neutrophils). In the brain, inflammation is mostly sustained by the mast cells, astroglia and microglia, though other cell types such as oligodendrocytes, neurons and perivascular macrophages, also have roles in initiating and sustaining inflammation. There is a degree of overlap in cell-types present between the peripheral immune landscape and neuroimmune landscape, with some immune cells present in both the periphery and brain, such as the adaptive immune CD4+ and CD8+ cells, though these are present in smaller numbers in the CNS.

In the brain, inflammation is mostly sustained by the mast cells, perivascular macrophages, astrocytes and microglia. Both the initiation and prolongation of inflammation involves interaction between these cells, along with neurons, endothelial cells and other immune cells (Dong, Zhang and Qian, 2014; Skaper, Facci and Giusti, 2014; Hendriksen et al., 2017; Prinz and Priller, 2017).

Original figure based off above literature.
Microglia, in particular, have been heavily implicated in neuroinflammation, with their main function considered to be localised neuro-immune surveillance. In later life, microglia appear to become sensitive to pro-inflammatory stimuli (Cherry, Olschowka and O’Banion, 2014; Mosher and Wyss-Coray, 2014; Hoogland et al., 2015; Gefen et al., 2019) transitioning more readily into their more ‘activated’ form (M1 phenotype), with increased presentation of surface receptors for inflammatory mediators and secretion of cytokines, chemokines and other inflammatory compounds – a phenomenon known as microglial priming (see Figure 1.9). Increased activation of microglia, and the sustained release of pro-inflammatory mediators, has been outlined as a key mechanism for damage to neurons and supporting glia that underlies neurodegeneration (Block, Zecca and Hong, 2007), but also indirect weakening of the BBB, further exacerbating infiltration of inflammatory mediators from the periphery (Haruwaka et al., 2019).

Figure 1.9 Mechanisms of neuroinflammation: microglial priming

Schematic illustrating how microglia can undergo changes in response to raised inflammatory mediators which can perpetuate the systemic-inflammation -> neuroinflammation cycle. An initial rise in inflammatory mediators (from either direct brain insult or influx from the periphery) can ‘prime’ resident brain cells (such as the resting microglia) into pro-inflammatory phenotypes. Alongside upregulated presentation of receptors, downregulation of anti-inflammatory mediators, and reduced signalling (chemotaxis) to other immune cells, these activated microglia release further inflammatory mediators, stimulating further glia to adopt pro-inflammatory phenotypes (1). The release of these mediators can also damage the

---

17 Figure based off literature detailed in section 1.6.1
Alongside an age-related shift to a heightened pro-inflammatory state, activated microglia down-regulate their production of anti-inflammatory cytokines (see Figure 1.9), hampering the resolution of inflammation. Aged microglia have also been shown to be less responsive to encounters with the aggregation-prone proteins considered at the heart of neurodegenerative aetiology, such as amyloid-β (Aβ) and α-synuclein (hallmarks of AD and PD respectively) – displaying reduced phagocytic capability and decreased chemotaxis to recruit the adaptive immune cells (CD4+ and CD8+ cells) (Hickman, Allison and El Khoury, 2008; Labzin, Heneka and Latz, 2018).

These changes result in an amplified reactive inflammatory response that is less effective at resolving the initial inflammatory insult and ultimately causes exaggerated neuroinflammation and subsequent neuronal damage. Neuroinflammation (as characterised by the presence of highly reactive glial cell subtypes) can be seen across a range of neurological diseases, which are in and of themselves often defined as either ‘acute’ or ‘chronic’ in nature. Acute neurological diseases are usually defined as those that arise from a sudden and critical inflammatory response – such as strokes, traumatic brain injury or bacterial meningitis. These can later transition into chronic states of neuroinflammation, which is why acute neurological insults increase an individual’s risk of developing subsequent neurodegeneration (Skaper et al., 2018). Neuroinflammation is also suspected to account for degrees of neurodegeneration or other structural alterations that occur with age (i.e., in the absence of a clear acute injury or pathological condition), and that this may relate to differences in cognitive functioning (Gorelick, 2010; Cribbs et al., 2012; Lin et al., 2018; Furman et al., 2019).

1.6.2 **Blood brain barrier**

For a long time, the brain was considered uniquely ‘immune-privileged’. Early experiments in the late 1800s and early 1900s by Ehlrich, Goldman,
Bouffard, Franke and Lewandowsky involved the injection of peripheral dyes into the circulation. When these dyes failed to stain brain tissue, it was understood that there was a biological interface between the blood and brain barring the direct passage of substances from the periphery to the brain. This interface (illustrated below) is now known as the blood brain barrier (BBB) though this exact term was not used until the early 1920s (Saunders et al., 2014). Since then, the BBB has been extensively studied, particularly in relation to neuroinflammation and neurodegeneration (Banks, 2009; Banks and Erickson, 2010; Galea, 2021). Our current understanding of this blood-brain interface is that it is bidirectional, with a variety of mechanisms via which chronic inflammation in the periphery can influence the CNS and vice versa.

**Figure 1.10 Blood brain barrier**

*Schematic demonstrating how the brain, unlike other organs in the body, has a unique structural interface called the blood brain barrier (BBB) that stops the passage of certain substances from the systemic circulation into the brain. Pictured to the right are the capillary endothelial cells, astrocytic end-feet, and pericytes embedded in the capillary basement membrane – together, these cells constitute the neurovascular unit (NVU) of the BBB that allows for bidirectional communication between the periphery and brain. This system permits the selective and active transport of different nutrients, ions, organic anions, and macromolecules like glucose and amino acids that are essential to brain function, as well as the passage of some tiny molecules by passive diffusion in both directions. At the same time, the BBB prevents the diffusion of big or hydrophilic molecules, such as pathogens, immune cells, antibodies, and larger inflammatory mediators into the brain parenchyma.*
1.6.3 Mechanisms of immune-crosstalk between periphery and brain

The BBB is not impenetrable, and in healthy states, permits the passage of different molecules and cells through a mix of active and passive mechanisms regulated by different proteins and junctions. While some inflammatory mediators can diffuse across the BBB, larger compounds (such as CRP) are not usually able to cross over from the blood to the brain. However, in cases of both acute and chronic inflammation, the BBB has been observed to become more permeable or ‘leaky’, permitting passage of previously excluded substances as illustrated in Figure 1.9. Of interest, CRP has been shown to enter the brain via damaged BBB in animal models (Slevin et al., 2015) – and even in cases of apparent BBB integrity, studies have found that elevated plasma CRP levels are nonetheless strongly correlated with CSF CRP levels \( r = 0.855, p < 0.001 \) (Felger et al., 2020).

Similar to what is observed in microglia described in section 1.6.1, raised levels of inflammatory mediators in the CNS can trigger other resident brain cells to adopt neuroinflammatory-roles. This is seen in the transition of astrocytes into pro-inflammatory reactive phenotypes (A1 astrocytes) which secrete a host of compounds which ramp up neuroinflammation (Figure 1.11a). Other astrocytes – there are other nine different classes of astrocyte with various supporting roles in the CNS – have important roles in regulating BBB integrity and form part of the neurovascular unit (NVU), with astrocytic-feet interacting with vascular endothelial cells to promote the transport of substances like water and glucose from the blood into the brain parenchyma. Inflammatory activation of these NVU-based astrocytes causes increased release of cytokines and vasodilators, such as prostaglandins (PGE2), neuromodulators (such as VEGFA, OSM and G-CSF) and metalloproteinases (MMP1, MMP9, MMP12) which increases the permeability of the BBB via a variety of cellular mechanisms (Figure 1.11b) – from extracellular matrix reorganisation to nitric oxide (NO)-mediated damage to endothelial cells (Argaw et al., 2012; Liu et al., 2013; Lan et al., 2022).

Other interactions, such as pericyte-mediated BBB breakdown, and oligodendrocyte-vascular interactions (Figure 1.11c), have also been linked
to overspill of inflammatory mediators from the periphery (Smyth et al., 2018; Dohgu et al., 2019; Niu et al., 2019; Anwar, Özkan and Gürsoy-Özdemir, 2021).

**Figure 1.11 Blood brain barrier disruption**

Illustration demonstrating how various inflammatory mediators can activate astrocytes, microglia and oligodendrocytes into pro-inflammatory phenotypes which result in damage not only to neurons (not illustrated) but also to the BBB. Different mechanisms of BBB disruption by raised inflammatory mediators are overviewed in a-c, in depth discussion of these mechanisms can be found in the following reviews (Colombo and Farina, 2016; Liu et al., 2018; Anwar, Özkan and Gürsoy-Özdemir, 2021; Lan et al., 2022). (a) Activated astrocytes (A1, purple) may secrete many cytokines, chemokines and proteases, such as IL-1, IL-6, TNF-α, CXCL1, IL-8, CXCL10, MCP-1 and G-CSF, which can directly damage the endothelial cell walls of the BBB. (b) astrocyte interaction (with astrocytic end feet; blue) at the BBB can
disrupt vasculature when primed by inflammatory stimuli, as characterised by increased release of various prostaglandins and proteases (c) oligodendrocyte (purple) and pericyte (peach) mediated breakdown of BBB integrity.

From a functional perspective, delirium is a good example of elevation of how inflammation in the periphery must ostensibly cause neuroinflammatory reactions in the brain with direct functional consequences. Delirium is considered a severe and sometimes transient type of impairment in capacity that is highly linked to peripheral infections in the elderly, such as urinary tract infections (UTIs), that has been considered both a consequence of and risk factor for neurodegenerative conditions (Cunningham, 2011). Epidemiological studies demonstrate that patients with delirium have raised inflammatory mediators and that in cases of infection treatment of acute inflammation causes delirium to subside (Cunningham and Skelly, 2012). Oft reported ‘brain-fog’ during illness, or post-illness, is another example of what is considered a neuroinflammatory after-effect of raised inflammatory mediators in the periphery (Brusaferri et al., 2022). In both of these examples, raised inflammatory mediators in the CSF have been noted (Cape et al., 2014; Yang et al., 2021), and both histopathology and animal models have shown systemic inflammation causes disruptions to astrocytes, glia and the brain’s neurovasculature (Hoogland et al., 2015; Rothhammer et al., 2016).

Finally, there are BBB-independent mechanisms of cross-talk between the periphery and brain including vagal nerve stimulation, modulation via the gut-brain axis, meningeal-glymphatic routes of entry and HPA-axis activation. These are discussed in several reviews in the context of chronic inflammation contributing to adverse brain health outcomes (Bettcher et al., 2021; Yu et al., 2022) and summarised in Figure 6 of Chapter 6.2.

Taken together, these findings have led to the modern perspective of the CNS having a dynamic immune landscape, with several interfaces allowing communication with the peripheral immune system.
1.7 Summary

Chronic inflammation (as distinct from acute inflammation) is more common at both ends of the life course. This type of inflammation is linked to increased risk of mortality and age-related diseases, a trend that has informed the theory of ‘inflammaging’. Chronic inflammation may have central role in driving differences in brain health outcomes, but extant approaches to examining this relationship are flawed.

CRP remains the most widely investigated marker of inflammation, and there is an extensive literature reporting its associations with interlinking aspects of lifestyle and disease. Chronic inflammation varies due to heritable and non-heritable (environment, lifestyle, age itself) factors, but non-heritable factors explain a greater degree of this variation. Because of this, I emphasise how accurate profiling of chronic inflammation is key to increase our understanding of the drivers and consequences of inflammation in relation to brain health outcomes. I explain that the phasic and rapid response of CRP is what makes it such a useful (and popular) marker of acute inflammation, but caveat its limitations in profiling sustained inflammation in specific contexts (e.g. preterm infants and population cohorts).

The preferential use of CRP is interesting when set against the other myriad inflammatory biomarkers that are reflective of inflammatory responses. Many of these mediators have overlapping roles and distinct expression and function in different tissues and contexts. Because of this, I suggest that assigning one biomarker as the ‘most important’ or ‘most reflective’ of chronic inflammation is inappropriate. As a potential solution to this, I discuss the concept of composite indices of inflammatory burden and the rise of multi-omic biomarkers of inflammation. Historically, the relative technical difficulty and cost of measuring multiple markers in biological samples meant this approach was uncommon, but the advent of high-throughput technologies has changed this landscape. I suggest that examining multiple facets of the
inflammatory response could enable a more comprehensive capture of inflammatory burden in relation to brain health outcomes, a concept further developed in Chapter 3 of this thesis.

Finally, I discuss how peripheral inflammation relates to neuroinflammation and brain health phenotypes in the context of neurodevelopment and cognitive decline. I outline the differences between the immune landscape of the systemic circulation and CNS and the mechanisms of crosstalk between them. I illustrate how elevations in the periphery can cause neuroinflammation in a positive feedback loop, where the persistent recruitment and activation of leukocytes and glial cells causes unrelenting inflammation, weakening of the BBB, and ultimately injury to neurons and macroscale damage.
2 Brain and cognitive ageing

2.1 Introduction

The overall aim of this thesis is to characterise how chronic inflammation relates to brain structure and function across the lifecourse. While I have discussed the mechanistic basis of how elevations in peripheral inflammation manifest as microstructural changes which drive brain health consequences, in order to examine this relationship in population cohorts, it is important to outline the ways in which these features (brain structure and cognitive ability) change with age.

This chapter will give an overview of cognitive ageing, including an explanation on why adopting a lifespan approach to brain ageing is important. The two main approaches used to investigate cognitive ageing (cognitive testing and neuroimaging) will be outlined: first, with an introduction to cognitive ability, its assessment, and relationship with age, followed by an outline of the neuroimaging techniques of structural MRI and diffusion MRI.

The neuroimaging substrates of cognitive and brain ageing are then discussed, including an outline of global, regional and tissue-related alterations to the brain with age. This examination of cognitive epidemiology research encompasses the lifespan: examining brain structural differences from neurodevelopment to old-age, to discuss the potential unifying features driving differences in cognitive ageing. This brings us back to the role of chronic inflammation in brain and cognitive ageing and the best approaches to assess this relationship.
2.2 Lifespan perspective of brain ageing

Cognitive ageing, the process that describes the changes in cognitive ability with age, is a complex and dynamic process. Individuals vary greatly in their initial levels of cognitive functioning, their overall rate of cognitive decline, and consequently on the degree of impact that this has on their day-to-day living. Experiencing steeper rates of cognitive decline has been linked with higher rates of morbidity, disability, and mortality (Deary et al., 2009; Corley, Cox and Deary, 2018). Cognitive ageing is a consequence of the physical changes in the brain that occur with age – so called ‘brain ageing’. These changes in brain structure (thought to partly underpin cognitive ageing) can arise years before any clinical expression of cognitive impairment, and can sometimes be too subtle to detect with standard neuroimaging approaches. This means that identifying the preclinical attributes of these processes – genetic, environmental, and lifestyle – is of significant public health interest.

Given that age-related cognitive decline and brain structural changes are evident from early adulthood (Salthouse, 2009, 2016, 2019; Bethlehem et al., 2022), and the emerging role of early life factors on shaping lifelong trajectories of brain and cognitive health (Case and Paxson, 2009; Deary et al., 2009; Walhovd et al., 2016; Corley, Cox and Deary, 2018; Vidal-Pineiro et al., 2021; de Rooij, 2022; Corley et al., 2023), adopting a lifespan perspective on cognitive ageing has gained increasing traction (see Figure 2.1). Evidence points towards how brain changes in midlife relate to later-life cognitive function, and the role of early life experiences in determining both the rate, and extent, of cognitive and brain ageing (de Rooij, 2022; Corley et al., 2023). In light of this evidence, it is important to establish how genetic, environmental and epigenetic factors affect brain structure and function across multiple periods of life in order to understand what influences cognitive trajectories, specific windows of vulnerability, and where intervention may be most effective. In focusing exclusively on determinants of cognitive ageing in older age, where cognitive decline is (albeit differentially) present and established, we may miss some of the earlier tipping-points to
accelerated ageing trajectories and the key factors that drive these. Sampling at multiple time points in longitudinal study designs is also preferable given this evidence, and the fact that cognitive ageing is a dynamic within-person process.

Figure 2.1 Lifespan perspective on brain and cognitive ageing

Schematic\(^\text{18}\) illustrating the importance of examining earlier time points when considering later life cognitive function owing to the dynamic changes in brain structure seen across the lifecourse. Part (a) is a schematic of general trends seen in aspects of cognitive ability (fluid and crystallised) with advancing age. Part (b) of the figure is informed by the findings of a recent multi-cohort study (\(n = 101,457\)) which profiled changes to brain structure across the lifespan, using MRI scans from subjects from under a year old to 100 years old (Bethlehem et al., 2022); original figure and findings presented and commented on in section 2.6.1.

\(^\text{18}\) Bottom panel of figure is an adaptation based off findings by Bethlehem et al., (2022)
2.3 Cognitive ability

2.3.1 Assessing cognitive ability

Cognitive ability is an all-encompassing term to describe the capacity to think, reason, plan, learn and understand. It can be assessed in a variety of ways, including self-report questionnaires, psychometric tests, neuropsychological tests and screening tools. Whereas there are multiple valid research traditions with a multiplicity of methods for measuring the cognitive capacity of the brain – and more than 100 distinct terms in use to describe facets of cognition (Lara et al., 2015) – I focus here on measures of cognitive ability primarily used in cognitive epidemiology research. There are two main approaches that carry extensive evidence of their external validity with respect to important health outcomes: cognitive screening tools and cognitive test batteries.

2.3.1.1 Cognitive screening tools

Cognitive screening tools are methods of rapid assessment of cognitive function. They are less comprehensive than neuropsychological assessments, and carry potential of misclassification when used in isolation (Roebuck-Spencer et al., 2017), but have the advantage of being easy to conduct and replicate, especially in clinical settings (for which many of them were primarily designed). The aim of cognitive screening is to identify cognitive impairment, where a predetermined threshold distinguishes normal vs cognitive impairment. Lots of work has been done to check the reproducibility of these tools and to determine an optimal ‘threshold’ of impaired ability (Laurin et al., 2009; Metti et al., 2014; Singh-Manoux et al., 2014). The consequence of this is that they are poor at capturing variation within cognitive ability, and may be less sensitive at capturing cognitive impairment than more comprehensive approaches (Roebuck-Spencer et al., 2017). Examples of these tools are the Montreal Cognitive Screening
Assessment (MoCA), the Mini Mental State Examination (MMSE) and the Modified Mini-Mental Status Exam (3MS) (McDowell et al., 1997; Nasreddine et al., 2005).

2.3.1.2 Cognitive domains in psychometric test batteries

Neuropsychological testing are methods used by cognitive researchers to assess the different areas of mental functioning that are involved in the processing of information. Commonly referred to as 'cognitive test batteries', these are groups of standardized (psychometric, validated) tests designed to measure various aspects of cognitive functioning which are typically administered in a series and can take several hours to complete. These tests cover multiple cognitive domains, such as memory, language, attention, executive functioning, visual-spatial processing, motor skills, and social-emotional functioning. Typically, multiple tests per domain will be administered. This has important psychometric and statistical advantages in parsing apart the test-specific error variance from the between-person differences in the underlying cognitive domain-of-interest. Individual cognitive tests are noisy: a participant could be fatigued before lunchbreak during a certain test, or be more suited or used to pen-and-paper than a computer-based task. Having a test battery that spans multiple diverse cognitive domains also has the added advantage for computing a measure of general intelligence (or $g$), as outlined in the next section. An in-depth discussion of different types of cognitive assessment that are used in the analyses of this thesis is provided in the methodology section of this thesis (Chapter 5).

2.3.1.3 Spearman's construct of general intelligence

The first person to propose that there was an underlying factor of intelligence was Charles Spearman (Spearman, 1904). He observed that those who performed well on some cognitive tests also tended to perform better across many tests; in other words, there was strong evidence for common variance across cognitive domains. Spearman's construct of general intelligence was a revolutionary idea at the time of its conception in 1904: he proposed that
intelligence represents a single, general ability that underpins all other
cognitive capacities, and that this general factor (known as ‘g’), was
hereditary and could be measured with a single test (Spearman, 1961).
Spearman’s theory of a single, general intelligence has played an important
role in modern theories of intelligence and though there have been (and are
ongoing) debates surrounding the precise nature of g, the cross-test
correlations (the so-called positive manifold) first observed by Spearman has
been widely replicated (Johnson, Nijenhuis and Bouchard, 2008; Lyons et al.,
2017).

A general factor of intelligence typically explains around 40% of the total
variance in individual cognitive tests (Carroll, 1993), and different measures
of ‘g’ (built from distinct tests testing similar domains) display high correlation
\((r = 0.7-1.0)\) (Johnson, Nijenhuis and Bouchard, 2008). Individual differences
in g have been found to be highly heritable (Plomin and Spinath, 2002), and
this genetic component appears to persist across the lifecourse (Deary,
2014; Lyons et al., 2017), with numerous genetic loci implicated in
contributing to variance in g (Davies et al., 2015, 2018). Importantly, there is
an extensive literature within cognitive epidemiology reporting robust
associations between differences in g both in early and later life and an array
of important life outcomes including, job performance, illness, independent
living, dementia, and death (Deary et al., 2021). All of these demonstrate the
value of g as a marker of cognitive health.

2.3.1.4 Fluid intelligence and crystallised intelligence

The next step change in intelligence research came from Raymond Cattell in
the 1940s, closely followed by his student, John Horn, 25 years later. Cattel
proposed that g itself was limited in its characterisation of cognitive
functioning, and proposed the idea that within g there was a more plastic,
fluid form of intelligence \((g_f)\) and a more robust, crystallised form \((g_c)\) (Cattell,
1940). This theory was an extension from Spearman’s findings that
intelligence could be assessed by a single, unitary construct, and proposed
that intelligence was composed of two partly distinct (but correlated) components that could be developed and improved independently of each other.

In 1965, Horn (Horn and Cattell, 1966) refined these constructs – suggesting that $g_f$ is the ability to think abstractly and solve novel problems (as measured by performance on cognitive tasks testing reasoning, processing speed, executive functioning and working memory). This capacity is considered to be less driven by experience and more related to innate capacity, which is subject to greater declines with age. Crystallized intelligence, on the other hand, is defined as the ability to use acquired knowledge and skills to solve problems – more related to vocabulary, verbal comprehension, and akin to ‘wisdom’ – thereby unlike $g_f$, an intelligence heavily dependent on experience and education, and less affected by age.

Since then, studies have repeatedly replicated this work (Smith, 2016), finding that $g_f$ is particularly vulnerable to decline with advancing age, whereas aspects of $g_c$ remain preserved in later life even in extreme cases of cognitive decline such as dementia.

2.3.1.5 Carroll’s Three Stratum theory of cognitive ability

A further advancement of the concept of $g$ came in the form of Carroll’s Three Stratum theory of cognitive ability, which proposed that cognitive ability is composed of three distinct strata: (1) an overarching general factor of cognitive ability (2) broad categories of cognitive domains (3) narrowly defined cognitive abilities, which are grouped by broad cognitive domains owing to their correlational structure (Carroll, 1993). That is, one can statistically model the patterns of test score covariances into three levels which describe that people’s scores might be relatively good because i) they perform generally well across all tests (i.e. $g$), ii) they are generally good at a particular type of test (e.g. memory) and iii) they are good at the specific test being administered. Within this latter point are also factors like having a bad
day, or being worse at the mode of input (e.g. computers); this so-called test-
specific error variance (noise) necessarily sits with the individual cognitive
test as it cannot (by definition) be shared across tests. When deriving $g$ from
statistical dimension reduction techniques (a methodology used in the
empirical work of this thesis, discussed in greater depth in Chapter 5),
individual narrow measures of cognitive ability such as matrix reasoning,
block design, digit symbol coding are used to construct the latent factors of
broad cognitive domains (processing speed, fluid intelligence, crystallised
intelligence etc.), which are themselves all correlated and indicate a general
(superordinate) factor ‘$g$’.  

2.4 Cognitive ageing

A general decline in cognitive function appears to be a ubiquitous and
universal feature of ageing, but the rate and extent at which people
experience this decline differs. This difference is partly influenced by
cognitive reserve (baseline cognitive ability), which is also different between
individuals (Deary et al., 2009). Cognitive epidemiology is the study of the
relationship between cognitive function and health. The goal of cognitive
epidemiology is to understand how cognitive function is related to both the
factors that drive it and the outcomes it relates to, and to identify potential
interventions that can improve cognitive function and health (Deary, 2010).

2.4.1 Pathological cognitive ageing (dementia)

The most recognised form of pathological cognitive ageing is dementia, an
umbrella term that describes a class of neurodegenerative disorders which
cause the progressive loss of cognitive ability and day to day functioning.
Alzheimer’s disease (AD) is considered the leading cause of dementia,
accounting for around 50-75% of all cases in the UK, followed by vascular
dementia ($\sim 20\%$), Lewy body dementia ($10$-$15\%$) and frontotemporal
dementia ($\sim 2\%$) (Wittenberg et al., 2020). Some of these diagnoses are
known to be comorbid, such as AD and vascular dementia, but have distinct
neuropathological hallmarks and presentations (Eldholm et al., 2018). There
are other types of neurogenerative disorder that are classified according to
different criteria: those, like dementias, which are characterised by their
clinical presentation (Parkinson’s disease, motor neuron disease,
Huntington’s disease; PD, MND , HD), those classified by the localisation of
neurodegeneration (frontotemporal dementia, extrapyramidal disorders,
spinocerebellar degenerations), and those with specific underlying
neuropathology (multiple sclerosis, cerebellar ataxia, amyotrophic lateral
sclerosis) (Dugger and Dickson, 2017). As this thesis primarily focuses on
normative or healthy brain ageing, detailed descriptions of neurodegenerative
conditions are omitted, but the importance of discussing them briefly is in
their relation to chronic inflammation. Much of the leading theories behind
‘inflammaging’ and chronic inflammation’s role in cognitive ageing came from
early studies that established a link between chronic inflammation and
neurodegeneration, as reviewed in in (Amor et al., 2010; Chitnis and Weiner,
2017) and covered in the following sections.

The microglia discussed in **Chapter 1.6** are one of the key links between
systemic inflammation, neuroinflammation and neurodegeneration.
Aggregated proteins found in AD, PD and HD elicit a pro-inflammatory
response from microglia, which try to phagocytose the amyloid, synuclein or
Huntington in the same way macrophages would in the rest of the body. As
discussed, in this attempt they spew out an array of pro-inflammatory
mediators, which in turn prime other brain cells. The cycle becomes self-
perpetuating: glia have transitioned to their highly strung, ‘activated’ forms,
and the sustained inflammation maims other neurons, furthering
neurodegeneration. Moreover, there are other cellular-effects that impair
brain-function: axons can demyelinate, making them less efficient at sending
information, and inflammation can alter synaptic maintenance, leading to
altered neuronal signalling and decreased neuroplasticity.
2.4.2 Non-pathological cognitive ageing

Healthy cognitive ageing, or ‘normal’ cognitive ageing, describes the process of changes to cognitive function with age in the absence of clear pathology\(^\text{19}\) (Harada, Love and Triebel, 2013). Cognitive decline is a universal process: ability across multiple aspects of cognition appears to deteriorate as people age, irrespective of whether an individual has a diagnosis of a neurogenerative condition. However, translationally, early differentiation of normal ageing from neurodegenerative diagnoses is paramount for timely treatment to stave off further cognitive decline in the cases of several neurodegenerative disorders – such as dopaminergic treatments for PD (Mizuno, 2014), or cholinesterase inhibitors and immunotherapy treatments for AD and MS (Chitnis and Weiner, 2017; van Dyck et al., 2022). Equally, a better characterisation of what is considered ‘typical’ ageing is what enables discovery of the specific drivers underlying differences between individuals, stratifying cases of accelerated cognitive ageing.

As alluded to in section 2.3.1.4, not all cognitive abilities are equally susceptible to ageing at the population level: the relative mean amount of mean decline between ‘crystallised’ vs fluid forms of intelligence differs, with latter ostensibly more vulnerable. However, those mean levels are derived from individual differences in levels and changes. Whereas the mean levels at the population level show divergent ageing effects, evidence for within-person cognitive ageing tendencies demonstrates that cognitive declines across domains are correlated (Tucker-Drob, 2011; Ghisletta et al., 2012; Ritchie et al., 2016). That is, those who decline in one cognitive domain increasingly appear vulnerable to declines across multiple domains as a function of advancing age – and this is increasingly true as the sample age increases (Tucker-Drob, Brandmaier and Lindenberger, 2019). This lends

\[^{19}\] This dichotomisation is in itself problematic, as will later be explored in the brain structural alterations that appear with age in both neurodegenerative disease and in individuals without such a diagnosis. In reality, pathological changes associated with neurodegenerative disease may exist on a continuum, but for the purpose of distinguishing the type of ageing that occurs individuals without ‘overt’ signs of pathology this distinction suffices.
further support to the utility of ‘g’ as an important measure for capturing substantial meaningful variance in cognitive status in older age.

2.4.3 Inflammation and cognitive ageing

Both later life cognitive ability and rates of change in cognitive ability have been linked to childhood intelligence, with higher educational attainment linked to higher cognitive ability in later life\textsuperscript{20} (Zahodne \textit{et al}., 2011; Marioni \textit{et al}., 2012). These findings align with the theory of ‘cognitive reserve’: the ability to maintain cognitive functioning in the wake of age-related changes or disease processes (Stern, 2002). However, only around 40-50% of variance is explained by these models which examine cognitive ability in early life vs later life (Deary \textit{et al}., 2009), indicating that a substantial proportion of age-related impact on cognitive ability is not established in childhood and is determined by other processes. So what’s driving the other 50%? Many factors – from general aspects of health and fitness (Chou \textit{et al}., 2019; Zammit \textit{et al}., 2019), to disease, to lifestyle (Tilvis \textit{et al}., 2004; Salthouse, 2009; Chou \textit{et al}., 2019; Lipnicki \textit{et al}., 2019) – have all been associated with cognitive ageing. This indicates that there are both modifiable and inherent age-related factors that drive cognitive ageing.

As outlined in \textbf{Chapter 1}, a factor that appears linked to ageing itself, cognitive decline, and elements of health and lifestyle is chronic inflammation. In the context of the most widely researched inflammatory biomarker, CRP, both cross sectional and longitudinal studies set in non-clinical populations have found that serum CRP correlates with poorer cognitive functioning: from general cognitive ability (g) (Yaffe \textit{et al}., 2003, 2014; Schram \textit{et al}., 2007; Wersching \textit{et al}., 2010), to the individual cognitive domains of verbal memory (Gimeno, Marmot and Singh-Manoux, 2008; Noble \textit{et al}., 2010; Bettcher \textit{et al}., 2012; Marsland \textit{et al}., 2015) spatial reasoning (Marsland \textit{et al}., 2015); processing speed (Heringa \textit{et al}., 2014; \textit{footnote 20} Education appears to act on the \textit{intercept} of the slope but not \textit{differences} in cognitive ageing trajectories, i.e. while education is related to better cognitive ability in later life, this does not reflect less steep rates of cognitive decline.)
Beydoun et al., 2018; and executive function (Ganguli et al., 2014; Schram et al., 2007; Tegeler et al., 2016; Wersching et al., 2010). Conflictingly however, these findings are contrasted by null findings in otherwise similar populations of comparative age profiles – the studies referenced here are predominantly set in cohorts of individuals > 60 years, as studies that profile cognitive consequences of inflammation in midlife are rarer (Laurin et al., 2009; Singh-Manoux et al., 2014; Yaffe et al., 2014; Walker et al., 2017). For example, there have been reports of non-significant associations between CRP with global aspects of cognitive functioning for measures of $g_f$ (Dik et al., 2005), $g$ (Weuve, Paul M Ridker, et al., 2006) as well as null associations between CRP and verbal memory (Dik et al., 2005; Weuve et al., 2006; Jordanova et al., 2007; Rafnsson et al., 2007; Luciano et al., 2009; Wersching et al., 2010; Tampubolon, 2016; Tegeler et al., 2016; Beydoun et al., 2018), spatial reasoning (Luciano et al., 2009), executive function (Noble et al., 2010; Beydoun et al., 2018; Hajjar et al., 2018) and processing speed (Ganguli et al., 2014). While the literature on the association between CRP and cognitive decline presents mixed findings, there is increasing evidence in support of systemic inflammation’s contribution to cognitive decline when other inflammatory mediators are taken into consideration (Schram et al., 2007; Cribbs et al., 2012; Singh-Manoux et al., 2014; Walker et al., 2019; Slaney et al., 2023).

In light of these findings, it is reasonable to conclude that peripheral inflammation plays a significant role in contributing to differences in cognitive ageing. The associations observed in longitudinal studies (Kuban et al., 2017; Sluiman et al., 2022), and higher likelihood of developing dementia (Tan et al., 2007; Darweesh et al., 2018; Cullen et al., 2021), rising profile of anti-inflammatory treatments and immunotherapy agents for ameliorating cognitive impairments (Mortada et al., 2021), along with the underlying biological mechanisms and experimental evidence (Aktas et al., 2007; Amor et al., 2010; Prinz and Priller, 2014, 2017; Leigh and Flatt, 2015; DiSabato, Quan and Godbout, 2016; Chitnis and Weiner, 2017; Sun, Koyama and Shimada, 2022), provide a compelling argument in support of this stance.
However, further research is needed to fully understand the complex interplay between peripheral inflammation, central nervous system processes, and cognitive decline, as well as to explore potential therapeutic interventions targeting inflammation to mitigate cognitive decline in ageing populations. In particular there still remains considerable debate on the ways in which chronic inflammation relates to cognitive ageing in terms of the vulnerability of specific cognitive abilities, the directionality of the association, and the global and regional brain structural substrates of this process (points of ambiguity that the empirical chapters of this thesis attempt to shine light on). Given that brain structural changes in part may drive functional deterioration, cognitive capability, and neurodegenerative disease, the neuroimaging literature of these relationships are covered next.
2.5 Neuroimaging

2.5.1 Structural Magnetic Resonance Imaging (MRI)

Structural magnetic resonance imaging (MRI) has become an invaluable tool in understanding how the brain ages. Figure 2.2 illustrates the fundamental principles underlying the utility of MRI.

**Figure 2.2 MRI principles: distinguishing tissue types**

Schematic illustrating how MRI allows the differentiation of tissue types (a) water is made up of two hydrogen atoms and one oxygen atom; the hydrogen nucleus contains one positive charge - a proton spinning around on its axis (b) protons in the body usually are unaligned. However, (c) in MRI scanner, the protons align with the application of the magnetic field ($B_0$). When a second radio frequency (RF) pulse is applied, the protons realign and emit signals (d) these tell us about different tissue types, which give off different amounts of energy (e) from MRI we can extract $T1$ and $T2$ weighted images.

The physics of this process are outlined thoroughly elsewhere (Farrall, 2006). Briefly, MRI involves the application of a strong magnetic field, which serves
to align the protons present in the water molecules of the body's tissues. When a subsequent radio frequency pulse is then applied, the protons in the water molecules (specifically within the hydrogen atoms) are disrupted and then realign themselves; this process is known as resonance. During the realignment process, the protons emit detectable signals that are contingent upon the specific characteristics of the encompassing tissue. These signals are captured by a receiver coil, and by employing mathematical transformations, three-dimensional information can be derived. This enables the differentiation of diverse tissue types, including grey matter, white matter, and cerebrospinal fluid.

By harnessing these principles, structural MRI empowers researchers to discern and delineate the distinct anatomical features and tissue compositions of the brain, facilitating the investigation of focal age-related changes and their impact on cognitive processes.

2.5.2 Diffusion MRI

Diffusion MRI (dMRI) is a method that exploits cross-tissue and regional differences in water molecular diffusion in the brain, providing novel information about the microstructural environment.

In free space, water is known to diffuse in a random manner according to Brownian motion (Le Bihan and Iima, 2015). There are parts of the brain where we see this largely unconstrained motion, such as in cerebrospinal fluid (CSF). However, there are many other types of structure in the brain which contain water in a more densely packed environment: from neurons, to the myelin sheaths that envelop axons, to other elements of neural cytoarchitecture that exist within both intracellular (cytoplasm) and extracellular (lymph, ventricles, interstitial) fluid. In general, the diffusion of water in brain tissue is hindered by cell membranes, and as such, aspects of brain structure that are highly cellular tend to exhibit lower diffusion gradients. Equally, water that moves within intracellular compartments (where its
diffusion is hindered by organelles), will have a lower diffusion gradient than extracellular fluid. In this way, dMRI becomes a useful method by which to detect differences and alterations in brain tissue structures; the difference in the way water diffuses between tissues provides a contrast from which dMRI can detect and be used to infer differences in brain tissue architecture.

While dMRI refers to the contrast of the acquired images, diffusion tensor imaging (DTI) is a specific type of modelling of the dMRI datasets, which I will discuss next.

2.5.3 Diffusion Tensor Imaging

Since the introduction of this methodology in 1994 (Basser, Mattiello and LeBihan, 1994), DTI has been increasingly used to study the white matter architecture of the brain. Basser and colleagues found that water molecules’ motion (or diffusion) appeared to be much faster along white matter fibres than perpendicular to them. The difference between these two motions (parallel and perpendicular, also termed diffusion anisotropy) is the basis of the physics behind how DTI can be used to detect microstructural differences in the brain’s white matter. DTI does not directly evaluate white matter: instead, it utilises the concept that diffusion is uniform (equal in all directions) in the CSF and cell bodies, but directional (higher in one direction than others) in the axons that constitute white matter. DTI is useful because it provides us with a quantifiable measure of both the extent and orientation of water diffusion within individual volume units in the magnetic resonance image, which gives us a picture about the health (or integrity) of white matter.

While DTI principles and applications have been extensively described and reviewed in the literature (Soares et al., 2013), the take-home concept behind DTI is that water diffuses differently across brain structures depending on different characteristics. The DTI model represents each voxel (a small cube of data within the brain image) as an ellipsoid, or ‘tensor’, which contains parameters (the eigenvectors $\varepsilon_1$, $\varepsilon_2$, and $\varepsilon_3$ and the eigenvalues $\lambda_1$, $\lambda_2$, and $\lambda_3$).
λ3) that characterise the diffusion of water inside of it. These properties are what describe the magnitude of diffusion across three principle orthogonal axes and are summarised in Figure 2.3 below:

![Diagram of the diffusion tensor model](image)

**Figure 2.3. Diffusion Tensor Model.**

Diagram\(^{21}\) of the diffusion tensor model representing water diffusion as an ellipsoid with three eigenvectors (ε1, ε2, and ε3) representing the 3 principal axes of diffusion and three eigenvalues (λ1, λ2, and λ3) representing the magnitude of diffusion. The value of fractional anisotropy (FA) varies between 0 and 1. For perfect isotropic diffusion, \(λ_1 = λ_2 = λ_3\), the diffusion ellipsoid is a sphere, and FA = 0. With progressive diffusion anisotropy, the eigenvalues become more unequal and the ellipsoid becomes more elongated, and FA tends towards 1.

The main parameters that can be derived from these parameters are outlined in Figure 2.4 – enabling derivation of measures of white matter microstructure in the form of scalar measures such as fractional anisotropy (FA) or mean diffusivity (MD), and more nuanced representations of fibre directionality with techniques such as tractography.

---

\(^{21}\) Adapted from Soares et al., (2013)
Figure 2.4 FA map, using anisotropy and isotropy in brain imaging

The FA map is a grey-scale display of FA values across the image. Brighter areas are more anisotropic than darker areas. A-D is a schematic demonstrating that individual axonal fibre tracts which are more densely packed together have higher anisotropic values, whereas sparsely packed together fibres or thinly myelinated axons display more isotropic values. In the brain, diffusion across white matter tends to be more restricted across axons and therefore anisotropic (directionally-dependent); in contrast, in the grey matter, diffusion is impeded in all directions by neuronal cell bodies and membranes and is therefore less anisotropic and will appear more grey on the FA-map. At the far extreme, molecules within CSF in the ventricles of the brain are free to travel any which direction they want, making the CSF isotropic.

These diffusion metrics can be summarised globally for the whole brain or ascribed to specific regions of interest within the brain. In the literature, DTI
metrics (e.g., FA and MD) are normally used to describe changes in the integrity of white matter, e.g., to highlight potential alterations in the structure of certain brain regions in neurodegeneration and disease, used as a proxy of potential axonal injury.

However, different aspects of brain architecture other than myelination can impede the diffusion of water molecules, affecting the coherence and magnitude of the diffusion signal. While reduced FA or increased diffusivity measures are often interpreted as indicative of compromised structural integrity, these metrics are influenced by various factors beyond axonal health such as changes in tissue water content, fibre crossing, and partial volume effects (Jones, Knösche and Turner, 2013) have been developed. Moreover, DTI provides an indirect measurement of microstructural properties and does not capture the complexity of cellular and molecular processes underlying white matter organisation.

There are also some limitations of the DTI model when it comes to investigating how white matter microstructure changes with age, and various methods for how to analyse complex white matter architecture, particularly for delineating individual tracts of interest (Jeurissen et al., 2013). For tract-based DTI methods, there are different approaches which can be broadly classified on how the units of white matter to-be-analysed are identified (voxel-based or tract-based) (O'Donnell, Westin and Golby, 2009) which I discuss directly below.

2.5.3.1 Tractography

Measures of white matter microstructure such as FA and MD can be measured within major white matter tracts (i) in a voxel-wise sense within the white matter skeleton with tract-based spatial statistics (TBSS) (R. Smith et al., 2020); (ii) averaged across fibres identified using tractography algorithms.
Tractography is the processing of tracking white matter tracts using dMRI data. Tractography allows 3D visualisation of individual white matter tracts based on estimates of the directionality of water diffusion within each voxel (Mori et al., 1999). There are two main types of tractography used in neuroimaging field: probabilistic and deterministic (Mukherjee et al., 2008). Both of these methods start with an initial starting point (or ‘seed’) for individual tracts of interest to then map the tract through voxels with similar diffusion orientations based on user-defined constraints. For more detail on the ways in which DTI data was analysed in the empirical analyses of this thesis, refer to Chapter 5 (an example of PNT derived white matter tract ROIs is provided in Figure 5.2) and for a discussion of their appropriateness, section 9.7.2.

Of relevance to this thesis, tractography has been applied in both neonatal and ageing cohorts to examine both the maturational and ageing profiles of white matter architecture in the human brain. This has revealed distinct patterns of development of individual white matter tracts (Geng et al., 2012) and variation in WM tract integrity with age (Sexton et al., 2014; Cox, Ritchie, et al., 2016; Beck et al., 2021), which are further discussed in section 2.6.1.

2.6 Assessing brain ageing with MRI

Looking at the brain with neuroimaging is useful in potentially revealing individual heterogeneity that may underpin clinical endpoints (e.g. dementia diagnoses) and complex phenotypes like cognitive ability.

Different cognitive capacities appear to decline at different rates, a process underpinned by various structural changes that happen in relevant regions of the brain. Of particular interest is that certain types of cognition are more vulnerable to decline with age, and certain brain regions exhibit more structural alterations (from volume loss, to changes in white matter integrity,
to white matter hyperintensity burden, to other neuroimaging hallmarks of brain-ageing), as well as faster rates of these changes. Both the rate at which brain regions mature (Sowell et al., 2003) and age differs across the brain (MacDonald and Pike, 2021; Cox and Deary, 2022).

As individual brain imaging metrics are by their nature only partially informative of function (not informing us directly about dynamic changes in the brain, like neurotransmission, synaptic functioning, etc), both a capture of the full range and inter-relatedness of these changes is needed to uncover the global and regional foundations of specific cognitive changes with age. Alongside this, sample sizes needed for brain imaging studies need to be sufficiently large to be robust and informative (Marek et al., 2022). In the literature to date, these sample sizes differ by age-group, with certain populations having smaller n numbers, such as neonatal neuroimaging cohorts which often contain less than 100 participants (Makropoulos, Counsell and Rueckert, 2018).

I focus here on presenting an overview of the main highlights of structural substrates of brain ageing: global brain changes with age, regional brain changes with age, and the rise of neuroimaging ‘brain-age’ methods to chart the changes to brain structure observed with age.

### 2.6.1 Global and regional brain changes with age and chronic inflammation

There are some general patterns in population studies of ageing brains: a general decrease in brain volume, both in the grey and white matter, a loss of white matter integrity, and an increase in white matter aberrations such as white matter hyperintensities (WMH) and other markers of small vessel disease (SVD).

Though much work in the neuroimaging field suffers from small sample sizes (Resnick et al., 2003; Raz et al., 2005), there are notable exceptions to this.
In particular, a recent effort pooled data from over 100 studies (n = 101,457) to examine changes to brain structure across the lifespan, where MRI scans were obtained from subjects from under a year old to 100 years old (Bethlehem et al., 2022).

Here, researchers charted how rapid accelerations in global brain volumes occur in early life, followed by gradual losses in brain volume starting as early as adolescence (a process illustrated in Figure 2.5). Their key findings were that the volume of grey matter increases rapidly from mid-gestation onwards, peaking in childhood, just before around six years of age, before gradually decreasing, with deeper brain matter (subcortical volumes) demonstrating slower maturation and later declines. Global white matter volume exhibits a broadly similar developmental trajectory, but peaks much later, at around 29 years of age, and declines rapidly after 50 years of age. While substantive, this study had noticeable under-representation of younger-aged participants, particularly early-life, given the lack of open access neonatal MRI data.

Figure 2.5. Trajectories of change in MRI phenotypes across the lifespan.

Adapted from (Bethlehem et al., 2022) a schematic depiction of non-pathological ageing trajectories of the median for different MRI phenotypes (grey matter volume, white matter volume, subcortical volume, ventricular volume, total cerebrum volume, mean cortical thickness, total surface area) with age. Circles demarcate peak rates in growth; triangles depict peak volumes.

While overall white matter volume appears to change with age, so too does aspects of WM microstructure. Both cross-sectional and longitudinal dMRI studies in population cohorts demonstrate general trends of FA decreasing with advancing age, and MD increasing (Sexton et al., 2014; Cox, Ritchie, et al., 2016; Vinke et al., 2018; Beck et al., 2021). This process happens globally but some white matter tracts appear more vulnerable to age-related alterations. In particular, the cortico-cortical and thalamo-cortical connections show greater associations with age than the projection fibres (the current reasoning for why this happens is discussed in the overview of early-life brain development in section 2.7, as this temporal trajectory of age-related white matter degradation inversely recapitulates what happens in the developmental maturation of white matter).

These regional, or more localised changes can be found across other metrics, such as reductions in subcortical volumes, increases in the size of the brain’s ventricles, and reductions in metrics such as regional cortical thickness and surface area (Frangou et al., 2022). With age, the decreases in volume are most pronounced in the frontal and temporal lobes, which are regions of the brain involved in higher level cognitive functions such as memory, language, and decision-making (McDonald et al., 2012). Age-related declines in these brain regions are frequently reported in cross-sectional neuroimaging studies (Sowell et al., 2003; Dickerson et al., 2004; Allen et al., 2005), and longitudinal studies (Fjell et al., 2009, 2014). There is substantial heterogeneity in the rate, patterning and progression of these losses in grey and white matter volume between individuals, with aspects of health and lifestyle considered key factors underpinning this variation (Raz et al., 2005; Yaffe et al., 2014; Cox et al., 2019; Zhao et al., 2019). Inflammation has been suggested as a particular factor of interest that may tie together other lifestyle and health determinants of brain ageing (Furman et al., 2019; Finger et al., 2022). In support of this, chronic inflammation has been linked to region-specific atrophy in grey and white matter (Marsland et al., 2015; Grabert et al., 2016; Raj et al., 2017), with evidence suggesting that immune
cell distributions, endothelial dysfunction and changes to cerebrovasculature may underlie regional vulnerability (the mechanisms of which are covered in section 1.6).

Studies have also indicated that divergences in brain ageing trajectories may be apparent in midlife or even set off kilter in early-life (Salthouse, 2009; Vidal-Pineiro et al., 2021; de Rooij, 2022), decades before the clinical manifestation of functional impairment. Longitudinal follow-up of individuals in midlife demonstrate that volume loss in these regions can be linked to increase risk of undiagnosed AD and MCI, highlighting the potential of neuroimaging approaches to detect sub-clinical neurodegeneration (Dickerson et al., 2004, 2011). Again, chronic inflammation has been suggested as a key driver to these earlier departures off course from normative ageing (Walker et al., 2019, 2022; O’Donovan et al., 2021).

A point of caution here is that cross-sectional neuroimaging has been shown to underestimate age-related changes to brain structure, particularly for the most dynamic periods of change in both childhood and older-age (Di Biase et al., 2023). This recent study also reinforced the importance of controlling for lifestyle factors that are suspected to influence brain health, where prediction accuracies in individual age trajectory estimates of grey matter volume were improved by accounting for factors linked to poor health (such as alcohol consumption and diet). While longitudinal studies obviously offer an advantage in this respect, the usefulness of cross-sectional studies should not be overlooked: cross-sectional neuroimaging data remain useful for examining group differences, regional brain structure variation and examining

23 In this study that pooled data from UK Biobank (UKB) and Adolescent Brain Cognitive Development (ABCD) cohorts (resulting in a N > 9,000 participants, age range 9 – 80ys), cross-sectionally inferred trajectories of different brain structural metrics like cortical volume, thickness and surface area appeared relatively stable relative to longitudinally measured metrics. In contrast, longitudinal estimates of metrics like brain surface area (SA) highlight substantial variance in step with age. Di Biase et al. (2023) demonstrated that by relying on the cross-sectional normative model, researchers may wrongly conclude that various neuroimaging metrics like SA are preserved in the ageing process (and in turn, may wrongly assume that other brain metrics are more reflective of changes to cognition and decline independently of SA).
structural associations that are potentially linked to individual variables of interest (e.g. inflammation).

An alternative hypothesis is the ‘frontal ageing’ hypothesis, in which anterior brain regions appear to exhibit greater vulnerability to ageing than other white matter fibres. Regional brain atrophy in frontal regions has been suggested particularly sensitive to the effects of advancing age in normative cognitive ageing (Fjell et al., 2009, 2017) and there is some evidence to suggest that chronic inflammation may be the driver of this effect in preferentially affecting these frontal regions. For example, a study by Marsland et al. (2015) found that higher levels of systemic inflammation were associated with reduced white matter integrity in the anterior corpus callosum. Other studies on cognitively normal adults have reported similar findings, with inflammation-related white matter changes predominantly located in the frontal regions of the brain. A recent study built a composite signature of 9 inflammatory-proteins and examined it in relation to brain connectivity, finding certain networks (default mode, limbic) were most sensitive to chronic inflammation (Markov et al., 2022). Further research is needed to fully elucidate the relationship between chronic inflammation, white matter changes, and the anterior-posterior gradient of aging.

2.6.2 ‘Brain-age’ methods

More recent efforts have turned towards machine learning (ML) methods to develop neuroimaging-biomarkers of brain ageing. Such ‘brain-age’ metrics are increasingly used to characterise inter-individual differences in the rate and progression of cognitive decline (Cole and Franke, 2017; Vidal-Pineiro et al., 2021). These measures attempt to capture ‘biological’ brain ageing by applying machine learning to MRI data to predict chronological age; the difference between these two measures (predicted brain age and actual age; often termed ‘brain-age delta’) is used to describe whether an individual’s
brain appears comparatively older or younger than their actual age. This
discrepancy between actual age and biological age is a concept we will also
see in epigenomics research with epigenetic clocks (discussed in Chapter
3.8.1). In Chapter 8 of this thesis, alongside structural MRI, DTI and
cognitive test scores, I examine brain age in relation to multiple inflammatory
biomarkers, to characterise whether systemic inflammation relates to
accelerated rates of brain ageing (distributions of brain age and chronological
age visualised below).

Figure 2.6 Brain-age vs chronological age distributions in STRADL

Brain-age delta has been shown to forecast risk of various outcomes from
mortality, to accelerated cognitive ageing, to dementia (Franke and Gaser,
2012, 2019; Cole et al., 2018; S. M. Smith et al., 2020; Elliott et al., 2021),
and associates with biologically relevant features, such as cardiovascular risk
factors (Beck et al., 2022). Studies that investigate this metric of brain-age in
relation to peripheral inflammation are scarcer: of those published, the
majority are set in neuropsychiatric populations with contrasting findings. For
example, a study on advanced brain age in depression and anxiety (Han et
al., 2021) found no significant relationship with inflammatory markers (n =
260). In a larger study however of over 1,400 participants, while serum
protein levels did not show strong relationships with brain age, individuals
with higher measures of an epigenetic proxy of inflammation (discussed further in section 3.8.3) did have higher (i.e. older) relative brain age (Green, Squillace, et al., 2021).

While these findings are promising for investigating the determinants of differences in brain ageing, the validity of ‘brain age’ as a measure of inter-individual differences in brain ageing trajectories has been called into question. Specifically, a recent study from the Center for Lifespan Changes in Brain and Cognition (Vidal-Pineiro et al., 2021) investigated whether larger brain age deltas reflect steeper rates of brain ageing induced from events in later life, early-life, or a combination of both. They illustrate that brain age variation in adulthood can be mostly attributed to genetic and early life factors, with both birth weight and a genetic risk score of brain age emerging as the main reasons for deviation in normative brain ageing trajectories. While differences in brain age had been previously been shown to have a genetic component (Brouwer et al., 2021), this study’s novel findings on the importance of birthweight on brain-age variation highlights the relevance of early-life factors on later life brain health trajectories which is the next point of discussion.

2.7 Early life experiences influence the brain throughout life

Just as there are both regional and global patterns of change to the brain in later life, there are maturational processes in early life. These are complex, non-linear, and can be considered both progressive (cell growth, myelination) and regressive (synaptic pruning, apoptosis) in early neurodevelopment (Silk and Wood, 2011). Many of these maturational stages are considered windows of increased vulnerability to environmental exposures, as illustrated by the associations of maternal health and lifestyle factors on outcomes in offspring. There is emerging support for the importance of early-life factors in later-life cognitive reserve (de Rooij, 2022). This is evidenced both in the influence of early life molecular processes which appear to influence life-long neurodevelopmental trajectories (Walhovd et al., 2016; Werling et al., 2020)
and in structural differences at birth relating to later-life brain atrophy (Wheater et al., 2021). Lending support to this theory are the findings that variation in head size at birth and birthweight, childhood cognitive ability and education are linked to variance in brain structure in early life (de Rooij, 2022). This effect also appears to persist into adulthood, both for individual neuroimaging outcomes such as grey and white matter volume (Muller et al., 2014; Cox, Dickie, et al., 2016; Walhovd et al., 2016; Wheater et al., 2021) and, as just covered, for brain-age measures (Vidal-Pineiro et al., 2021).

Genetic influences have also been attributed to the relationship between cognition in infancy and childhood determining cognitive trajectories throughout the lifespan, where associations between higher cognitive ability in youth correlate with better ageing outcomes decades later (Deary, 2012; Walhovd et al., 2016; Ferguson, Brunsdon and Bradford, 2021). Epigenetic regulation, the bridge between these two influences (genetic and environment), holds significant promise in shedding light on this relationship, and is discussed further in Chapter 3.6.

Preterm birth is particularly illustrative example of how early life exposures can result in lifespan differences in brain structure and function, with chronic inflammation as a key mediator of these effects.

2.7.1 Preterm birth

The neurodevelopmental outcomes of preterm birth (PTB; infants born before 37 weeks gestational age) are stark: ~15% of infants born preterm (<32 weeks gestation) develop cerebral palsy, 30-50% experience intellectual disability, and preterm birth increases the risk of problems with educational achievement, language, mental-health, and many other morbidities (Blencowe et al., 2013). These problems present with lifelong challenges, and preterm birth has one of the highest numbers of disability-adjusted life years of any childhood condition. These neurological, cognitive, and
psychiatric outcomes are considered to be a consequence of brain structural changes that we will discuss next.

### 2.7.2 Encephalopathy of Prematurity: neuroimaging substrates of PTB

Impaired neurodevelopment of the preterm infant is associated with a wide range of brain structural differences termed the 'encephalopathy of prematurity' (EoP) phenotype (Volpe, 2009). EoP is a term used to describe the constellations of brain structural alterations seen in PTB compared to infants born full term, including white matter injury, increased ventricle volume, decreased cortical surface area and a reduction in certain regional brain volumes, cortical folding and gyrification. Here, I will discuss the neuroimaging characteristics of EoP, how inflammation may be a key arbitrator of these perturbations, although for in depth characterisation of these neuroimaging hallmarks of EoP consult the following reviews (Boardman et al., 2010; Boardman and Counsell, 2020).

There are global differences observed in brain structure when infants are born too early. Compared to term infants, preterm infants have both decreased global grey and white matter volume (Makropoulos, Counsell and Rueckert, 2018; Volpe, 2019) an effect that appears more pronounced the earlier the infant is born (Bouyssi-Kobar et al., 2016). These changes are correlated, meaning that infants with reduced grey matter tend to also have decreased white matter (Boardman et al., 2006). What's more, many of these brain structural alterations appear to predict neurodevelopmental outcomes in later life, with longitudinal studies illustrating the predictive power of neonatal neuroimaging in anticipating cognitive ability in childhood and adulthood (Beauchamp et al., 2008; Boardman et al., 2010; Meng et al., 2016). Volume reductions in specific areas of the brain in preterm infants have been documented across numerous brain regions, disparities that persist decades after birth. Many of these regional volume changes, particularly those in the temporal, parietal and frontal areas of the cortex have been found to mediate the effects of preterm birth on both early-life and
later-life cognitive ability (Nagy et al., 2009; Nosarti et al., 2014). Persisting differences in regional brain volume include the hippocampus (Cheong et al., 2013), caudate nuclei (Nosarti et al., 2008, 2014), thalamus (Giménez et al., 2006), corpus callosum (Taylor et al., 2011), internal capsule, insula, putamen (Allin et al., 2004) and cerebellum (Allin et al., 2001; Taylor et al., 2011). Other measures of brain structure, such as cortical thickness and cortical surface area, have also been found to be reduced (long-term) in the prefrontal, temporal and parietal regions of the brain of preterm infants (Bjuland et al., 2013; Skranes et al., 2013).

White matter injury is a particularly researched topic in relation to preterm birth. White matter injury can either be localised or diffuse, and there are accounts of both global diffuse changes and focal areas of white matter loss in preterm infants. A particular patterning of white matter loss that occurs near the lateral ventricles of the brain is referred to as periventricular leukomalacia (PVL) and is considered a neurostructural hallmark of preterm birth. PVL in itself can be either macroscopic, microscopic or diffuse: the former involves the degeneration of multiple cell types (glia, neuronal axons) and can be visualised from MRI (Leviton, Allred, et al., 2013). At the other end of the scale, MRI fails to capture most cases of cystic microscopic PVL: these are areas of white matter loss that are <1mm in diameter, and are more commonly identified post-mortem, where histological analysis has revealed axonal degeneration and elevated macrophage activity. The final type, diffuse PVL, is the most common presentation of white matter injury: a non-specific loss of integrity across different brain regions, characterised at the histopathological level by loss of pre-myelinating oligodendrocytes, hypomyelinated axonal fibres, astrogliosis and high microglial burden.

Alongside histopathology the use of DTI can quantify these microscopic differences noted between infants born at term vs preterm, with a large body of evidence indicating that premature infants display reductions in FA and increases in MD at global and regional levels. These neuroimaging characteristics reflect a dysmaturation profile in white matter development
with trends of fewer subcortical-cortical neural connections in preterms (Meng et al., 2016; Thomason et al., 2017), reduced complexity (Scheinost et al., 2016) and disrupted connections between the thalamus and cortex (Ball et al., 2015). As with the reductions in brain volume observed in preterm infants, this altered neonatal neural connectivity is associated with poorer neurodevelopmental outcomes (Rathbone et al., 2011), an effect that can persist into adulthood (Meng et al., 2016).

As alluded to earlier, the neonatal neuroimaging literature suffers from smaller population sizes compared to adult neuroimaging studies, and as such these early-life trajectories of brain-ageing are less well defined. Nevertheless, the picture that emerges from the research to date is that early-life is associated with macroscale and microscale brain structural differences, many of which are persisting and linked to functional consequences in later life.

2.7.3 Inflammation, white matter tract development, and contribution to EoP

To understand why some areas of the brain, in particular the brain’s white matter, appear to be more vulnerable to altered development in preterm compared to term infants, the developmental programming of white matter should be considered.

White matter tracts develop at different rates both in utero and postnatally, with the latest longitudinal neuroimaging studies indicating that in utero and the first year of postnatal life are when the most dynamic changes happen (Stephens et al., 2020). White matter development is a complex sequence of programmed events: from proliferation, migration, cellular specification, differentiation of dendrites, synaptogenesis, gliogenesis, myelination and synaptic pruning. In humans, the long projection fibres develop first (a process that occurs largely during the second trimester of pregnancy), with sequential and partly simultaneous development of the thalamocortical,
commissural, and associative pathways (Kostović et al., 2014). The process of this refinement is initial proliferation, followed by synaptic pruning to remove less used connections. Thereafter, myelination stabilises connections that are functionally important. Though many of these growth patterns overlap, there are distinct periods of plasticity observed for each class of white matter fibre tract; these windows of increased growth are also windows of increased vulnerability to perturbation (Ment et al., 2009).

This increased vulnerability of white matter during certain growth periods is consistent with many studies examining white matter injury in the third trimester. Several studies have outlined that there are dysregulated temporal-spatial patterns of white matter development in preterm infants (Rose et al., 2014). In longitudinal cohorts of preterm infants, diffusion metrics have been shown to mature in a central–peripheral and occipital–frontal manner – whereby frontal and peripheral white matter appears to mature much more slowly (Partridge et al., 2004; Kersbergen et al., 2014). This pattern of myelination (from deep to superficial brain regions) was first outlined in post-mortem studies (Yakovlev, 1967; Kinney et al., 1988) and was consistent with the ‘first-in-last-out’ hypothesis which posits that the brain regions slowest to mature are also the ones that are most vulnerable to decline. In preterm infants, these peripheral connections are also more vulnerable to aberration (Batalle et al., 2017). There are several theories as to why the later-developing regions are the first to degrade, with the leading being that brain regions which exhibit significant plasticity over prolonged periods are more susceptible to inflammatory exposure (Casey, Giedd and Thomas, 2000).

Mechanistic evidence for this comes from work that finds that regions of the brain that are later to develop have an imbalanced ratio of immature to mature oligodendrocyte populations, something that may also be influenced by inflammatory processes (van Tilborg et al., 2018; Jäkel et al., 2019). Oligodendrocytes produce the myelin that insulate neuronal axons to allow for efficient conduction and are a heterogeneous cell type depending on
when they differentiate to produce myelin (Power et al., 2002). Those that mature later tend to ensheath smaller axons (Bartzokis et al., 2004) – at mesoscale level, this presents as and more thinly myelinated axonal fibres in these later-to-develop brain regions. This can be detected by DTI, where more thinly myelinated fibres display with lower FA values. This reduction in integrity in anterior brain regions is potentially what confers enhanced vulnerability of these regions to age-related neurodegeneration (Nagy et al., 2009; Nosarti et al., 2014) – an effect that may be mediated by the role inflammation plays in the disruption of white matter.

Both the brain and the immune system are not mature at birth; as discussed in Chapter 1.4.1, the immune system in early life is highly reactive, with higher levels of pro-inflammatory receptor expression and decreased levels of anti-inflammatory mediators. The number of inflammatory hits experienced (such as multiple postnatal morbidities related to inflammation, recurrent episodes of infection or sepsis, or combined fetal-inflammatory episodes, such as histologic chorioamnionitis alongside preeclampsia) has also been shown to correlate with the degree of brain structural aberrations (Bassler et al., 2009; Korzeniewski et al., 2014; Glass et al., 2018; Humberg et al., 2020), emphasising the disruptive, dose-dependent role of inflammation in early brain development.

The picture that emerges from this research is that genetics, early life exposures, and epigenetic changes influencing neurodevelopment can prime brain structure to either confer resilience or vulnerability to later-life cognitive decline. Peripheral inflammation is shown to associate with a range of EoP characteristics and later life neurodevelopmental outcomes, though there is a lack of robust profiling of these associations and accounting for confounding factors in these relationships. The sum of these findings suggest that early-life can dysregulate brain development in infancy and can lead to substantial long-lasting effects on brain structure and function in later-life.
2.8 Summary

This chapter reinforces the importance of adopting a lifespan perspective on brain and cognitive maturation and ageing, particularly in terms of identifying the determinants of age-related cognitive decline. It also highlights key methodological approaches for measuring both cognitive functioning and brain structural measures.

With respect to cognitive function, both individual differences in levels and age-related changes across an array of tests can be usefully understood at test-specific, domain-specific (e.g., memory, processing speed, spatial reasoning) and global ('g') levels of explanation. Modelling cognitive performance in this way has clear advantages for capturing both mean level trends at the population level as well as differentiating between individual differences therein. I discuss the importance of the field of cognitive epidemiology in identifying the correlates of cognitive ageing.

I then outline the neuroimaging substrates of cognitive ageing, by first describing how neuroimaging is used to inform us about aspects of brain function. I cover the main types of neuroimaging metrics used in the work in this thesis: structural and diffusion MRI, before discussing the global and regional changes seen with age, as well as newer methods that attempt to chart differences in brain ageing trajectories.

I conclude by drawing back to the lifespan perspective of brain and cognitive ageing, using the example of preterm birth as an example of how early-life exposures, such as sustained inflammation, are linked to brain dysmaturation. I discuss the literature of how aberrations in brain structure in early life are linked to persisting brain structural and functional differences in later life.
3 DNA methylation

3.1 Introduction

Epigenetics most commonly describes chemical changes that help to regulate gene expression without altering the underlying DNA sequence. While epigenetic modifications include changes in proteins around which DNA is packaged (histones) and the expression of non-coding RNAs, the most widely studied epigenetic mechanisms is DNA methylation (DNAm) – the addition of methyl groups to DNA which changes the expression of genes. There is increased awareness of the importance of epigenetics in healthy programmed development, and that alterations in epigenetic mechanisms can contribute to a range of diseases and differences in health trajectories.

This chapter begins with an overview of the history of DNAm research, the molecular mechanisms governing DNAm, its measurement, and the approaches used to examine DNAm in epigenetic epidemiological research. The current literature on the association of DNAm in health and disease is then discussed, with a particular focus on brain and cognitive associations and inflammatory-related disease correlates. To conclude, I introduce the concept of aggregate DNAm measures (DNAm signatures – analogous to polygenic risk scores) to study phenotypes, traits and exposures and how this method can be applied to index chronic inflammation and examine its impact on brain and cognitive ageing.

3.2 Discovery

DNA methylation (DNAm) describes a direct chemical modification to DNA that can influence whether a gene is silenced or expressed. Following the serendipitous discovery of DNAm in 1948 (Hotchkiss, 1948), the importance
of DNAm in development was realised, particularly in relation to cellular differentiation and programmed gene regulation (Holliday and Pugh, 1975; Compere and Palmiter, 1981).

Following the initial interest of DNAm in relation to development, work since has shown that DNAm has myriad roles on cellular functioning throughout the lifecourse (Moore, Le and Fan, 2013), with roles in regulating transcription (Maunakea et al., 2010), chromosomal stability (Vilain et al., 2000) and influencing health and disease from early development onwards (Kofink et al., 2013; Nugent and Bale, 2015; Li et al., 2022). More recently, attention has turned towards the potential of DNAm to track or delineate differences in age and health trajectories (Bock, 2009; Bakulski and Fallin, 2014; Levine et al., 2018) and the value of DNAm-based predictors of these differences.

In direct contrast to genetic alterations, epigenetic patterns, like DNAm, have been shown to change dynamically throughout the lifecourse (Bjornsson et al., 2008; Horvath, 2013; Richmond et al., 2015), and can be influenced by environmental factors and changes in health status. DNAm can therefore be viewed as an interface by which the genotype and environment relate to phenotype. Because of this, aberrant DNAm can signpost differences in how individuals respond to both endogenous and exogenous events, and underscore how risk factors contribute to the onset and progression of disease.

3.3 Molecular mechanisms

DNA methylation is classified as the addition of a methyl group (CH₃) to the fifth carbon atom of a cytosine (C) base, resulting in a 5-methylcytosine (5mC) molecule (Figure 3.1).
DNA is modified by the addition of a methyl group (CH₃) at the 5-carbon position of cytosine by DNMTs. This reaction is potentially reversible by DNA demethylases.

Abbreviations: DNMT: DNA methyltransferase; 5mC: 5-methylcytosine

This process is reversible, catalysed by enzymes known as DNA methyltransferases (DNMTs), and in humans almost always occurs between cytosine and guanine bases separated by a phosphate. Because of this, specific ‘sites’ of methylation in the genome are therefore referred to as ‘CpG’ dinucleotides (Song et al., 2011).

Identifying biologically relevant CpG sites is therefore at the heart of epigenomic studies in terms of identifying differences in DNAm for particular health outcomes and creating DNAm-based biomarkers to signpost risk of disease (Bock, 2009). Thanks to advances in genotyping technology (outlined next), we can now measure the relative proportion of methyl groups

---

Figure 3.1 The process of DNA methylation (cytosine to 5-methylcytosine)

24 DNA is modified by the addition of a methyl group (CH₃) at the 5-carbon position of cytosine by DNMTs. This reaction is potentially reversible by DNA demethylases. Abbreviations: DNMT: DNA methyltransferase; 5mC: 5-methylcytosine

24 adapted from Relton & Davey Smith (2010)
at individual CpG sites and examine how differential DNAm patterns relate to health and disease (Yousefi et al., 2022). Research has demonstrated that these changes in DNAm levels appear to be the result of both innate biological factors like CpG site density (Baubec and Schübeler, 2014) cellular aging over time (Johnson et al., 2012), cells’ unique developmental lineage (Kim and Costello, 2017; Lappalainen and Greally, 2017), how the body responds to disease (Petronis, 2010; Cavalli and Heard, 2019) and exogenic ones like environmental exposures (Feil and Fraga, 2012; Hou et al., 2012; Cavalli and Heard, 2019). DNAm patterns are also influenced by genetic sequence variations – called methylation quantitative trait loci (mQTL) – which associate with both cis (local loci) or trans- (distal loci) effects on methylation (Lemire et al., 2015).

Because DNAm varies dynamically in response to differences in health, age, and disease, it has been researched both as (1) a mediator of how risk factors contribute to disease and (2) as a biomarker of these health differences in epidemiological research (these epidemiological frameworks are expanded on in section 3.6).

3.4 Measurement

Multiple methods have been developed for measuring DNA methylation. Advances in high throughput molecular techniques have led to a dominance of microarray-based technologies, which allow for the simultaneous measurement of methylation levels at multiple sites in the genome (Laird, 2010). Microarray-based technologies involve hybridizing DNA fragments to a microarray chip, which are labelled with fluorescent dyes that are specific to either methylated or unmethylated DNA. The chip is then scanned and the fluorescence intensity of each spot is measured, allowing methylation levels to be determined at each site in the genome. The current gold-standard in the field is the use of microarray-based bisulfite sequencing (Villicaña and Bell, 2021; Yousefi et al., 2022). This technique involves treating the DNA with sodium bisulfite, which converts unmethylated cytosines to uracils, while
methylated cytosines remain unchanged; the DNA is then amplified with primers that are specific to either methylated or unmethylated DNA, and these amplified fragments are then hybridized to the microarray chip and scanned, allowing methylation levels to be determined at specific CpG sites.

There are different platforms available to measure DNAm. To date, Illumina bead-chip platforms have been most popular, where probes target bisulfite-converted DNA (Bibikova et al., 2009). The Infinium HumanMethylation27 BeadChip (27K) was one of the earliest models used to measure DNAm, targeting around 27,000 sites (0.1% of all possible CpGs) mainly in CpG islands (CGIs) within promoters (Bibikova et al., 2009). This was followed by the widely-used Infinium HumanMethylation450 array (450K), which targeted ~480,000 sites (1.7% of total CpGs), consisting of the 27K sites and increased coverage. The latest version is the Infinium MethylationEPIC BeadChip (EPIC), which targets ~850,000 sites (3% of total CpGs), including almost all of the 450K sites, with additional CpG sites in enhancer gene regions – i.e., the latest array is capable of quantifying DNAm at almost double the number of CpG sites than its predecessor. This evolution illustrates the rapid development of technology in this field and increased resolution power to examine epigenetic modifications – for a more detailed overview of the current strategies, platforms, pipelines and differences in DNAm measurement, see the review by (Campagna et al., 2021).

3.5 Analysis: EWAS vs Candidate Gene approaches

The ability to measure DNAm at scale has resulted in a boom of epigenetic research which examines DNAm variation on a parallel with population-based genetic research (Yousefi et al., 2022). There are two main approaches adopted to examine methylation profiles: epigenome-wide association studies (EWAS) and candidate gene approaches. An EWAS is a type of study that looks at the entire epigenome, or the complete set of epigenetic modifications, across a population of individuals – an EWAS compares epigenetic marks (such as DNAm – these are sometimes also
referred to as methylation-wide association studies or MWAS) in these populations to capture epigenetic variation association with a particular outcome measure. Differentially methylated CpG sites that are associated with a phenotype of interest can provide insight into the mechanisms of disease, the biological pathways involved in regulating that disease and highlight specific sites for further investigation. In this way, an EWAS can catalogue a reference list of CpG sites that are robustly associated with a particular outcome measure. In contrast, a candidate gene approach focuses on a specific set of genes that are thought to be related to the trait or disease of interest. This approach allows researchers to examine the methylation profiles of these genes in more detail but does not provide an overall picture of the epigenome. Compared to candidate gene approaches, an EWAS approach is often preferred owing to its unbiased assessment of DNAm in relation to study outcomes. Though EWAS shares many parallels with its genetic predecessor (genome-wide association studies; GWAS) – with shared caveats of multiple testing burden, etc. – results from EWASs have distinct biological implications which are discussed in Chapter 9.2.

Both EWASs and candidate gene approaches have been used to predict circulating plasma protein levels (Hillary et al., 2020) as well as specific disease outcomes, with severable publicly available databases that catalogue published CpG-trait associations (Battram et al., 2022).

3.6 Epigenetic epidemiology

Epidemiology is the study of disease in populations – its causes, consequences, and how to identify, ameliorate and prevent these adverse outcomes. Epigenetic epidemiology is the study of how epigenetic changes relate to these outcomes. It is now well established that DNAm is dynamic in response to different environmental and biological events throughout the lifecourse – showing aberrant patterns with differences in nutrition (Heijmans et al., 2008), pollution (Christensen et al., 2009; Ladd-Acosta and Fallin,
2019), stress (Perera and Herbstman, 2011; Rijlaarsdam et al., 2016; Cao-Lei et al., 2020; Zhang and Liu, 2022), with substantial evidence that DNA methylation (DNAm) profiles change with increasing age and contribute to a large array of age-related diseases (Bocklandt et al., 2011; Hannum et al., 2013; Heyn, Méndez-González and Esteller, 2013; Horvath, 2013; Ballestar, Sawalha and Lu, 2020). Consistent with this theory are studies that show that monozygotic twins begin life with highly congruent epigenomic profiles, and that these profiles diverge with age and are predominantly driven by lifestyle and health differences (Fraga et al., 2005; Li et al., 2022).

While some epigenetic changes with age appear to be stochastic and without observable effect on phenotype – a process known as ‘epigenetic drift’, which describes the increase in inter-individual DNA methylation variability with age – (Egger et al., 2004; Shah et al., 2014), it is clear that some environment-mediated changes can influence gene expression to increase disease likelihood (Feinberg, 2007; Bock, 2009; Hirst and Marra, 2009) and that differences in epigenetic marks (such as DNA methylation levels) can provide information on the phenotype of individuals.

This plasticity is how the epigenome can be considered as a biological archive of exposure (Relton, Hartwig and Davey Smith, 2015) – one that amasses changes over the lifecourse in response to life events, injury, illness and lifestyle. This is potentially useful for understanding the drivers of differences in brain structure and function with age. While efforts need to be taken to avoid non-biologically relevant, spurious associations (and particular care needs to be taken in handling of DNA methylation data, such as cell-subtype heterogeneity), reliable non-causal associations can still be useful as predictive or prognostic markers when it comes to understanding brain structural and functional variance, and the reasons for differences between individual brain age trajectories.
Figure 3.2 illustrates how epigenetic epidemiology has evolved from the study of epigenetic modifications such as DNAm as a direct mechanism of disease, to encompass the relationships between genotype, epigenotype and outcome, as well as the indirect impact of DNAm in relation to disease and the potential of DNAm as a biomarker to signpost risk of disease and disease-risk factors. These different frameworks of conducting epidemiological research have informed our understanding of differences in health between individuals. By leveraging DNAm data to profile the molecular characteristics of disease populations with greater resolution, we are better placed to identify those at greater risk of developing a disease.

3.6.1 Association of DNAm with health and disease

As outlined in Figure 3.2, DNAm may represent a mechanism of disease pathophysiology, a modifier of disease risk, and biomarker of disease incidence. Because of DNAm’s role in developmental programming introduced in section 3.2, initially DNAm was heavily researched in relation
to disorders of developmental origin such as neural tube defects (Price et al., 2016) and congenital disorders (Elhamamsy, 2017). Since then, there have been numerous lines of evidence supporting aberrant DNAm profiles in populations of individuals with neurodevelopmental disorders (Aref-Eshghi et al., 2018, 2020; Hüls et al., 2021), such as ADHD and autism (Cecil and Nigg, 2022), and the incidence and associated morbidities of preterm birth (Parets et al., 2013; Mohandas et al., 2018; Everson et al., 2020; Hüls et al., 2021).

DNAm from cord blood in preterm populations has indicated that that differential methylation at many CpG sites is associated with low birth weight, gestational age at birth and head size (Knight et al., 2016; Küpers et al., 2019; Merid et al., 2020). Longitudinal studies have also demonstrated that preterm birth status significantly associates with differential methylation in adulthood (Cruickshank et al., 2013). The question of whether DNAm alterations that occur in utero or at birth influence lifecourse health trajectories is considered in studies that work off the Developmental Origins of Disease Hypothesis (Barker, 2004; Ozanne and Constância, 2007), with epigenetics lending new insights into the degree to which early-life predisposes later life disease risk (Agarwal et al., 2018; Felix and Cecil, 2019; Shanthikumar et al., 2020).

As outlined in Chapter 1.5.2.4, there is a strong connection between inflammation, cancer, and chronic innate immune activation (Coussens and Werb, 2002), with more than 30% of cancers associated with chronic inflammation linked to different aspects of oncogenesis (Suzuki et al., 2009; Das, Karthik and Taneja, 2021). Aberrant DNAm is detectable in early-stage tumours, has shown promise in predicting treatment response (Laird, 2003), cancer-stage (Van Neste et al., 2016), and has been suggested as a premalignant biomarker of certain cancer subtypes (Weisenberger et al., 2006; Suzuki and Bird, 2008; Levine et al., 2015; Hashimoto, Zumwalt and Goel, 2016; Baglietto et al., 2017). Because of this, DNAm has become a promising candidate for aspects of cancer diagnostics and therapeutics, a
subject covered in several reviews (Pan et al., 2018; Locke et al., 2019). However outside of cancer and developmental disorders, increasing research has investigated the association of differential methylation in relation to multiple aspects of health and disease throughout the lifecourse, from aetiologies of childhood (Lima et al., 2020; Luo et al., 2021; Xu et al., 2021; Caramaschi et al., 2022), to those of old-age (Bell et al., 2012; Johnson et al., 2012; De Jager et al., 2014; Lunnon et al., 2014; Chouliaras et al., 2018; Henderson-Smith et al., 2019). This includes studies examining disorders of metabolism (Chambers et al., 2015), cardiovascular disease (Cappozzo et al., 2022) and autoimmune conditions (Chen et al., 2017; Hedrich et al., 2017).

Aberrant DNA methylation in relation to cognition and brain health outcomes has also been studied, both for neurodevelopmental disorders as discussed above and for cases of neurodegeneration such as Alzheimer’s disease (Lu et al., 2013; De Jager et al., 2014; Lunnon et al., 2014; Smith et al., 2019), Parkinson’s disease (Smith et al., 2019) and multiple sclerosis (Celarain and Tomas-Roig, 2020). However there are less studies that examine DNA methylation differences in cases of normative ageing or cognitive decline in the absence of specific pathology (Chouliaras et al., 2018).

3.6.2 Association of DNA methylation with lifestyle and environmental exposure

Thus far, aside from studies of biological and chronological age (which dominate the literature – discussed further in section 3.8.1) analysis of DNA methylation data has mostly attempted to determine whether variation in DNA methylation patterns has mechanistic relevance in contributing to endpoint outcomes (e.g. role of DNA methylation in contributing to certain diseases).

However, increasingly the potential of DNA methylation for indexing both exogeneous and endogenous risk factors that predispose or exacerbate health risks has
been explored, particularly if these are difficult or unreliable to measure via other approaches (Ladd-Acosta, 2015; Ladd-Acosta and Fallin, 2019).

The most seminal example of an environmental factor associating with alterations in DNA methylation (DNAm) is exposure to cigarette smoke, with stark differences in DNAm profiles observed between smokers and non-smokers. For example, a meta-analyses of 16 population cohorts (totalling over 15,000 individuals) found that 18,760 CpG (mapping to over 7,000 genes) were related to smoke exposure (Joehanes et al., 2016). This is a robust association: differentially methylated DNAm in smokers has been replicated across many populations (Breitling et al., 2011; Sun et al., 2013; Elliott et al., 2014); and it is non-restricted to tissue-type, with replicable findings of smoking’s impact on peripheral DNAm, regardless of whether DNAm is sampled from whole blood, saliva or cord blood (Monick et al., 2012; Dawes et al., 2019). Smoking-related DNAm variance has also been linked to an array of age-related diseases, from cardiovascular health (Zhang Yan et al., 2016; Maas et al., 2020), to cancer (Zhang et al., 2016; Baglietto et al., 2017; Battram et al., 2019), to neurodegeneration (Corley et al., 2019) as well as accelerated biological ageing (Lei et al., 2020; Cardenas et al., 2022; Klopack et al., 2022) and increased mortality risk (Bojesen et al., 2017).

This association is also one that persists even after an individual stops smoking, with DNAm signatures able to proxy non-smokers, former smokers and current smokers (McCartney, Stevenson, et al., 2018; Sugden et al., 2019; Langdon et al., 2021). Second-hand exposure to cigarette smoke also associates with DNAm differences (Kim, 2019), and some smoking-linked DNAm alterations appear to be passed from mother to infant (Suter et al., 2010; Murphy et al., 2012; Richmond et al., 2015; Joubert et al., 2016). In a large scale EWAS meta-analysis (PACE consortium – 13 mother-infant cohorts across US and Europe) prenatal smoke exposure in mothers was associated with aberrant methylation patterns in offspring (Joubert et al., 2016), many of which have been shown to persist later in life in different cohorts (Ladd-Acosta et al., 2016; Rauschert et al., 2019).
The specific mechanisms of how smoking may give rise to epigenetic alterations has been covered comprehensively elsewhere (Lee and Pausova, 2013), but briefly the role of tobacco, nicotine and inflammation have all be put forward as modulators of DNAm. Pertinent to this point is that the rise of smoking alternatives, such as e-cigarettes and vaping – which lack tobacco but are sources of nicotine – and their impact on the epigenome and related health consequences are subject to wider debate, with some studies demonstrating a link between these alternatives with aberrant DNAm (Martin et al., 2016; Xie et al., 2021) but with considerably less reproducible findings compared to the robust association of smoking with DNAm. Those that are shared between cigarette smoking vs vaping have been linked to inflammatory-related genes (Besingi and Johansson, 2014; Richmond et al., 2021; Andersen et al., 2022).

Many other exogenous exposures that are linked to chronic inflammation have been associated with DNAm differences, including factors like alcohol, diet, medication, physical activity levels and pollution (Niculescu and Zeisel, 2002; García-Calzón et al., 2017; Liu et al., 2018; Hibler et al., 2019). Overall, these findings support the hypothesis central to this thesis: that peripheral-tissue-derived DNAm measurements can serve as useful tools for examining the health complications arising from an exposure, as well as potentially acting as a more robust biomarker of exposure in the first place (Corley et al., 2019).

3.7 The case for using DNAm to characterise chronic inflammation

Many of the diseases associated with aberrant DNAm discussed so far share a common trait in that they are all associated with inflammatory processes. As illustrated in Figure 1.5, there are numerous conditions linked to chronic inflammation.
One of the reasons for looking at DNAm in relation to inflammation is that the proportions of different inflammatory mediators may be epigenetically regulated, either directly (e.g. influencing immune cells, or upregulating production of cytokines) or indirectly (e.g. influencing ancillary processes that relate to inflammation, such as genes that have roles in blood vessel formation, tumour pathways, or reducing effectiveness of dampeners of inflammation). Understanding the epigenetic regulation of loci related to circulating inflammatory-protein levels may therefore help to uncover targets for attenuating inflammatory responses, leading to potential therapies for diseases that originate from, or are accompanied by, inflammatory-related dysregulation. Studying differential DNAm in relation to chronic inflammation could also elucidate pathways that link the risk factors that predispose individuals to chronic inflammation and the health outcomes that ensue, such as neurodegeneration and cognitive decline. Finally, if chronic inflammation can be better captured by alterations in the DNA methylome, more accurate profiling of inflammation in individuals could be conducted, with implications for personalised medicine.

Before covering composite epigenetic signatures of inflammation, the epigenetic epidemiological literature on the association of DNAm with inflammation needs to be addressed. This literature can be divided into two main categories: (1) studies that examine associations of DNAm with inflammatory mediators (2) studies that examine DNAm in relation to inflammatory conditions.

3.7.1 DNAm and CRP: Candidate Gene and EWAS studies

As outlined in Chapter 1.5.2, the acute phase protein CRP is one of the most common ways to proxy chronic inflammation. Studies have linked raised levels of CRP to genetic factors, lifestyle factors and health outcomes, but the molecular mechanisms of underlying these relationships are still not well defined. Epigenetic epidemiological research has, however, offered some further clarity on this subject. While epigenetic studies of other inflammatory
proteins have been performed (Ahsan et al., 2017; Hillary et al., 2020; Zaghlool et al., 2020), in this next section I focus on reviewing the current literature of epigenetic regulation of CRP.

The majority of inflammation-specific EWASs find a global trend of hypomethylation at genes involved in regulating inflammation in cases of higher circulating inflammatory protein levels (Ligthart et al., 2016; Gonzalez-Jaramillo et al., 2019; Wielscher et al., 2022). The EWASs to date have implicated over 1,500 differentially methylated CpG sites in regulating blood levels of CRP, corresponding to a host of different genes responsible for regulation of vasculature as well as innate and adaptive immune responses (see Appendix 11.1, eTable 2 for the top CpG sites and relevant genes pertaining to circulating CRP that were used to examine the relationship between DNA methylation (DNAm), inflammation and brain ageing in the empirical Chapters 6 and 7 of this thesis).

Table 3.1 summarises the EWAS studies on serum CRP in population cohorts to date which have highlighted CRP’s putative role in age-related disease outcomes. These studies have been conducted across a variety of populations, with varying degrees of statistical-power, consisting of different background demographics and age-ranges. It should be noted that only two studies to date have examined CRP-related DNA methylation (DNAm) variance in a cohort of children (Guénard et al., 2013; Yeung et al., 2020), with the majority conducted in peripheral blood samples taken from adult cohorts. In studies that examined DNA methylation (DNAm) in African population cohorts, levels of serum CRP levels were notably higher (Cronje et al., 2020; Chilunga et al., 2021), a finding consistent with other epidemiological research on demographic disparities in inflammation levels (literature reviewed in section 1.5.2.4), and potentially related to the higher prevalence of chronic infections in these populations (HIV, hepatitis C) (Escadafal et al., 2020).
<table>
<thead>
<tr>
<th>Author</th>
<th>Outcome</th>
<th>Tissue type</th>
<th>Sample size</th>
<th>Age/ Country</th>
<th>Methylation sites/ method</th>
<th>Adjustment</th>
<th>Main findings</th>
</tr>
</thead>
<tbody>
<tr>
<td>Guineb et al., 2013</td>
<td>CRP</td>
<td>VB</td>
<td>n = 93</td>
<td>12.35±5.77/Canada</td>
<td>Illumina HumanMethylation 450K BeadChip</td>
<td>age + sex</td>
<td>Significant correlations between gene methylation and plasma CRP levels were found for 16 involved in IL-6 signaling (8 showed inverse correlation and 7 positive); 13/16 remained significant after adjustments.</td>
</tr>
<tr>
<td>Sun et al., 2013</td>
<td>CRP</td>
<td>PBL</td>
<td>n = 999</td>
<td>69.27±7.36/USA GENDOA</td>
<td>illumina HumanMethylation 27K BeadChip</td>
<td>age + sex, BMI + smoking</td>
<td>201/4 of 221 CRP-associated CpG sites showed an inverse correlation of greater methylation with lower level of CRP; 24 of the top 30 CpG sites remained significant in both replication subsets with IL-6R, IL-30 and TNF459 as top genes (p=5.8x10⁻¹², p=1.9x10⁻¹⁰ and p=2.0x10⁻⁹, respectively).</td>
</tr>
<tr>
<td>Lighart et al., 2016</td>
<td>CRP</td>
<td>VB</td>
<td>n = 8,883</td>
<td>mean age between 60 and 87/1 Consortia</td>
<td>illumina HumanMethylation 27K and 450K BeadChip, technical covariates + smoking + BMI</td>
<td>age + sex + WBC cell composition + technical covariates + smoking + BMI</td>
<td>218 CpG sites were significantly associated with CRP. Of these, 123 CpG sites were positively associated and 95 were negatively associated. 88 CpG sites, in 67 genes, were still significantly associated in the replication cohort (p&lt;4.11E⁻⁶). The top CpG sites were located in AIMP1 (&lt;−3.32E⁻⁷), P4HB (P = 2.53E⁻¹⁰), 2.09E⁻¹⁰ and 4.57E⁻¹⁰, respectively.</td>
</tr>
<tr>
<td>Mard et al., 2016</td>
<td>CRP</td>
<td>PB</td>
<td>n = 1,741</td>
<td>60.9±8.8/89/Germany</td>
<td>illumina HumanMethylation 450K BeadChip</td>
<td>age + sex, BMI + smoking + WBC composition</td>
<td>4 CpG sites were hypermethylated with elevated CRP (mapping to AQP3, BCL3, SODC3, and intergenic at chromosome 19q13.3). In validation panels, 24 CpG sites remained significant: AIMP1 and SODC3</td>
</tr>
<tr>
<td>Ride et al., 2019</td>
<td>CRP and 160 protein levels from the Otaki Oncology II and Otaki Inflammation panels</td>
<td>PBL</td>
<td>n = 152</td>
<td>mean age; 50 (reverted at 56) Valdeorobor Intervention Programme</td>
<td>Illumina HumanMethylation Epic BeadChips</td>
<td>age + sex + sex + smoking + WBC composition</td>
<td>44 CpG sites showed a consistent direction of association as had been previously reported by Lighart et al. (2016)</td>
</tr>
</tbody>
</table>
Table 3.1 Epigenome Wide Association Studies of C-Reactive Protein

<table>
<thead>
<tr>
<th>Study</th>
<th>CRP</th>
<th>Blood Type</th>
<th>n</th>
<th>Mean Age ± SD</th>
<th>Study Type</th>
<th>Age-Related Variables</th>
<th>Controls/Replication</th>
<th>CRP Sites Associated with CRP</th>
</tr>
</thead>
<tbody>
<tr>
<td>Young et al., 2020</td>
<td>CRP</td>
<td>Cord blood</td>
<td>392</td>
<td>Mean age &lt;1 year</td>
<td>EWAS</td>
<td>Illumina Infinium MethylationEPIC</td>
<td>Maternal age + smoking + insurance + pre-pregnancy BMI + platelet counts</td>
<td>7 Cpg sites associated with CRP</td>
</tr>
<tr>
<td>Corigli et al., 2020</td>
<td>CRP</td>
<td>WB</td>
<td>120</td>
<td>Mean age 54 range (36-76)</td>
<td>EWAS</td>
<td>Illumina Infinium MethylationEPIC</td>
<td>Age + BMI + cell counts + smoking</td>
<td>no genome-wide significant or novel CRP-DNAm associations reported</td>
</tr>
<tr>
<td>Chiang et al., 2021</td>
<td>CRP</td>
<td>WB</td>
<td>560</td>
<td>Mean age: 51 ± 18 years</td>
<td>EWAS</td>
<td>Illumina HumanMethylation450K</td>
<td>Age + sex + alcohol consumption + smoking + estimated cell types + technical effects (hybridization batch and array position)</td>
<td>14 Cpg sites showed genome-wide significant associations with CRP. Top hits were in the genes PCG (cg14593260: P = 2.55x10^-14), FAM124B (cg122558447: P = 8.05x10^-15) and DNAJC23 (cg04795131: P = 1.35x10^-13). Only 3 of the 14 sites have been previously linked to inflammation and immune functioning (CD81 and LRRRC14 genes). In a sensitivity analysis of participants without evidence of acute inflammation (serum CRP &gt; 15mg/L), only DNA levels at one Cpg site (cg06511882 in TSS1502 of MORC2 gene) showed genome-wide significant associations with CRP</td>
</tr>
<tr>
<td>Aldrich et al., 2022</td>
<td>CRP</td>
<td>WB</td>
<td>23,774</td>
<td>Mean age — varies by cohort (36 – 76 years)</td>
<td>EWAS</td>
<td>Illumina HumanMethylation450K</td>
<td>Age + sex + estimated blood cell count + technical covariates; BMI; sensitivity analysis</td>
<td>1,511 independent uncorrected loci passed genome-wide significance, 729 of which were considered a potential consequence of altered blood CRP levels. Of these 706, 218 of the sites identified by Liang et al. (2016) were replicated. MR analysis additionally found that 6 of these loci (cg0111103, cg0366660, cg06470551, cg14708331, cg00039639, cg17069616, cg00318407) were causally linked to increased blood CRP levels.</td>
</tr>
</tbody>
</table>

One EWAS has particular relevance to the work presented in the empirical chapters of this thesis. Ligthart and colleagues examined the association of circulating CRP and DNAm in relation to health outcomes, using whole blood samples from 12,974 individuals. They found 218 CpG sites that were associated with circulating CRP levels in a validation cohort (Ligthart et al., 2016). Of these 218 CpG sites, 56 of which were then replicated in a longitudinal cohort that examined CRP levels in individuals at 12.974 individuals. They found 218 CpG sites that were associated with circulating CRP and DNAm in relation to health outcomes, using whole blood samples from 12,974 individuals. They found 218 CpG sites that were associated with circulating CRP levels in a validation cohort (Ligthart et al., 2016). Of these 218 CpG sites, 56 of which were then replicated in a longitudinal cohort that examined CRP levels in individuals at 12.974 individuals.
time points 10 years apart (Myte et al., 2019) and 57 were used in the most recent multi-cohort CRP EWAS of 22,774 individuals (Wielscher et al., 2022), which found 1,511 CpG sites associated with serum CRP.

In Ligthart’s 2016 study, the top differentially methylated regions related to circulating CRP levels pertained to the genes AIM2 and SOCS3. AIM2 (Absent in Melanoma 2) is an inflammasome receptor (Rottenberg and Carow, 2014) from the interferon family, upstream of CRP, and is involved in the processing of IL1β (Hornung et al., 2009) that is associated with numerous chronic and acute inflammatory conditions (Dinarello, 2011). Hypomethylation at AIM2 had been previously linked to higher circulating CRP levels in a candidate gene study (Miller et al., 2018), backing up the relevance of this site in the regulation of inflammation. Alongside Myte et al., (2019)’s replication of differential methylation at this site, a more recent study of 22,000 participants also replicated Ligthart et al., (2016)’s work, finding that increased DNAm at AIM2 (downregulation of AIM2 gene expression) was linked to lower serum CRP levels, and additionally concluded that this effect is predominantly driven by higher BMI (where BMI → CRP → DNAm; supplementary data 9 from (Wielscher et al., 2022)).

Hypomethylation at the other top gene hit, SOCS3, was also reported in four of these EWAS studies (Ligthart et al., 2016; Marzi et al., 2016; Myte et al., 2019; Wielscher et al., 2022), and remained significant in external validation cohorts in three of these (Marzi et al., 2016; Myte et al., 2019; Wielscher et al., 2022). In Wielscher’s study (2022), this association was similarly explained by the BMI → CRP → DNAm mediation model. SOCS3 (suppressor of cytokine signalling 3), is considered part of the JAK/STAT signalling cascade and can aggravate inflammation, with differential methylation of SOCS3 associated with inflammation in cases of cancer, atherosclerosis and obesity (Carow and Rottenberg, 2014; Wang et al., 2014, p. 3; Chen et al., 2015; Xu et al., 2018, p. 3).
Though most of these EWASs are cross-sectional, and offer little insight into the directional causality of the DNAm ~ CRP relationship (Walton et al., 2018), a few recent studies have found putative causal relationships between inflammatory-related DNA methylation (DNAm) and the inflammatory-related outcome measures of BMI and Chron’s disease (Wahl et al., 2017; Somineni et al., 2019). For CRP specifically, Wielscher et al. (2022) used 2-sample Mendelian Randomization (MR) approach to find that 8 loci (cg04111102, cg14099685, cg26470501, cg14702231, cg02039839, cg17580616, cg00138407) were causally linked to increased blood CRP levels, a topic discussed later in section 9.7.9.

In the multi-ethnic EWASs of CRP (Ligthart et al., 2016; Wielscher et al., 2022), many of the replicated CpG sites reported were associated with lifestyle factors related to inflammation (e.g. BMI and smoking) as well as health outcomes (type 2 diabetes, obesity and CVD) reinforcing the widespread network of epigenetic regulation of inflammation across various conditions and pleiotropic epigenetic network across various phenotypes. The sites identified in these EWASs have since been linked to an array of inflammatory related morbidities including alcoholism (Scholl et al., 2022), COPD (Lee et al., 2022) and other clinically-relevant phenotypes (Zaghlool et al., 2020). Equally, CRP and related DNA methylation differences have been shown to associated with accelerated epigenetic age measures (Verschoor et al., 2018) (a concept discussed further in section 3.8.1). Combined, this literature provides further support for the ‘inflammaging’ hypothesis and indicates that DNA methylation may be a useful tool for tracking trends towards sustained inflammatory states with advancing age and the related increased risk of age-related disease.

Overall, the interrelationships between inflammatory-risk factors, CRP, DNA methylation and disease outcomes identified in these EWASs point towards shared epigenetic modifications with raised inflammatory mediators that may underscore differences in health outcomes. This provides strong rationale to investigate DNA methylation and inflammation in relation to further clinical outcomes.
with potential inflammatory pathophysiology, such as brain and cognitive
ageing.

3.8 DNAm signatures (methylation risk scores)

Studies are beginning to create composite measures capturing epigenetic
variation, one example being DNAm signatures which aggregate genome-
wide epigenetic information pertaining to a specific trait or exposure of
interest.

This approach is a development of a standard practice in genetic research:
the use of polygenic risk scores (PRS) to predict health and disease risk,
based on the combined effects of multiple genetic variants across the
geno
me. PRS are made up of individual’s risk alleles of a pre-selected
number of single nucleotide polymorphisms (SNPs), which are then weighted
and summed. Weights are typically estimated from a GWAS for the trait of
interest in an independent discovery sample (Hüls et al., 2017). For complex
conditions which are highly polygenic, PRS are an attractive approach to
examine an individual’s susceptibility to a particular phenotype, as many
causal SNPs individually have small effect sizes that fail to reach genome
wide significance (Martin et al., 2019). With the advent of multiplexed
detection of large numbers of CpG sites via microarray techniques, the same
approach has now been adapted to DNAm data to create methylation risk
scores (MRS), which I refer to throughout this thesis as DNAm signatures
(Hüls and Czamara, 2020).

These can be made by quantifying methylation levels at relevant CpG sites
(see Figure 3.3 for approaches for selecting sites of relevance), multiplying
the methylation value at these given CpGs by the effect size from either a
previous EWAS or training sample set, and then summing the values (where
the linear combination of CpG sites provides an estimate for the trait /
exposure of interest). Further details on different ways CpG sites are
selected to build DNAm signatures is covered in the methodology section of this thesis (Chapter 5), including an overview of the popular penalised regression methods used to identify CpG sites related to a particular trait.

DNAm signatures offer an advantage over PRSs in their shared genetic and environmental regulation. If an endogenous or exogenous factor results in an epigenetic modification, but then desists, certain epigenetic markers will return to their previous state (Eckstein, Rea and Fondufe-Mittendorf, 2017). In other instances, epigenetic marks will persist as a marker of historical exposure, particularly if that exposure was chronic (Relton, Hartwig and Davey Smith, 2015). As emphasised in section 3.6, this property of epigenetic modifications is a particularly attractive one from an epidemiological perspective as it may help to index a history of exposure of certain experiences and environmental factors.

In certain circumstances proxying an exposure with a DNAm signature is very useful. Let’s take the example of smoking: if DNAm changes accrue with increased smoking exposure (and attenuate with cessation or less exposure), DNAm signatures could act as useful biomarkers from which to both classify exposure levels and examine associated health risks. This is particularly relevant for lifestyle traits and factors that are hard to reliably estimate (largely self-reported, or if other proxied measures, e.g. cotinine for smoking or alanine transaminase (ALT) for alcohol consumption, are also less than ideal) and complex environmental exposures that have phasic biological correlates (such as inflammatory proteins), as well as instances where direct analogues are difficult to measure (such as during pregnancy) or unethical (venepuncture in preterm infants).

Figure 3.3 below demonstrates the process of determining appropriate weights for DNAm signatures (Bakulski and Fallin, 2014; Odintsova et al., 2021).
Flow chart for appropriate methods for creating DNAm signatures

DNAm signatures are created by taking a weighted sum of the methylation levels at CpG sites pertinent to the trait or exposure of interest. This is done by first examining observed CpG associations in a discovery cohort, before then validating the composite in a second independent cohort. From here, the estimated proportion of variance in the exposure or trait of interest accounted for by these DNA-methylation profiles can be calculated, as well as their associations with other outcomes of interest such as health and disease (the methodological considerations of this approach, to generate inflammatory-related DNAm signatures, will be discussed in further depth in Chapter 5 Study Cohorts and Methods).

These DNAm-based predictors of health, disease and exposure are now rising, with many demonstrating their additive potential to explain outcomes compared against baseline models with clinical or demographic factors, PRSs, or even against phenotypic measures of the trait/exposure of interest themselves (McCartney, Hillary, et al., 2018; Corley et al., 2019; Hamilton et al., 2019; McCartney et al., 2022).

The most well-known example of this prediction model approach are epigenetic clocks, which will be discussed next.
3.8.1 DNAm signatures of age (epigenetic clocks)

The field of epigenetic epidemiology experienced a significant upsurge in interest after finding that analysis of inter-individual variation in DNAm alone could reliably predict chronological age (Horvath, 2013). This led to the distinction between biological and chronological ageing, and the idea that patterns of DNAm variation could reveal divergences from healthy ageing trajectories or signpost disease risk.

Epigenetic clocks are proxies of chronological age built by finding a set of CpG sites that are robustly associated with ageing. They are typically generated by taking the linear combination of weighted methylation levels for certain CpG sites via penalised regression models, examples of which include Ridge, least absolute shrinkage selection operator (LASSO) or elastic net (further details on these approaches in relation to DNAm-based predictors of exposures and traits is expanded on in Chapter 5.3.6). These methods allow for selection of key features that optimally predict chronological age, and have been replicated across populations, tissues, and associate with the risk of numerous age-related disease outcomes (Hannum et al., 2013; Levine et al., 2015; Hillary et al., 2019).

The residuals from a linear regression model of epigenetic age on chronological age (i.e., the discrepancy between ‘epigenetic’ vs phenotypic ageing) are often referred to as a measure of ‘biological’ age acceleration, on a parallel to what was described in Chapter 2.6.2 for neuroimaging estimates of ‘brain-age’. Accelerated biological age as estimated by these measures is associated with mortality and morbidity, including brain-health outcomes (Cardenas et al., 2022), and several studies have demonstrated that low-level chronic inflammation is associated with increased rates of epigenetic age (Sundermann et al., 2019; Zhu et al., 2021). The development of these predictors of ageing (Table 3.2 summarises the most popular epigenetic clocks) has brought into question the concept of ‘healthspan’ over ‘lifespan’, with the most recent and reproducible epigenetic age predictors including the integration of multiple clinical measures of phenotypic ageing (Levine et al.,
A discussion of the utility and accuracy of these clocks is neatly reviewed by Bell et al. (2019), including a commentary on the specific challenges of this evolving field and recommendations for further development.

### Table 3.2 Epigenetic clocks

<table>
<thead>
<tr>
<th>Reference</th>
<th>Epigenetic clock</th>
<th>Number of CpG sites</th>
<th>Tissue-type</th>
<th>Prediction (outcome of interest)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Hannum et al., (2013)</td>
<td>Hannum</td>
<td>71</td>
<td>Blood</td>
<td>Biological age</td>
</tr>
<tr>
<td>Horvath et al., (2013)</td>
<td>DNAmAge</td>
<td>353</td>
<td>Multi-tissue</td>
<td>Biological age</td>
</tr>
<tr>
<td>Levine et al. (2018)</td>
<td>PhenoAge</td>
<td>513</td>
<td>Multi-tissue</td>
<td>Morbidity and mortality</td>
</tr>
<tr>
<td>Lu et al., (2019)</td>
<td>GrimAge</td>
<td>1030</td>
<td>Blood</td>
<td>Biological age (lifespan) and healthspan</td>
</tr>
</tbody>
</table>

*Note this is not an exhaustive list of epigenetic clocks to date and is shown to provide an overview of DNA methylation-based predictors of age.*

Table 3.3 below provides a brief overview of different DNA methylation-based exposures, including inflammation, and the methods (EWAS, candidate gene, LASSO) used to build them.

#### 3.8.2 DNAm signatures of exposures and risk factors

DNAm signatures can also be used to provide proxy measures of various exposures. There are several advantages of this approach: (1) the potential of DNAm signatures to overcome the limitations of self-reported or incomplete data – something particularly relevant for stigmatised lifestyle habits such as smoking and drinking, where underreporting is common (2) DNAm is a continuous measure (from 0-100%), meaning that categorical phenotypes (such as disease status, smoker/non-smoker status, alcoholic consumption status; binary classification of “high” vs “low” inflammation) can be represented as continuous outcome measures, affording greater granular
indexing of certain exposures. Recent examples of DNAm signatures that appear to account for degree of exposure include a DNAm classifier that accurately distinguished between current, former and never smokers (Langdon et al., 2021) (discovery n = 1063; validation cohort n = 717); a DNAm signature that distinguished heavy alcohol consumption over moderate consumption (Liu et al., 2018) and a blood-based DNAm signature of cumulative lead exposure (Colicino et al., 2021).

To date, the most well-characterised composite methylation biomarker of a lifestyle factor / environmental exposure is smoking. As introduced in section 3.6.2, extensive research examining smoke exposure, DNAm and related health and disease outcomes has been conducted in adult cohorts where DNAm signatures of smoking index both direct and indirect exposure to cigarette smoke, but also correlate with the cumulative dose of smoking and time since quitting smoking (Elliott et al., 2014; Wilson et al., 2017; Sugden et al., 2019). DNAm signatures of smoking have leveraged to classify the degree to which people smoke, the risk of health outcomes related to smoking, and indirect exposure to smoke. A summary of recent DNAm signatures used to index smoking behaviour, including the number of CpG sites used in the composite score and method used to aggregate this information, is provided in Table 3.3.

<table>
<thead>
<tr>
<th>Reference</th>
<th>Number of CpG sites</th>
<th>Tissue-type</th>
<th>Prediction (outcome of interest)</th>
<th>Methodology to derive DNAm signature</th>
</tr>
</thead>
<tbody>
<tr>
<td>(Elliott et al., 2014)</td>
<td>187 CpG sites</td>
<td>Peripheral blood</td>
<td>To distinguish current smokers from non-smokers (former, never)</td>
<td>Summed weighted DNAm score, weights derived from a previous EWAS (Zeilinger et al., 2013) (n = 1793 discovery cohort; n = 479 validation cohort)</td>
</tr>
<tr>
<td>(Bierut et al., 2008)</td>
<td>4 CpG sites</td>
<td>Peripheral blood</td>
<td>To distinguish current from never smokers, and former from never smokers</td>
<td>EWAS followed by stepwise logistic regression with forward selection</td>
</tr>
<tr>
<td>(McCartney, Stevenson, et al., 2018)</td>
<td>5 - 90 CpG sites</td>
<td>Peripheral blood</td>
<td>To distinguish smokers from non-smokers and time</td>
<td>Summed weighted DNAm score, weights derived from a</td>
</tr>
</tbody>
</table>
This study examined 4 separate DNAm signatures of smoking behaviour in adults. They used 4 CpGs for the candidate-gene “current smokers vs former smokers” classifier; 9 CpGs for candidate-gene “ever smokers vs never smokers” classifier; 13 CpGs for weighted score where weights were derived by previous EWAS by Maas et al. (2019); 29 CpGs used for LASSO-derived “ever smokers vs never smokers”; 21 CpGs for agonist LASSO “current vs former smokers” model.

This study examined 3 separate DNAm signatures of prenatal smoking. The first was in older children, where a DNAm score was derived from 19 CpG sites. The second was prediction of prenatal smoking (in newborns) = DNAm score derived from 568 CpG sites; both of the 19 and 568 scores used weights from the EWAS by Joubert et al. (2016) which used cord blood DNA methylation values of CpGs. The third was prediction of smoking (adults) = DNAm score derived from 2,623 CpG sites, based of EWAS by Joehanes et al (2016). 12 sites overlapped across all 3 scores.

<table>
<thead>
<tr>
<th>Study (year)</th>
<th>CpG Sites</th>
<th>Sample Type</th>
<th>Smoking Status</th>
<th>Methodology</th>
</tr>
</thead>
<tbody>
<tr>
<td>Langdon et al., 2021</td>
<td>4 – 29 CpGs</td>
<td>Peripheral blood</td>
<td>To distinguish between current, former and never smokers and assessment of previous smoking</td>
<td>Summed weighted DNAm score, weights derived from a previous EWAS by Maas et al. (2019); (6 populations, N = 3764); plus a candidate gene approach following LASSO, plus weighted scores from a LASSO model, as well LASSO supplied with genome-wide 450K data (agnostic LASSO model)</td>
</tr>
<tr>
<td>Reese et al., 2017</td>
<td>28 CpG sites</td>
<td>Cord blood</td>
<td>Prediction of prenatal smoke exposure</td>
<td>EWAS followed by LASSO regression</td>
</tr>
<tr>
<td>Richmond et al., 2018</td>
<td>19 – 2,623 CpG sites</td>
<td>Peripheral blood</td>
<td>Prediction of prenatal smoke exposure</td>
<td>Weighted DNAm score using methylation values of CpGs identified by newborn cord blood DNA methylation values of CpGs based EWAS (Joubert et al., 2016) (n = 9389) and adult score informed by EWAS by Joehanes et al. (2016)</td>
</tr>
<tr>
<td>Sugden et al., 2019</td>
<td>2,623 CpG sites</td>
<td>Peripheral blood</td>
<td>Out of sample performance prediction of smoking DNAm score in two longitudinal cohorts</td>
<td>Summed weighted DNAm score, weights derived from a previous EWAS by Joehanes et al. (2016)</td>
</tr>
<tr>
<td>Rauschert et al., 2019, 2020</td>
<td>204 CpGs</td>
<td>Peripheral blood</td>
<td>Prediction of prenatal smoke exposure</td>
<td>EWAS followed by elastic net regression</td>
</tr>
<tr>
<td>Blostein et al., 2022</td>
<td>6,073 CpG sites</td>
<td>Saliva</td>
<td>Out of sample performance prediction of</td>
<td>Weighted DNAm score using methylation values of CpGs [5,666]</td>
</tr>
</tbody>
</table>

25 This study examined 4 separate DNAm signatures of smoking behaviour in adults. They used 4 CpGs for the candidate-gene “current smokers vs former smokers” classifier; 9 CpGs for candidate-gene “ever smokers vs never smokers” classifier; 13 CpGs for weighted score where weights were derived by previous EWAS by Maas et al. (2019); 29 CpGs used for LASSO-derived “ever smokers vs never smokers”; 21 CpGs for agonist LASSO “current vs former smokers” model.

26 This study examined 3 separate DNAm signatures of prenatal smoking. The first was in older children, where a DNAm score was derived from 19 CpG sites. The second was prediction of prenatal smoking (in newborns) = DNAm score derived from 568 CpG sites; both of the 19 and 568 scores used weights from the EWAS by Joubert et al. (2016) which used cord blood DNA methylation values of CpGs. The third was prediction of smoking (adults) = DNAm score derived from 2,623 CpG sites, based of EWAS by Joehanes et al (2016). 12 sites overlapped across all 3 scores.

27 Outperforming the previously developed score by Reese et al., (2017).
Various prenatal smoke exposure DNAm signatures including those by (Reese et al., 2017; Richmond et al., 2018; Rauschert et al., 2020) identified by an earlier genome-wide consortium meta-analysis of newborn cord blood DNAm (Joubert et al., 2016)---

<table>
<thead>
<tr>
<th>Alcohol</th>
<th>(Liu et al., 2018)</th>
<th>114 CpG sites</th>
<th>Peripheral blood</th>
<th>Prediction of heavy alcohol consumption(^{28})</th>
<th>EWAS followed by LASSO regression</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>(Yousefi et al., 2019)</td>
<td>114 CpG sites</td>
<td>Peripheral blood</td>
<td>Out-of-sample performance(^{29}) prediction of DNAm alcohol signature for drinking at different time points across the lifecourse</td>
<td>Weighted DNAm score (DNAm-Alc) using coefficients made available from the lasso models estimated by (Liu et al., 2018)</td>
</tr>
<tr>
<td></td>
<td>(Dugué et al., 2021)</td>
<td>114 CpG sites</td>
<td>Peripheral blood</td>
<td>Out-of-sample performance prediction of how DNAm alcohol signature associates with cancer incidence.</td>
<td>Weighted DNAm score (DNAm-Alc) using coefficients made available from the lasso models estimated by (Liu et al., 2018)</td>
</tr>
</tbody>
</table>

| Inflammation | (Barker et al., 2018) | 7 CpG sites | Peripheral blood and cord blood | To predict inflammatory burden (CRP) and examine its association with adolescent mental health outcomes | Summed weighted DNAm score (DNAm CRP), weights derived from a previous EWAS (Ligthart et al., 2016) (n = 8,863 discovery sample; n = 4,111 validation cohort)\(^{30}\) |

---

\(^{28}\) heavy alcohol consumption coded as \(\geq 42\) g per day in men and \(\geq 28\) g per day in women. Defined by self-reported measures of consumption of alcoholic beverages consumption in the Health and Lifestyle Questionnaire, which asks about weekly frequency and amount of alcohol consumption of different alcoholic drinks in the past year. This data was then converted into units per week, and then converted to grams of ethanol per day (where one unit of alcohol in the UK is defined as 7.9g).

\(^{29}\) Validation of DNAm signature of alcohol consumption derived by (Liu et al., 2018) in two independent population cohorts (ALSPAC-ARIES; n = 1049, HN5000 cohort: N = 281, mean age = 58.4 ± 9.9 SD) to phenotypic (self-report) alcohol consumption, plus comparison with Alcohol Use Disorders Identification Test (AUDIT) score (a scale of alcohol use disorder)\(^{30}\)

\(^{30}\) CpG site selection was based on the CRP-EWAS findings from Ligthart et al. (2016), where 58 CpG sites were found to associate with plasma CRP levels in both a discovery meta-analysis (9 cohorts, n = 8,863) and a replication meta-analysis (4 cohorts, n = 4,111). Of these 58, those that associated with both (a) whole-blood gene expression levels; and (b) at least one cardiometabolic phenotype considered to be of relevance to CRP (e.g., body mass index, coronary heart disease, etc.) were used to stringently select a small number of CpG sites most predictive of inflammatory-burden.
<table>
<thead>
<tr>
<th>Authors</th>
<th>CpG sites</th>
<th>Sample Type</th>
<th>Methodology Description</th>
<th>Summed weighed DNAm score (DNAm CRP), weights derived from a previous EWAS (Ligthart et al., 2016)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Stevenson et al., 2020</td>
<td>7 CpG sites</td>
<td>Whole blood</td>
<td>Out-of-sample performance prediction of how DNAm CRP signature associates with cognitive ability.</td>
<td>(n = 8,863 discovery sample; n = 4,111 validation cohort)</td>
</tr>
<tr>
<td>Stevenson et al., 2021</td>
<td>35 CpG sites</td>
<td>Whole blood</td>
<td>Creation of DNA methylation score for circulating IL6 and association with cognitive ability</td>
<td>EWAS followed by Elastic net</td>
</tr>
<tr>
<td>Zhao et al., 2021</td>
<td>54 CpG sites(^{31})</td>
<td>Whole blood</td>
<td>Out-of-sample performance prediction of how DNAm CRP signature associates with risk of lung cancer-specific death.</td>
<td>Summed weighted DNAm score (DNAm CRP), weights derived from a previous EWAS by Ligthart et al. (2016)</td>
</tr>
<tr>
<td>Castagné et al., 2020</td>
<td>61 CpG sites</td>
<td>Whole blood</td>
<td>Assessment of DNAm signature of inflammation and socioeconomic status across the lifecourse</td>
<td>Summed 61 CpG sites (derived via candidate gene approach) normalised z-scores to generate a composite inflammatory methylome z-score</td>
</tr>
<tr>
<td>Wielscher et al., 2022</td>
<td>1511 CpG sites</td>
<td>Whole blood</td>
<td>Investigated the impact of CRP-associated CpGs on clinically relevant phenotypes (BMI, COPD)</td>
<td>Calculated a beta weighted risk score using the coefficients from the multi-ethnic discovery analysis</td>
</tr>
</tbody>
</table>

**Table 3.3 DNAm signatures of exposure (smoking, alcohol consumption and inflammation)**

*Table displays DNAm-based predictors of various exposures, the number of CpG sites used to generate a DNAm signature, and the methodology used to generate composite DNAm signature.*

Epigenetic changes post exposure appear to be dynamic and dependent on the level of dose experienced. In McCartney et al., (2018)’s study, the degree to which someone smoked previously was found to affect the persistence of DNAm changes following cessation. In people who used to smoke, but were

---

\(^{31}\) 54/ 58 CpG sites Ligthart et al. (2016) found to associate with circulating CRP levels, 54 CpGs were used here as 4 were not available on the 850 K array used by this study.
considered ‘light’ smokers (classified by number of cigarettes per day), within a year of stopping, their DNAm profiles resembled that of non-smokers. By contrast, in individuals who were considered heavy smokers before quitting, it took up to 9 years for DNAm profiles to revert to that of non-smokers. These findings illustrate well how DNAm signatures may offer an augmented means to characterise cumulative dose of exposure with finer granularity.

Related to this, DNAm signatures have also been used to examine exposures and experiences in the perinatal period and their impacts on offspring. Because smoking is stigmatised in pregnancy, the reliability of self-reported measures has been called into question. Recent epigenomic studies have looked at how DNAm-proxied smoke exposure influences health in pregnancy, the persistence of DNAm signatures postpartum both for mothers and infants in indexing prenatal smoke exposure (Blostein et al., 2022), and how DNAm signatures of prenatal smoke exposure relate to risk of health outcomes postpartum for both parents and infants (Joubert et al., 2016; Reese et al., 2017; Hannon et al., 2019; Rauschert et al., 2019). The best performing DNAm signature of prenatal smoke exposure in pregnancy to date comes from (Rauschert et al., 2020) who built a DNAm signature from 204 smoking-related CpG sites using an elastic net approach. This DNAm signature outperformed the previous prenatal-smoking DNAm signature derived by (Reese et al., 2017), who by contrast used LASSO following EWAS approach, and the top identified CpG sites of this signature were also hits in previous findings of prenatal smoke exposure with DNAm (Joubert et al., 2016; Rzehak et al., 2016; Richmond et al., 2018). However, a notable limitation of Rauschert’s study design is that they did not control for blood cell counts in their models, a practice somewhat contested in epigenomic research (see Chapter 9 for further discussion). Overall, the results of these studies suggest that maternal exposures (such as smoking) are associated with changes in DNA methylation that persist postnatally both in mothers and in offspring.
This examination of DNAm signatures in mother-infant cohorts has led to resurgence in research into how perinatal experiences and exposures influence offspring’s susceptibility to disease across the lifecourse, including brain health outcomes. In addition to smoking, DNAm signatures of nutritional exposures, such as prenatal folate exposure during pregnancy, have been investigated in cord blood (Bakulski et al. 2020; Joubert et al. 2016a), as well as environmental exposures such as pollution (Bakulski et al., 2021), prenatal alcohol consumption (Abrishamcar et al., 2022) and prenatal stress (Cao-Lei et al., 2020) and susceptibility to allergy in childhood (Kilanowski et al., 2022).

I outline the above as evidence for how, and why, DNAm has been used as an index of prior exposure. Examining composite biomarkers of inflammation, however, are rare, despite this being an ideal exposure to model from multiple sources owing to the complex nature of the inflammatory response and the limitations of assessing inflammation in utero.

3.8.3 DNAm signatures of inflammation

As outlined in section 3.8.3, EWAS and penalised regression approaches have been leveraged to predict protein levels in the systemic circulation. In Chapter 1.5.3, I discussed how sampling the levels of single inflammatory mediators may lead to misclassifications of baseline inflammation levels in population cohorts. This is particularly relevant for acute-phase reactants like CRP which exhibit substantial variation in their concentration levels within short timespans. Because of this, I emphasise the importance of alternative approaches to measuring chronic inflammation in population cohorts. Though the approach of composite metrics (literature reviewed in Chapter 1.5.5) that use aggregate measures of multiple inflammatory-related proteins has been applied to risk-prediction, the concept of leveraging DNAm data across multiple CpG sites that are robustly associated with inflammation to index baseline inflammation levels is more nascent. As a result, only a few studies to date have built DNAm-based predictors of inflammation and examined
them in relation to incident health outcomes. These are outlined in Table 3.3, alongside DNAm-based signatures of smoking and alcohol consumption.

Fewer studies still have extended this paradigm to investigate more disparate consequences of chronic inflammation such as those concerning the brain (brain structural differences, psychiatric phenotypes, cognitive decline and neurodegenerative diagnoses). However, there is evidence supporting that this may be a valid method of enquiry, given that one of these inflammatory-related DNAm signatures, built from CpG sites that were associated with circulating CRP levels in Ligthart et al. (2016)’s multi-ethnic EWAS (detailed in Chapter 5 of this thesis, and illustrated in Figure 5.5), exhibited larger associations with early life mental health outcomes (Barker et al., 2018), and in a cohort of elderly adults, with cognitive ability (Stevenson et al., 2020) than serum CRP levels themselves. In Chapter 6 of this thesis, this signature is examined in relation to structural neuroimaging phenotypes for the first time to attempt to explain the association observed between DNAm CRP and cognitive function (Conole et al., 2021), and in Chapter 7 it is examined in relational to neonatal neuroimaging phenotypes and postnatal morbidities (Conole et al., 2023). I also discuss the validation of this signature a cohort of individuals in midlife (Green, Shen, et al., 2021) in the context of neuroimaging outcomes and assess its utility for signposting cases of depression. Chapter 8 extends this pipeline to investigate over 100 DNAm signatures (built via elastic net) with multi-modal neuroimaging metrics and other aspects of brain and cognitive ageing.

The DNAm CRP signature examined throughout these chapters has also recently been applied to demonstrate potential of early-life inflammation in anticipating risk of later life mental health symptomatology (Edmondson-Stait et al., 2022, p.) and brain-age outcomes in LBC1936 (Green, Squillace, et al., 2021). Of interest, a recent DNAm signature of 44 CpG sites was built for assessing COVID-19 severity (Castro de Moura et al., 2021) via an EWAS followed by elastic-net approach, able to distinguish between patients with mild symptomology, patients who’d received oxygen supplementation, and
ICU-bound patients with respiratory failure; two of the main loci in this signature are part of the DNAm CRP score used in Chapters 6 and 7.

These studies demonstrate how DNAm signatures can reflect variance in the circulating plasma proteome, and increasing work is focusing on how DNAm proxies of proteins relate to health and disease (Zaghlool et al., 2020; Danni A Gadd et al., 2022; Danni A. Gadd et al., 2022). In the empirical chapters of this thesis, I investigate the hypothesis of whether DNAm signature of inflammatory-related proteins can reveal previously missed influence of chronic inflammation on aspects of brain and cognitive ageing.

3.8.4 Application of DNAm signatures to examine brain health outcomes

Various exposures and experiences may compromise health by way of epigenetic modifications. In this way, DNAm signatures of exposures (e.g. alcohol, smoking or inflammation) could provide insight into the relative risk of an associated disease or phenotype and also a means to profile, track, and assess chance of reversion of phenotypic manifestations of these changes in line which changes in lifestyle, treatment or health. As highlighted, they may also provide a good account of cumulative dose of past exposure (such as heavy drinking or heavy smoking), which is a useful index when assessing disease susceptibility. Because of this, studies are turning towards using validated signatures of various exposures to anticipate risk or explain variance in other aspects of health. For alcohol and smoking, DNAm signatures have been used to anticipate CVD risk (Chamberlain et al., 2022).

This is particularly relevant for assessing the impact of chronic inflammation on aspects of health. While the level of inflammation (e.g. binary high or low classification of inflammatory protein levels, or continuous assessment of inflammatory protein levels) has been previously used to explore risk of cognitive decline or MRI correlates, what is more relevant from a disease pathophysiology perspective is the extent of damage incurred at a molecular
level from elevated inflammatory mediators. An epigenetic surrogate (that accrues information about DNAm levels across different inflammatory-related sites of the genome) may be better placed to capture this. This feature is likely of more direct relevance to the development of deviations in brain health trajectories. This measure may also vary across people depending on shared genetic and environmental factors, so may be a more informative tool for examining aspects of brain health and explain variance in hallmarks such as white matter loss, reduced white matter integrity and other neuroimaging aberrations.
3.9 Summary

This chapter introduced one of the main tools used to characterise chronic inflammation in this thesis: DNAm. I outlined how DNAm lies at the interface between the genome, environment, and disease, and as such may be investigated as a direct mechanism of disease aetiology, a modifier of disease risk, and a biomarker of exposure or disease outcome. I discussed the key findings from the field of epigenetic epidemiology and outlined that a persistent theme among diseases associated with DNAm differences, including those related to brain health outcomes, is chronic inflammation.

I introduced the concept of DNAm signatures, which are an approach to assimilate epigenome-wide information into a single value. I discussed how DNAm signatures, like PRSs, can be used to proxy exposures or disease / phenotypes. I evaluated the progress in DNAm-based predictors of health and disease, comparing examples of DNAm signatures of lifestyle exposures (such as smoking and alcohol consumption) and disease risk (such as CVD, brain health outcomes and various cancer types). I also discussed how validated DNAm signatures of exposure have then been applied to out-of-sample population cohorts to additionally examine incident risk related health outcomes. I discussed how while previous studies have investigated differential methylation in relation to inflammation and inflammatory diseases, very few studies attempt to aggregate these results into a direct predictor of inflammatory burden and related disease-risk. Of those that do, few extend this paradigm to examine brain health outcomes.

The next Chapter summarises the main concepts introduced in Chapters 1-3 that have informed the rationale for this thesis. I conclude by demarcating the gaps in the literature that this thesis sets out to address.
4 Research gaps and aims of thesis

4.1 Introduction

Chronic low-grade inflammation is known to be involved in multiple conditions, including brain health phenotypes, across the lifecourse. However, the traditional approaches used to measure inflammation have drawbacks in population epidemiological research contexts, and studies that examine the association of inflammation and brain health are held back by their limited characterisation of brain morphology and cognitive ability, and often omit key confounders of these relationships in their analyses.

Related to this (and potentially because of these factors), evidence linking inflammation levels to brain health outcomes from studies to date is often conflicting. There is a lack of consensus surrounding how peripheral inflammation relates to brain structure and function with age, with inconsistent evidence for causality, or clarity on how elements of health and lifestyle influence this association. Further exploration of the epigenetic architecture and downstream molecular pathways of chronic low-grade inflammation is therefore warranted, with integrated assessment of lifestyle, cognitive and neuroimaging data and novel approaches to indexing inflammatory burden. Moreover, examination of the relationship between inflammation and brain health needs to be conducted across the lifecourse to better characterise this relationship. Deeper phenotyping and integration of such data across different ages could enable greater understanding of system-level immunological alterations, elucidation of both focal and broad putative immune-related pathophysiologic mechanisms that relate to brain structural differences, and signpost those at risk of certain phenotypes such as neurodevelopmental impairments or accelerated cognitive ageing.

In this Chapter, I review the gaps and weaknesses of the literature to date and how these have informed the overarching aims of this thesis and individual hypotheses driving the analyses presented in Chapters 6-8.
4.2 Gaps in the literature:

In Chapter 2, I discussed both the clinical functional presentation (cognitive ability) and macroscopic presentation (neuroimaging correlates) that are associated with both brain and cognitive ageing and preterm birth. These structural alterations are underpinned by a host of molecular and cellular events including functional alterations in brain cells (neurons and supporting glia) leading to reductions in synaptogenesis, neurite outgrowth, demyelination, gliosis, mitochondrial dysfunction and many others (Mattson and Arumugam, 2018; Gonzales et al., 2022). A key link between many of these molecular changes is dysregulated immune activation and resultant inflammation, as discussed in Chapter 1.6. Alongside epidemiological evidence that elevations in peripheral inflammation associate with poor neuroimaging and cognitive outcomes (reviewed in Chapter 2.4.3), there are various animal and in vitro studies that give mechanistic evidence of the direct role of inflammatory-processes on aspects of cerebrovasculature and brain tissue that are considered to give rise to these phenotypes.

However, at both scales (macroscale and microscale), a significant proportion of the literature that assesses the association of systemic inflammation and brain health focuses on neurodegenerative disease states (Vos et al., 2003; Beecham et al., 2013; Chiang et al., 2017; Chitnis and Weiner, 2017; Swardfager et al., 2017; Darweesh et al., 2018; Newcombe et al., 2018; Rissanen et al., 2018; Brown et al., 2019; Matthews, 2019). By comparison, the relationship between systemic inflammation and neurocognitive function in non-pathological ageing is less studied. This is true both for studies that focus on the functional or behavioural manifestations of change (such as studies that examine cognitive outcomes or psychiatric conditions) as well as the neuroimaging literature that examines the association of elevated inflammatory markers in relation to brain structural metrics. The same is true at the other end of the lifecourse, with most studies that examine the role of inflammation in brain development focusing on cases
of overt brain injury such as neonatal stroke, severe cerebral haemorrhage, cystic periventricular leukomalacia or cerebral palsy, rather than examining neurostructural differences in preterm children without severe neonatal brain injury (Giraud et al., 2020).

When it comes to such non-clinical studies that examine inflammatory biomarkers in relation to either neuroimaging metrics, cognitive ability, or both, there remains dissent on the brain health consequences of elevated CRP. This literature (reviewed in Chapter 2) is heterogenous on multiple levels: in the way studies are designed (cross-sectional/longitudinal/mixed), and in the related features of these protocols (e.g. for longitudinal studies, follow-up time periods); in the ways researchers index cognitive ability (from use of cognitive screening tools like the MMSE or 3MS, to cognitive assessment batteries, to examining associations of inflammatory markers with individual cognitive test scores) or aspects of brain structure (MRI, fMRI, DTI). There is also variation in the participant demographics of the sample itself (age, ethnicity, sex); in the type of assay used to quantify CRP; in the profiling of CRP (assessing continuous levels vs binary splitting of CRP into ‘high’ or ‘low’ categories); in the inclusion/exclusion criteria for the sample and in the analytical approaches to assess these relationships. Obviously, each of these differences have a bearing on the findings obtained and the interpretation of the results.

Alongside these features, there is a lack of consistent accounting for mediating and confounding factors in inflammation-brain health relationships in both the inflammation-cognition literature and inflammation-neuroimaging literature. This is a key weakness of studies given that there is substantial evidence that an array of environmental and lifestyle-related factors across the lifecourse – from stress, adversity, to smoking and alcohol consumption – have been shown to associate with brain-health related phenotypes.

Many of the exposures that have been linked to brain health are complex and difficult to characterise, with most studies relying on self-reported estimates
to quantify degrees of exposure. Epigenetic modifications are a promising way to examine these relationships given that (1) there is a heritable component of both aspects of brain structure and function (endophenotypes), and potentially to the exposures/traits of interest themselves (e.g. predisposition to certain diseases) and (2) biologically, environmental factors may contribute to brain health outcomes via silencing or activating various genes through epigenetic modifications (in the case of DNAm, hyper or hypo methylation at specific sites). In this way, DNAm levels offer an ideal way to examine the influence of environmental stimuli on brain-related phenotypes (Lancaster et al., 2018) and to tease apart the individual contribution of chronic inflammation in the context of other aspects of health and lifestyle.

Despite this clear use-case, very few studies integrate neuroimaging, DNAm, lifestyle and inflammation data. Those that do tend to be restricted to specific clinical pathology cases (Cela rain and Tomas-Roig, 2020) rather than normative brain ageing. What’s more, while there is increasing general interest on the association of DNAm with early-life exposures in relation to brain and cognitive outcomes (Provenzi et al., 2016; Everson et al., 2020; Camerota et al., 2021; Luo et al., 2021), there is a distinct lack of research on the association between DNAm, early life exposures, and neuroimaging features in these earlier developmental time periods (Wheater et al., 2020) or even in earlier time periods of adulthood (early adulthood, midlife), with the vast majority of studies focusing on later-life (> 60 years) cognitive consequences. We close this research gap in Chapter 7 of this thesis (a neonatal cohort) and in Chapter 8 (cohort of individuals between the ages of 28 – 81), in an effort to compare how inflammation relates to brain health outcomes at different stages of the lifecourse.

Finally, one of the major biases of the literature to date is that singular measures of inflammatory protein such as CRP are used to index chronic inflammation. As discussed in Chapter 1.5.3, this approach may provide an unreliable capture of baseline inflammation levels. This factor alone could account for the discrepancies seen in the inflammation-brain health literature.
to date, particularly when paired with the inconsistencies in study design. This includes the varied implementation of cut-offs of elevated CRP reflective of acute inflammation (e.g. infection); range within these thresholds if applied (<40mg/l, <29mg/l, <10mg/L and <3mg/L all reported); differences in measurement of CRP (fasting / non fasting levels; quantification via low or high sensitivity assays) and statistical approaches to indexing inflammation (continuous outcome measures vs binary classification). The validity of CRP as an index of inflammation has already been brought into question in certain populations such as preterm infants where underdeveloped immune and organ systems may render this marker less appropriate (see Chapter 1.5.3).

One way to circumvent the common pitfalls of observational epidemiology is by proxying predictor measures with epigenetic surrogates, via the creation of DNAm signatures (as outlined in Chapter 3.8). In contrast to the transitory nature of serum CRP, DNAm is relatively stable in the short-term, and may be able to capture a biological archive of inflammatory burden that is more reflective of baseline inflammation levels, or a cumulative account of inflammatory insults. Support for this approaches comes from the predictive utility of DNAm signatures of other complex traits and exposures above and beyond that captured by baseline models of PRS indices of risk (Shah et al., 2015; McCartney, Stevenson, et al., 2018; McCartney et al., 2022; Thompson et al., 2022), and the persistence of epigenetic aberrations relating to exposure levels (McCartney, Stevenson, et al., 2018; Chamberlain et al., 2022). Comparatively, inflammatory-related DNAm signatures are very rare, with less than 10 published studies to date that have built composite scores of inflammation and examined them in relation to incident health outcomes (Barker et al., 2018; Michaud et al., 2020; Stevenson et al., 2020, 2021; Zhao et al., 2021; Gadd et al., 2022; Edmondson-Stait et al., 2022). This PhD represents the first time these signatures have been utilised to proxy chronic inflammation to examine the relationship between inflammation and neuroimaging outcomes. Where protein levels are obtained at the same

\[32\] Note many of these have been published after the analyses presented in Chapters 6-8 of this thesis were conducted and / or published
sampling time-point, these DNAm signatures of inflammation are compared against equivalent proteomic signatures, and where cognitive data is available, this is also assessed alongside brain structural metrics. In Chapter 1.5.5 I argue the case for how profiling a greater number of inflammatory-markers, or adopting composite approaches to characterising inflammatory responses, could be advantageous in this research context.

Overall, there is a lack of clear demarcation of age-related neurostructural alterations that are a consequence of chronic inflammation in terms of (1) which brain regions inflammation disproportionately affects (2) how inflammation-related neurostructural changes map to aspects of cognitive impairment, particularly at the level of individual subdomains of cognitive functioning (3) how these relationships present at different stages of the lifecourse (4) the driving lifestyle factors that influence these associations and the directionality of these relationships. Furthermore, there is a call for studies that integrate different data modalities (multi-omic approaches) to delineate determinants, mechanisms and trajectories of how inflammation relates to brain and cognitive ageing (Nikolova and Hariri, 2015; Lancaster, Morris and Connelly, 2018; Wheater et al., 2020). These key gaps form the basis for the rationale of following empirical chapters of this thesis, which assess how chronic inflammation relates to brain structure and function at different stages of the human lifecourse with an epigenetic epidemiological framework.

4.3 Aims of thesis:

Informed by these key omissions and limitations of the literature to date, the aims of this thesis are as follows: (1) to examine how DNAm signatures of inflammation compare with circulating inflammatory protein levels in relation to variance in brain structure (to include: global, cortical and subcortical metrics; white matter integrity; white matter pathology) and aspects of cognitive functioning (general factors of cognitive ability as well as cognitive subdomains); (2) to specifically characterise how inflammatory-related DNAm
associates with aspects of brain white matter across the life-course, and how alterations in brain structure mediate the association between inflammation and cognitive ability; (3) to investigate the impact of various aspects of health and lifestyle on inflammatory-related DNAm and its association with brain structure, connectivity and function.

These objectives were investigated in three studies that examined epigenetic signatures of inflammation across the lifecourse in relation to brain health outcomes.

4.3.1 Chapter 6 Aims and Hypotheses:

**Inflammatory-related DNAm in relation to neuroimaging and cognitive outcomes in old age.** The following hypothesis is predicated upon the following three observations: i) DNAm changes are relatively stable (Lancaster, Morris and Connelly, 2018); ii) aggregate DNAm signatures may provide a biological archive of certain exposures (as reviewed in Chapter 3.8.2); and iii) inflammatory-related DNAm signatures exhibit higher test-retest reliability, more consistent associations with cognitive ability, and more stable longitudinal dynamics than circulating serum protein levels (Stevenson et al., 2020). I hypothesised that inflammatory-related DNAm may be better placed to capture variance in brain structure associated with chronic inflammation in later life, in line with the ‘inflammaging’ theory of brain ageing. To test this hypothesis, I compared association magnitudes between serum high sensitivity C-Reactive Protein (hsCRP) concentrations and composite DNAm signature of CRP (built from 7 CpG sites) in relation to multiple global and regional structural brain imaging measures and to various aspects of cognitive function in a well-characterised community-dwelling sample of healthy adults in older age (521 healthy older adults, ~ 73 years from the Lothian Birth Cohort 1936). I hypothesised that a DNAm signature of CRP, created from an out of sample EWAS, would show significantly stronger associations with brain health than serum hsCRP levels across a range of detailed neuroimaging and cognitive measures. I additionally
hypothesised that elements of brain structure would mediate the association between elevated inflammation and poor cognitive functioning, highlighting proximal biological pathways of the inflammation-cognition relationship. Finally, I hypothesised that variation in aspects of health and lifestyle would influence associations between DNAm, brain structure and cognitive ability. The rationale for this work was based on clarifying the impact of chronic inflammation on aspects of brain-ageing in later life, given that a historical reliance on phasic protein levels may have clouded previous inference and, in part, account for the observed heterogeneity of previous epidemiological research linking elevated inflammatory markers with both cognitive and structural brain ageing phenotypes.

4.3.2 Chapter 7 Aims and Hypotheses: Inflammatory-related DNAm in relation to neuroimaging and postnatal health outcomes in early-life: I hypothesised that there would be a significant difference in DNAm CRP between preterm infants sampled at TEA and term control subjects. I additionally anticipated that DNAm CRP would be able to capture an allostatic load of inflammatory exposure in the preterm infants. To assess this hypothesis, I investigated how DNAm CRP scores associated with various inflammatory exposures in the perinatal period such as maternal risk factors (smoking in pregnancy, preeclampsia), fetal risk factors such as intrauterine infection (histologic chorioamnionitis) and postnatal inflammatory conditions (such as sepsis) in both term and preterm infants. I additionally anticipated that the DNAm CRP might be able to capture the cumulative impact of these perinatal exposures in line with the multiple-hit hypothesis of inflammation (discussed in section 2.7.3) in a dose-dependent manner, with number of inflammatory episodes associating with the DNAm CRP score. Finally, I also hypothesised that variance in DNAm CRP would be associated with aspects of neonatal MRI metrics characteristic of EoP in the preterm infants to include lower deep grey matter volume, lower white matter volume and altered connectivity as assessed by dMRI parameters. In particular, that a higher inflammatory load as indexed by
DNAm CRP would associate with altered white matter integrity in the white matter tracts that are late to develop in pregnancy, owing to the enhanced vulnerability of this period – as informed by the mechanistic basis of this outlined in Chapter 1.6 and the epidemiological evidence for white matter injury and dyssmaturation in prematurity reviewed in Chapter 2.8.3.

4.3.3 Chapter 8 Aims and Hypotheses:

**Inflammatory-related DNAm signatures in relation to neuroimaging and cognitive outcomes in midlife** (709 adults, age range 28-81 years, Stratifying Longitudinal Resilience and Depression Cohort). Bolstered by the findings presented in Chapters 6 and 7, this Chapter extended the research paradigm of using inflammatory-related DNAm signatures of inflammation to examine differences in brain structure and function by generating over 100 inflammatory-related DNAm signatures reflective of different circulating protein levels involved in immune signalling and other inflammatory-processes. I hypothesized the research to date may be underestimating the influence that dysregulated inflammation has on incident health outcomes and that the underlying assumption that a very small subset of inflammatory markers is sufficient to profile chronic inflammation (and is therefore informative of inflammation-brain health relationships) needs to be reconsidered. The hypothesis here is that a number of individual CpG sites (with a continuum of effect sizes) are reflective of aberrant inflammatory responses in response to shared genetic and environmental factors, but that brain-health relevant signals of these may be missed as differential methylation at individual CpG sites may not pass threshold significance. Hence, aggregating associative effects of all biomarkers (through the creation of DNAm signatures) may be able to capture the wider profile of inflammation, and provide improved assessments of the relationship between inflammation and brain health outcomes. This chapter aimed to comprehensively capture how aspects of brain health (including multi-modal neuroimaging metrics, cognitive outcomes and brain-age scores) associate
with these DNAm signatures of inflammation, anticipating that DNAm signatures related to pro-inflammatory processes would associate with less favourable brain and cognitive health outcomes. In particular, this study aimed to examine the regional impact of DNAm signatures on brain cortical and subcortical volumes as well as regional variance in white matter tract integrity. Again, this study examined these relationships in the context of contributions from aspects of health and lifestyle (smoking, alcohol consumption, BMI and hypertension). Finally, this chapter aimed to explore interrelationships between inflammatory-related DNAm signatures, brain structure and cognitive ability through mediation modelling.

4.4 Summary

I outlined the aims and hypothesis of the empirical chapters of this thesis, how these relate to the major aims of this thesis, and their novelty with respect to existing studies on the relationship between chronic inflammation with DNAm, neuroimaging and cognitive ability. In the following chapter, I describe the three cohort studies these analyses are set in (the Lothian Birth Cohort 1936, Theirworld Edinburgh Birth Cohort, Generation Scotland) and the major methodologies used throughout Chapters 6-8.
5 Study cohorts and methods

In this chapter, I provide an overview of the main protocols for the three cohort studies used in this thesis which correspond to the empirical studies presented in Chapters 6, 7 and 8: The Lothian Birth Cohort 1936 (Chapter 6), Theirworld Edinburgh Birth Cohort (Chapter 7), and Generation Scotland: the Scottish Family Health Study (Chapter 8). I outline the measurement of neuroimaging, DNAm, inflammatory protein and lifestyle data within these cohorts. I additionally outline the rationale for the main methods used in my empirical work.

5.1 The Lothian Birth Cohort 1936

The LBC1936 is a sample of generally healthy, community-dwelling older adults, who were born in 1936 in the Lothian area of Scotland. Many of these took the Scottish Mental Survey in 1947 (SMS1947) at around 11 years of age, which included a test of cognitive ability (the Moray House Test; MHT). Years later, these individuals were recruited for a range of cognitive, neuroimaging, sociodemographic and health assessments to capitalise on this catalogue of early life cognitive ability, details of which are provided in the papers by Deary et al., (2007; 2012). There have been six ‘waves’ of longitudinal follow-up of these individuals where repeat assessments were performed as they progressed into later-life. The first of these (Wave 1) included 1,091 individuals who were around 70 years of age. Three years later, Wave 2 of assessments took place (n = 866), which included the addition of brain MRI.

The following sections refer to data from the LBC1936 cohort relevant to analyses presented in Chapter 6 of this thesis. This study was cross-sectional in design and set in Wave 2 of the LBC1936 study, so the following
data discussed pertains to data collected at this time point only (unless specified otherwise).

5.1.1 Ethics and funding

The LBC1936 is supported by Age UK (Disconnected Mind program) and the Medical Research Council (MR/M01311/1). DNA methylation typing was supported by the Centre for Cognitive Ageing and Cognitive Epidemiology (Pilot Fund award), Age UK, The Wellcome Trust Institutional Strategic Support Fund, The University of Edinburgh, and The University of Queensland.

Ethical approval for the Lothian Birth Cohort studies was obtained from the Multi-Centre Research Ethics Committee for Scotland (MREC/01/0/56) and the Lothian Research Ethics committee (LREC/1998/4/183; LREC/2003/2/29). All participants provided written informed consent. The research was carried out in compliance with the Helsinki Declaration.

5.1.2 Clinical assessment

This was a study on a cohort of elderly adults (~ 73 years), participant characteristics are reported in Chapter 6.2. Of note, as this study’s aim was to investigate chronic inflammation in instances of ‘normative’ cognitive ageing to address the paucity of these studies, participants were excluded if they had a self-reported history of stroke, Parkinson’s, or dementia or had an MMSE < 24, indicative of MCI.

5.1.3 Inflammatory protein data in the Lothian Birth Cohort 1936

To assess inflammation, CRP (mg/L) was measured from whole-blood samples, quantified using both a low-sensitivity assay (lsCRP) and a high-sensitivity assay (hsCRP). As the low sensitivity assay cannot quantify CRP values < 3 mg/L, the hsCRP values were used for the analyses presented in
Chapter 6. This assay was performed at the University of Glasgow using an enzyme-linked immunosorbent assay (ELISA; R&D Systems, Oxford, UK).

In Chapter 6 for the 521 participants with complete cognitive, neuroimaging and lifestyle data recorded at Wave 2, mean (± SD) hsCRP levels were 2.02 ± 1.92 mg/L. This final n screened out participants who had a recorded CRP level of > 10mg/L as this is generally suggestive of infection or illness at the time of sample collection (see discussion of reference CRP values for epidemiological studies in Chapter 1.5.2.2). To account for skewed distributions, CRP levels were log-transformed (natural log) as part of the data-cleaning process before analyses were performed.

5.1.4 Cognitive data in LBC1936

In Chapter 6 of this thesis cognitively normal participants (exclusion criteria screened out participants who had a self-reported history of a neurodegenerative disorder, stroke, or MMSE < 24) were assessed via their scores from a comprehensive battery of the following cognitive tests:

**Digit symbol coding:**
Digit symbol coding is a subset of the Wechsler Adult Intelligence Scale-III UK (WAIS-III UK) (Wechsler, 1997) used to measure *processing speed*. This paper-based task requires participants to assign symbols to numbers based on a code key. Participants are given 120s to assign as many symbols to numbers as possible, with the final score recorded as the number of correct answers (max 93) within the strict time-limit.

**Digit span backwards:**
Digit span backwards assesses *short-term working memory* from the Wechsler Memory Scale-III UK (WMS-III UK) (Weschler, 1998). This involves auditory processing, as the participant must repeat back to the examiner a

---

33 In an earlier draft of the Neurology paper, both ls-CRP and hs-CRP were used and compared against DNAm CRP; after suggestion from reviewers, only the hs-CRP measure was used.
series of lengthening numerical sequences (2-8 digits) that have been read to them by the examiner. When the participant fails to correctly report the sequence or when the maximum string length (8 digits) is appropriately reached, the test is over.

**Block design:**
Block design assesses *visuospatial reasoning* and is a subtest from the Wechsler Adult Intelligence Scale-III (Wechsler, 1997). Participants are shown a 2D design for a 9-block model in a book and asked to assemble the design using red and white coloured blocks within two minutes. Scoring is based on speed of assembly and accuracy of reconstruction.

**Symbol search:**
Symbol search is a subtest of the Wechsler Adult Intelligence Scale-III assessing *processing speed*. This is a pen and paper based task where participants are asked to view rows of symbols and target symbols, and indicate when the target symbols appear in each row with a Yes / No response. Scores are calculated based on how many correct answers are given within a two-minute time frame.

**Letter number sequencing:**
Letter-number sequencing from the Wechsler Adult Intelligence Scale-III, assesses *working memory*. Participants are asked to listen and repeat back to the examiner a series of numbers in increasing order and letters in alphabetical order. Scores are based on the number of correct responses (max 21).

**Matrix reasoning:**
Matrix reasoning is a subtest of the Wechsler Adult Intelligence Scale-IIIUK used to assess *non-verbal reasoning*. Participants are asked to examine a grid of pictures with one missing square, where all the items within the grid are following an unspecified rule. Participants are asked to select the picture that fits the missing item from five options. The max score for this test is 26.
Global factors (gf) and latent scores of cognitive domains:
In Chapter 2.3.1, I outlined how scores derived from tests of different aspects of cognitive function generally display correlation structures. Because of this, the use of dimensionality reduction methods is often applied to derive latent scores pertaining to specific aspects of cognitive ability. As explained in Chapter 2.3.1.4, fluid intelligence is a relevant latent measure to construct when asking questions about inter-individual differences in cognitive ageing, as greater declines in fluid intelligence are seen with age than crystallised forms. A principal component analysis (PCA) was conducted on the above mentioned raw test scores, where the first un-rotated component was used as a measure of gf (explaining ~ 50% of the variance in the cognitive data). Latent scores of the individual cognitive domains of processing speed, visuospatial ability and verbal memory were also derived by the same method (see Appendix 11.1, eTable3 for individual tests used and loadings).

5.1.5 Neuroimaging data in LBC1936

The specific protocol for each imaging metric used in Chapter 6 (global brain volume measures; white matter tract integrity; white matter hyperintensity burden; cortical ROIs) is briefly outlined in the methods section of 6.2 and accompanying supplemental methods (Appendix 11.1). For comprehensive details on image acquisition and subsequent processing, see the protocol paper by Wardlaw et al. (2011). Here I will give a brief overview of the different types of neuroimaging metric assessed.

5.1.5.1 Global brain volume metrics

Briefly, LBC1936 participants at Wave 2 were scanned on a GE Signa Horizon 1.5 Tesla HDxt clinical scanner (General Electric, Milwaukee, WI, USA). Acquisition comprised T2-, T2*- and FLAIR-weighted axial scans, and a high-resolution 3D T1- weighted volume sequence acquired in the coronal
plane. Cortical reconstruction and volumetric segmentation of T1-weighted images were performed with FreeSurfer v5.3 (http://surfer.nmr.mgh.harvard.edu/) using the Desikan-Killiany atlas (Desikan et al. 2006). White matter connectivity data for individual tracts – FA and MD – were derived, with tracts segmented using the BEDPOSTX/ProbTrackX algorithm in FSL (https://fsl.fmrib.ox.ac.uk) and Tractor (https://www.tractor-mri.org.uk). Total brain (TB), grey matter (GM), normal-appearing white matter (NAWM) and white matter hyperintensity (WMH) volumes were segmented using a semi-automated multi-spectral technique (Hernández et al., 2012). WMH volumes were log-transformed prior to analyses as they had a skewed distribution; all volume measures were controlled for ICV in analyses. For further details see Chapter 6 methods and accompanying supplemental information (Appendix 11.1).

5.1.5.2 Vertex-wise cortical metrics

Cortical regions of interest (ROIs) – cortical thickness, volume and surface area – were examined in Chapter 6 as previous work has shown that these may be particularly sensitive to age-related differences (as discussed in Chapter 2.5). Example plots of cortical thickness for two LBC1936 participants are presented below.

Figure 5.1 LBC1936 cortical thickness data

Two randomly selected LBC1936 participants. These plots were created using the fsbrain package; script available at EleanorSC github.
Chapter 6 outlines how cortical volume, thickness and surface area vertex-wise analyses were performed across the average surface with linear models to investigate the effect of inflammation (serum hsCRP and DNAm CRP) on focal aspects of the cortex. “Vertex” here refers to the perpendicular distance between grey and white matter surfaces. Previous studies have reported the various QC steps of this (Ritchie et al., 2018), but briefly cortical surface maps from each LBC1936 participant were manually examined, with participants removed if there was evidence of low-quality scan or motion artifacts (listwise deletion of general segmentation failure and casewise elimination of skull stripping issues / boundary errors).

Thereafter, cortical vertex-wise regression analyses were performed using MATLAB SurfStat toolbox (http://www.math.mcgill.ca/keith/surfstat) for Matrix Laboratory R2018a (The MathWorks Inc., Natick, MA). The resulting statistical maps (t-maps) were corrected for multiple comparisons using FDR with a q-value of 0.05 across all 327,684 vertices on the cortical surface.

5.1.5.3 Regional white matter microstructure

As outlined in Chapter 2.4.3, DTI can be used to infer information about white matter microstructure. Tractography techniques, such as probabilistic neighbourhood tractography (PNT) can then be used to derive average FA and MD per white matter tract of interest. This was performed for each LBC1936 participant and is illustrated in Figure 5-2.
Figure 5.2 PNT derived white matter tract ROIs in LBC1936 participants

Glass brain plot displaying lateral and superior views of MRI segmented white matter tract ROIs using probabilistic tractography in one LBC1936 participant

5.1.5.4 General factors of white matter integrity (gFA, gMD)

‘Global’ measures of white matter structure can be generated owing to the shared variance in diffusion parameters in the major white matter tracts of the brain. The methods to generate general factors include PCA and factor analysis. Both methodologies produce highly colinear measures; a PCA approach captures maximal variance across all tracts, whereas a factor analysis can exclude tract-specific error variance and instead models reliable shared variance across tracts. Both methods have been used previously to examine changes in white matter architecture with age (Penke et al., 2010; Cox, Ritchie, et al., 2016; Vaher et al., 2022).

In Chapter 6, gFA and gMD were derived for each participant by taking the first un-rotated principal component from a PCA of the average FA / MD measures of the tracts highlighted in Figure 5.2; participants with up to 2
missing values from specific tracts had data replaced with the mean value for that tract.

5.1.6 DNA methylation data in the Lothian Birth Cohort 1936

DNAm data was obtained from whole blood using standard methods at the Edinburgh Clinical Research Facility at the Western General Hospital. DNA methylation was measured using Illumina Infinium HumanMethylation450K BeadChips in the LBC1936 at Wave 2, with QC reported previously (Shah et al., 2014).

5.1.7 Inflammatory-DNAm signature

As explained in Chapter 3.8, DNAm signatures can be used to aggregate genome-wide information about a particular trait or exposure. In practical terms, after pre-processing of DNAm data is performed, the degree (%) of methylation present at any given CpG site is supplied as a β-value – simply, this is a continuous measure between 0 and 1 which represents the ratio of hyper/hypomethylation: the methylated signal divided by the total of the methylated and unmethylated signals (Elliott et al., 2014).

For each individual where DNAm data was available at Wave 2 of LBC1936, a weighted score was obtained by multiplying this normalised methylation value (the β-value) at relevant CRP-related CpG sites – identified as per a large EWAS of CRP by Ligthart et al. (2016); details of which are covered in Chapter 3.8.4 and summarised in Figure 5.5– by their respective weights, then summing these values:

\[ b_1cpg_1 + b_2cpg_2 + \ldots + b_7cpg_7 \]

Where “cpg” is the normalised methylation value for the LBC1936 participant at a given site and “b” is the effect size from Ligthart et al., (2016).
5.2 Theirworld Edinburgh Birth Cohort (TEBC):

The following subsections refer to data relevant to analyses presented in Chapter 7 of this thesis. This study was cross-sectional in design and set in neonates in the TEBC study who had DNAm data available. The following data discussed pertains to data collected from both term born and full term infants and clinical details about various perinatal exposures they experienced (inflammatory episodes or risk factors experienced by mothers in pregnancy and fetal episodes).

5.2.1 Ethics and funding

Preterm and term born infants delivered at the Royal Infirmary of Edinburgh, UK were recruited to the Theirworld Edinburgh Birth Cohort, a longitudinal study designed to investigate the effect of preterm birth on brain development (Boardman et al., 2020). Cohort exclusion criteria were major congenital malformations, chromosomal abnormalities, congenital infection, overt parenchymal lesions (cystic periventricular leukomalacia, hemorrhagic parenchymal infarction) or post-hemorrhagic ventricular dilatation.

Ethical approval was obtained from the National Research Ethics Service, South East Scotland Research Ethics Committee (11/55/0061, 13/SS/0143 and 16/SS/0154). Informed consent was obtained from a person with parental responsibility for each participant. The study was conducted according to the principles of the Declaration of Helsinki.

5.2.2 Clinical assessment (health and lifestyle data)

Previous work in this cohort and others has indicated that compared to full term infants, infants born before 37 weeks GA have an altered immune profile leading to higher susceptibility to sustained inflammation. To assess the degree to which this burden of inflammation influenced by antenatal factors, shared experiences at the fetal-interface, or early postnatal exposures and interventions, data that was considered relevant to maternal,
fetal or neonatal inflammatory-exposures was recorded; these are defined in Table 5.1.

<table>
<thead>
<tr>
<th>category</th>
<th>variable</th>
<th>definition</th>
</tr>
</thead>
<tbody>
<tr>
<td>antenatal</td>
<td>corticosteroids</td>
<td>betamethasone 12mg given before delivery</td>
</tr>
<tr>
<td></td>
<td>magnesium sulphate</td>
<td>antenatal administration of 4g MgSO4 (magnesium sulphate heptahydrate [8 mmol magnesium ions])</td>
</tr>
<tr>
<td></td>
<td>smoking</td>
<td>defined as self-report of current smoker status during pregnancy</td>
</tr>
<tr>
<td></td>
<td>preeclampsia</td>
<td>diagnosed in pregnancy</td>
</tr>
<tr>
<td>fetal</td>
<td>Histologic chorioamnionitis (HCA)</td>
<td>placental pathology report which classifies the presence of HCA as an inflammatory response in the placental membranes of any grade or stage</td>
</tr>
<tr>
<td>neonatal</td>
<td>Bronchopulmonary dysplasia (BPD)</td>
<td>defined as an oxygen requirement at 36 weeks corrected gestation</td>
</tr>
<tr>
<td></td>
<td>Necrotising enterocolitis (NEC)</td>
<td>characterised as Bell Stage 2 or above</td>
</tr>
<tr>
<td></td>
<td>Retinopathy of prematurity</td>
<td>diagnosed following birth</td>
</tr>
<tr>
<td></td>
<td>Early or late onset neonatal sepsis</td>
<td>detection of bacterial pathogen from blood culture, or physician decision to treat for ≥5 days in the context of growth of coagulase negative staphylococcus from blood or a negative culture</td>
</tr>
</tbody>
</table>

Table 5.1 Clinical data of inflammatory exposures

5.2.3 Neuroimaging data in Theirworld Edinburgh Birth Cohort

The specific protocol for each imaging metric used in Chapter 7 (global brain volume measures; white matter tract integrity; white matter hyperintensity burden; cortical ROIs) is briefly outlined in the methods section of 7.2 and accompanying supplemental methods (Appendix 11.2). For comprehensive details on image acquisition and subsequent processing, see the protocol paper by Boardman et al. (2020). Of note, this study used data from two MRI acquisition sessions (Phase 1 and Phase 2), hence scanner variable was then included in all following analyses; in both Phases, MRI was obtained at the same time as saliva sample collected (for DNAm data collection) at term equivalent age (TEA). In Phase 1, structural and dMRI data were performed on 93 infants using a MAGNETOM Verio 3T clinical MRI scanner (Siemens...
Healthcare GmbH, Erlangen, Germany). For the second phase (n=121), structural and dMRI were performed on 121 infants using a MAGNETOM Prisma 3T clinical MRI scanner (Siemens Healthcare GmbH, Erlangen, Germany). Further details are included in Chapter 6.2 methods, and a flow diagram of study recruitment and attrition is provided in the supplemental document attached to this paper in Appendix 11.2.

Another important methodological note is that all the subjects were registered to the Edinburgh Neonatal Atlas (ENA50) (Blesa et al., 2020) as opposed to the Desikan Killany atlas (for adults) which was used in Chapters 7 and 8.

![Figure 5.3 Representative neonatal brain maps](image)

**Figure 5.3 Representative neonatal brain maps**

*Figure from Sullivan et al., (2022) reproduced with permission. Averaged maps adapted to neonatal brain template for (A) cortical segmentation, (B) axial diffusivity, (C) mean diffusivity, (D) radial diffusivity, (E) fractional anisotropy, (F) neurite density index, (G) orientation dispersion index.*

5.2.3.1 Peak width skeletonised metrics

Alongside the methods introduced in Chapter 2.5.2 another tactic employed to examine white matter microstructure is using peak width skeletonised metrics (Beaudet et al., 2020). This approach is based on skeletonization of dMRI data (most commonly performed using TBSS), where histogram analysis of global white matter microstructure is performed to generate general measures of microstructure such as peak width skeletonised mean diffusivity (PSMD) and peak width skeletonised general anisotropy (PSFA).
These metrics are obtained by calculating difference between 5th percentile and the 95th percentiles of the distribution of the voxel FA or MD values within the brain white matter skeleton (Baykara et al., 2016). In contrast to single measures of FA and MD, PSMD or PSFA are classified as dispersion statistics, and may offer value in capturing diffuse differences in the brain, such as diffuse white matter injury that can be part of the EoP presentation in preterm infants. This makes it a choice methodology to employ when examining structural alterations in preterm birth.

5.2.3.2 General factors of white matter integrity in TEBC

gFA and gMD were also constructed in a similar fashion to that described in section 5.1.5.4, where the first unrotated component of a PCA on 16 white matter tracts was extracted for each dMRI metric (gFA ~ 61% variance, gMD ~ 72%); example scree plot for FA below.

![FA PCA scree plot](image)

**Figure 5.4 Scree plot demonstrating a strong single factor capturing common variance across white matter tracts**

*Based on individuals from the TEBC study sample described in Chapter 6.2. Y axis displays the eigenvalue against the number of components (x axis) for FA in the study sample.*
5.2.3.3 Neurite orientation dispersion and density imaging (NODDI)

One of the issues raised in Chapter 2.5.3 is the issue of crossing fibres in DTI analysis. Newer methods, such as neurite orientation and dispersion density imaging (NODDI), may overcome these limitations in modelling the space between axons (intra-extra neurite space) found in grey and white matter as well as the CSF within voxels. NODDI gives the CSF volume fraction (ISO), which stands for the isotropic water fraction, as well as the orientation dispersion index (ODI) and neurite density index (NDI).

These extra diffusion-based measures of tissue microstructure are often assessed in neonatal cohorts and have been discussed in various papers (Zhang et al., 2012; Parker et al., 2021). In Chapter 6, PSNDI (where NDI = a capture of the intra-neurite volume fraction within a given voxel) is examined alongside PSAD, PSRD, PSFA and PSMD.

5.2.4 DNA methylation data in Theirworld Edinburgh Birth Cohort

DNAm data was obtained from neonatal saliva samples. The saliva was collected at TEA in Oragene OG-575 Assisted Collection kits, by DNA Genotek, and DNA extracted using prepIT.L2P reagent (DNA Genotek, Ontario, Canada). DNA was bisulfite converted and methylation levels were measured using Illumina HumanMethylationEPIC 450K BeadChip (Illumina, San Diego, CA, USA) at the Edinburgh Clinical Research Facility (Edinburgh, UK). The arrays were imaged on the Illumina iScan or HiScan platform and genotypes were called automatically using GenomeStudio Analysis software version 2011.1 (Illumina). DNAm was analysed in two batches (related to the two phases of neuroimaging acquisition). Details on subsequent pre-processing are provided in the supplementary material that accompanies the paper presented in Chapter 7 (Appendix 11.2).

DNAm data was obtained from 311 infants. Of these, 29 did not pass DNAm pre-processing QC criteria and were excluded. Infants with congenital
abnormality were excluded (n = 1) and cases where sex from DNAm did not match recorded sex were also removed (n = 3). This group of 278 neonates included 20 sets of twins, where random removal of one twin from each set was performed, leaving the final sample number at 258 (155 preterm, 103 full term controls).

5.2.5 DNAm signature pipeline

![Figure 5.5 DNAm CRP signature pipeline](image-url)
Figure 5.6 Distributions of DNAm CRP in TEBC participants

5.3 Stratifying Resilience and Depression Longitudinally cohort (STRADL)

Chapter 8 presents findings in the ‘Stratifying Resilience and Depression Longitudinally’ (STRADL) which is a subset of the larger Generation Scotland: Scottish Family Health Study (GS). Details of the study protocol are provided by Habota et al. (2019).

5.3.1 Ethics and funding

Ethical approval for STRADL was formally obtained from the NHS Tayside committee on research (reference 14/SS/0039), and all participants provided their written informed consent.

5.3.2 Proteomics data in STRADL

Inflammatory-related proteins were measured from blood samples from STRADL participants and quantified using the SOMAscan® platform (which is an aptamer-based assay that measures numerous proteins from blood
samples concurrently). Of the 4,235 proteins measured, the 109 that matched the DNA methylation (DNAm) signatures were used in analyses. Relative distributions of these in the 778 STRADL participants that had DNAm data are shown in Appendix 11.3, Supplementary Figure 2.

5.3.3 Cognitive data in STRADL

Cognitive testing assessments within STRADL were face-to-face administration of common validated neuropsychological tests, the protocol of which has been previously reported (Habota et al., 2019).

For the data presented in Chapter 8, seven measures of cognitive function were assessed in analyses: general cognitive ability ($g$), general fluid cognitive ability ($gf$) and individual domains of matrix reasoning, processing speed, executive function, vocabulary, and verbal declarative memory in line with the approach adopted by previous studies in this cohort (Whalley et al., 2016; Habota et al., 2019). The cognitive tests pertaining to these metrics are outlined in the succeeding subsections.

**Mill Hall Vocabulary:**
The Mill Hill Vocabulary (MHV) is a test of *verbal cognitive ability* that consists of 44 items where participants are asked to find words that are synonyms of other words. The total correct score of this was used in analyses.

**Verbal Fluency:**
Verbal Fluency assess *executive function*. In this exercise, participants have one minute to say as many words that begin with a particular letter as they can. Participants are not permitted to use proper nouns, numbers, or the same word more than once with a different ending (e.g. catch succeeded by catching). This is repeated for the letters C, F, and L. The total number of words is then added to determine the final score.
Digit-Symbol Coding:
This is the same test from the Wechsler Adult Intelligence Scale-III UK (Wechsler, 1997) described in full in section 5.1.4.

Logical Memory:
The Logical Memory Story test is one of the tests of the Wechsler Memory Scale-III that tests short and long-term verbal declarative memory (Weschler, 1998). Participants are instructed to read a short passage which contains 25 items to memorise. Their ability to recall these items immediately after finishing is scored as immediate recall and their ability to recall items after half an hour is delayed recall. The total score (immediate recall + delayed recall), out of 50, is the variable of interest in Chapter 8.

General cognitive ability (g) general fluid cognitive ability (gf) and latent cognitive domain metrics
The above tests were used to generate latent scores pertaining to ability for different aspects of cognitive function:

<table>
<thead>
<tr>
<th>Cognitive metric</th>
<th>Cognitive test</th>
<th>N</th>
<th>Mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td>processing speed</td>
<td>Digit Symbol score (number correct)</td>
<td>772</td>
<td>67.9</td>
<td>14.8</td>
</tr>
<tr>
<td>verbal declarative reasoning</td>
<td>summed measure of immediate and delayed scores from the recall section of one story of the Logical Memory test</td>
<td>776</td>
<td>31.4</td>
<td>7.4</td>
</tr>
<tr>
<td>non-verbal reasoning</td>
<td>Matrix Reasoning (number correct)</td>
<td>776</td>
<td>8.2</td>
<td>2.4</td>
</tr>
<tr>
<td>vocabulary</td>
<td>Vocabulary score (number correct)</td>
<td>774</td>
<td>31.7</td>
<td>4.1</td>
</tr>
<tr>
<td>executive function</td>
<td>Vocabulary score (number correct)</td>
<td>774</td>
<td>42.6</td>
<td>11.8</td>
</tr>
<tr>
<td>general cognitive ability (g)</td>
<td>1st unrotated component from a PCA of the matrix reasoning test, verbal fluency test, Mill Hill vocabulary test, logical memory and digit-symbol coding tests</td>
<td>768</td>
<td>0</td>
<td>1</td>
</tr>
<tr>
<td>general fluid cognitive ability (gf)</td>
<td>1st unrotated component from a PCA of the matrix reasoning test, verbal fluency test, logical memory and digit-symbol coding tests</td>
<td>770</td>
<td>0</td>
<td>1</td>
</tr>
</tbody>
</table>

Table 5.2 STRADL Cognitive test score summary
Data presented is for individuals from STRADL who had complete protein and DNAm data available.
Figure 5.7 Distributions of cognitive test metrics in STRADL

Histogram distribution plots for cognitive metrics from the STRADL dataset where participants had complete DNAm data. Metric means are illustrated in vertical lines, SD presented as dotted lines, n numbers reported alongside. Variables were trimmed to remove outliers (>3.5 SD from mean) before plotting.

5.3.4 Neuroimaging data in STRADL

Neuroimaging data used in Chapter 8 involves MRI acquisition at different sites: Aberdeen and Dundee (scanner site was included as a covariate in subsequent analyses). Details of this are provided in full by Habota et al. (2019).

Briefly, for Aberdeen MRI, data were acquired at the Aberdeen Royal Infirmary with a Philips Achieva 3T TX-series scanner (Philips Healthcare, Best, Netherlands). At the Dundee site (Ninewells Hospital), a Siemens 3T Prisma-FIT (Siemens Healthineers, Erlangen, Germany) scanner was used. Full details are provided in Chapter 8.2 methods section and accompanying supplemental document (Appendix 11.3), and have also been described in detail by (Habota et al., 2019).

The following data was obtained: 3D T1-weighted fast gradient echo with magnetisation preparation; 3D T2-weighted fast spin echo; 3D Fluid
Attenuation Inversion Recovery (FLAIR); Diffusion Tensor Imaging (DTI); and Susceptibility Weighted Imaging (SWI) or T2*-weighted gradient echo (Habota et al., 2019).

5.3.4.1 Global MRI metrics

Six main global neuroimaging metrics were derived in the current study: global grey matter, global white matter, whole brain volume as well as global measures of cortical thickness, volume and surface area.

Global grey matter was calculated by summing the volume of cortical, subcortical and cerebellar grey matter within both hemispheres of the brain. Global white matter was calculated in the same fashion, excluding any tissue that was not white matter. Total brain volume summed both grey and white matter volume but excluded the brainstem, ventricles, cerebrospinal fluid and choroid plexus. Global cortical measures were derived by summing metrics from both left and right hemispheres.

FLAIR images were examined for the presence of WMH, which were defined by punctuate, focal or diffuse lesions in the deep or periventricular white matter, basal ganglia or brainstem, visible as areas of hyperintensity in respect to normal appearing white or grey matter. The amount of WMH was graded according to the Fazekas scale; periventricular and deep WMH were graded separately on a scale from 0-3; these grades were then summed to provide a total Fazekas grade (range 0-6), which is treated as continuous measure of global WMH burden. Distributions of this Fazekas score, alongside other main neuroimaging measures assessed in Chapter 8, can be seen in Figure 5-7 below.
5.3.4.2 Derivation of brain age

Brain age was derived using the software package brainageR (Version 2.1; available at https://github.com/james-cole/brainageR). This estimates brain age by applying machine learning to voxel-wise grey, white matter and CSF volumetric data derived from structural T1-weighted MRI as described by Cole et al. (2017). This estimate was then regressed on chronological age and the residuals from this model were used to index ‘relative brain age’. Relative distributions and sample sizes for both brain age and residualised brain age (brain acceleration) are presented in Figure 5-7.

5.3.4.3 Derivation of regional brain MRI metrics

7 subcortical structures (nucleus accumbens, amygdala, caudate nucleus, hippocampus, pallidum, putamen and thalamus) and 34 cortical regions were segmented. Further details are provided in the methods section of Chapter 8.
5.3.4.4 DTI data (white matter tract metrics) in STRADL

FA and MD was available across 24 ROI tracts (Table 5-5). As per described in 5.2, a global factor of FA and MD was determined by applying PCA on relevant white matter tracts to extract a latent measure. Scores of the first unrotated principal component of FA/MD were extracted and set as the dependent variable (proportion of variance explained by gFA and gMD was 45.2% and 61.9% respectively).

<table>
<thead>
<tr>
<th>White matter tract</th>
<th>Abbreviation</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cingulum (hippocampus)</td>
<td>CGH</td>
</tr>
<tr>
<td>Cingulum (cingulate gyrus)</td>
<td>CGC</td>
</tr>
<tr>
<td>Fornix*</td>
<td>FX</td>
</tr>
<tr>
<td>FX Fornix (cres) / Stria terminalis</td>
<td>FX.ST</td>
</tr>
<tr>
<td>Inferior fronto-occipital fasciculus</td>
<td>IFOF</td>
</tr>
<tr>
<td>Superior fronto-occipital fasciculus</td>
<td>SFOF</td>
</tr>
<tr>
<td>External capsule</td>
<td>EC</td>
</tr>
<tr>
<td>Superior longitudinal fasciculus</td>
<td>SLF</td>
</tr>
<tr>
<td>Sagittal striatum</td>
<td>SS</td>
</tr>
<tr>
<td>Uncinate fasciculus</td>
<td>UNC</td>
</tr>
<tr>
<td>Body of corpus callosum*</td>
<td>BCC</td>
</tr>
<tr>
<td>Genu of corpus callosum*</td>
<td>GCC</td>
</tr>
<tr>
<td>Splenium of corpus callosum*</td>
<td>SCC</td>
</tr>
<tr>
<td>Corpus callosum*</td>
<td>CC</td>
</tr>
<tr>
<td>Corona radiata</td>
<td>CR</td>
</tr>
<tr>
<td>Internal capsule</td>
<td>IC</td>
</tr>
<tr>
<td>Anterior corona radiata</td>
<td>ACR</td>
</tr>
<tr>
<td>Posterior corona radiata</td>
<td>PCR</td>
</tr>
<tr>
<td>Superior corona radiata</td>
<td>SCR</td>
</tr>
<tr>
<td>Corticospinal tract</td>
<td>CST</td>
</tr>
<tr>
<td>Anterior limb of internal capsule</td>
<td>ALIC</td>
</tr>
<tr>
<td>Posterior limb of internal capsule</td>
<td>PLIC</td>
</tr>
<tr>
<td>Posterior thalamic radiation</td>
<td>PTR</td>
</tr>
<tr>
<td>Retrolenticular limb of internal capsule</td>
<td>RLIC</td>
</tr>
</tbody>
</table>

Table 5.3 STRADL white matter tract ROIs (DTI data)
5.3.5 DNA methylation data in STRADL

Peripheral blood DNAm was sampled from individuals in the STRADL dataset, these were quantified using the Illumina EPIC array; details of this and QC information are outlined in (Gadd et al., 2022) and are provided in Supplementary Information (Appendix 11.3).

5.3.6 Machine learning approaches for selecting relevant CpG sites

Feature selection is an important aspect of dealing with high dimensional data. With DNAm data, this applies to selecting pertinent CpG sites, a process illustrated in Figure 5-10. Common machine learning methods for building DNAm proxies of age, disease or exposures, as discussed in Chapter 3.8.1, include Ridge, LASSO and Elastic net, with the latter being a regularisation method that combines both the LASSO and Ridge regression methods. Regularisation is a way to fit a model, get better estimates of effect sizes, and perform feature selection simultaneously – it’s efficient, quick, and can provide researchers with more informative set of features, and by restricting the magnitude of coefficients, can give us a better (and more stable) estimate of the outcome of interest. Cross-validation is a technique that assess the generalizability models to unseen datasets – this takes the form of training and test splits, and is useful for tuning model hyperparameters (in this context – the ‘penalty’ applied in the penalised regression model). For more details on the statistical theory behind these machine learning techniques, refer to the following primer (James et al., 2013).

Alongside the use of out of sample EWAS or candidate gene study, these different approaches in the context of DNAm predictors of exposures or disease have advantages and disadvantages, a process illustrated in Figure 5.9. Because of an automatic variable selection, where LASSO selects one feature out of a set of highly correlated features, LASSO approaches are optimal only when a few sites (CpGs) are expected to be selected for
prediction compared to the total number of sites in the data (the $p > n$
problem). In instances where large fraction of CpGs are anticipated to be
associated with a given trait or exposure of interest, Ridge regression where
no variable selection is performed is more appropriate, but runs the risk of
overfitting the data. As a trade-off between the two, elastic net is particularly
well-suited for selecting CpG sites for DNA methylation predictors of health and disease
because it allows for the selection of multiple features with minimal loss of
accuracy and reduced multicollinearity. For more in depth details about both
the statistics and applicability of these approaches in the context of
epigenomics research, consult the following reviews (Engebretsen and
Bohlin, 2019; Hu, Zhou and Li, 2022; Yousefi et al., 2022).
Figure 5.9 Ridge, LASSO and elastic net

Flow diagram of CpG selection approaches in the context of building DNA methylation (DNAm) signatures. Alongside conducting (a) an EWAS or (b) a candidate gene study to identify CpG sites associated with a particular trait or exposure, penalised

\[
\left( \sum_{i=1}^{N} y_i - X_i \beta \right) + \lambda (\alpha \| \beta \|_1 + (1 - \alpha) \| \beta \|_2)
\]

Information sourced from https://carpentries-incubator.github.io/high-dimensional-stats-r/03-regression-regularisation/index.html
regression approaches such as (c) Ridge regression, (d) LASSO and (e) elastic net can be used. Ridge and LASSO differ in how they handle correlated variables: while Ridge regression shrinks correlated variables toward each other, LASSO has a sparse feature selection approach and will only identify one feature within a set of correlated ones. Therefore, Ridge regression tends to perform better than LASSO when the predictors are highly correlated, but suffers from issues of overfitting. Elastic net is considered an intermediary between these two techniques, which provides some regularization to help prevent overfitting. These regularization parameters of elastic net (the $\alpha$ and $\lambda$ described by the equation) can be tuned to achieve a balance between the accuracy of prediction and the number of variables selected, which gives it an advantage over other methods. Finally, this is an attractive approach because the outputs of elastic net are easy to interpret: results can be presented as regression coefficients.

5.3.7 Inflammatory-related DNAm signature pipeline

As discussed, penalised regression approaches can be used to select CpG sites of interest. For the work presented in Chapter 8, an elastic net approach was used as outlined by Gadd et al., (2022) to generate 109 DNAm signatures reflective of variance in various circulating protein levels. This pipeline is outlined in Figure 5.10 below.
Diagram overview of the protocol used for building the inflammatory-related DNAm signatures that are examined in relation to brain and cognitive ageing in Chapter 8 of this thesis. Full protocol summarised by Gadd et al. (2022). Using two population cohorts (LBC1936 and KORA), elastic net models were used to train DNAm

---

© Eleanor Conole 2023, outline of protocol of Gadd et al. (2022) ‘Epigenetic scores for the circulating proteome as tools for disease prediction’
signatures that explain variance in circulating protein levels. In the final 109 signature, there was overlap for 6 signatures (GZMA, MMP.1, CXCL10, NTRK3, CXCL11, EN.RAGE) between the Olink and SOMAScan trained samples.

Of note, the specific DNAm CRP signature that examined in Chapters 6 and 7 was also applied to examine neuroimaging outcomes in the STRADL cohort (Green, Shen, et al., 2021), although DNAm was quantified using the in this population sample, one of the 7 CpG sites (cg06126421) was unavailable, meaning the score was composed of 6 CpG sites. In Chapter 8, a different DNAm signature of circulating proteomic CRP trained on CRP quantified by SOMAmer protein measurements in the KORA cohort is used, a composite score built from 100 CpG sites. Only one loci was present in both scores (the DNAm CRP score examined in Chapters 6-7 vs the score used in Chapter 8): cg18181703, a site that regulates the expression of the gene SOCS3.

5.4 Summary

This Chapter provided an overview of each of the population cohorts used in the following chapters (LBC1936, TEBC, STRADL), including the initial protocol and the data collection process for the variables analysed in this thesis and relevant methodologies.
6 Chronic inflammation and brain health in old age

6.1 Introduction

The literature reviewed in Chapters 1-2 highlights that while numerous studies find associations between chronic inflammation and aspects of brain and cognitive ageing, questions remain about the extent of inflammation’s association with brain structure and function. I point out that previous work predominantly relies on single time point sampling of inflammatory proteins from blood samples – the problem with this approach is the phasic and volatile nature of these inflammatory proteins (section 1.5.3), a feature that renders them less useful for measuring inflammation that is chronic in nature and ‘baseline’ inflammation levels in population cohorts. I suggest that composite metrics that assimilate inflammatory information may overcome this issue to provide a better capture of chronic inflammation. I also outline that very few studies examine detailed inflammation, neuroimaging and cognitive data concurrently, or attempt to account for confounding aspects of lifestyle and health on these relationships.

With these gaps in mind, this chapter presents the results of an empirical research study that examines the associations of two markers of inflammation (serum protein hsCRP and DNAm CRP) in relation to various aspects of brain structure and function in a cohort of individuals in their 70s. The use of this DNAm signature has been validated across several studies (reviewed in Chapter 3.8), and 7 CpG sites that constitute this score were selected based off the findings by the largest scale EWAS to date on CRP. Previous work from this cohort has found that DNAm CRP was more stable than serum protein levels of CRP, and by comparison DNAm CRP increased

---

36 at the time of investigation; since the publication of this work there has been a larger EWAS by Wielscher et al., (2022), the findings of which were discussed in Chapter 3.7.1.
steadily with age, had higher test-retest reliability and longitudinally demonstrated a greater association with poor cognitive functioning than serum CRP (Stevenson et al., 2018, 2020). The rationale for examining how DNAm CRP compared against serum CRP for brain structural phenotypes was therefore driven by the evidence that the methylation marker may provide a better readout of inflammatory exposure than traditional approaches, and may therefore give a more accurate reflection of how baseline inflammation levels at the population level affect brain and cognitive ageing phenotypes.

This paper was published in Neurology (Conole et al., 2021) in December 2021 and is included in full in section 6.2. Supplementary material is provided in Appendix (11.1).
6.2 DNA methylation and protein markers of chronic inflammation and their associations with brain and cognitive ageing
DNA Methylation and Protein Markers of Chronic Inflammation and Their Associations With Brain and Cognitive Aging

Eleanor L.S. Conole, BSc, MRes, Anna J. Stevenson, PhD, Susana Muñoz Maniega, PhD, Sarah E. Harris, PhD, Claire Green, MSc, Maria del C. Valdés Hernández, PhD, Mathew A. Harris, PhD, Mark E. Bastin, DPhil, Joanna M. Wardlaw, MD, Ian J. Deary, PhD, Veronique E. Miron, PhD, Heather C. Whalley, PhD, Riccardo E. Marioni, PhD, and Simon R. Cox, PhD

Neurology® 2021;97:e2340-e2352. doi:10.1212/WNL.0000000000012997

Abstract

Background and Objectives
To investigate chronic inflammation in relation to cognitive aging by comparison of an epigenetic and serum biomarker of C-reactive protein and their associations with neuroimaging and cognitive outcomes.

Methods
At baseline, participants (n = 521) were cognitively normal, around 73 years of age (mean 72.4, SD 0.716), and had inflammation, vascular risk (cardiovascular disease history, hypertension, diabetes, smoking, alcohol consumption, body mass index), and neuroimaging (structural and diffusion MRI) data available. Baseline inflammatory status was quantified by a traditional measure of peripheral inflammation—serum C-reactive protein (CRP)—and an epigenetic measure (DNA methylation [DNAm] signature of CRP). Linear models were used to examine the inflammation–brain health associations; mediation analyses were performed to interrogate the relationship between chronic inflammation, brain structure, and cognitive functioning.

Results
We demonstrate that DNAm CRP shows significantly (on average 6.4-fold) stronger associations with brain health outcomes than serum CRP. DNAm CRP is associated with total brain volume (β = −0.197, 95% confidence interval [CI] −0.28 to −0.12, pFDR = 8.42 × 10^{-6}), gray matter volume (β = −0.200, 95% CI −0.28 to −0.12, pFDR = 1.66 × 10^{-5}), and white matter volume (β = −0.150, 95% CI −0.23 to −0.07, pFDR = 0.001) and regional brain atrophy. We also find that DNAm CRP has an inverse association with global and domain-specific (speed, visuospatial, and memory) cognitive functioning and that brain structure partially mediates this CRP–cognitive association (up to 29.7%), dependent on lifestyle and health factors.

Discussion
These results support the hypothesis that chronic inflammation may contribute to neurodegenerative brain changes that underlie differences in cognitive ability in later life and highlight the potential of DNAm proxies for indexing chronic inflammatory status.
This study provides Class II evidence that a DNSm signature of CRP levels is more strongly associated with brain health outcomes than serum CRP levels.

Low-level systemic chronic inflammation has emerged as a hallmark and potential driver for individual differences in brain aging.\(^1\)\(^-\)\(^5\) Yet while chronic inflammation has been consistently linked to dementia,\(^6\)\(^-\)\(^8\) studies investigating peripheral inflammatory markers in nonclinical groups show disparity with respect to cognitive outcomes\(^9\)\(^-\)\(^14\) and have not yet clarified the magnitude and regional extent of brain structural associations.\(^13\)\(^-\)\(^18\)

One reason for this inconsistency is that there are no standard biomarkers for chronic inflammation, and to date many studies have relied upon blood biomarkers of acute inflammation such as C-reactive protein (CRP). A significant caveat of this approach is assuming baseline inflammation status from highly phasic protein levels, which are subject to swift and rapid concentration changes in blood plasma.\(^19\) This introduces significant noise at the epidemiologic level\(^20\)\(^,\)\(^21\) (Figure 1D) and few studies take repeat measures of serum CRP\(^12\) or attempt to correct for within-person fluctuations. A more accurate reflection may come from an epigenetic approach: DNA methylation (DNSm) profiles have been identified in inflammatory diseases\(^22\)\(^,\)\(^23\) and inflammation-related disease outcomes\(^24\)\(^,\)\(^25\) and are theorized to provide more stable reflections of inflammatory exposure.\(^24\)\(^,\)\(^26\)\(^-\)\(^28\)

In the same cohort as in the present study, a DNSm proxy of CRP exhibited greater longitudinal stability and stronger associations with cognitive functioning than serum CRP levels.\(^11\)\(^,\)\(^28\)

Here, we predict that a DNSm signature of CRP will show significantly stronger associations with brain health outcomes than its serologic counterpart. Our objective is to examine the

**Figure 1** Chronic Inflammation Increases With Age and May Contribute to Variance in Cognitive Ability

A Schematic demonstrating lifespan curves for chronic inflammation. Inflammatory load tends to increase with age; lifestyle, genetics, and health conditions can all influence susceptibility to chronic inflammation and account for variance in inflammation levels between individuals.\(^1\)\(^,\)\(^3\) (B) Chronic inflammation can be measured by inflammatory proteins taken from a blood sample, such as serum levels and DNA methylation (DNSm) proxies of C-reactive protein (CRP). (C) Lifespan curves for cognitive ability outlining how there is considerable interindividual heterogeneity in rate and timing of cognitive decline, with some people on more accelerated cognitive aging trajectories than others.\(^2\) (D) Trajectories of LBC1936 participants respective of inflammation scores over age, as outlined in Stevenson et al.,\(^11\) illustrating comparative stability of DNSm inflammation marker compared to serum CRP.

---

**Glossary**

AD = Alzheimer disease; BMI = body mass index; CFI = comparative fit index; CI = confidence interval; CRP = C-reactive protein; CVD = cardiovascular disease; dMRI = diffusion MRI; DNSm = DNA methylation; EWAS = epigenome-wide association studies; FDR = false discovery rate; gFA = general fractional anisotropy; GM = gray matter; gMD = general mean diffusivity; ICV = intracranial volume; NAWM = normal-appearing white matter; QC = quality control; RMSEA = root mean squared error approximation; SRMR = standardized root mean square residual; TB = total brain; TLI = Tucker–Lewis index; VRF = vascular risk factor; WM = white matter; WMH = white matter hyperintensity.
The relationship between chronic inflammation, brain structure, and cognitive aging in a large community-dwelling sample of older adults.

Methods

Participants

The Lothian Birth Cohort 1936 (LBC1936) comprises individuals who were surviving members of the Scottish Mental Survey 1947, born in 1936, and who were living in Edinburgh and the surrounding area (the Lothians) when the study began in 2004. Full details of the recruitment procedures and protocols have been published. Participants took part in 4 waves of testing in later life (at mean ages 70, 73, 76, and 79 years) as part of an investigation into the determinants of cognitive aging. At each wave, participants were interviewed and tested individually by a trained psychologist and a research nurse during a visit to the Wellcome Trust Clinical Research Facility (wtcrf.ed.ac.uk), Western General Hospital, Edinburgh, United Kingdom. This visit included cognitive and other psychological assessments, physical examinations, extensive history taking, and blood analyses. From wave 2 onwards, neuroimaging data are also available.

The current study on chronic inflammation is cross-sectional (all variables described here were collected in 2007, at wave 2 of the LBC1936 study) and addresses the following primary research questions:

1. Does an epigenetic inflammation measure (DNAm CRP) show stronger associations with brain structure and function than serum CRP levels? (Class II evidence)
2. Does an epigenetic inflammation measure (DNAm CRP) show stronger associations with white matter (WM) microstructure than serum CRP levels? (Class II evidence)
3. To what extent can alterations in brain structure explain the association between inflammation and cognitive ability? (Class II evidence)

Participants were free from neurodegenerative diagnoses at baseline and were excluded if they had a self-reported history of stroke, Parkinson disease, or Alzheimer disease or had a Mini-Mental State Examination score <24, indicating mild cognitive impairment. We also excluded participants with serum CRP level >10 mg/L, suggestive of acute infection or illness at the time of blood draw. After exclusions, a total of 521 participants had complete inflammation, cognitive, neuroimaging, and relevant health data. For further details on data availability and attrition, see Figure 2 and eTable 1 (links.lww.com/WNL/B629).

Brain Imaging Data

Structural and diffusion tensor imaging MRI acquisition and processing in LBC1936 were performed according to an open-access protocol. A 1.5T GE Signa HDx clinical scanner (General Electric) was used to collect structural T1 (voxel size 1 × 1 × 1.3 mm), T2 (voxel size 1 × 1 × 2 mm), T2* (voxel size 1 × 1 × 2 mm), and fluid-attenuated inversion recovery–weighted images (voxel size 1 × 1 × 4 mm); for full details on MRI sequence measures, refer to Table 1 in the open access protocol article. Local processing and quality control (QC) of cortical reconstruction and segmentation was performed using FreeSurfer v5.1 on T1-weighted volumes. Full information on brain imaging acquisition, QC, and variables used in analyses is detailed in the supplementary eMethods (links.lww.com/WNL/B629).

CRP Data

Serum CRP was measured from whole blood samples using a high-sensitivity assay (ELISA; R&D Systems).

DNAm Preparation and DNAm CRP Score

Genome-wide DNAm was measured in blood samples using the Illumina Human MethylationEPIC BeadChip at the Edinburgh Clinical Research Facility Genetics Core; the epigenetic measure of chronic inflammation was calculated for each participant as described previously. Briefly, a DNAm CRP score was assembled for each participant in wave 2 of LBC1936; this was created by means of a weighted composite score, based on a discovery meta-analysis (9 cohorts, n = 8,863) and a replication meta-analysis (4 cohorts, n = 4,111) of CRP–epigenome-wide association studies (EWAS). Methylation beta values were derived for the 7 CpG sites shown to have the strongest association with serum CRP levels, then multiplied by their standardized regression weights and added together. Given that all regression weights from the EWAS were negative, a higher DNAm CRP score (i.e., closer to 0) corresponds to a higher inflammatory profile. Relative weights for the 7 CpGs are included in the supplementary document (eTable 2, links.lww.com/WNL/B629).

Cognitive Ability Data

All participants in the LBC1936 underwent a detailed battery of standardized cognitive tests. From these, participant scores for 3 distinct cognitive domains (visuospatial ability, processing speed, and verbal memory) alongside a general fluid-type cognitive ability score (g) were created based upon well-fitting, hierarchical structural equation models tested in our previously published work; relevant cognitive tests and individual weightings can be found in the supplementary document (eTable 3, links.lww.com/WNL/B629).

Lifestyle Variables

Building from previous work that looked at the impact of vascular risk factors (VRFs) on cognitive aging, we selected the most pertinent variables available to us in the LBC1936 cohort that may influence or confound the relationship between inflammation, brain health, and cognitive aging. Lifestyle variables included body mass index (BMI; kg/m²), calculated from height and weight at the time of interview (see eMethods for details), alongside variables relating to self-reported health and disease history: cardiovascular disease...
history (CVD), hypertension, diabetes, smoking status (coded as current smoker [1] vs ex/nonsmoker [0]), and alcohol use (coded as drinker [1] vs nondrinker [0]). Regular anti-inflammatory drug use was also collected at baseline and coded as on medication [1] or not on medication [0].

**Statistical Analyses**

Statistical analyses were performed in R version 3.6.1 (r-project.org). Alpha was 0.05 for all analyses and results were corrected for multiple comparisons using the false discovery rate (FDR). Standardized coefficients are reported throughout to facilitate comparison of associations. Serum measures of CRP were log-transformed to correct a positively skewed distribution. White matter hyperintensity (WMH) volume was log transformed, after which it showed an approximately normal distribution. All global MRI volumetric measures (total brain [TB], gray matter [GM], normal-appearing WM [NAWM], WMH) were corrected for intracranial volume (ICV) and expressed as a ratio of ICV. For volumetric brain associations, differences between association magnitudes (serum CRP vs DNAm CRP associations) were assessed using the Williams test for dependent groups with overlapping correlations (cocor.indep.groups.overlaps) as implemented in the “cocor” R package (cran.r-project.org/web/packages/cocor/cocor.pdf). We ensured that models showed acceptably low multicollinearity (variance inflation was ascertained using “vif” in the “car” package in R, cran.r-project.org/web/packages/car/car.pdf). Pairwise bivariate associations were assessed between markers of inflammation, neuroimaging, and lifestyle covariates using Pearson correlation.
correlation. All models were adjusted for age and sex. Details of individual analyses are as follows.

**Volumetric Brain Associations With Inflammation**

Linear regression models were used to identify the proportion of phenotypic variance explained by DNAm CRP and to determine whether this was independent of the serum CRP signal for each brain health phenotype. Logistic regressions were conducted for self-reported disease history variables with binary outcomes (disease/no disease).

**Regional Brain Analyses**

Localized associations between DNAm CRP score and vertex-wise cortical volume, area, and thickness were performed using linear regression, controlling for age, sex, and ICV. We used the SurfStat MATLAB toolbox (math.mcgill.ca/keith/surf-stat) for Matrix Laboratory R2012a (The MathWorks, Inc.). The resulting statistical maps (t-maps) were corrected for multiple comparisons using FDR with a q value of 0.05 across all 327,684 vertices on the cortical surface.

**Sensitivity Analyses**

Ancillary mixed-effects models including interaction terms (e.g., inflammation × age, inflammation × sex, inflammation × anti-inflammatory drug use) investigated whether the association of chronic inflammation with brain health outcomes was modified or confounded by age, sex, or the use of anti-inflammatory medication (eTable 4, links.lww.com/WNL/B629). Similarly, we included lifestyle and health covariates in a fully adjusted model (alongside age and sex) to determine whether individual aspects of health and lifestyle had an impact on the association of inflammation with brain health phenotypes (eTable 5, links.lww.com/WNL/B629).

**Mediation Analyses**

We ran mediation analyses in a structural equation modeling (SEM) framework using the R "lavaan" package (cran.r-project.org/web/packages/lavaan/lavaan.pdf). This simultaneously characterized associations among CRP, brain, and cognitive metrics, and also specifically tested the hypothesis that brain structure would partly and significantly mediate associations between measures of CRP and cognitive ability. Both single and multiple mediator models were specified (see Figure 5, A–C, as example). Single mediator models provided information on the proportion of CRP–cognitive associations attributable to individual neuroimaging metrics. By contrast, in multiple mediator models, brain structural variables were entered simultaneously as covarying mediators (see path diagram, Figure 5C). This allowed us to quantify the proportion of variance in CRP–cognitive associations uniquely explained by each facet of brain structure (GM, NAWM, WMH, general fractional anisotropy [gFA], general mean diffusivity [gMD]). The primary estimates of interest in this study are the degree of change (mediation) in the direct path (c to c’ between inflammation measures (DNAm CRP or serum CRP) and cognitive ability (g or processing speed or visuospatial ability or verbal memory) when the indirect path from inflammation to cognitive ability via brain structure (a × b) is included. A significant mediation of the c path (to c’) is denoted by the

### Table 1 Cross-sectional Associations Between Serum CRP and DNAm CRP With Neuroimaging and Cognitive Outcomes

<table>
<thead>
<tr>
<th>Phenotype</th>
<th>Serum CRP (model 1)</th>
<th>DNAm CRP (model 2)</th>
<th>Δ Association magnitudes, p value</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>β</td>
<td>SE</td>
<td>p Value</td>
</tr>
<tr>
<td><strong>Neuroimaging</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TB</td>
<td>−0.033</td>
<td>0.040</td>
<td>0.579</td>
</tr>
<tr>
<td>GM</td>
<td>−0.026</td>
<td>0.042</td>
<td>0.670</td>
</tr>
<tr>
<td>NAWM</td>
<td>−0.027</td>
<td>0.042</td>
<td>0.657</td>
</tr>
<tr>
<td>WMH</td>
<td>0.018</td>
<td>0.042</td>
<td>0.740</td>
</tr>
<tr>
<td>gFA</td>
<td>−0.055</td>
<td>0.045</td>
<td>0.347</td>
</tr>
<tr>
<td>gMD</td>
<td>0.025</td>
<td>0.045</td>
<td>0.677</td>
</tr>
<tr>
<td><strong>Cognitive</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Visuospatial ability</td>
<td>−0.082</td>
<td>0.038</td>
<td>0.069</td>
</tr>
<tr>
<td>Processing speed</td>
<td>−0.088</td>
<td>0.038</td>
<td>0.054</td>
</tr>
<tr>
<td>Verbal memory</td>
<td>−0.046</td>
<td>0.038</td>
<td>0.347</td>
</tr>
<tr>
<td>gF</td>
<td>−0.098</td>
<td>0.038</td>
<td>0.027 a</td>
</tr>
</tbody>
</table>

Abbreviations: CRP = C-reactive protein; DNAm = DNA methylation; gF = general cognitive ability; gFA = general fractional anisotropy; GM = gray matter; gMD = general mean diffusivity; NAWM = normal-appearing white matter; TB = total brain; WMH = white matter hyperintensity.

Cross-sectional associations of inflammation measures and neuroimaging and cognitive phenotypes. All p values reported are false discovery rate corrected. *p < 0.05.
statistical significance of this indirect effect. Bootstrapping was used to calculate standard errors. Multiple comparisons were corrected for by FDR correction. These mediations were re-run when accounting for self-reported health variables as covariates: in model 1, age and sex were covariates; in model 2, they were age, sex, BMI, hypertension, diabetes, smoking status, and alcohol use. To account for missing data bias, we took account of all available data, using full information maximum likelihood estimation. Model fit was evaluated based on root mean squared error approximation (RMSEA), the comparative fit index (CFI), and the Tucker–Lewis index (TLI). We considered a model an acceptable fit when it respected the following thresholds: RMSEA ≤0.05; SRMR ≤0.06; CFI ≥0.97; and TLI ≥0.95, as recommended.

Data Availability
The data analyzed in this study are not publicly available as it contains data that could compromise participant consent and confidentiality, but can be requested via a data access request to the Lothian Birth Cohorts research group.

Standard Protocol Approvals, Registrations, and Patient Consents
Ethical permission for the LBC1936 was obtained from the Multi-Centre Research Ethics Committee for Scotland (MREC/01/0/56) and the Lothian Research Ethics Committee (LREC/2003/2/29). Written informed consent was obtained from all participants. All necessary patient/participant consent has been obtained and the appropriate institutional forms have been archived.

Results

DNAm CRP Is Associated With Global and Regional Brain Volume
We studied 521 eligible older adults (aged ~73 years; refer to Figure 2 and eTable 1, links.lww.com/WNL/B629) and looked at epigenetic vs serum inflammation associations across a range of neuroimaging and cognitive measures (Table 1). To index chronic inflammation, an epigenetic measure of CRP (DNAm CRP) was assembled for each participant (see Methods). The correlation between the DNAm CRP score and serum log(CRP) was moderate ($r = 0.29, 95\%$ confidence interval [CI] 0.28–0.4), and the DNAm CRP score showed a stronger correlation with serum CRP than any one of its composite CpGs.

We found that higher inflammatory burden, indexed by DNAm CRP scores, associated with poor cognitive and neuroimaging brain health outcomes (Table 1). DNAm CRP exhibited significantly larger (6.4-fold, on average) associations with brain structural MRI metrics (including global GM and WM atrophy, poorer WM microstructure, and increased WMH burden) than serum CRP. These DNAm CRP-associated brain structural changes were independent of anti-inflammatory drug use, age, or sex (eTable 4, links.lww.com/WNL/B629).

Participants with a higher inflammatory burden on average had greater overall brain atrophy, with higher DNAm CRP associating with lower total brain volume ($\beta = -0.197, 95\%$ CI $-0.28$ to $-0.12$, $p_{\text{FDR}} = 8.42 \times 10^{-6}$), GM volume ($\beta = -0.200, 95\%$ CI $-0.28$ to $-0.12$, $p_{\text{FDR}} = 1.66 \times 10^{-5}$), and WM volume ($\beta = -0.150, 95\%$ CI $-0.23$ to $-0.07$, $p_{\text{FDR}} = 0.001$). Models that included additional health and lifestyle covariates (BMI, smoking, alcohol consumption, hypertension, diabetes, and cardiovascular disease history) attenuated the relationship between DNAm CRP and brain health outcomes by up to 40% (eTable 5, links.lww.com/WNL/B629). Of these, the associations between DNAm CRP with WM measures (WMH, NAWM) were the most attenuated (34%–40%). Out of the lifestyle and health factors accounted for, smoking appeared to have the greatest influence on the attenuation (as illustrated in supplementary eFigure 2, links.lww.com/WNL/B629).

After examining global brain structural alterations, we looked at specific regional cortical brain associations with higher inflammation levels. We found regional heterogeneity in the patterning of associations between CRP measures and cortical metrics: atrophy in frontal, anterior lateral, and medial temporal lobes was associated with higher DNAm CRP (Figure 3B); inflammation associations with brain cortical thickness are presented in the supplementary document (eFigure 1, links.lww.com/WNL/B629). Overall, these results emphasize that the DNAm-CRP score associates with lower cortical volume of specific brain regions (lateral and medial temporal regions of the brain), which show overlap with those of serum CRP and unique variance (Figure 3F), with DNAm CRP reflecting atrophy beyond the serum CRP score.

DNAm CRP Is Associated With White Matter Microstructure in Specific White Matter Tracts
Next, we investigated whether higher DNAm CRP was related to lower WM microstructure based on global and regional diffusion MRI (dMRI) measures by looking at inflammation associations with WM tract fractional anisotropy and mean diffusivity. Whereas serum CRP–dMRI associations were null in all cases (all $p_{\text{FDR}} >0.089$) (eTables 6–7, links.lww.com/WNL/B629), higher DNAm CRP scores predicted overall lower gFA ($\beta = -0.162, p_{\text{FDR}} = 6.94 \times 10^{-9}$) and higher gMD ($\beta = 0.124, p_{\text{FDR}} = 0.010$). For specific WM tracts, the strongest associations were seen for the arcuate fasciculus and uncinate fasciculus, with lower FA and higher MD with higher DNAm CRP (see Figure 4; eTables 6–7, links.lww.com/WNL/B629). For global measures of WM tract integrity (gFA, gMD), accounting for health and lifestyle covariates did not substantially alter the magnitude or significance of these associations (eTable 5, links.lww.com/WNL/B629); however, at the level of individual WM tracts, the relationship between DNAm CRP and FA and MD was attenuated when lifestyle factors were included in the models (eTables 8–9, links.lww.com/WNL/B629); this is illustrated in supplementary eFigure 3 (links.lww.com/WNL/B629).
Brain Structure Partly Mediates the Association of DNAm CRP With Cognitive Ability

As higher DNAm CRP levels were associated with lower cognitive performance both here (Table 1) and previously, we quantified the degree to which brain structural differences contribute to the inflammation–cognition association, and which facets show the strongest unique contributions to this relationship using an SEM framework. Bivariate associations between all variables (inflammation, brain structure, cognitive ability, and lifestyle measures) are provided in eTable 10 (links.lww.com/WNL/B629). Whereas TB volume, GM volume, NAWM volume, and WMH volume all emerged as significant mediators in single SEM models (percentage attenuation 14%–21%; eTable 11, links.lww.com/WNL/B629), multiple mediator models were used to test the degree to which each global MRI metric contributed uniquely to mediation of the same association (Figure 5D; eTable 12, links.lww.com/WNL/B629). Here, the sum total of MRI measures significantly mediated the association between DNAm CRP and general cognitive ability ($\beta = -0.047 [-0.076 to -0.018], p_{FDR} = 0.002; \text{percentage attenuation } 29.7\%$). The unique contributions to this variance were largest for NAWM volume ($\beta = -0.03 [-0.053 to -0.023], p_{FDR} = 0.012$), indicating that the loss of WM may contribute to inflammation-associated differences in cognitive functioning in older age. Out of the individual cognitive domains, processing speed was the most significantly mediated by the sum total of MRI metrics ($\beta = -0.058 [-0.090 to -0.027], p_{FDR} = 0.001; \text{percentage attenuation } 41\%$). Again, NAWM emerged as the largest unique contribution to this variance ($\beta = -0.037 [-0.063 to -0.010], p_{FDR} = 0.013; \text{percentage attenuation } 37\%$), with NAWM accounting for the largest unique contribution to this effect ($\beta = -0.030 [-0.063 to -0.010], p_{FDR} = 0.012$). While verbal memory was significantly mediated by the sum total of MRI metrics ($\beta = -0.031 [-0.057 to -0.007], p_{FDR} = 0.026; \text{percentage attenuation } 33\%$), there were no significant contributions from individual MRI metrics (eTables 12–13, links.lww.com/WNL/B629).

Finally, with the addition of lifestyle and health covariates to our models, no aspect of brain structure remained a significant mediator of the associations between DNAm CRP and...
general cognitive ability ($\beta_{\text{mediation}} = -0.023 \ [ -0.049 \text{ to } 0.003], p_{\text{FDR}} = 0.167$ (see Figure 5, eTables 11–12, links.lww.com/WNL/B629) or any of the individual cognitive domains (eTable 12, links.lww.com/WNL/B629).

Discussion

Only recently has there been a push for integrated multi-omics approaches to better characterize chronic inflammation.\textsuperscript{3,26} DNAm profiles may act as promising peripheral biomarkers for cognitive-aging differences at the population level, given their relative stability in the short term, and their joint modulation by both genetic and lifestyle traits. Elsewhere, DNAm markers of inflammation have proved informative in predicting a range of age-related health outcomes, from cardiovascular disease to depression,\textsuperscript{23,24,36} but few studies have applied this same approach to cognitive aging differences in healthy cohorts. As chronic inflammation is considered to be an insidious, cumulative, and often undetected contributor to cognitive aging,\textsuperscript{1,3,5,14} the importance of such epigenetic markers may be their utility to index inflammatory load with greater reliability than phasic protein measures. In this study, DNAm CRP was more robustly associated with a range of cognitive and neuroimaging metrics than serum CRP, supporting our original hypothesis. We discovered that DNAm CRP shows consistently stronger associations with brain structure than serum CRP (on average, 6.4-fold greater), that these associations are not regionally homogeneous across the brain’s cortex, and that specific aspects of brain structure partly mediate (up to 29.7%) associations between an epigenetic signature of CRP and cognitive
functioning. Our results highlight the potential of epigenetic approaches to indexing in inflammation in population cohorts and suggest that chronic inflammation may contribute to both focal and global brain structural changes that underlie differences in cognitive aging.

We found regional heterogeneity in the patterning of associations between CRP measures and cortical metrics, indicating differential regional vulnerability to chronic inflammation.

Reductions in brain cortical volume and thickness in frontal, anterior lateral, and medial temporal lobes were associated with increased DNAm CRP. Consistently, previous studies report structural changes associated with inflammatory markers in the temporal and frontal cortices.17,18 Atrophy in these regions is implicated in cognitive decline,37 and differential patterns of proinflammatory receptor distribution may underlie why some brain regions are more vulnerable to inflammation than others. For example, in patients with Alzheimer disease (AD), proinflammatory cytokine receptor density and expression are increased in regions of neurodegeneration, including the medial frontal and temporal cortices.38 Higher inflammation levels have also been related to progression of atherosclerosis, with evidence for differential effects of CRP in different beds of the arterial brain supply.39,40 These findings suggest that raised levels of inflammatory mediators may contribute to localized brain atrophy via their differential expression in brain tissue and cerebrovasculature.

Overall, the results of our mediation analyses indicate that chronic inflammation’s detrimental effect on WM beyond other brain structural features may underlie the inflammation-associated differences in cognitive functioning in older age. Whereas numerous studies have found associations between reduced WM volume and raised inflammatory markers in...
healthy cohorts\textsuperscript{14,17,41} and those with chronic inflammatory conditions,\textsuperscript{42} few have looked at inflammation, brain structure, and cognitive function concurrently.\textsuperscript{13} Fewer still have attempted to more robustly characterize chronic inflammation beyond assessing serum inflammatory protein profiles—although a notable exception comes from recent work, where the same DNAm CRP signature was found to be significantly associated with widespread reductions in WM integrity beyond that of serum CRP.\textsuperscript{36} In agreement with this study, but in an older cohort, we found that higher DNAm CRP related to increased WMH burden, reduced NAWM volume, and ostensibly poorer WM microstructure (lower FA and higher MD). In particular, the WM tracts of arcuate fasciculus and uncinate fasciculus showed the most consistent significant relationships with DNAm CRP levels (across both FA and MD), alongside significantly lower FA in the anterior thalamic radiation, which are consistent with studies assessing the effects of vascular risk on microstructure with advanced age.\textsuperscript{31}

In agreement with longstanding findings from neurocognitive studies—where, consistently, inflammation is more strongly associated with declines in processing speed than other cognitive domains\textsuperscript{9,10}—our results indicate that some cognitive domains (processing speed) may be more mediated by the brain structural consequences of chronic inflammation than others (verbal memory, visuospatial ability). Processing speed has been strongly linked to WM integrity at both global and regional levels\textsuperscript{43} and many of the downstream effects of proinflammatory processes directly affect WM integrity (to include demyelination, de-afferentation and gliosis; see Figure 6 and eFigure5, links.lww.com/WNL/B629).\textsuperscript{14} As such, chronic inflammation’s contribution to diffuse and global WM loss may disproportionately affect cognitive

\textbf{Figure 6} Mechanisms of Neurodegeneration via Increased Systemic Chronic Inflammation

(A) Chronic inflammation is pertinent to brain aging in that inflammatory mediators in the periphery can damage the blood–brain barrier (BBB), permitting entry into the brain where they go on to disrupt neurons and glia and perpetuate a chronic inflammatory state. This directly contributes to various neurodegenerative pathways (illustrated) that lead to brain cell death. (B) Suggested mechanisms by which the causes of inflammaging (immunosenece, lifestyle, clinical health) and related consequences may drive brain health (structural and cognitive) outcomes. (C) Study model: chronic inflammation is a key driver of cognitive decline through its effects on brain structure. Left shows generic directed acyclic graph for mediation analysis (left panel) and for the study example (right panel). A = exposure; BDNF = brain-derived neurotrophic factor; C = confounder; CRP = C-reactive protein; IGF-1 = insulin-like growth factor 1; IL1\textbeta = interleukin-1\textbeta; IL6 = interleukin-6; M = mediator; PGE2 = prostaglandin E2; ROS = reactive oxygen species; SASP = senescence associated secretory phenotype; TNF-\alpha = tumor necrosis factor-\alpha; Y = outcome. Created with BioRender.com.
functions that require the coordination of brain regions (e.g., processing speed), compared to more functionally localized ones (e.g., verbal memory). However, we have not formally compared the magnitude of these attenuations, and judge that this greater degree of attenuation is likely to be a general shared process plus some degree of noise, given that variance across cognitive domains is shared at the general level.

The attenuation seen in inflammation–brain health associations when lifestyle factors were accounted for is to be expected given what is known about inflammation and VRFs on brain–health outcomes. Vascular inflammation is considered to be a shared mechanism linking cardiometabolic factors (to include hypertension and smoking) with poor cognitive outcomes. Upstream in serum CRP levels, which may explain why some VRFs show no association, such as in the Rotterdam Study,6 Whitehall II longitudinal cohort study,12 and Framingham Study.8 Similarly, while we have identified a range of health and lifestyle variables that could influence inflammatory load (BMI, diabetes, CVD history, smoking, alcohol consumption, hypertension), there are many nonmodeled variables that could contribute to this effect, as discussed in depth elsewhere.1,3

Finally, a clear limitation of the study is that our epigenetic surrogate of inflammation was measured in blood rather than brain tissue. While brain-based biomarkers are the optimal choice for investigating cognitive outcomes, it is impractical to profile such methylomes in brain tissue in living humans. Furthermore, the use of postmortem brain tissue samples has its own problems (in particular, the stability of global DNAm following death) and cannot reliably reflect the plastic state of methylomes in vivo. Future studies should consider examining a wider range of DNAm inflammatory markers (DNAm levels of interleukins, prostaglandins, and neurotrophins); DNAm inflammatory markers in younger participants (where there is likely greater variation in baseline inflammation levels); DNAm inflammatory markers in specific brain pathology cases (e.g., multiple sclerosis); as well as how peripheral inflammatory and neuroinflammatory DNAm patterns equate, and how each relates to cellular differences within the brain to give rise to the structural alterations we observe here.

Our findings do not establish causality but support the hypothesis that chronic systemic inflammation may contribute to neurodegenerative brain changes that underlie differences in cognitive ability in later life. Previous studies exploring this relationship may underestimate the brain and cognitive sequelae of chronic inflammation by relying on single measurements of phasic serum proteins. By using an epigenetic inflammation measure, which integrates information from multiple immune-related CpG sites, we may provide a more reliable measure of chronic inflammation and thus a more comprehensive overview of the consequences of chronic inflammation on brain structure and function. Reliable monitoring of inflammatory exposure could enable clinicians to review the efficacy of drug and lifestyle interventions to attenuate inflammation levels with a view to improving cognitive outcomes.

Acknowledgment
The authors thank the Lothian Birth Cohort 1936 members who took part in this study and the Lothian Birth Cohort 1936 research team members who collected, entered, and checked data used in this article. The Lothian Birth Cohort 1936 study
acknowledges the financial support of NHS Research Scotland through Edinburgh Clinical Research Facility.

**Study Funding**
This research was funded in whole or in part by the Wellcome Trust (grant number 108890/Z/15/Z). For the purpose of open access, the author has applied a CC-BY public copyright license to any author-accepted manuscript version arising from this submission. The LBC1936 and this research are supported by Age UK (Disconnected Mind project) and by the UK Medical Research Council (G0701120, G1001245, MR/M013111/1, MR/K026992/1). Methylation typing was supported by the Centre for Cognitive Aging and Cognitive Epidemiology (Pilot Fund award), Age UK, The Wellcome Trust Institutional Strategic Support Fund, The University of Edinburgh, and The University of Queensland. The Olink Proteomics assays were supported by a NIH research grant R01AG054628. This work was in part conducted in the Centre for Cognitive Aging and Cognitive Epidemiology, which is supported by the Medical Research Council and Biotechnology and Biological Sciences Research Council (MR/K026992/1) and which supports IJD. E.L.S.C. and A.J.S. are supported by funding from the Wellcome Trust 4-year PhD in Translational Neuroscience (108890/Z/15/Z to E.L.S.C., 203771/Z/16/Z to A.J.S.). M.E.B., I.J.D., and S.R.C. are also supported by an NIH research grant R01AG054628. S.R.C is supported by a Sir Henry Dale Fellowship jointly funded by the Wellcome Trust and the Royal Society (grant number 221890/Z/20/Z).

**Disclosure**
The authors report no disclosures relevant to the manuscript. Go to Neurology.org/N for full disclosures.

**Publication History**
This manuscript was prepublished in MedRxiv (doi: doi.org/10.1101/2020.10.08.20205245). Received by Neurology May 11, 2021. Accepted in final form October 15, 2021.

### Appendix (continued)

<table>
<thead>
<tr>
<th>Name</th>
<th>Location</th>
<th>Contribution</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sarah E. Harris, PhD</td>
<td>Lothian Birth Cohorts group, Department of Psychology, University of Edinburgh, UK</td>
<td>Drafting/revision of the manuscript for content, including medical writing for content; major role in the acquisition of data</td>
</tr>
<tr>
<td>Claire Green, MSc</td>
<td>Division of Psychiatry, Royal Edinburgh Hospital, University of Edinburgh</td>
<td>Drafting/revision of the manuscript for content, including medical writing for content</td>
</tr>
<tr>
<td>Maria del C. Valdés Hernández, PhD</td>
<td>Lothian Birth Cohorts group, Department of Psychology, University of Edinburgh, UK</td>
<td>Drafting/revision of the manuscript for content, including medical writing for content; major role in the acquisition of data</td>
</tr>
<tr>
<td>Matthew A. Harris, PhD</td>
<td>Division of Psychiatry, Royal Edinburgh Hospital, University of Edinburgh, UK</td>
<td>Drafting/revision of the manuscript for content, including medical writing for content; major role in the acquisition of data</td>
</tr>
<tr>
<td>Mark E. Bastin, DPhil</td>
<td>Centre for Clinical Brain Sciences, University of Edinburgh, UK</td>
<td>Drafting/revision of the manuscript for content, including medical writing for content; major role in the acquisition of data</td>
</tr>
<tr>
<td>Joanna M. Wardlaw, MD</td>
<td>Lothian Birth Cohorts group, Department of Psychology, University of Edinburgh, UK</td>
<td>Drafting/revision of the manuscript for content, including medical writing for content; major role in the acquisition of data</td>
</tr>
<tr>
<td>Ian J. Deary, PhD</td>
<td>Lothian Birth Cohorts group, Department of Psychology, University of Edinburgh, UK</td>
<td>Drafting/revision of the manuscript for content, including medical writing for content; major role in the acquisition of data; Study concept or design</td>
</tr>
<tr>
<td>Veronique E. Miron, PhD</td>
<td>UK Dementia Research Institute, Edinburgh Medical School, University of Edinburgh, UK</td>
<td>Drafting/revision of the manuscript for content, including medical writing for content; analysis or interpretation of data</td>
</tr>
<tr>
<td>Heather C. Whalley, PhD</td>
<td>Division of Psychiatry, Royal Edinburgh Hospital, University of Edinburgh, UK</td>
<td>Drafting/revision of the manuscript for content, including medical writing for content; analysis or interpretation of data</td>
</tr>
<tr>
<td>Riccardo E. Marioni, PhD</td>
<td>Lothian Birth Cohorts group, Department of Psychology, University of Edinburgh, UK; Centre for Genomic and Experimental Medicine, Institute of Genetics and Cancer, University of Edinburgh, UK</td>
<td>Drafting/revision of the manuscript for content, including medical writing for content; study concept or design; analysis or interpretation of data</td>
</tr>
<tr>
<td>Simon R. Cox, PhD</td>
<td>Lothian Birth Cohorts group, Department of Psychology, University of Edinburgh, UK</td>
<td>Drafting/revision of the manuscript for content, including medical writing for content; study concept or design; analysis or interpretation of data</td>
</tr>
</tbody>
</table>

### References


6.3 Conclusion

In summary, the results presented in section 6.2 demonstrate that a DNA methylation signature of inflammation exhibits larger associations with aspects of brain structure and cognitive function than serum protein CRP levels in a cohort of older individuals in their 70s and may be better placed to capture differences in brain and cognitive ageing related to inflammaging. This study builds on previous work that indicated that DNAm CRP trajectories are more stable than protein serum CRP levels overtime (Stevenson et al., 2020), and therefore may provide a more reliable capture of baseline inflammation levels. The results here suggest that this may offer relevance to examining the impact of chronic inflammation on the brain, both in terms of stratifying cognitive and brain structural differences and highlight the consequences of subtle elevations of inflammation for brain health outcomes in older age. In particular, several aspects of brain structure were found with higher DNAm CRP where serum CRP conferred no significance, such as reductions in total grey matter volume, white matter volume and increased burden of white matter hyperintensities. Health and lifestyle factors such as smoking, BMI and hypertension attenuated these associations by up to 40%, with smoking emerging as the largest driver of these. Finally, when the relationship between DNAm CRP and fluid intelligence ($g_f$) was examined with structural mediation modelling, normal appearing white matter volume was a significant mediator of the association between elevated inflammation and poor cognitive function, indicating that inflammation’s impact on white matter specifically may be a key driver of inflammation-associated cognitive decline.

This chapter focused on investigating the interrelationship between brain health outcomes, inflammation and DNA methylation in older age. The next chapter pivots to the opposite end of the lifecourse and focuses on examining these relationships in early life, looking at how chronic inflammation relates to neurodevelopment in a mixed cohort of term and preterm infants.
7 Chronic inflammation and brain health in early life

7.1 Introduction

Building on work outlined in Chapter 6.2, which demonstrated that DNAm signatures of inflammation show stronger associations with incident brain and cognitive health outcomes in later life, in this Chapter I applied this framework to an early-life cohort to better characterise inflammatory burden and its impact on neurodevelopment. Here, I hypothesised that variance in the neonatal methylome could reflect a convergence of amassed inflammatory burden and provide clarity on the relationship between inflammation, exposures in the perinatal period, and brain dysmaturation in relation to preterm birth. I use an inflammatory-related DNAm signature that has been validated previously in other cohorts at different stages of the human lifecourse (Barker et al., 2018; Stevenson et al., 2020; Conole et al., 2021; Green, Shen, et al., 2021), with this study sample being independent from the sample that generated the weights. To date, no DNAm signature of inflammation has been examined in neonatal populations, and few studies examine neuroimaging, inflammation and confounding maternal and infant health data simultaneously. This study addressed this research gap by measuring DNAm in 258 infants from Theirworld Edinburgh Birth cohort, deeply phenotyped maternal and infant health data and comprehensive neuroimaging metrics.

This study was published in Brain Behaviour and Immunity (May 2023) and is included in full in section 7.2. Supplementary material for this paper is available in the Appendix (11.2). All code used for these analyses are available at the following Github repository: https://github.com/EleanorSC/TEBC_DNAmCRP.
7.2 Immuno-epigenetic signature derived in saliva associates with the encephalopathy of prematurity and perinatal inflammatory disorders
Immuno-epigenetic signature derived in saliva associates with the encephalopathy of prematurity and perinatal inflammatory disorders

Eleanor L.S. Conole, Kadi Vaherd, Manuel Blesa Cabez, Gemma Sullivan, Anna J. Stevenson, Jill Hall, Lee Murphy, Michael J. Thrippleton, Alan J. Quigley, Mark E. Bastin, Veronique E. Miron, Heather C. Whalley, Riccardo E. Marion, James P. Boardman, Simon R. Cox

Keywords: Perinatal Inflammation DNA methylation Epigenetics Encephalopathy of Prematurity White Matter Preterm Birth Necrotizing Enterocolitis Diffusion Tensor Imaging Neurodevelopment Multiomics

ABSTRACT

Background: Preterm birth is closely associated with a phenotype that includes brain dysmaturation and neurocognitive impairment, commonly termed Encephalopathy of Prematurity (EoP), of which systemic inflammation is considered a key driver. DNA methylation (DNAm) signatures of inflammation from peripheral blood associate with poor brain imaging outcomes in adult cohorts. However, the robustness of DNAm inflammatory scores in infancy, their relation to comorbidities of preterm birth characterised by inflammation, neonatal neuroimaging metrics of EoP, and saliva cross-tissue applicability are unknown.

Methods: Using salivary DNAm from 258 neonates (n = 155 preterm, gestational age at birth 23.28 – 34.84 weeks, n = 103 term, gestational age at birth 37.00 – 42.14 weeks), we investigated the impact of a DNAm surrogate for C-reactive protein (DNAm CRP) on brain structure and other clinically defined inflammatory exposures. We assessed i) if DNAm CRP estimates varied between preterm infants at term equivalent age and term infants, ii) how DNAm CRP related to different types of inflammatory exposure (maternal, fetal and postnatal) and iii) whether elevated DNAm CRP associated with poorer measures of neonatal brain volume and white matter connectivity.

Results: Higher DNAm CRP was linked to preterm status (-0.0107 ± 0.0008, compared with -0.0118 ± 0.0006 among term infants; p < 0.001), as well as perinatal inflammatory diseases, including histologic chorioamnionitis, sepsis, bronchopulmonary dysplasia, and necrotising enterocolitis (OR range |2.00| to |4.71|, p < 0.001). Preterm infants with higher DNAm CRP scores had lower brain volume in deep grey matter, white matter, and hippocampi and amygdalae (β range [0.185] to [0.218]). No such associations were observed for term infants. Association magnitudes were largest for measures of white matter microstructure among preterms, where elevated epigenetic inflammation associated with poorer global measures of white matter integrity (β range |0.206| to |0.371|), independent of other confounding exposures.

Conclusions: Inflammatory-related DNAm captures the allostatic load of inflammatory burden in preterm infants. Such DNAm measures complement biological and clinical metrics when investigating the determinants of neurodevelopmental differences.

1. Introduction

Preterm infants are at an increased risk of elevated inflammation, related health complications, and adverse neurodevelopment compared to infants born at term (Back, 2015; Bennet et al., 2018; Hagberg et al., 2015; Inomata et al., 2014; Shah et al., 2008; Stoll et al., 2004; Humberg et al., 2020). While the aetiology of these outcomes is multifactorial, inflammation is considered to be a key component linking preterm birth
and poor neurodevelopmental and mental health outcomes via its effects on cerebral maturational processes (Malaez and Dammann, 2009; Reiss et al., 2022; Kelly et al., 2016; Favrais et al., 2011). Neonatal neuroimaging has identified neurostructural hallmarks of preterm birth commonly referred to as Encephalopathy of Prematurity (EoP), including dysmaturation of cortical and deep grey matter, atypical white matter development and disrupted connectivity (Boardman and Counsell, 2020). Recent advances in epigenetics may permit new ways to characterise sustained inflammation and reveal new insights into the relationship between inflammatory exposures, inflammation and neonatal brain and health outcomes.

Preterm infants are more susceptible to sustained inflammation than term infants and can be subject to multiple inflammatory stimuli during the perinatal period (Leviton et al., 2012; Dammann and Leviton, 2014). Alongside maternal lifestyle-related exposures (Chahal et al., 2017), various complications during pregnancy such as preeclampsia and histologic chorioamnionitis (Sullivan et al., 2021; Dammann et al., 2016) can induce both maternal and fetal inflammatory responses and increase the risk of a sustained pro-inflammatory state postnatally (Sullivan et al., 2021; Dammann et al., 2016; Yoon et al., 1997; Anblagan et al., 2016; Han et al., 2019). Preterm infants are additionally at higher risk for developing severe inflammatory conditions in the first few weeks of life, which may in turn perpetuate inflammation (Humberg et al., 2020) – including bronchopulmonary dysplasia (BPD), necrotizing enterocolitis (NEC), severe retinopathy of prematurity (ROP) and episodes of sepsis (Bassler et al., 2009). Preterm infants often present with multiple chronic conditions at once (Singh et al., 2019), putting them at higher risk of a greater allostatic load of inflammation (Leviton et al., 2012; Singh et al., 2019; Korzeniewski et al., 2014; Barnett et al., 2018).

Though numerous studies report associations between inflammation and cognitive outcomes in preterm populations (Humberg et al., 2020; Sulieri et al., 2022; Kuhn et al., 2017; O’Shea et al., 2013), research linking inflammatory biomarkers with neurostructural measures yield inconsistent findings (Inomata et al., 2014; Shah et al., 2008; Kelly et al., 2016; Yoon et al., 1997; Anblagan et al., 2016; Travis et al., 2015; Lee et al., 2021; Sullivan et al., 2020; Wu et al., 2021). Gaining a clearer understanding of the pathways via which sustained inflammation in very early life may precipitate well characterised cognitive and neurostructural outcomes requires novel approaches. The relative inconsistency of work to date is likely due to heterogeneity in study design: there is substantial variation in the demographic characteristics of study samples; the degree to which residual confounding factors are controlled; and the presence or absence of term control groups. Moreover, the relative nascentcy of the neonatal neuroimaging field (Korom et al., 2022) contributes to substantial variation in the acquisition and selection of brain outcome measures, and there is marked anatomic variation in early life, which can confound investigation of structural-function relationships (Dimitrova et al., 2020).

Above these factors, we argue that the measures used to characterise inflammation in the first place may account for the greatest source of ambiguity in the inflammation-brain structure literature. In both clinical and research settings, there is a historical reliance on sampling phasic inflammation-related protein measures from blood to signpost inflammation. Of these, C-Reactive Protein (CRP) is the most widely adopted (Brown et al., 2019), although there are criticisms of this approach (Bower et al., 2012; DeGoma et al., 2012). Accurate characterisation of sustained (and not transient or acute) inflammation arguably requires repeated sampling over long foliations, which few studies venture to profile (Macallister et al., 2019). Single recordings of CRP levels, such as those typically obtained for clinical purposes, may lead to misclassifications of baseline inflammation levels when used in population research contexts due to the high variation in baseline CRP within individuals over time (Bower et al., 2012; DeGoma et al., 2012; Bogaty et al., 2013; Nash et al., 2013). In the case of PTB, CRP levels are considered a less than ideal capture of inflammatory processes because the hepatic enzymatic machinery that generates CRP is maturation-dependent over the 2nd and 3rd trimesters of pregnancy, with studies suggesting that reference levels of CRP indicative of elevated inflammation should be dependent on gestational age (Macallister et al., 2019; Borowski et al., 2022; Chiesa et al., 2011; Chiesa et al., 2001; Hofer et al., 2012; Hofer and Resch, 2011; Matoba et al., 2009). Drawing blood from preterm infants (for research purposes) also has ethical implications, as this is an intrusive procedure which would require repeated cannulations to achieve a baseline reading. Because of these drawbacks, there is a precedent to find alternative ways to capture inflammatory burden to fully characterise its impact.

Our previous work demonstrated that DNA methylation (DNAm) markers of inflammation may provide more stable readouts of cumulative inflammatory exposure (Stevenson et al., 2018; Stevenson et al., 2020) and shed greater insight into the consequences of inflammation on brain structure (Conole et al., 2021; Green et al., 2021). DNAm is an epigenetic mechanism that can act as an interface by which environmental exposures influence gene function. DNAm is dynamic during fetal development, both in terms of the developing immune system (Martino et al., 2011) and brain (Spier et al., 2015) and may mediate the impact of maternal, fetal and postnatal exposures on brain development (Ozanne and Constancia, 2007). In the context of preterm birth, only a limited number of studies have investigated DNAm changes (Konwar et al., 2018; Liu et al., 2013; Merid et al., 2020; Sparrow et al., 2016; Winchester et al., 2022; Fumagalli et al., 2018; Chen et al., 2015; Wheeler et al., 2022; Camerota et al., 2021; Eversen et al., 2020) – of these, few examine DNAm in relation to neonatal neuroimaging metrics (Sparrow et al., 2016; Chen et al., 2015; Wheeler et al., 2022). Additionally, though some of these studies have examined DNAm in relation to postnatal health outcomes (Eversen et al., 2020; Massaro et al., 2021), no study to date has examined inflammation, DNAm, and neuroimaging concurrently in the neonatal period.

Here, using a cohort of 258 infants (103 term, 155 preterm), we examine (Back, 2015) how a salivary DNAm signature of the inflammatory marker CR-protein (DNAm CRP) relates to preterm birth (Bennet et al., 2018) how this signature associates with maternal, fetal and postnatal inflammatory exposures both individually and in aggregate, and (Hagberg et al., 2015) how variance in this measure relates to global measures of MRI brain volume, diffusion MRI (dMRI) correlates of connectivity, and regional variation in individual white matter tracts.

2. Methods

2.1. Study population

Preterm (gestational age at birth < 37 weeks) and term born infants delivered at the Royal Infirmary of Edinburgh, UK were recruited to the Theirworld Edinburgh Birth Cohort, a longitudinal study designed to investigate the effect of preterm birth on brain development (Boardman et al., 2020). Cohort exclusion criteria were major congenital malformations, chromosomal abnormalities, congenital infection, overt parenchymal lesions (cystic periventricular leukomalacia, hemorrhagic parenchymal infarction) or post-hemorrhagic ventricular dilatation. Ethical approval has been obtained from the National Research Ethics Service, South East Scotland Research Ethics Committee (11/55/0061, 13/SS/0143 and 16/SS/0154). Informed consent was obtained from a person with parental responsibility for each participant. DNAm data were available from 258 neonates, 214 of whom also had successful structural and diffusion MRI acquisition.

2.2. Study variables

Inflammatory exposures were coded as binary variables (1 = present, 0 = absent) and were grouped as follows: maternal (pertaining to mother / maternal exposure), fetal (affecting placenta or fetus) or neonatal (affecting infant after birth). Table 1 presents participant characteristics of these categories. Histologic chorioamnionitis (HCA)
Edinburgh Clinical Research Facility (Edinburgh, UK). The arrays were made and reagent (DNA Genotek, ON, Canada). DNA was bisulfite converted and previously (Wheater et al., 2022); for full details, refer to called automatically using GenomeStudio Analysis software version.

Table 1

Demographic and clinical features of study sample (n = 258). P values denote significant difference between term and preterm groups.

<table>
<thead>
<tr>
<th>Characteristics &amp; clinical features</th>
<th>Term infants (n = 103)</th>
<th>Preterm infants (n = 155)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Sex: Female (%)</td>
<td>44 (43)</td>
<td>75 (48)</td>
<td>0.2166</td>
</tr>
<tr>
<td>Gestational age at birth/weeks (range)</td>
<td>39.7 (37.00 – 42.14)</td>
<td>34.84</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Gestational age at scan/weeks (range)</td>
<td>42.27 (39.84 – 47.14)</td>
<td>40.56</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Birth weight/g (range)</td>
<td>3482 (2346 – 4670)</td>
<td>1177 (500 – 2100)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Birth weight z-score (range)</td>
<td>0.43 (2.30 – 2.96)</td>
<td>–0.19 (3.13 – 1.58)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>DNAm CRP (mean, SD)</td>
<td>–0.012 (0.001)</td>
<td>–0.011 (0.001)</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>

Maternal / fetal

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Term infants (n = 103)</th>
<th>Preterm infants (n = 155)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Maternal age (years)</td>
<td>33.7 (19 – 48)</td>
<td>31.1 (17 – 44)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Antenatal corticosteroid administration in pregnancy, n (%)</td>
<td>2 (1.9)</td>
<td>148 (95.5)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>MgSO4 administration in pregnancy, n (%)</td>
<td>0 (0)</td>
<td>112 (72.3)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Smoked during pregnancy (%)</td>
<td>2 (1.9)</td>
<td>29 (19.0)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Histologic choorioamnionitis, n (%)</td>
<td>6 (5.8)</td>
<td>49 (31.6)</td>
<td>&lt;0.001</td>
</tr>
<tr>
<td>Preeclampsia, n (%)</td>
<td>7 (6.8)</td>
<td>22 (14.2)</td>
<td>&lt;0.001</td>
</tr>
</tbody>
</table>

Neonatal

<table>
<thead>
<tr>
<th>Characteristics</th>
<th>Term infants (n = 103)</th>
<th>Preterm infants (n = 155)</th>
<th>P value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Necrotizing enterocolitis, n (%)</td>
<td>0 (0)</td>
<td>10 (6)</td>
<td>N/A</td>
</tr>
<tr>
<td>Bronchopulmonary dysplasia, n (%)</td>
<td>0 (0)</td>
<td>48 (31)</td>
<td>N/A</td>
</tr>
<tr>
<td>Retinopathy of prematurity, n (%)</td>
<td>0 (0)</td>
<td>8 (5)</td>
<td>N/A</td>
</tr>
<tr>
<td>Sepsis, n (%)</td>
<td>0 (0)</td>
<td>36 (23)</td>
<td>N/A</td>
</tr>
</tbody>
</table>

was defined via placental histopathology, as reported previously (Sullivan et al., 2021; Anblagan et al., 2016). Incidence of any neonatal sepsis (either late onset or early onset sepsis) was defined as detection of bacterial pathogen from blood culture, or physician decision to treat for ≥ 5 days in the context of growth of coagulase negative staphylococcus from blood or a negative culture. Necrotising enterocolitis (NEC) was defined as stages two or three according to the modified Bell’s staging for NEC. Bronchopulmonary dysplasia (BPD) was defined by the requirement for supplemental oxygen at 36 weeks gestational age. Birthweight z-scores were calculated according to International Fetal and Newborn Growth Consortium for the 21st Century (INTER-GROWTH-21st) standards. Further details on variable selection and classification are provided in supplementary methods.

2.3. DNA extraction and methylation measurement and pre-processing

Saliva was obtained on the same day of MRI acquisition which was at term equivalent age (TEA) for preterm infants and shortly after birth for term infants (Table 1). Saliva was collected in Oragene OG-575 Assisted Collection kits, by DNA Genotek, and DNA extracted using prepIT.L2P reagent (DNA Genotek, ON, Canada). DNA was bisulfite converted and methylation levels were measured using Illumina Human-Methylation EPIC BeadChip (Illumina, San Diego, CA, USA) at the Edinburgh Clinical Research Facility (Edinburgh, UK). The arrays were imaged on the Illumina iScan or HiScan platform and genotypes were called automatically using GenomeStudio Analysis software version 2011.1 (Illumina). Details of DNAm pre-processing have been outlined previously (Wheater et al., 2022); for full details, refer to supplementary methods.

2.4. Inflammatory-related methylation signature

For each individual (n = 258), a weighted linear signature (DNAm CRP) was obtained by multiplying the methylation proportion at a given CpG by the effect size from a previous epigenome wide association study (EWAS) of CRP (Ligthart et al., 2016) (supplementary Table 1), and then summing these values (see equation (1) below).

where “cpg” is the normalised methylation value for the LBC1936 participant at a given site and “b” is the effect size from Ligthart et al., (2016). The original CRP EWAS examined peripheral blood DNA methylation profiles in relation to circulating CRP levels across multiple adult cohort studies (see Fig. 1). This method to generate an inflammatory-related DNAm score has been used previously in various population cohorts to index cumulative inflammation (Stevenson et al., 2020; Conole et al., 2021; Green et al., 2021; Barker et al., 2018).

2.5. MRI acquisition

This study incorporates data from two phases of MRI acquisition which is reflected in the flowchart of the study sample (supplementary Figure S3). The data acquisition of this study has been reported previously (Wheater et al., 2022).

In the first phase (n = 93), structural and dMRI were performed in neonates using a MAGNETOM Verio 3 T clinical MRI scanner (Siemens Healthcare GmbH, Erlangen, Germany) and 12-channel phased-array head coil. For dMRI, A protocol consisting of 11 baseline volumes (b = 0 s/mm² [b0]) and 64 diffusion-weighted (b = 750 s/mm²) single-shot spin-echo planar imaging (EPI) volumes acquired with 2 mm isotropic voxels (TR/TE 7300/106 ms) was used; 3D T1-weighted (T1w) MPRA GE (TR/TE 1650/2.43 ms) with 1 mm isotropic voxels was acquired.

For the second phase (n = 121), structural and dMRI were performed neonates using a MAGNETOM Prisma 3 T clinical MRI scanner (Siemens Healthcare GmbH, Erlangen, Germany) and 16-channel phased-array pediatric head and neck coil. This was used to acquire dMRI in two separate acquisitions: the first consisted of 8 b0 and 64 volumes with b = 750 s/mm², the second consisted of 8 b0, 3 volumes with b = 200 s/ mm², 6 volumes with b = 500 s/mm² and 64 volumes with b = 2500 s/ mm². An optimal angular coverage for the sampling scheme was applied (Caruyer et al., 2013). In addition, an acquisition of 3 b0 volumes with an inverse phase encoding direction was performed. All dMRI volumes were acquired using single-shot spin-echo planar imaging (EPI) with 2-fold simultaneous multi-slice and 2-fold in-plane parallel imaging acceleration and 2 mm isotropic voxels; all three diffusion acquisitions had the same parameters (TR/TE 3500/78.0 ms). Images acquired by motion artifact were re-acquired multiple times as required; dMRI acquisitions were repeated if signal loss was seen in 3 or more volumes. 3D T2-weighted SPACE images (T2w) (TR/TE 3200/409 ms) with 1 mm isotropic voxels and 3D T1w MPRA GE (TR/TE 1970/4.69 ms) with 1 mm isotropic voxels were also acquired.

Infants were fed and wrapped and allowed to sleep naturally in the scanner without sedation. Pulse oximetry, electrocardiography and temperature were monitored. Flexible earplugs and neonatal earmuffs (MiniMuffs, Natus) were used for acoustic protection. All scans were supervised by a doctor or nurse trained in neonatal resuscitation. Structural images were reported by an experienced pediatric radiologist (A.J.Q.), and each acquisition was inspected contemporaneously for motion artefact and repeated if there had been movement while the baby was still sleeping; dMRI acquisitions were repeated if signal loss was seen in 3 or more volumes.

As details on dMRI pre-processing have been previously outlined (Blesa et al., 2021) please refer to supplementary methods for specifics. T2w images from phase 2 were processed using the dHCP pipeline.
The T1w images from phase 1 were processed using specific software for brain skull-stripping and tissue segmentation (Doshi et al., 2013). The phase 1 pipeline relies on some atlases, for these purposes, 10 subjects from the phase 2 that have both T1w and T2w were selected. The volumes extracted include cortical grey matter, deep grey matter, white matter, hippocampi and amygdalae, cerebellum, brainstem, cerebrospinal fluid (CSF) and ventricles.

From the diffusion images we calculated the tensor – fractional anisotropy (FA), mean diffusivity (MD), axial diffusivity (AD), and radial diffusivity (RD) – and the NODDI (intracellular volume fraction [NDI] maps) (Zhang et al., 2012). All the subjects were registered to the Edinburgh Neonatal Atlas (ENA50) using DTI-TK (Zhang et al., 2012; Blesa et al., 2020). The diffusion tensor derived maps of each subject (FA and MD) were calculated after registration; NDI was then propagated to the template space using the previously calculated transformations. The data was skeletonized using the ENA50 skeleton and then multiplied by

---

**Fig. 1.** DNAm CRP signature pipeline. The inflammatory-related DNAm signature used in this study is comprised of immune-related CpG sites identified from a multi-cohort EWAS of CRP (Ligthart et al., 2016) which examined DNAm in relation to circulating CRP levels in peripheral blood. Relative weights for the CpG sites of interest are reported in supplementary Figure 1.
a custom mask. Finally, the peak width of the histogram of values computed within the skeletonised maps was calculated as the difference between the 95th and 5th percentiles. Global values of white matter microstructure reported in this study are the peak width of skeletonised from the same pipeline previously used to characterise brain structural differences between preterm and term infants (Blesa et al., 2020).

2.6. Tract segmentation and extraction of tract-averaged dMRI metrics

As above, details of individual white matter tract segmentation and subsequent extract of tract-averaged dMRI metrics have been outlined previously from infants from this study sample (Vaher et al., 2022). Briefly, FA and MD were derived for the left and right hemispheric tracts of the arcuate fasciculus (AF), anterior thalamic radiation (ATR), cingulum cingulate gyrus (CCG), corticospinal tracts (CST), inferior fronto-occipital fasciculus (IFOF), inferior longitudinal fasciculus (ILF), uncinate fasciculus (UNC) and genu and splenium of the corpus callosum (CC).

2.7. Statistical analysis

All statistical analyses were performed in R (version 4.0.5) (R Core Team, 2020).

2.7.1. Selection of covariates and confounding variables

In comparing participant demographics between term infants (n = 103) and preterm infants (n = 155), p-values derived from the t-test (for continuous variables) and Chi-square test or Fischer’s Exact test if the count was below 5 (categorical variables) (Table 1) (Kim, 2017). In all models examining the association between DNAm CRP with health outcomes and brain MRI metrics, we adjusted for infant sex, gestational age at birth, gestational age at scan, and birthweight Z score. Scanner variable was included when data included two phases of MRI acquisition. We also tested for an interaction between gestational age × DNAm CRP and infant sex × DNAm CRP in sensitivity analyses, where a significant interaction would indicate differences in association magnitudes at different gestational ages / between males and females. To quantify the amount of variance in each brain imaging biomarker accounted for by DNAm CRP, for all neuroimaging associations we report both the adjusted R² and the incremental R² (Tzoulaki et al., 2009), the latter of which was calculated by comparing the R² of each model with that from a baseline model (reported as Model H₀) R² in which the MRI measure was modelled with covariates only (e.g., Model H₀ = MRI metric ~ sex + birthweight + gestational age + gestational age at scan + scanner).

After examining global associations, we wanted to control for variables that could either cause the raised DNAm CRP (the exposure), variance in MRI metrics (the outcome), or both. We ran bivariate correlations and from these we included all common correlates of the exposure and the outcome in our second regression model (Model H₂) to eliminate alternative explanations of the outcome due to confounding (supplementary Figure 2). In these models, neither antenatal treatment of corticosteroids and MgSO₄ for anticipated preterm birth were included to circumvent issues of multicollinearity, since they were given to the majority of mothers (96 % and 72 % respectively) in the preterm group and were highly correlated with preterm status (r = 0.71 – 0.93, p < 0.001). The magnitude of effects are classified as small, medium, or large when the standardized coefficients are 0.1, 0.3, or 0.5, respectively as classified by Cohen (Cohen, 1992).

Finally, accounting for the fact that inflammatory risk factors are positively correlated, we included all inflammatory risk factors in one multiple linear regression alongside DNAm CRP signature for each MRI variable of interest. This allowed us to account for unique contribution of DNAm CRP in the context of inflammatory-related exposures to variance in brain structural outcomes. For further details on the selection of covariates in this study, see supplementary methods.

2.7.2. Multiple inflammatory hits and DNAm CRP

We next performed investigations to assess whether DNAm was related to number of inflammatory episodes experienced. Due to small numbers of individual inflammatory risk factors and the frequent overlap of episodes experienced in the preterm group, we created binary outcome measures based on combinations of inflammatory risk factors or conditions experienced, combining infants that experienced three or more morbidities into a single group, resulting in four possible levels for the risk score of 0, 1, 2, or 3 + alongside a term control reference (0 inflammatory episodes). Results are presented firstly unadjusted (model H₀), then adjusted for gestational age at birth, infant sex and birthweight z-score (model H₂), and then adjusted for gestational age at birth, infant sex and birthweight z-score as well as administration of MgSO4 and corticosteroids in pregnancy (model H₃), given these have been identified as potential confounders of the relationship between inflammation and health outcomes in previous studies (Lingam and Robertson, 2018; Odufalu et al., 2022). Other potential covariates, such as gestational diabetes and maternal age, were not found to significantly associate with inflammation (supplementary Figure 1), so were not controlled for in these models. Results are presented as odds ratios (OR) and 95 % confidence intervals (CI) for categorical outcome measures.

2.7.3. Dnam CRP and global brain structure associations

To determine the effect of inflammation (DNAm CRP) on neuroimaging outcomes, data were analysed using regression models, controlling for factors considered relevant to inflammation and EoP outcomes (see supplementary methods for details on selection procedure of covariates). For baseline models, these controlled factors were: gestational age at birth, gestational age at scan, infant sex, MRI scanner, and birthweight Z score; visual inspection of diagnostic plots suggested no regression assumptions were violated (an example is provided in supplementary figure 4). We aimed to contextualise these associations with clinical health data. Inflammatory risk factors were added simultaneously as covariates into a second model (models H₂ and H₃) in addition to the standard covariates of gestational age at birth, gestational age at scan, birthweight Z score and sex (baseline models, H₀ and H₁). As no term infants had postnatal inflammatory episodes (sepsis, NEC, ROP, BPD), analyses were stratified according to term or preterm status; an overview of these models is provided in supplementary figure 5. Brain volume metrics were corrected for ICV. In models testing global brain structural metrics such as brain volumes and PSMD and PSFA, MRI scanner was included as a binary covariate as MRI data from both phases of data collection were included (refer to supplementary figure 3, study sample flowchart). All continuous variables were standardised using z-score scaling to obtain standardised effect sizes (β). P-values were corrected for multiple testing using the false discovery rate (FDR) method and significance was deemed FDR corrected p-value (pFDR) < 0.05. 95 % CIs are reported throughout.

2.7.4. Dnam CRP and dMRI white matter tract associations

dMRI measures of white matter appear to be highly correlated (e.g. high FA in an individual tract such as the arcuate fasciculus is often accompanied by high FA across all other white matter tracts in that individual), a property that persists from early infancy through to older age (Vaher et al., 2022). As a result of this, it is common to derive general factors (g-factors) of white matter microstructure to characterise global white matter microstructure. One PCA was conducted for FA and MD parameters across the 16 tracts to quantify the proportion of shared variance between them; in each analysis, each subject was described by 16 features, computed as the tract-averaged values of FA or MD across each tract (supplementary Figure S6). The first unrotated principal component (PC) scores were extracted as the single-metric g-factors, gFA and gMD (scree plot and PCA variable contributions illustrated in supplementary Figure S7).
Fig. 2. Multiple inflammatory hits associate with raised DNAm CRP (A) distributions of DNAm CRP according to number of inflammatory episodes experienced by infant (B) Venn diagram showing the overlap postnatal inflammatory morbidities in study sample (C) Odds ratios and 95% confidence intervals for contribution of DNAm CRP to inflammatory exposures, asterisks (*) indicate statistically significant (FDR-corrected p < 0.05) (D) Scatter plots of the relationships between gestational age and birthweight, coloured according to number of inflammatory episodes/exposures (top panel) and DNAm CRP (bottom panel). Models are based on full sample (n = 258); for further details see supplementary Figure 8.
As with global brain structural metrics, two models were used:

Model H$_1$: dMRI metric $\sim$ DNAm CRP + gestational age at scan + gestational age at birth + infant sex + birthweight

Model H$_2$: dMRI metric $\sim$ DNAm CRP + gestational age at scan + gestational age at birth + infant sex + birthweight + all inflammatory risk factors

In comparison to global MRI volumetric metrics and PSFA, PSMD, PSAD and PSRD, all individual tract associations, gFA, gMD and PSNDI were limited to neuroimaging data from phase 2 of the study, hence no scanner variable was included in these analyses.

### 2.8. Data and code availability

Requests for original image and anonymised data will be considered through the BRAINS governance process (https://www.brainsimagelbank.ac.uk). Raw DNAm data are available upon request from Theirworld Edinburgh Birth Cohort, University of Edinburgh (https://www.tebc.ed.ac.uk/2019/12/data-access-and-collaboration), while DNAm and metadata are not publicly available, generated DNAm CRP signatures are included alongside scripts for data analysis. All brain volumetric metrics were obtained using the scripts provided in https://github.com/amaakropoulos/structural-pipeline-measures. The segmented tracts in the ENA50 template space are available online: https://git.ecdf.ed.ac.uk/jbrl/ena. Code for primary data analysis and figures are available at https://github.com/EleanorSC/TEBC_DNAmCRP and code for tract propagation and average calculation are available at https://git.ecdf.ed.ac.uk/jbrl/neonatal-gfactors.

### 3. Results

#### 3.1. Participant characteristics

The study group consisted of 258 neonates: 155 participants were preterm and 103 were controls born at full term, see Table 1 for participant characteristics and supplementary figure 3 for a flowchart of data acquisition. Among the preterm infants, 48 (31 %) had bronchopulmonary dysplasia, 10 (6 %) developed necrotising enterocolitis, 8 (5 %) developed ROP, 49 had HCA (32 %), 22 (15 %) were born to women whose pregnancy was complicated by preeclampsia, and 36 (23 %) had an episode of postnatal sepsis. Of the 258 participants with DNAm data, 214 also had MRI data. Correlations between all variables are provided in supplementary figure 1.

#### 3.2. Multiple inflammatory hits increase risk of elevated inflammation

Preterm Infants in the sample for whom DNAm data and composite neonatal inflammatory risk scores were available (n = 155), had high prevalence (n = 112, 72 %) of experiencing at least one of the documented inflammatory exposures (Fig. 2B), which included incidence of smoking during pregnancy, preeclampsia, HCA, sepsis, BPD, NEC or ROP. A small subset of these infants experienced three or more of these exposures (n = 24, 15 %).

There was an association between number of inflammatory episodes and the inflammatory-related DNA methylation signature, with higher DNAm CRP in infants who had experienced greater exposure to inflammation. DNAm CRP was associated with higher odds of several perinatal morbidities including HCA, sepsis, BPD, and NEC. These relationships remained significant following adjustment for gestational age at birth, birthweight, and infant sex as well as perinatal variables of administration of corticosteroids and MgSO$_4$ in pregnancy (Fig. 3C, supplementary Table S3). The association of DNAm CRP with ROP was no longer significant after controlling for MgSO$_4$ and corticosteroid administration (model H$_2$). DNAm CRP was also associated with three or more inflammatory episodes. There was no significant association of DNAm CRP with maternal smoking in pregnancy or preeclampsia. Infants with increasing numbers of complications were more likely to be gestationally younger at birth and have lower birthweights (Fig. 2D). Furthermore, preterm infants had significantly higher DNAm CRP (Table 1, p < 0.001). When examining DNAm CRP alongside clinical inflammatory exposures (supplementary Table S4), there was no significant difference between term infants with no inflammatory episodes vs those with one (a breakdown of the 14 infants who had an inflammatory exposure record is provided in supplementary Figure SS; of note, term infants only had antenatal inflammatory exposures, as no term infants went on to develop a postnatal inflammatory condition, or experience neonatal sepsis). The largest difference was found between term infants with no inflammatory episodes and preterm infants with 3 or more inflammatory risk-factors (p < 0.001).

#### 3.3. DNAm CRP and brain volumes

Overall, magnitudes of associations between DNAm CRP and global MRI brain volumes were modest, explaining a small amount of additional variance beyond covariates (of infant sex, gestational age at birth, birthweight z-score, gestational age at scan, scanner variable). After examining inflammation-brain structure associations across all infants (n = 214; supplementary Table SS), we stratified analyses into term (n = 87) and preterm (n = 127) subgroups to examine group differences (Fig. 3B). Term infants displayed no significant brain structural associations with DNAm CRP, whereas preterm infants with higher DNAm CRP displayed brain volume reductions in deep grey matter, white matter, and hippocampi and amygdalae (Fig. 3B).

Analyses were repeated to include interactions between DNAm CRP and both sex and gestational age (supplementary Figure S9). While null findings were observed with the former (p > 0.05; supplementary Table S7), within the preterm cohort there was evidence for interactions with gestational age at birth (supplementary Table S8). Higher DNAm CRP was consistently associated with lower brain volume in infants of lower gestational ages (i.e. extremely preterm infants tended to have higher DNAm CRP and correspondingly smaller global brain volume measures). In contrast, there was no significant interaction between gestational age and DNAm CRP within the term sub-group (p > 0.05; supplementary Table S9).

In fully adjusted models (Fig. 4, supplementary Table S11), where analyses were conducted separately for a term control model (n = 87) and preterm subgroup (n = 127) where aggregate inflammatory risk factors were examined separately (models H$_2$-H$_3$) there remained a significant association of DNAm CRP with deep grey matter volume ($\beta$ = -0.206, p = 0.008), white matter volume ($\beta$ = -0.346, p = 0.0006), and cerebellum volume ($\beta$ = -0.201, p = 0.013). For most brain metrics, the strength of the association between DNAm CRP and MRI metric was increased when additional inflammatory covariates were included in the model (percentage increase for deep grey matter volume = 6 %, white matter volume = 39 %, and cerebellum volume = 27 %). Individually modelling risk factors revealed that this increase was mostly driven by controlling for incidence of sepsis, whereas brain structural associations were most attenuated by controlling for incidence of BPD (Fig. 4).

#### 3.4. Global white matter microstructure associations with DNAm CRP

Preterm infants with higher DNAm CRP had poorer measures of white matter tract integrity. This was seen at the global level for all peak width of skeletonised white matter microstructure metrics (Table 2); PSFA, PSMD, PSRD, PSAD; $\beta$ range $[0.186]$ to $[0.341]$, incremental $R^2$ 2.7 - 9 %. In term infants, there were no significant associations (supplementary Table S10). As with global brain volumetric measures, there were significant interactions between gestational age and DNAm CRP across measures of white matter integrity excepting PSRD: PSFA (interaction $\beta$ = -0.225; main effect $\beta$ = -0.212), PSMD (interaction $\beta$ = -0.257; main effect $\beta$ = -0.212).
0.371) and PSAD (interaction $\beta = -0.271$; main effect $\beta = 0.232$), indicating infants at younger gestational ages were more likely to have poor white matter integrity with high DNAm CRP. When controlling for inflammatory risk factors, DNAm CRP associations between PSRD and PSAD were no longer significant.

Within a smaller subgroup of this sample, individual tract FA and MD as well as neurite density index data was available (Phase 2, refer to supplementary figure 3 for study flow diagram). PCA-derived single-metric g-factors, gFA and gMD (scree plot and PCA variable contributions illustrated in supplementary Figure S7) were almost exactly correlated with those previously reported in a larger sample of Their-world Edinburgh Birth cohort infants (Vaher et al., 2022). When examining the association between DNAm CRP and global white matter measures in this subsample, the most striking association was seen with differences in a general factor of fractional anisotropy, gFA ($\beta = -0.52$ [95% CI −0.304, −0.736], $p = 1.48 \times 10^{-5}$, incremental $R^2 = 22 \%$) and mean diffusivity, gMD ($\beta = 0.423$ [95% CI 0.661, 0.191], $p = 7.79 \times 10^{-4}$, incremental $R^2 = 14 \%$). These effect sizes were attenuated by controlling for additional inflammatory risk factors (model H2) but remained significant ($\beta$ range |0.35| to |0.37|, $p < 0.05$). No significant associations were found between DNAm CRP with PSNDI.

3.5. Individual white matter tract associations with DNAm CRP

We next examined associations between DNAm CRP and individual tract-averaged FA and MD. In all models, term infants displayed no significant tract associations with DNAm CRP (supplementary Figure S10). In preterm infants, altered FA was present in both hemispheric tracts of the AF, CST, IFOF, ILF, UNC and CCG (Fig. 5 shows tract-averaged fractional anisotropy for each of the 16 tracts for the term and preterm neonates). Some tract associations were specific to hemisphere such as decreased FA in the left (but not right) ATR. Equally, altered FA and MD was present in only the genu (and not splenium) of the corpus callosum. A breakdown of all dMRI results is reported in supplementary Table S12. However, after adjusting for additional inflammatory risk factors, (in order of effect size) only FA in the right corticospinal tract (15.4 %), right AF (10 %), and right CCG (8.7 %) remained significant. For tract MD, bilateral increases in MD were observed in the AF, CST, IFOF, ATR and CCG. Hemispheric specific associations were found for the left ILF, left ATR and genu of the corpus callosum. Of these associations, AF and ILF were no longer significant when accounting for additional inflammatory exposures.

4. Discussion

In this study, we integrated data from placenta, saliva and brain MRI in a large cohort of 258 infants to characterise the association of inflammation with brain structure. Previous research has typically relied on C-Reactive Protein (CRP) to measure inflammation, yet CRP’s fluctuating nature, variability within short timeframes (which can lead to
individuals displaying transiently elevated or lowered CRP levels in population studies), and the developmental-dependant physiology of hepatocytes, renders it an inadequate tool for profiling inflammation in preterm populations. We demonstrate that a composite buccal-cell DNA methylation measure of inflammation trained in adult peripheral blood samples associates with comorbidities of preterm birth that are characterised by a pro-inflammatory state and widespread differences in brain structure among preterm infants. The inflammatory-related DNAm signature was particularly associated with white matter dysmaturation at term-equivalent age both globally and at the level of individual white matter tract microstructure, associations that largely remained significant when accounting for inflammatory exposures. These data motivate further research into the potential of immune-DNAm markers for translational medicine in the neonatal period as diagnostic tools for identifying those at risk for inflammatory-related morbidities and neurodevelopmental impairment.

Fig. 4. Associations between DNAm CRP and brain volumes and the impact of inflammatory risk factors on associations. Top panel displays schematic of different perinatal inflammatory exposures controlled for, including smoking in pregnancy, preeclampsia, histologic chorioamnionitis (HCA), bronchopulmonary dysplasia (BPD), necrotizing enterocolitis (NEC), severe retinopathy of prematurity (ROP) and neonatal sepsis. Bottom panel displays standardized regression coefficients for DNAm CRP associations between brain volumetric measures, where points show standardized coefficients and 95% confidence intervals. The alpha of points denotes whether associations were significant, with pFDR < 0.05 outlined by dark points and non-significant associations faded points. The two main models, H1 and H3, represent baseline models run on term (n = 87) and preterm (n = 127) subgroups respectively; these, and all subsequent models, controlled for infant sex, gestational age at birth, gestational age at scan and birthweight Z score and scanner variable. Models H2 and H4 control for all inflammatory risk factors in aggregate in term and preterm infants respectively. The additional shapes show the degree to which these associations, in preterms, were influenced by accounting for allied inflammatory exposures individually (H3-H12, corresponding with supplementary Table S11).
### 4.1. Inflammatory-related DNA methylation biomarkers

There has been increasing interest in using methylation data to advance our understanding of the causes and consequences of preterm birth as outlined in several reviews (9, 79, 80). Only recently has attention turned to exposures in the perinatal period, the role of epigenetics in utero for neurodevelopment, and the potential for peripherally sampled DNA to capture the impact of environmental exposure in relation to brain and cognitive outcomes (Barker et al., 2018). Among these developments are the use of methylation risk scores of exposure, which integrate information from multiple CpG sites to provide a record of exposure or to capture a complex trait (Bakulski and Fallin, 2014; Bakulski et al., 2021; Suarez et al., 2020, Yousefi et al., 2022). This approach has been used to examine maternal smoking (Odintsova et al., 2021; Reese et al., 2017; Richmond et al., 2015; Richmond et al., 2018) glucocorticoid, and prenatal folate exposure during pregnancy (Bakulski et al., 2021; Suarez et al., 2020), alongside environmental exposures such as pollution (Suarez et al., 2020). Examining DNA methylation of inflammation is uncommon, but given the health associations with inflammatory-related DNA in adult cohorts (Conole et al., 2021; Green et al., 2021; Somineni et al., 2019), and the shared nature of epigenetic changes between mothers and infants (Camerota et al., 2021; Sasaki et al., 2022), we hypothesised that variance in the neonatal methylome could reflect a convergence of amassed inflammatory burden from different perinatal origins.

### 4.2. DNA methylation with gestational age and multiple inflammatory exposures

Our findings suggest that epigenetics offers a solution to the traditional limitations of assessing inflammatory burden in infancy (which include the phasic nature of CRP responses, particularly in preterm infants where CRP may be an inappropriate marker owing to the development-related physiology of the liver, and unethical serial sampling in vulnerable neonates).

Preterm infants displayed higher DNA methylation than term infants, and associations between DNA methylation and postnatal health and brain outcomes were restricted to the preterm infants. This novel finding that gestational age at birth correlates strongly with inflammatory-related DNA methylation (r = -0.62, p < 0.001) aligns with previous observations of elevated inflammatory protein concentrations with lower gestational ages and prematurity (Humberg et al., 2020; Dammann and Leviton, 2014), lending further weight to the validity of this measure for carrying clinical significance. Equally, finding that inflammation-related alterations in brain structure were more pronounced in preterms who were gestationally younger (supplementary Figure S9) highlights the dose-dependent effect on brain structure, whereby increasing prematurity (lower gestational age) is associated with higher inflammatory burden and related structural consequences. These findings are in line with previous studies which demonstrate that extremely preterm infants are at enhanced risk of neurodevelopmental impairment, with sustained postnatal inflammation significantly increasing this risk (Leviton et al., 2012; Korzeniewski et al., 2014; Barnett et al., 2018; Glass et al., 2018). We speculate that the enhanced vulnerability of extremely preterm infants is due to rapid developmental changes during the second and third trimester of pregnancy for both the developing immune system and brain – in particular, the disruptive impact of inflammation on neurogenesis, neuronal migration, synaptogenesis and myelination (Hagberg et al., 2015). As these processes are highly dynamic during these periods and early postnatal life, preterm infants are both more susceptible to sustained inflammation and neurodevelopmental disruption. Relative differences in immune system maturity and neurodevelopmental processes between extremely preterm infants, preterm and term infants likely explain these findings.

Our finding that multiple inflammatory hits contributed to higher DNA methylation strengthens the hypothesis that DNA methylation may index the allostatic load of inflammation during neonatal intensive care. In both preclinical studies and cohort groups, preterm infants with multiple inflammatory episodes or morbidities display an increased risk for brain structural abnormalities compared to infants who had only one inflammatory episode or condition recorded (Glass et al., 2018; Yanni et al., 2017; Fleiss et al., 2015). We also find stronger relationships between DNA methylation and postnatal inflammatory factors (NEC, BPD, sepsis) than antenatal factors (preeclampsia, maternal smoking), indicating early-life exposures contribute to greater variance in DNA methylation as compared to maternal inflammation (as depicted in Fig. 2C). The multi-hit hypothesis of sustained inflammation (Leviton et al., 2012; Barnett et al., 2018; Yanni et al., 2017) suggests that postnatal health complications related to preterm birth can perpetuate a chronic inflammatory state, with timing of insults a key factor for why preterm infants are more susceptible than term infants to sustained inflammation (Ophelders et al.,

### Table 2

| associations between DNA methylation (DNAm) CRP and global white matter microstructure and the impact of inflammatory risk factors on associations in preterm infants; standardized regression coefficients for DNA methylation (DNAm) CRP associations between global white matter microstructure metrics for preterm infants. Betas (standardized coefficients) and 95% confidence intervals are reported. Bold text indicates statistically significant association (FDR-corrected p < 0.05). Model H1 controls for infant sex, gestational age at birth, gestational age at scan birthweight Z-score and scanner variable; model H2 additionally controls for inflammatory risk factors and associated morbidities (maternal smoking in pregnancy, preeclampsia, HCA, sepsis, BPD, NEC and ROP). |
|----------------|----------------|----------------|-----|-----|-----|-----|-----|-----|
| Model H1       | beta           | lower CI       | upper CI      | p   | r²  | additional r² | n   |
| PSFA           | -0.186         | -0.324         | -0.048        | 0.009 | 0.540| 0.027         | 127 |
| PSMD           | 0.341          | 0.166          | 0.517         | 2.17E-04 | 0.256| 0.090         | 127 |
| PSRD           | 0.312          | 0.122          | 0.501         | 0.002 | 0.130| 0.075         | 127 |
| PSAD           | 0.201          | 0.030          | 0.372         | 0.023 | 0.294| 0.031         | 127 |
| gFA            | -0.520         | -0.736         | -0.305        | 1.48E-05 | 0.441| 0.216         | 64  |
| gMD            | 0.426          | 0.191          | 0.661         | 0.001 | 0.333| 0.145         | 64  |
| PSNDI          | 0.089          | 0.321          | 0.142         | 0.452 | 0.354| 0.006         | 64  |
| Model H2       | beta           | lower CI       | upper CI      | p   | r²  | additional r² | n   |
| PSFA           | -0.215         | -0.375         | -0.055        | 0.009 | 0.577| 0.026         | 127 |
| PSMD           | 0.206          | 0.009          | 0.403         | 0.042 | 0.357| 0.024         | 127 |
| PSRD           | 0.175          | 0.041          | 0.392         | 0.115 | 0.222| 0.017         | 127 |
| PSAD           | 0.093          | 0.095          | 0.280         | 0.336 | 0.414| 0.005         | 127 |
| gFA            | -0.371         | -0.617         | -0.124        | 0.005 | 0.573| 0.073         | 64  |
| gMD            | 0.355          | 0.082          | 0.627         | 0.014 | 0.477| 0.067         | 64  |
| PSNDI          | -0.123         | -0.419         | 0.173         | 0.420 | 0.385| 0.008         | 64  |

*a* scanner variable is controlled for when examining PS metrics but not gFA, gMD and PSNDI (single-scanner sample).
Previous work has shown that a paediatric buccal-cell derived epigenetic age acceleration measure (PedBE) is associated with adverse neonatal brain growth and neurodevelopmental outcomes among children born very preterm with a neonatal infection (Gomaa et al., 2022). Of interest here is that none of the CpG sites which constitute the DNAm CRP used in this study were identified as top hits within the previous EWAS in this cohort examining methylation associated with gestational birth (Wheater et al., 2022) or the PedBE (Gomaa et al., 2022) gestational clock (built from 94 CpG sites), indicating that the DNAm CRP score is capturing something unique over residual gestational age-related differences, and further affirming our hypothesis that sustained inflammation may be driving differences in brain dysmaturation above and beyond preterm status itself.

The CpG sites which constitute the score are involved in vascular and immune function (supplementary Table S1). In the original EWAS by Ligthart et al (2016), the top CpG hits relating to circulating CRP levels mapped to the genes AIM2 and SOCS3; the former of which is an

Fig. 5. DTI-tract associations with DNAm CRP. Standardized regression coefficients for DNAm CRP associations between tract fractional anisotropy (FA) for preterm infants (squares), preterm infants in models controlling for additionally inflammatory risk factors (circles) and term infants (triangles). Filled shapes are left tracts and open shapes are right hemispheric tracts, except in the case of the CC where filled shapes are the splenium and open shapes are the genu of the corpus callosum. Points show standardized coefficients and 95% confidence intervals. All models are controlled for sex, gestational age at birth, gestational age at scan and birthweight Z score. Model H2 (circles) additionally controls for inflammatory risk factors (maternal smoking during pregnancy, preeclampsia, HCA, neonatal sepsis, BPD, NEC, ROP). For MD associations see supplementary Figure S10.
inflammase receptor, upstream of CRP, involved in the processing of interleukins responsible for the induction of CRP from hepatocytes. The importance of this locus in inflammation has been reinforced by candidate gene studies and further EWAS work (Miller et al., 2018; Myte et al., 2019), including the latest work on 22,000 participants (Wielscher et al., 2022). The second of these top hits, SOCS3 (cg18181703) has been significantly associated with a general factor of cognitive ability (Stevenson et al., 2020), as well as processing speed (Marioni et al., 2018) and brain structural characteristics such as WMH burden and global grey and white matter atrophy (Conole et al., 2021), suggesting the importance of this locus in driving cognitive changes. How this maps to neurodevelopmental testing scores in preterm infants in early infancy (van Beek et al., 2022), with some evidence for standard neurodevelopmental testing paradigms underestimating impairment and developmental delay following preterm birth (Anderson et al., 2010; Vohr et al., 2012). SOCS3 is involved in pro-inflammatory cytokine pathways (Rottenberg and Carow, 2014), is implicated in demyelination via oligodendrocyte disruption (Emery et al., 2006), has been previously implicated in a murine-model of preterm-related WM injury (Boccazzi et al., 2021) and has been found to be upregulated in the microglia of AD patients (Walker et al., 2015). This evidence suggests that aberrant methylation of cg18181703 may play an important intermediary role linking peripheral inflammatory processes with white matter maturation in the preterm neonate. This locus may also be particularly susceptible and responsive to differences in health and lifestyle as this site has also been linked to metabolic differences across the lifecourse such as BMI (Ali et al., 2016) and incident type 2 diabetes (Chambers et al., 2015), conditions themselves which have been linked to elevated CRP levels. More studies that conduct EWAS of inflammation in neonatal cohorts is needed, as the largest multi-cohort EWASs to date have been conducted in adult populations (Ligthart et al., 2016; Wielscher et al., 2022). Overall, our findings indicate that epigenetic modifications are an essential mechanism by which inflammatory risk factors could lead to long-term disruptions in both immune and brain development (Ozanne and Constancia, 2007; Fleiss and Gressens, 2012), and this work highlights the utility of profiling such changes alongside other clinical and biological data.

4.3. White and deep grey matter dysmaturation in preterms with elevated inflammation

The most striking finding from this study is the association of DNAm CRP with widespread variances in brain structure in preterm but not term infants (Figs. 3-5). We observe larger effect sizes for associations of DNAm CRP with global white matter microstructure (gFA and gMD) than white matter volume in a sub-population of these infants. This may be because diffuse white matter injury antedates overall reductions in white matter volume, with DNAm CRP-DTI associations capturing a more subtle dysmaturation of programmed development (Skiodl et al., 2010). Both global white matter volume, microstructure and regional white matter integrity were lower in preterms with elevated DNAm CRP, with infants at younger gestational ages more prone to elevated inflammation and related poor white matter integrity. These findings echo the results of prior studies (Korzheniwska et al., 2014; Volpe, 2019; Favrais et al., 2011; Dubner et al., 2019), and are overall consistent with the theory that alterations in white matter microstructure are largely a consequence of dysregulation of white matter development driven by inflammation (Dubner et al., 2019).

In addition to white matter, volume reductions were observed in the hippocampi and amygdala and deep grey matter with increased DNAm CRP, though the former did not remain significant after accounting for additional inflammatory risk factors (supplementary Table S11). The association of elevated DNAm CRP with lower deep grey matter volume is consistent with previous research that finds that preterms infants exhibit deep grey matter loss relative to term infants (Padilla et al., 2015; Inder et al., 2005; Boardman et al., 2006). Given the relationship we outline here between preterm birth and inflammatory load, inflammation may be a key driver of these differences, both via its direct effects on brain structure and its contribution to related damage such as sensitisation to hypoxia ischaemia, excitotoxic insults and other early-life stressors (Bennet et al., 2018; Lammertink et al., 2022). These widespread alterations in brain structure are particularly interesting given the evidence base for inflammation relating to cognitive impairment, as studies have shown that both hippocampal volume and thalamic volume loss accompanying white matter microstructural alterations are linked to neurodevelopmental outcomes in early childhood (Boardman et al., 2010; Beauchamp et al., 2008; Ball et al., 2013). Inflammation-related grey matter loss is considered a consequence of dysregulated neuronal development, with inflammatory mediators disrupting processes such as dendritic arborization and cortico-thalamic connectivity (Favrais et al., 2011). As consolidation of thalamocortical connections happens in the third trimester of pregnancy, deep grey matter structures may be vulnerable to inflammatory stimuli (Volpe et al., 2011).

We also observed regional variance in how DNAm CRP associates with white matter tracts, a finding consistent with previous studies that indicate that certain white matter tracts are more vulnerable to inflammatory-adjacent events such as hypoxia ischaemia (Kostovic et al., 2014) traumatic brain injury (Malabe and Dammann, 2009), intraventricular haemorrhage (Levito et al., 2013) and cerebral palsy (Lin et al., 2010). Different white matter tracts develop at different rates in utero and display distinct transient growth periods of increased axonal development. These windows of plasticity have been outlined as particularly vulnerable to perturbation (Levito and Gressens, 2007; Ment et al., 2009), with inflammation disrupting the developmental lineage of oligodendrocytes, resulting in hypomyelinated axons (Back, 2015; Favrais et al., 2011; Back et al., 2001; Majnemer et al., 2000). Developmental growth periods of certain white matter tracts may therefore underscore regional vulnerability to elevated inflammation, with younger tracts likely to have higher proportions of pre-myelinating oligodendrocytes vulnerable to inflammatory mediators. However, we caution that we lack the statistical power to reliably detect differences between the magnitude of associations in regional white matter structure, and instead interpret these findings as evidence of the pervasive and widespread impact of inflammation on the development of white matter. Correspondingly, though there were differences between the association significance for the left and right hemispheres for several of the delineated tracts, these unilateral findings are in keeping with previous studies of similar sample size (Inomata et al., 2014; Alexandrou et al., 2014); as the magnitudes were similar (with overlapping confidence intervals), this did not indicate a strong basis for laterality of effects.

4.4. Strengths and limitations

To our knowledge, this is the first time an epigenetic measure of inflammation has been examined in a preterm cohort in relation to brain health outcomes. The effect sizes reported in this study are consistent with that of previous epidemiologic studies of DNAm and early life outcomes (Breton et al., 2017). The sample size (n = 258 for inflammatory exposure, and n = 121–214 for neuroimaging associations), is akin to that of previous work examining inflammation and brain structure in preterm infant populations (Shah et al., 2008; Kelly et al., 2016; Sullivan et al., 2020; Glass et al., 2008), and in many cases more substantial, with the vast majority of prior work conducted in sample sizes of<100 infants (Inomata et al., 2014; Yoon et al., 1997; Anblagan et al., 2016; Travis et al., 2015; Lee et al., 2021; Volpe, 2019; Alexandrou et al., 2014; Basu et al., 2015). There is a distinct scarcity of detailed methylation alongside multi-modal neuroimaging data (Wheater et al., 2020; Lancaster et al., 2018), particularly in neonatal
cohorts (Fumagalli et al., 2018; Chen et al., 2015; Sparrow et al., 2016), making this a valuable contribution to the DNAm-neuroimaging field.

Although the weights for the predictor were trained in adult blood samples, we observed similar associations between DNAm CRP and brain structural outcomes to those in previous studies of adults (Conole et al., 2021; Green et al., 2021). Given we have now applied this method to buccal-cell DNAm, it is encouraging to see similar associations between DNAm CRP and brain structural outcomes, especially given that DNAm is highly tissue specific (Davies et al., 2012), and previous research has reported on cross-tissue differences in magnitude and direction of effects for other traits (Walton et al., 2016). This cross-tissue approach (where weights were originally created from blood-based DNAm, and later applied to saliva-based composite signatures) has also been adopted in other studies (Suarez et al., 2020; Blostein et al., 2022). While future studies would ideally measure CRP directly from serum or blood spots in infants to enable direct cross-tissue comparisons, saliva has the advantage of being one of the most accessible tissue samples for infant populations, and may be more suitable than other peripheral samples (such as blood) when examining brain and cognitive outcomes owing to the brain and buccal cell shared ectodermal origins (Lowe et al., 2013; Braun et al., 2019). While we suggest that future studies consider a larger range of DNAm signatures (based off other inflammatory mediators implicated in PTB, such as IL6), we do not consider the lack of direct comparison between DNAm CRP and serum CRP a serious limitation of this study. We have previously shown that DNAm CRP outperforms serum CRP in anticipating brain structural and cognitive outcomes when derived from peripheral blood samples in adult population samples (Stevenson et al., 2020; Conole et al., 2021; Green et al., 2021). Additionally, within preterm populations, the use of CRP as a reliable measure of inflammation, or predictor of infection, is debated. A significant advantage of DNAm is that it may provide a historical archive of exposure that can be leveraged in instances where other clinical or biological data was not originally collected. To obtain useful baseline inflammation levels, serial sampling of CRP is encouraged; however, this is both impractical and ethically disputed in preterm neonates (given the number of needling episodes and volume of blood required to obtain average readings). There is evidence that preterm infants display dysregulated serum CRP responses compared to term infants, with studies cautioning its reliability as a biomarker of inflammation in this population group (Macallister et al., 2019; Borowski et al., 2022). The threshold values for raised CRP levels are dependent on gestational age at birth, with infants born too early displaying more phasic responses in serum CRP profiles which are less reflective of severity of infection. A recent systematic review (20 studies; n = 1,615 infants) concluded that CRP in neonates was an unreliable measure of inflammatory responses and should be avoided with regards to profiling sustained inflammation, or for subsequent guiding treatment (Brown et al., 2019). Furthermore, irrespective of preterm status, the use of sampling CRP postnatally is complicated by the non-specific rise in CRP levels related to the stress of delivery itself (Chiesa et al., 2001; Hofer et al., 2012). In the absence of direct comparison with inflammatory mediators from blood samples, the strong correlates with clinical inflammatory conditions (both fetal and postnatal) is affirming, and we have taken steps to account for possible sources of confounding (supplementary Fig. 2).

Neuroimaging studies are notoriously heterogenous in their design given the array of different MRI acquisition techniques, processing pipelines and choice of outcome measures. The choice of neuroimaging features is even more relevant in the context of preterm birth to adequately address the motivating research questions (Korom et al., 2022). Here, our choice of neuroimaging features was guided by established characterizations of EoP in preterm infants, namely water content and dendritic/axonal complexity and dysmaturation within the white matter, and grey matter volume (Blesa et al., 2020). While we consider this comprehensive characterisation of brain structure from NODDI and dMRI data a significant strength of this study, we acknowledge that microstructure measures such as FA and MD in older cohorts are commonly considered surrogates of white matter integrity or myelination, the white-matter pathways in this study are still developing at the time of gestational age at scan (range 37.70–45.14 weeks), and as such may not reflect permanent differences. Longitudinal follow up is therefore encouraged for future studies designed to examine the implications of sustained inflammation in preterms for neurodevelopmental outcomes and life course brain health.

We do not attempt to discuss the causality of the relationship between DNAm CRP with brain structure, though the causality of such associations is a persistent topic of debate in epigenetic epidemiological research and has been discussed in depth in reviews (Yousefi et al., 2022; Cecil et al., 2022; Birney et al., 2016). Future work examining transcriptomic changes on the same peripheral samples from which DNAm data is collected, as well as statistical approaches like two-step mendelian randomization, are important developments to unpick causality of these relationships. Studies investigating whether these differences in DNAm remain, amplify, or attenuate with age are advised, as well as how sensitive these signatures are in the context of intervention (anti-inflammatory medications and treatments, as well as lifestyle interventions such as the cessation of smoking in pregnancy). As infants born before 37 weeks gestation display reduced protection against immune-mediated disturbances (e.g., absent placental trophic and hormonal factors, immature responses to free radicals), examining these analytes in relation to DNAm signatures could demarcate specific pathways that confer added vulnerability for preterm brain dysmaturation. There is also precedent to examine whether composite methylation proxies of inflammation differ across psychopathology or specific neurological cases such as Cerebral Palsy or neurodevelopmental disorders such as Fragile X syndrome, autism and ADHD, given examples of other poly-epigenetic signatures of psychiatric disorders (Cecil et al., 2022). Future work that focuses on such DNAm dynamics in relation to these outcomes in ongoing longitudinal studies of infants born preterm is therefore of interest, as well as replication in different population samples.

Finally, DNAm was sampled in neonates postnatally. While this is rational when examining variances in DNAm and brain structural differences in infants, future studies that examine both maternal and infant DNAm could examine the degree to which exposures are shared or specific to parent and offspring. Equally, multiple DNAm sampling during pregnancy could elucidate key critical periods of susceptibility to inflammation by parsing out exposures specific to trimester or months of pregnancy, affording new insights into the spatiotemporal patterning of brain development in relation to dynamic immune changes in the perinatal period.

4.5 Future directions.

The use of DNAm biomarkers in neonatology has the potential to provide a more comprehensive picture of the health risks associated with preterm birth and can potentially be used to inform the clinical management of at-risk pregnancies, preterm infants and later-life health outcomes. The increasing evidence for the association between DNA methylation and clinical features of patients, such as the ability to stratify patients on the basis of disease subtypes (Somini et al., 2019; Aref-Eshghi et al., 2020; Good et al., 2021; He et al., 2020), reflect dose-dependent exposure to certain substances such as smoking and alcohol (Langdon et al., 2021; Colicino et al., 2021; Yousefi et al., 2019) demonstrates the general utility of DNA methylation signatures as biomarkers on various disease-endpoints, diagnoses and exposures. The clinical potential of DNAm is further supported by the stability and robustness of DNA methylation and the convenience of DNAm sampling. Extending this paradigm to investigate how DNAm can be leveraged to investigate interactions at the maternal-fetal interface, exposures in the perinatal period, and early-life outcomes has lagged behind this progress (Cecil et al., 2022), though there are recent studies that demonstrate the utility of DNAm in indexing prenatal exposures and early-life outcomes (Camerota et al., 2021; Everson et al., 2020; Bakulski et al., 2021;
Abrishamcar et al., 2022).

In the context of neonatology, DNA methylation-based biomarkers may offer an opportunity to develop predictive and prognostic tools to optimise neonatal care. Inflammation-associated DNA methylation signatures such as those explored in this study could provide insight into the relative risk of adverse outcomes and act as a means to monitor how intervention, treatment and lifestyle effectivemly ameliorate the risk of these. The potential of profiling DNAm to track health trajectories in pregnancy and anticipate PTB is of particular translational value for identifying the relatively ‘silent’ and sudden onset triggers for preterm labour, such as preeclampsia, HCA, premature rupture of the membranes (PRROM) or a signpost risk of pregnancy-related disease (such as hyperemesis gravidarum, acute fatty liver of pregnancy, cholestasis) that can cause complications to both mother and infant. Overall, robust indexes inflammatory burden would enable better obstetric management to anticipate preterm birth, mitigate morbidity risk (through timely administration of prenatal steroids, magnesium sulphate, tocolytics and optimal delivery procedures), and thus improve the health and long-term outcomes for many children. Furthermore, the identification and validation of further inflammation-associated methylation markers, particularly those connected to specific pregnancy complications and exposures and postnatal complications, may provide insights into the developmental pathways that affect risk of sustained inflammation.

5. Conclusion

Inflammatory-related DNAm is associated with risk of postnatal health outcomes and brain dysmaturation. Our results indicate that multiple inflammatory-related hits from different origins (pertaining to maternal, fetal, and postnatal exposures) may be captured by changes to the DNA methylation profiles of infants and may help to explain variations in brain structure in preterm populations, circumventing limitations of traditional measures of inflammation. As early birth is associated with sudden change in immune-related risks, which coincide with the developing immune system and windows of neurodevelopmental plasticity, it is theorised that preterm infants are at greater risk of inflammation-related disruption of white matter. Our work here provides new layers to this theory, with epigenetic markers of inflammation associating with diffuse and global brain – and particularly white matter – alterations in preterm but not term infants, indicating that sustained inflammation may be a key driver of neurodevelopmental disruption. In summary, the association of an DNA methylation signature with inflammatory outcomes, and inflammation-related neural phenotypes, supports the use of methylation data in integrative, multimodal approaches toward disease stratification in the perinatal period.

Funding

This research was funded in whole, or in part, by the Wellcome Trust (Grant No 108890/Z/15/Z and 221890/Z/20/Z). For the purpose of open access, the authors have applied a CC BY public copyright licence to any Author Accepted Manuscript version arising from this submission.

Declaration of Competing Interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: R.E.M has received a speaker fee from Illumina and is an advisor to the Epigenetic Clock Development Foundation and Optima Partners.

Data availability

Data will be made available on request.

Acknowledgements

Some of the participants were scanned in the University of Edinburgh Imaging Research MRI Facility at the Royal Infirmary of Edinburgh which was established with funding from The Wellcome Trust, Dunhill Medical Trust, Edinburgh and Lothians Research Foundation, Theirworld, The Muir Maxwell Trust and other sources. We thank Thorsten Feiweier at Siemens Healthcare for collaborating with dMRI acquisitions (Works-in-Progress Package for Advanced EPI Diffusion Imaging). The authors are grateful to the families who consented to take part in the study and to all the University’s imaging research staff for providing the infant scanning.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.bbi.2023.03.011.

References


7.3 Conclusion

This Chapter investigated the influence of inflammatory-related DNAm on neuroanatomic variation and postnatal clinical inflammatory outcomes in infants using brain MRI linked to biosamples, integrating lifestyle and health data from both mothers and infants. Examining inflammation and its associated impact on brain health in early-life is challenging from multiple angles: (1) inflammation as an exposure is typically difficult to measure in preterm neonates, where underdeveloped immune and organ systems can give rise to noisy serum protein readings (2) shared maternal exposures and fetal inflammation is difficult to characterise (3) the impact of inflammation on early neurodevelopment may only manifest as meaningful cognitive health outcomes years later. Thus, there is an incentive to develop tools that can accurately index past exposure, particularly in the prenatal and perinatal period, and assess their relationship with quantifiable neuroimaging outcomes. DNAm signatures hold promise in helping researchers delineate the sequence of events leading from health to aberrant neurodevelopment.

Here, we demonstrate that a salivary-based DNAm measure of inflammation (DNAm CRP) associates gestational age, with preterm infants displaying higher DNAm CRP than term infants. The concept that DNAm may assess the allostatic load of inflammation during neonatal critical care is strengthened by our discovery that aggregate inflammatory hits associated with higher DNAm CRP. Comorbidities of preterm birth that are characterised by a pro-inflammatory state (neonatal sepsis, histologic chorioamnionitis, necrotising enterocolitis and bronchopulmonary dysplasia) related to a higher risk of inflammation. These associations remained when accounting for medication usage that is considered protective for adverse postnatal outcomes (antenatal corticosteroids and magnesium sulfate administration) alongside birthweight and sex, indicating an independent contribution of inflammation to these phenotypes outside of preterm birth status itself. Postnatal experiences appear to have a greater weight on impacting
inflammatory burden than antenatal factors (smoking in pregnancy, preeclampsia).

In addition to this, we find that higher inflammation as indexed by DNAm CRP is linked to widespread neurostructural features including lower white matter volume and deep grey matter volume. Higher inflammatory burden was particularly related to poorer measures of white matter microstructure (FA and MD) both globally and for certain white matter tracts; this may be capturing an early developmental disruption, given that diffuse white matter injury precedes overall reductions in white matter volume. Our findings are consistent with previous work on the association between inflammation and EoP metrics (Dammann and Leviton, 2004, 2014; Inomata et al., 2014; Bennet et al., 2018; Humberg et al., 2020). Equally, though we did not look at cognitive outcomes in this cohort, previous longitudinal studies have shown that sustained inflammation is related to neurodevelopmental differences in preterm infants at 2 and 5 year follow up periods (Stoll et al., 2004; Shah et al., 2008; Schlapbach et al., 2011; Inomata et al., 2014; Bennet et al., 2018). Our findings may offer insight into these observations, as many of the white matter tracts which are associated with DNAm CRP in this study are tracts that have been linked to language, such as the genu of the corpus callosum, which supports language processing, and the arcuate fasciculus, which is associated with syntactic processing, prosody and semantics (Sket et al., 2019). Moreover, a previous study found that higher levels of this DNAm signature in childhood associated with lower cognitive test scores and higher mental health burden in adolescence (Barker et al., 2018).

This is the first time that an epigenetic signature of inflammatory burden has been examined in relation to brain structural differences in a neonatal cohort and highlights how white matter disruption in particular might be one of the reasons for later-life cognitive and psychiatric phenotypes. Relative to term infants, preterm newborns are thought to be more susceptible to inflammation-related alteration of white matter because early birth is linked to a sudden change in immune-related hazards, which coexist with both an
immature immune system and windows of myelination. Aspects of development during the 3rd trimester of pregnancy appear to be critical windows of perturbation where infection (such as HCA) and ensuing inflammation can confer brain structural consequences.

This supports the use of methylation data in integrative, multimodal approaches toward disease stratification in the perinatal period to identify infants at risk of postnatal complications. This study also has generated several questions for future investigation, particularly regarding further evaluation of clinical risk factors that result in sustained inflammation in the perinatal period, the effect of early-life stress on the epigenome, and the impact of protective factors that were unaccounted for in these analyses (e.g. breastmilk supplementation).
8 Chronic inflammation and brain health in midlife

8.1 Introduction

As introduced in Chapter 3, DNA methylation signatures of exposures are an emerging area of epigenetic epidemiology to provide new ways of indexing difficult exposures. Chapters 6 and 7 demonstrated that a DNAm signature of CRP shows significant associations with aspects of brain structure both in early life and in older age, as well as associating with multiple inflammatory hits. Here, I extend this rationale to numerous other inflammatory-proteins, examining 109 inflammatory-related DNAm signatures in relation to multiple brain and cognitive health outcomes in a cohort of individuals aged 28 - 81 years.

While previous studies have shown that blood-based DNAm signatures have utility in augmenting protein-trait association analyses in relation to electronic health-linked health outcomes (Gadd et al., 2022; Thompson et al., 2022), very few studies have examined DNAm signatures in relation to neuroimaging IDPs. As such, this study represents the most comprehensive evaluation of how inflammatory-related DNAm signatures associate with aspects of brain structure and function to date. This study aimed to outline which immune-related DNAm signatures associate across multiple metrics and are particularly pertinent to brain health, (ii) clarify whether DNAm offers advantages over protein levels in explaining variance in relative brain age, global neuroimaging metrics, and aspects of cognitive functioning, (iii) characterise regional patterning of chronic inflammation as indexed by DNAm signatures on brain volumes and white matter integrity, (iv) highlight the influence of aspects of health and lifestyle on these relationships, and (v) examine the interrelationships between DNAm, brain and cognition.
8.2 Examining multi-omic signatures of inflammation in relation to brain structure, cognitive ability, and brain ageing
Examining multi-omic signatures of inflammation in relation to brain structure, cognitive ability, and brain ageing

Eleanor L.S Conole\textsuperscript{1,2,3*}, Danni A. Gadd\textsuperscript{2}, Robert F. Hillary\textsuperscript{2}, Aleks Stoilcyn\textsuperscript{3}, Xueyi Shen\textsuperscript{3}, Mark E. Bastin\textsuperscript{1,3}, Stephen M. Lawrie\textsuperscript{4}, Andrew M. McIntosh\textsuperscript{2,4,5}, Veronique E. Miron\textsuperscript{6}, Heather C. Whalley\textsuperscript{3,4,5}, Riccardo E. Marioni\textsuperscript{1,2†}, Simon R. Cox\textsuperscript{11*†}

\textsuperscript{1}Lothian Birth Cohorts group, Department of Psychology, University of Edinburgh, Edinburgh EH8 9JZ, UK; \textsuperscript{2}Centre for Genomic and Experimental Medicine, Institute of Genetics and Cancer, University of Edinburgh, Edinburgh EH4 2XU, UK; \textsuperscript{3}Centre for Clinical Brain Sciences, University of Edinburgh, Edinburgh EH16 4SB, UK; \textsuperscript{4}Department of Psychiatry, University of Edinburgh, Edinburgh, UK; \textsuperscript{5}Scottish Imaging Network, A Platform for Scientific Excellence Collaboration (SINAPSE), Edinburgh, UK; \textsuperscript{6}MRC Centre for Reproductive Health, Queen’s Medical Research Institute, Edinburgh BioQuarter, University of Edinburgh, Edinburgh EH16 4TJ, UK

\textsuperscript{*}For correspondence: (ELSC); (SRC)

\textsuperscript{†}These authors contributed equally to this work

Abstract  Chronic inflammation is a hallmark of many age-related diseases, including cognitive decline, but is rarely well characterised in population-based studies. In a community-based sample of adults (n = 709, age range 28-81 years), we examined the relationship between inflammation and cognitive ageing by profiling multi-modal brain imaging, cognitive, clinical and lifestyle data with over 100 proteomic and DNA methylation (DNAm) signatures of inflammatory markers. We uncover associations between DNAm signatures and numerous aspects of global brain structure and cognitive ability ($\beta$ range $|0.097|$ to $|0.200|$), alongside regional impact the brain’s cortex ($\beta$ range $|0.087|$ to $|0.260|$) and focal vulnerability of specific white matter tract microstructure ($\beta$ range $|0.103|$ to $|0.185|$), as well relationships between inflammation, brain structure and cognition. Our findings offer new insights into the complex relationship between chronic inflammation and brain health, and the potential of DNAm signatures as tools for precision medicine.
Introduction

Chronic inflammation is considered a major contributor to morbidity, mortality, and age-related cognitive decline (Furman et al., 2019). However, uncertainties remain around the specific regions of the brain affected by inflammation, the age at which such structural changes occur, and the nature and extent of cognitive impairment (Conole et al., 2021). The approach to measuring chronic inflammation may be in part responsible for this ambiguity: traditionally, inflammation has been assessed through one-off measurements of inflammatory proteins in the blood such as C-Reactive Protein (CRP) or interleukin 6 (IL6). Measurements of these proteins only represent a trivial subset of the diverse molecules that constitute chronic inflammation, and display considerable variability both within and between individuals over time (DeGoma et al., 2012; Bogaty et al., 2013; Bower et al., 2012). In population cohorts, where repeat measures are rarely taken, this variability can cause substantial noise (Bogaty et al., 2013; Walker et al., 2022), potentially obscuring the true consequences of chronic inflammation on brain health outcomes. To address this, there have been increasing calls for integrating multiple inflammatory biomarkers across -omic data layers to explore relationships with incident health outcomes (Furman et al., 2019; Morrisette-Thomas et al., 2014; Wielscher et al., 2022).

While the proteome provides information about the proteins that are expressed in the body, the methylome gives insight into how these proteins are regulated. Because DNA methylation (DNAm) is involved in both gene expression and gene-environment interactions, differences in DNAm levels can potentially provide dynamic information capturing chronic inflammation and how it relates to aspects of brain health. In support of this, methylation at specific cytosine-phosphate guanine (CpG) sites has been linked to both circulating inflammatory proteins and various diseases characterised by a chronic inflammatory state (Wielscher et al., 2022; Ligthart et al., 2016; Kalla et al., 2023; Somineni et al., 2019; Mäki-Nevala et al., 2021). Moreover, DNAm signatures of inflammatory proteins (weighted linear sum of methylation levels at individual CpG sites) have shown promise in indexing amassed exposure to inflammation (Wielscher et al., 2022; Stevenson et al., 2020, 2021). Such DNAm signatures (also referred to as epigenetic scores - ‘EpiScores’ - or ‘methylation risk scores’), may help stratify individual risk of disease (Thompson et al., 2022; Gadd et al., 2022b), anticipate survival outcomes (Villanueva et al., 2015), or act as biomarkers of exposure (Yousefi et al., 2022).

Given the complex nature of inflammatory signalling cascades, and the difficulty of capturing a ‘baseline’ level of inflammatory exposure, the use of DNAm signatures of inflammatory proteins is particularly compelling for investigating brain health outcomes. However, in the few examples of extending this paradigm in relation to brain and cognitive outcomes to date, studies have focused on a limited pool of inflammatory markers, and examined brain differences at extreme ends of the lifespan (Conole et al., 2021, 2023). More generally, few studies looking at inflammation and brain ageing are set in midlife (Laurin et al., 2009; Schmidt et al., 2002; Walker et al., 2019), despite the evidence that this period is when cognitive changes are thought to diverge among individuals (Salthouse, 2016, 2019). This is also around the time when signs of neuropathology such as amyloid beta begin to arise (Sperling et al., 2011), and when
biological markers might be exhibiting early subtle changes that precede overt cognitive symptomatology (Jack et al., 2013).

To address this gap, we adopted a multi-omic approach to examine over 100 proteomic and methylomic signatures of inflammatory markers and their associations with brain health in a large community-based population of adults in mid to late adulthood (Box1). Our results reveal DNAm signatures of proteins linked to both favourable and detrimental brain and cognitive phenotypes, as well as regional impacts on the brain's cortex and focal vulnerabilities of specific white matter tract microstructure. These associations are broadly independent of aspects of confounding from immune cell proportions and related clinical risk factors. Through mediation modelling, we elucidate the interrelationships between inflammatory-related DNAm signatures, brain structure, and cognitive ability. Our findings provide new insights into the complex relationships between chronic inflammation, brain aging, and cognitive decline, and offer potential targets for intervention and prevention.
Box 1. Study design

(a) DNA methylation scores were trained on 953 circulating plasma protein levels in the KORA and LBC1936 cohorts as outlined previously by Gadd et al. (2022b). There were 109 DNAm signatures selected based on performance ($r > 0.1$, $p < 0.05$) in independent test sets. For the present study, the selected DNAm signatures were projected into sub-cohort of Generation Scotland where participants had neuroimaging data collected, the STRatifying Longitudinal Resilience & Depression (STRADL) cohort (b) In 709 individuals from this cohort, we examined associations between multi-omic signatures of inflammation with brain and cognitive health outcomes. 109 DNAm signatures were examined in relation to 112 brain and cognitive health phenotypes. Baseline models controlled for age and sex, sensitivity analyses controlled for age + sex + immune cell proportions and fully-adjusted models controlled for the above plus aspects of lifestyle (BMI + smoking + drinker status + hypertension). *In instances where a brain MRI metric was an outcome of interest, models additionally controlled for imaging site, batch, number of edits and ICV (c) study key findings: 73 DNAm signatures were identified as relevant to brain health, the top 10 DNAm signatures for poor and favourable brain health outcomes are displayed, with bars demonstrating the number of significant (PFDR < 0.05) associations they displayed with individual brain health metrics.
Results

Previously, DNAm signatures for circulating protein levels have shown to associate with various major health outcomes, indicating their potential as useful tools for precision medicine Gadd et al. (2022b). Here, we examine whether DNAm signatures complement and augment measured protein associations with 112 brain health outcomes (including multi-modal neuroimaging and cognitive assessment) from 709 participants of The Stratifying Resilience and Depression Longitudinally (STRADL) cohort. The 109 protein DNAm signatures described by Gadd et al. (2022b), including various inflammatory-related biomarkers, were examined alongside measured protein levels with multiple brain and cognitive health outcomes in a healthy group of adults in mid-to-late adulthood (age range 28 – 81 years). Participant characteristics are displayed in Supplementary Data 1.

We report 470 statistically significant associations (β range |0.053| to |0.260|, pFDR < 0.05) between 73 DNAm signatures and brain and cognitive ageing outcomes. DNAm signatures displayed more significant associations with brain and cognitive ageing metrics than measured protein levels (Fig.1; matched associations between proteomic and DNAm signatures are reported in full in Supplementary Data 11), but had small effect sizes explaining ≤ 6.7% (incremental $R^2$) of the variance in brain structure or cognitive outcomes. These consisted of 139 global brain imaging outcome associations, 29 cognitive test score associations, 100 regional cortical volume associations and 202 associations with aspects of regional white matter tract microstructure. Correlations of the 73 DNAm signatures that were associated with brain health outcomes (pFDR < 0.05) with covariates suggested interlinked relationships with estimated white blood cell proportions (monocyte, B-cell, CD4T, CD8T, and natural killer cells; Fig.1b, haemoglobin, platelet and FBC-related correlations presented in Supplementary Fig.6). These covariates were therefore added incrementally to fully-adjusted models as predictors (model 2; Fig.1c). Equally, as lifestyle habits such as smoking, alcohol consumption and cardiovascular risk factors (BMI, hypertension) have been linked to both the predictor variables and outcomes of interest in question, these were further controlled for in a fully-adjusted model (model 3). Of the original 470 significant associations in baseline models, 408 remained statistically significant after adjustment for immune cell proportions, and 391 remained statistically significant after further adjustment for lifestyle factors (results of all three models are supplied in Supplementary Data 12). Fig.1d displays a comparison of the effect sizes for matched protein levels and DNAm signatures for the baseline model (see Supplementary Data 11 for matched WBC-adjusted and fully adjusted model comparisons). Fig.1e illustrates this effect comparison for one example metric, relative brain age, which is an estimate of an individual’s brain age derived by applying machine learning to structural MRI data (Cole and Franke, 2017).
Figure 1. Comparison of DNAm signatures and circulating plasma proteins in predicting brain health outcomes. (a) correlation coefficients between projected DNAm signatures in the GS neuroimaging sample for both Olink (red points) and SomaScan (blue points), with error-bars set as 95% confidence intervals. (b) Correlation heatmaps between DNAm signatures and immune cell proportions. (c) Regression model summary: linear regression models tested relationships between 109 DNAm signatures and 112 brain health outcomes. (d) Comparison of protein vs DNAm signature association magnitudes with each brain health outcome metric, brain age. (e) Effect size differences between DNAm signatures and proteins visualised for one brain health outcome metric, brain age.
In the following sections, we breakdown these results in terms of whether associations were positive (i.e., DNAm signatures that associated with favourable health outcomes such as greater grey matter volume, higher cognitive ability test scores, decreased brain age) or negative (i.e. poor brain health outcomes, lower brain volumes, lower cognitive test scores), first outlining cognitive and global neuroimaging associations then regional brain volume and WM microstructure associations. We primarily focus on the associations between DNAm signature and brain health outcomes, providing protein-brain associations for illustrative purposes (to illustrate the extent to which composite DNAm signatures confer any advantage over proteins in predicting brain and cognitive phenotypes). Full reporting of matched proteomic-brain health associations is presented in Supplementary Data 11.

**Association of DNAm Signatures with Global Aspects of Brain and Cognitive Health**

There were 168 significant associations ($\beta$ range $[0.053]$ to $[0.200]$, $p$-FDR $< 0.05$) between DNAm signatures and global neuroimaging outcomes, cognitive metrics, and relative brain age (see Supplementary Figs 6-7 for outcome measure distributions). Of these, we focus first on the 102 associations where DNAm signatures were linked to poorer brain and cognitive aging outcomes (e.g., lower brain tissue volume measures, poorer cognitive test scores, and higher relative brain age), composed of 33 unique DNAm signatures ($\beta$ range $[0.053]$ to $[0.200]$, $p$-FDR $< 0.05$). Fig.2 displays which aspects of brain and cognitive health were associated with individual DNAm signatures (Fig.2a), their performance relative to equivalent proteomic associations (Fig.2b), and cases where DNAm scores associated across multiple phenotypes (Fig.2c).

There were a greater number of significant associations between DNAm signatures and brain health phenotypes than proteomic-brain health associations (Supplementary Data 11, DNAm: $n = 470$, protein: $n = 27$), and generally effect sizes were larger for DNAm-brain health associations (Fig.2b). For example, for global grey matter volume, 23 DNAm signatures significantly associated with alterations in global grey matter volume (average $\beta = 0.112$, $p$-FDR $< 0.05$) vs 4 protein levels (average $\beta = 0.082$, $p$-FDR $< 0.05$). Across all of these global neuroimaging metrics and cognitive outcome measures, there were only five instances where proteomic signatures displayed numerically larger effect sizes than DNAm equivalents: the association of protein levels of PIGR with processing speed ($\beta_{\text{protein}} = -0.124$, $p = 2.2 \times 10^{-4}$ vs $\beta_{\text{DNAm}} = -0.11$, $p = 1.23 \times 10^{-3}$), RARRES2 with processing speed ($\beta_{\text{protein}} = -0.12$, $p = 5.5 \times 10^{-4}$ vs $\beta_{\text{DNAm}} = -0.11$, $p = 2.98 \times 10^{-3}$), PRSS2 with global cortical thickness ($\beta_{\text{protein}} = -0.115$, $p = 3.27 \times 10^{-4}$ vs $\beta_{\text{DNAm}} = -0.093$, $p = 0.076$), and LGALS3BP with relative brain age ($\beta_{\text{protein}} = 0.083$, $p = 1.17 \times 10^{-4}$ vs $\beta_{\text{DNAm}} = 0.069$, $p = 1.63 \times 10^{-3}$).

In sensitivity analyses, we found that these results were largely robust to the inclusion of immune-cell proportion and health/lifestyle covariates. When adding immune-cell proportions incrementally as predictors, 141/168 DNAm associations remained significant. Controlling for aspects of health and lifestyle (BMI, hypertension, smoking, and alcohol consumption) attenuated associations by an average of 13.6% ($\beta$ range $[0.047]$ to $[0.167]$, Supplementary Data 12), leaving 135/168 significant, composed of 51 unique DNAm signatures linked to global brain health outcomes. For proteomic-DNAm associations, a similar attenuation (17%) was observed (Supplementary Data 11).
Figure 2. DNA methylation (DNAm) signatures associate with poor brain and cognitive outcomes. (a) Number of associations that were significant (p < 0.05) is highlighted by the coloured bar; number that pass FDR significance are in overlayed dark grey bars. (b) Comparison of DNAm vs protein associations between different brain and cognitive health outcomes with individual signatures (pFDR < 0.05). Points represent individual betas for associations for significant DNAm signatures (circles) and protein signatures (diamonds). (c) UpSet plot of the intersection of DNAm signatures with brain and cognitive ageing metrics, displaying the degree to which individual DNAm signatures associate across multiple poor brain and cognitive health outcomes.
Of interest, many DNAm signatures that had significant associations with brain health phenotypes overlapped across multiple aspects of global neuroimaging and cognitive phenotypes. The inflammatory-DNAm signatures of MMP12, CRP, and PIGR displayed the highest number of significant ($p_{FDR} < 0.05$) associations (Supplementary Data). Table 13 displays the total number of significant associations per 73 DNAm signatures across all 112 brain health outcome measures, with PIGR emerging as the top hit at $n = 32$ associations with poor brain health outcomes. Increased relative brain age and lower global grey matter volume displayed the largest number of associations with individual DNAm signatures (Fig.2c), whereas some brain and cognitive metrics examined, such as APOE e4 status and aspects of cognitive functioning (vocabulary, verbal fluency, executive function, and general fluid intelligence scores), did not display any significant associations with signatures following multiple testing correction.

We now focus on the 66 significant associations where DNAm signatures were associated with favourable brain and cognitive ageing (for example higher brain volume measures, better cognitive test scores, and lower relative brain age; $\beta$ range $|0.06|$ to $|0.160|$, $p_{FDR} < 0.05$). These composed of 23 unique DNAm signatures, 12 of which (SEMA3E, NCAM1, NTRK3, NOTCH1, CNTN4, SELL, WFIKKN2, SLITRK5, OMD, GP1BA, GDF.8) were significantly associated with multiple brain health metrics. Fig.3 displays which aspects of brain and cognitive health were affected by individual DNAm signatures, including cases where DNAm scores associated across multiple phenotypes (Fig.3c). DNAm displayed stronger associations with brain health than serological protein equivalents, where no significant associations were observed (Fig.3b).
Figure 3. DNAm signatures associate with favourable brain and cognitive outcomes. (a) Number of associations that were significant (p < 0.05) indicated by coloured bar; number that pass FDR significance indicated by overlayed dark grey bars. (b) DNAm vs protein associations between brain and cognitive health outcomes (pFDR < 0.05). Points represent individual betas for associations for DNAm (circles) and protein signatures (diamonds). (c) UpSet plot of the overlap of DNAm signatures with various brain and cognitive ageing metrics.
We report 100 significant associations (13 subcortical, 87 cortical) between DNAm signatures and regional neuroimaging metrics. In almost all instances, DNAm signatures explained more variance in regional brain volumes than matched protein measurements. There were only three exceptions to this: protein IGFBP4 showed a stronger association with lower brainstem volume than DNAm IGFBP4 ($\beta_{\text{protein}} = -0.117, p = 1.5 \times 10^{-4}$ compared to $\beta_{\text{DNAm}} = -0.100, p = 1.1 \times 10^{-3}$).

Similarly, protein PIGR showed a stronger association with lower insula volume than DNAm PIGR ($\beta_{\text{protein}} = -0.115$, $p = 2.1 \times 10^{-5}$ compared to $\beta_{\text{DNAm}} = -0.091, p = 1.5 \times 10^{-3}$) and lower superior temporal volume ($\beta_{\text{DNAm}} = -0.130$, $p = 5.28 \times 10^{-6}$ compared to $\beta_{\text{protein}} = -0.091, p = 1.5 \times 10^{-3}$). These findings were robust to further correction for WBC-proportions and lifestyle factors (96/100 retained $p < 0.05$ in fully-adjusted models, average $\beta$ attenuation 10.3%). The entire list of all associations, for matched DNAm and proteomic signatures, across all of the three models (baseline, WBC-adjusted, and fully-adjusted) is reported in Supplementary Data 11.

There were 13 significant associations between subcortical structures and DNAm (Supplementary Data 12), of which five were cases of signatures associating with higher subcortical volumes, and 8 were cases of DNAm signatures linked to lower regional subcortical volumes.
Figure 4. DNAm signatures associate with favourable brain and cognitive outcomes. (a) Number of associations that were significant (p < 0.05) is highlighted by the coloured bar; number that pass FDR significance are outlined in dark grey (b) ROIs examined, visualisation of subcortical structures created using ggseg3d package; FDR-significant associations where elevated DNAm signatures associated with variance in subcortical volumes.

For the eight associations where DNAm signatures associated with lower subcortical volumes, there were five unique DNAm signatures that associated with lower subcortical volumes: PIGR, CRP, FGF.21, IGFBP4, and SERPIND1 (β range |0.091| to |0.112|, pFDR < 0.05). No one DNAm variable/signature was significant across all subcortical structures, with IGFBP4 linked to the most individual subcortical ROIs with consistent direction, showing negative associations with volume in the grey matter of cerebellum, ventral diencephalon, and brainstem. Of the five positive associations, there were four unique DNAm signatures which associated with greater subcortical volumes: NTRK3, NCAM1, CNTN4, and VCAM1 (β range |0.091| to |0.142|, pFDR < 0.05) (displayed in Fig.4b).

There were 87 significant (pFDR < 0.05) unique associations between regional cortical volumes and DNAm signatures (β range |0.086| to |0.260|). Fig.5 displays the number of DNAm signatures that associated with cortical volumes. There were 20 unique DNAm signatures that associated with lower cortical volumes (β range |0.087| to |0.260|, pFDR
< 0.05) and 13 unique DNAm signatures which associated with greater cortical volumes (β range |0.086| to |0.150|, pFDR < 0.05); results presented in Fig.5b.

Some cortical regions had a higher number of significant associations with DNAm signatures, such as the supramarginal and fusiform regions. Equally, the same signatures tended to associate with reductions or increases among multiple regions. For example, PIGR was associated with lower cortical volumes in six cortical regions (β range |0.089| to |0.115|, pFDR < 0.05). RARRES2, VEGFA, and THBS2 across four regions (β range |0.088| to |0.115|, pFDR < 0.05); CRP, IGFBP4, and MMP12 across three regions (β range |0.088| to |0.156|, pFDR < 0.05), and SERPIN1 across two regions (β range |0.091| to |0.109|, pFDR < 0.05). In contrast, for signatures that associated with greater cortical volumes, DNAm NTRK3, NCAM1, NOTCH1, SEMA3E, GZMA, and GDF.8 all associated with higher volume across numerous cortical regions; the overlap in signatures that associated with multiple features of the cortex is illustrated in Fig.5b.

Figure 5. DNAm signatures associate with favourable brain and cognitive outcomes. (a) Number of associations that were significant (p < 0.05) is highlighted by the coloured bar; number that pass FDR significance are in overlayed dark grey bars (b) FDR-significant associations where elevated DNAm signatures associated with reduced regional cortical volumes, associations that remained significant (pFDR < 0.05) in models additionally controlling for aspects of lifestyle (hypertension, smoking, alcohol consumption and BMI) are highlighted in yellow. Y axis and cortical regions panel displays the 34 cortical ROIs in each hemisphere defined by using the Desikan-Killiany atlas (brain is inflated to allow a better view of gyri and sulci created using ggseg package). From top to bottom brain maps are: the medial right hemisphere, lateral right hemisphere, medial left hemisphere, lateral left hemisphere.
DNAm associations with regional white matter microstructure

In addition to an examination of how multi-omic signatures associated with an account of global white matter integrity (gFA and gMD, as presented in Fig. 2-3, Supplementary Data 12; 20 associations with 16 unique DNAm signatures), we examined individual white matter tract FA and MD across 24 individual white matter tracts to uncover whether there was regional specificity between inflammation and WM integrity.

Results of all 202 significant (pFDR < 0.05) associations between individual tract FA and MD for baseline models, alongside WBC-adjusted and lifestyle-adjusted models, are presented in Supplementary Data 12. In fully-adjusted models, the inclusion of lifestyle factors in models attenuated the majority of associations (n = 176) by an average of 21.1%, and increased associations in 26 instances by an average of 5.6%. Overall, this left 179/202 associations still significant, indicating a robust relationship between DNAm signatures and regional tract microstructure. Fig. 6 displays which DNAm signatures had the largest number of significant associations with poorer regional white matter tract integrity (Fig. 6), and which white matter tracts emerged as most vulnerable to DNAm signatures (Fig. 6b-d). Twenty tract ROIs displayed a significant association with one or more DNAm signatures. The only tracts that displayed no significant associations with signatures were the corticospinal tract (CST), inferior-fronto-occipital fasciculus (IFO), the retrolenticular part of the internal capsule (RLIC), and the hippocampal-cingulum (CGH).

The same DNAm signatures that associated with poorer global white matter microstructure (gFA and gMD) arose again: with DNAm CRP, MMP12, SERPIND1, PIGR, VEGFA, and FGF.21 associating with the largest number of significant alterations in regional WM integrity (Fig. 6a). DNAm SERPIND1 had the highest number of significant associations with reduced white matter integrity across white matter tracts, associating with either decreased FA or increased MD across 9 individual tracts (FA β range |0.103| to |0.185|, MD range |0.111| to |0.170|, pFDR < 0.05).

While Fig. 6a highlights that the majority of associations (n = 126) were between 19 DNAm signatures that appeared detrimental for regional white matter integrity, 76 of the 202 associations were instances of either increased FA or decreased MD reflective of putatively ‘healthier’ white matter (SEMA3E, IL19, NTRK3, CD209, WFIKKN2, CNTN4, NCAM1, SLITRK5, INSR, ESM1, OMD, ACY1, SPOCK2, LTA.LTB).

For the vast majority of associations, DNAm signatures outperformed proteins in associating with white matter tract integrity across regional tracts (results reported in Supplementary Data 11): for fractional anisotropy, in contrast to the 93 significant associations found with DNAm and regional FA, protein signatures were only significant in two instances: the association with the protein SMPD1 with EC FA (β = -0.135, pFDR = 0.013) and the association with ADIPOQ and UNC FA (β = -0.176, pFDR = 0.002). After adjustment for lifestyle factors, these protein-WM associations were attenuated by an average of 24% and were no longer significant.
Figure 6. Number of significant associations between DNAm signatures and poor regional white matter microstructure. (a) DNAm signatures that associate with poor white matter integrity metrics (decreased FA and increased MD), where number of associations that were significant (p < 0.05) is highlighted by the coloured bar; number that pass FDR significance are in overlayed dark grey bars. (b-d) visualise which WM tracts had the highest number of significant associations with unique DNAm signatures; to the left of panels, white matter tracts are visualised in the glass brain plots for each broad group of fibre tract; (b) projection (c) association, and (d) commissural fibres. To the right, number of significant associations (p < 0.05) is highlighted by the coloured bar and number that pass FDR significance are in overlayed dark grey bars. Abbreviations: anterior corona radiata (ACR); anterior limb of internal capsule (ALIC); body of corpus callosum (BCC); corpus callosum (CC); cingulum (cingulate gyrus) (CGG); Cingulum (hippocampus) (CGH); corona radiata (CR); corticospinal tract (CST); external capsule (EC); fornix (column and body of fornix) (FX); fornix / Stria terminalis (FX.ST); genu of corpus callosum (GCC); internal capsule (IC); inferior fronto-occipital fasciculus (IFO); posterior corona radiata (PCR); posterior limb of internal capsule (PLIC); posterior thalamic radiation (include optic radiation) (PTR); retrolenticular part of internal capsule (RLIC); splenium of corpus callosum (SCC); superior corona radiata (SCR); superior fronto-occipital (SFO); superior longitudinal fasciculus (SLF); Sagittal stratum (SS); and the uncinate fasciculus (UNC).
Inflammation-associated DNA methylation (DNAm) signatures in a biological and clinical context

We contextualised these findings with known epigenome-wide association study (EWAS) associations between CpG sites and health outcomes (Supplementary Data 8) alongside incident disease associations with composite DNAm signatures reported previously (Gadd et al., 2022b) in the wider Generation Scotland cohort, of which our population sample is a subset. The number of significant associations with brain health outcomes did not directly relate to degree of multi-morbidity (see Fig. 7), though associations were largely consistent in direction of effect (i.e., DNAm signatures that associated with poor brain health outcomes associated with higher incidence of various diseases; signatures that associated with favourable brain health outcomes correspondingly associated to decreased risk of incident disease).

Figure 7. Brain-related DNA methylation signatures identified in this study previously associated with incident disease outcomes.

Disease outcomes studied in Gadd et al. (2022b) are listed in the centre of the network, where positive associations between 73 DNA methylation signatures are in green and negative associations are in red. Nodes are DNA methylation signatures and are coloured according to which panel they were derived from, SomaScan (turquoise) or Olink (red). To the right of the plot, DNA methylation signatures that associated with poor brain and cognitive ageing outcomes are listed in order of number of associations (e.g., DNA methylation PIGR had n = 32 associations with poor aspects of brain health, followed by SERPIND1 with n = 26, etc.) To the left of the chord diagram, DNA methylation signatures pertaining to favourable brain health outcomes are listed in a similar fashion (where DNA methylation SEMA3E had n = 32 associations with markers of favourable brain health metrics, followed by DNA methylation NCAM1 with n = 23). Points are filled grey if they had < 5 (pFDR < 0.05) associations with brain-health outcomes. Abbreviations: Chronic obstructive pulmonary disease (COPD); ischaemic heart disease (IHD); rheumatoid arthritis (RA); inflammatory bowel disease (IBD).
Finally, we conducted mediation analyses to investigate if any of the neuroimaging phenotypes mediated the association between DNAm signatures and cognition (Supplementary Data 14). Of individual cognitive test scores, processing speed had the highest number of significant associations with individual DNAm signatures ($n = 18$) and so mediation modeling was performed to determine which aspects of brain structure drove this relationship; 11 DNAm signatures associated with lower processing speed and 7 with higher processing speed.

Mediations via global brain structural metrics were investigated first (Supplementary Data 14). Here, global grey matter volume was found to significantly and partially mediate the association between 12 DNAm signatures (MMP12, PIGR, SKR3, FGF.21, RARRES2, THBS2, CRP, NTRK3, NOTCH1, SEMA3E, NCAM1) and processing speed (percentage attenuation between $0.3\%$ and $9.9\%$; Supplementary Fig.9, Supplementary Data 14). Processing speed was the main cognitive outcome of interest as of the 109 DNAm signatures, while 20 had nominally significant ($p < 0.05$) associations with a general factor of cognitive ability ($\gamma$), only DNAm CRP remained significant after multiple testing ($p_{FDR} < 0.05$), global GM volume mediated the association (attenuation $9.8\%$; indirect effect size $\beta = -0.014$) only (Supplementary Fig.10).

In instances where global cortical volume emerged as a significant mediator of DNAm-cognitive associations (MMP12, PIGR, SKR3, RARRES2, THBS2, NOTCH1, NTRK3, SEMA3E, NCAM1), analysis of regional volumes was performed to examine whether alterations in specific cortical regions were driving cognitive changes (Supplementary data 15). Fig.8 displays how for all of these 9 DNAm signatures, regional volume in the superior temporal and fusiform cortex emerged as the unifying significant mediators of the association between inflammation and processing speed (attenuation $6.9\% − 13\%$). For DNAm signatures that associated with poorer processing speed test scores, there was mediation via lower superior temporal volume ($\beta$ range $−0.009$ to $−0.015$); those that associated with higher processing speed scores were mediated by higher superior temporal volumes ($\beta$ range $0.009$ to $0.016$).
Figure 8. Regional cortical volumes mediate the association between DNAm signatures and processing speed. (a) Significant associations between DNAm signatures and processing speed are mediated by various regional cortical volumes neuroimaging metrics; points show indirect effect size, where turquoise points represent significant mediation and grey circles non-significant associations. (b) Visualisation of cortical volumes (ggseg3d package) partially mediating the relationship between DNAm MMP12 and processing speed. (c) Path diagram of relationship between DNAm MMP12, superior temporal volume and processing speed (controlling for age, sex, neuroimaging site, batch and edits), n = 704.
Discussion

This is the first study to examine multiple immune-related DNAm signatures in relation to brain and cognitive ageing outcomes in broadly healthy individuals in mid to late adulthood. Our findings suggest that individual protein DNAm signatures capture inflammation-specific biomarker signals relevant to brain health and are of interest for risk-stratification of inflammation-related morbidities. Many of these DNAm-brain associations were larger than the analogous proteome-brain associations, broadly independent of immune-cell proportions, clinical risk factors, and had previously been linked to various age-related diseases, reinforcing the central role that inflammation plays in health trajectories. Finally, we explored relationships between inflammatory-related DNAm signatures, brain structure and cognitive ability through mediation modelling, finding support for the hypothesis that both global reductions in grey matter volume and atrophy in AD-related regions drive inflammatory-DNAm-associated declines in processing speed.

Prior work has indicated that the DNA methylome reliably indexes variance in circulating inflammatory proteins (Zaghlool et al., 2020), and can capture an immune-mediated relevant feature to aspects of brain ageing (Gadd et al., 2022a). Here, we find 470 associations between 73 distinct DNAm signatures and brain and cognitive phenotypes, lending weight to the theory that DNAm associated with circulating levels of plasma proteins reflect common hallmarks of inflammation-associated brain ageing. Stratifying associations by direction of effect revealed that the top DNAm signatures linked to poor brain health outcomes were PIGR, SERPIND1, CRP, MMP12, VEGFA, IGFBP4, FGF.21, SKR3, PRSS2, THBS2, CCL17, RARRES2, CCL11, ICAM5, SIGLEC1 and TGF.alpha (each associating with ≥ 5 markers of poor brain health in consistent directions, Supplementary Data 13).

These proteins are heavily associated with the innate and adaptive immune response and inflammatory signalling cascades (Supplementary Data 4-6). Moreover, many of these DNAm signatures were for proteins not heavily expressed in the brain (Supplementary fig.12), reinforcing the potential of peripheral-based surrogates to anticipate inflammation-related brain and cognitive ageing risk. In contrast, the top signatures that displayed associations with favourable brain health had higher expression in the brain (Supplementary fig.13) and were related to proteins that had functional roles in neurogenesis, neural development, cellular signalling, and recruitment of adaptative immune cells (SEMA3E, NCAM1, NTRK3, CNTN4, GDF.8, NOTCH1, WFIKKN2, SLITRK5, OMD, INSR; ≥ 5 markers of favourable brain health). Throughout, effect sizes were small according to Cohen (1992), but comparable to previous findings examining the association of composite inflammatory related DNAm signatures with brain structure (Conole et al., 2023, 2021; Green et al., 2021), protein-brain associations (Gadd et al., 2022a; Harris et al., 2020) and the association of vascular risk factors (VRFs)(Cox et al., 2019). Our results reinforce the utility of blood-based DNAm signatures as biomarkers of health (Thompson et al., 2022; Yousefi et al., 2022; Cheng et al., 2023) and highlight their conceivable utility to capture variance in systemic inflammation that contributes to subtle differences in brain health at both global and focal levels, potentially detectable prior to significant cognitive impairment or diagnosis of neurodegenerative disease.
The cortical regions most strongly associated with differences in DNAm signatures were the fusiform, supra-marginal, superior parietal and temporal regions of the brain's cortex (Fig. 5). Of these, lower cortical volume in fusiform gyrus had the highest number of significant hits with various pro-inflammatory DNAm signatures (PIGR, SERPIND1, CRP, MMP12, RARRES2, VEGFA, IGFBP4, THBS2; Fig. 5b), indicating that this structure may be more susceptible to inflammatory insult. This is consistent with previous findings of circulating inflammatory protein's association with brain structure (Wersching et al., 2010) as well as inflammatory-DNAm surrogates (Conole et al., 2021; Green et al., 2021), and more widely is similar to patterns of cortical atrophy linked to neurodegenerative disease and mild cognitive impairment (MCI) (Steenwijk et al., 2015; Edmonds et al., 2020), as well as vascular risk factor (VRF) associations seen in cognitively normal adults (Cox et al., 2019). Our findings pertaining to subcortical atrophy and increased WMH burden for these pro-inflammatory signatures – with the strongest seen for DNAm MMP12 and WMH – also concur with the literature on brain structural determinants of cognitive ageing (Hilal et al., 2015), and inflammation-related risk (Marsland et al., 2015). When assessing the degree to which alterations in brain structure mediated the associated between DNAm signatures and aspects cognitive function, regional cortical volumes in the superior temporal and fusiform cortex emerged as the unifying significant mediators of the association between inflammation and processing speed (attenuation 6.9-13%), where inflammatory-related DNAm signatures that were associated with poorer test scores (DNAm MMP12, PIGR, SKR3, RARRES2, THBS2) appeared to mediate this effect via atrophy in these regions (Fig. 8). These findings are in line with previous work which has established a link between temporal lobe atrophy and declines in processing speed both in non-clinical populations and studies investigating cognitive impairment in dementia (Zhao et al., 2018; Pelkmans et al., 2021).

Multiple inflammation-related DNAm signatures associated with both diffuse and regional white matter differences. While we report 20 associations between global white matter integrity (gFA or gMD; \( \beta \) range |0.091| to |0.137|, pFDR <0.05) with 16 unique DNAm signatures (PIGR, CRP, MMP12, IGFBP4, FGF.21, VEGFA, SERPIND1, ACY1, CCL25, SLITRK5, PRSS2, FAP, AFM, MMP.1, NCAM1, NTRK3), we also found that increased DNAm signature scores were associated with widespread regional white matter changes, particularly in the association fibre tracts (with higher in MD in the external capsule, and lower FA in the fornix, superior fronto-occipital fasciculus and sagittal striatum) and thalamic radiata (PTR, ALIC, PLIC). These tracts connect some of the more metabolically active regions of the brain that require a greater blood supply (Bartzokis et al., 2004), which may go towards explaining how peripheral inflammation may exert a more deleterious impact in these regions. In line with this, our patterning of alterations in tract microstructure is similar to reports of the relationship between vascular risk and WM integrity (Cox et al., 2019). Given that many of these associations remained when various VRFs were accounted for (e.g., smoking, hypertension, BMI), our findings add to the evidence base that there may be a chronic inflammatory driver of widespread alterations in white matter integrity that either shares similar mechanisms to, or underpins, that of vascular-mediated risk.

We also highlight that DNAm-based markers of inflammation are better at capturing this association with white matter microstructure than the circulating proteome: we report similar global alterations in WM integrity for a wider panel of immune-related DNAm markers (\( \beta \) range |0.091| to |0.137|, n = 680) vs weak associations for equivalent
protein levels ($\beta$ range |0.005| to |0.073|, n = 680). In the three examples to date examining DNAm-based signatures of inflammation with DTI metrics, global measures of WM tract integrity have been linked to a DNAm-based score of CRP ($\beta$ range |0.077| to |0.115|, n = 565, age ± 59.7) in individuals in mid-to-late adulthood STRADL (32), older age LBC1936 ($\beta$ range |0.124| to |0.162|, age 72.4) (Conole et al., 2021), and in a neonatal cohort ($\beta$ range |0.130| to |0.236|, n = 214, age < 1 year) (Conole et al., 2023). Outside of the neuroimaging-epigenetics literature, these have previously been considered as tracts that may be particularly vulnerable to modifiable risk factors associated with neurodegeneration (Wassenaar et al., 2019) and inflammation (Marsland et al., 2015; Bettcher et al., 2015; Walker et al., 2019), and reinforce the evidence that sustained inflammation may negatively affect white matter integrity in midlife (O'Donovan et al., 2021). Given that both diffuse white matter injury and regional alterations can in some cases precede wider grey matter injury or signs of cognitive impairment (Araque Caballero et al., 2018; Parker et al., 2022), further characterisation of the specific white matter degeneration associated with inflamming may be particularly useful for identifying cases of accelerated brain-ageing prior to significant cognitive impairment.

Our principal finding is that DNAm signatures displayed stronger and more significant associations with brain health outcomes than that of matched protein-levels. We found consistent effect directions between protein levels and DNAm signatures and many reported associations are in line with previously reported brain-proteomic associations (Gadd et al., 2022a; Harris et al., 2020). However, of 470 associations, there were only 10 instances where protein levels displayed stronger and statistically significant associations with various brain health outcomes than matched DNAm signatures (see Supplementary Data 11). One explanation for this finding is that DNAm patterns can relay information about gene expression regulation and the functional state of cells or tissues, which may be more informative about systemic inflammatory processes than phasic protein levels alone. Methylation at different CpG sites react to the environment, and may reflect both the external (e.g., exposure cigarette smoke) and internal (e.g., metabolic state, age-related functioning) context of an individual, as well as biological adaptation to disease states (Nabais et al., 2023). Numerous CpG sites present within the 73 DNAm signatures that we found to be relevant to brain health have been used in other composite methylation indices of inflammation (Barker et al., 2018; Myte et al., 2019; Zhang et al., 2017; Wielscher et al., 2022) and have been linked to inflammatory-related disease risk (Kalla et al., 2023) and severity (Somineni et al., 2019), with previous DNAm signatures demonstrating their potential as more stable markers of chronic inflammation longitudinally (Stevenson et al., 2020) and ability to index degrees of inflammatory multi-morbidity (Conole et al., 2023).

More widely, studies have illustrated the additive value of DNAm biomarkers for capturing dose-related to various exposures such as cigarette smoke and alcohol consumption (Langdon et al., 2021; Colicino et al., 2021; Yousefi et al., 2022; Ladd-Acosta and Fallin, 2019). In our example, DNAm signatures of proteins may capture a cumulative impact of sustained effects from raised peripheral inflammation beyond measuring protein levels themselves at a single time point. This is particularly relevant when investigating the impact of chronic inflammation and incident disease outcomes, where acute markers of inflammation are often relied on to characterise inflammation levels. As these proteins rise and fall within short time periods, in the absence of average readings to establish a baseline,
there is substantial noise to overcome which may mask the true impact of chronic inflammation on incident health outcomes of a chronic, rather than acute, inflammatory-nature. By contrast, DNAm signatures may provide a more stable measure of inflammation and reflect the amassed collateral impact of exposure to inflammation overtime. This would go towards explaining why we observe stronger associations between DNAm signatures and neural-phenotypes here and in previous studies (Stevenson et al., 2020; Green et al., 2021; Edmondson-Stait et al., 2022; Conole et al., 2021).

Given that differences in DNAm levels can be both as a result of modification via lifestyle and disease states, as well as stochastic changes with age, it is possible that the DNAm signatures here are capturing an environmental variance of chronic inflammation between individuals that is linked to worse or better brain health outcomes. Many CpGs within DNAm signatures included sites previously associated with lifestyle factors. Among these was a site mapped to the AHRR gene (cg05575921) which is strongly linked to smoking (present in n = 25 DNAm signatures; Supplementary Data 3). Equally, 23 DNAm signatures contained a site strongly tied to alcohol consumption (cg06690548; SLC7A11). Another site (cg02650017; PHOSPHO1 gene) that was in four of the DNAm signatures (NTRK3, NCAM1, B2M, OMD) has previously been found to vary as a consequence of BMI's impact on circulating CRP levels (Wielscher et al., 2022). There was also overlap with 24 sites that are used to determine the DNAm age metrics (Hannum et al., 2013), particularly DNAm MMP12 which contained 5 loci previously linked to age.

Several other loci common across multiple signatures (cg17501210, cg16411857, cg09349128, cg07190917, cg12054453, cg12992827, cg26470501, cg01059398) have been shown to associate with diseases directly linked to systemic inflammation such as IBD, lupus and arthritis (Supplementary Data 8, http://www.ewascatalog.org). Our findings suggest that these DNAm signatures could be useful for stratification of diseases that have an inflammatory pathophysiology, and reinforces previous findings looking at the association of these signatures with both classical inflammatory conditions such as IBD, rheumatoid arthritis and COPD as well as those pertaining to brain health outcomes such as stroke and depression (Gadd et al., 2022b). While many of the DNAm signatures described here proxy for proteins that have obvious and direct roles in systemic inflammation – such as the pro-inflammatory cytokines and chemokines (OSM, CCL11, CXCL9, CXCL10, CXCL11, IL19) and complement cascade components (CRP, C5), or JAK-STAT, NF-κB and PI3K-Akt signalling cascades (Supplementary Data 6) – others are proteins that rise in tandem with activation of pro-inflammatory mediators as part of the wider epiphenomena of chronic inflammation. Peripheral based immune-DNAm changes have been previously linked to age-related cognitive dysfunction in patients with mild cognitive impairment (Chouliaras et al., 2018), WMH burden (Yang et al., 2022), and incidence of different neurodegenerative disorders (Nabais et al., 2021), indicating that common inflammation-related pathophysiology is shared across both neurodegenerative disease and normative cognitive decline. The future applications of DNAm in this context are promising as inflammatory-related DNAm signatures could enable clinical monitoring, and demonstrate (or anticipate) response to treatment. Though multi-time point sampling of DNAm to assess these objectives is rare, a few recent studies have demonstrated differential DNAm signatures associated with treatment progression and therapeutic intervention over interval periods (Nair et al., 2020; Julià et al., 2022; Somineni et al., 2019). From a precision
medicine perspective, this is particularly compelling as our findings suggest that DNAm differences associated with plasma protein levels have the potential to improve current clinical risk prediction tools for accelerated brain ageing.

The results of our current study should be interpreted in the context of the following limitations: first, DNA methylation measured in peripheral whole blood was used both for the original development of the 109 DNAm signatures and our current projection of these signatures in STRADL in examining brain health relationships. Blood may not be the tissue most directly relevant for assessing the impacts of inflammation on the brain, though is arguably more clinically useful in being less invasive more readily accessible (Yousefi et al., 2022). Second, blood contains many cell types: these may not be the same as matched proteins (many of which, though circulating in the system circulation, are derived from the liver, thymus and CNS) and using whole blood DNAm may mask some cell-specific epigenetic variation. Though we had a sensitivity analysis that adjusted for white blood cell contributions (where 408/470 associations retained significance following incremental inclusion of monocyte, granulocyte, B-cell, CD4T, CD8T, and natural killer cells), as we did not have full blood count (FBC) related markers measured at the same time point as neuroimaging we did not include these in models (though we do report the associations of these from the baseline assessment in Supplementary Fig.6). Several recent reports have emphasised the importance for adjusting for immune-cell proportions in examining DNAm differences associated with disease heterogeneity and our limited cell counts may miss key immune subpopulations. Future work that examines these signatures alongside comprehensive immune and blood cell phenotyping is therefore optimal for investigating the utility of DNAm signatures in reliably indexing inflammatory exposure. This is particularly important because of age-specific changes in immune-ageing within immune cell subpopulations which may in part drive the predisposition towards a chronic inflammatory state in older age (Bergstedt et al., 2022). A new panel that deconvolves cell proportions in peripheral blood has significantly broadened the scope of DNAm-immune phenotyping in line with this aim. This approach parses out subtypes (providing a reference library of 56 immune cell profiles) which could assist in distinguishing between extrinsic (differences in cell-type proportions) and intrinsic (change in DNAm that is directly associated with the trait across cell types) signals to this effect (Salas et al., 2022).

Third, neither the Olink nor the SomaScan assays used to measure protein levels quantify absolute measures of proteins: aptamers potentially have more than one target in plasma samples, and alterations to the protein structure due to oligomerisation, post-translational alterations or genetic polymorphisms may significantly but unpredictably affect binding affinity and measurement (Joshi and Mayr, 2018). Fourthly, though we examined the influence of the lifestyle factors such as smoking, alcohol consumption, BMI, and hypertension to our associations, there are other non-modelled confounding factors that may influence these observations. For example, as we did not control for depression diagnosis in this study (a key selection feature of the STRADL population group) the associations observed between DNAm signatures and brain health outcomes may have been influenced by neuropsychiatric cases, prescription medications, and other comorbidities. Finally, as study is a cross-sectional design, our ability to draw any causal conclusions from correlational analyses is limited. Longitudinal studies are needed to further clarify whether the DNAm signatures related to brain health outcomes identified here play a causal role in driving inflammation-associated struc-
tural changes, whether they represent a direct downstream consequence of these changes, or whether they are induced by certain factors associated with these but not necessarily driven by chronic inflammation itself. Mendelian randomisation studies that integrate genomic and epigenomic drivers of these changes are one such method to delineate directionality in these relationships.

Conclusion

In summary, this study leverages a large number of DNA methylation signatures to quantify the chronic inflammatory component to individual differences in brain and cognitive ageing. Using proteomic, DNA methylation, and clinical data from 709 individuals in the STRADL cohort, we identified inflammatory-related DNA methylation signatures associated with a range of cognitive and multi-modal neuroimaging markers of brain health – including relative brain ageing, global brain structure, regional cortical patterning and individual white matter tract integrity, with inflammation-linked structural changes driving observed associations with cognitive functioning. This work demonstrates the value of multi-omic characterisation of inflammation to capture highly disease-relevant functions and justifies further efforts to collect DNA methylation in larger samples. Following further testing and validation in clinically-ascertained samples, the use of DNA methylation biomarkers of inflammatory exposure may have precision-medicine applications for cognitive risk stratification.
Methods

The study sample consisted of individuals recruited from The Stratifying Resilience and Depression Longitudinally (STRADL) cohort, which is a subset of individuals from Generation Scotland: The Scottish Family Health Study (GS). Generation Scotland constitutes a large, family-structured, population-based cohort of >24,000 individuals from Scotland, who were recruited between 2006-2011 Smith et al. (2013). Of this cohort, $N = 1188$ completed additional health assessments and biological sampling ~ 5 years after GS baseline Habota et al. (2019). During part of the follow-up assessment, participants were sent study packages that included questionnaires where BMI was calculated using height (m) and weight (kg). Participants were also asked whether they were currently consuming alcohol and whether they were current, former, or non-smokers, and self-reported whether they had hypertension.

Of these, $N = 1065$ individuals had proteomic data available and $N = 778$ of these had complete general demographic, DNAm and proteomic data available, and cognitive data and neuroimaging data were available for a subset of the 778 participants ($n = 709$ neuroimaging, $n = 728$ cognitive; Supplementary Data 1, accessed at https://github.com/EleanorSC/Inflammatory-DNAm_STRADL/blob/main/Supplementary_Data/Supplementary_document_Conole2023.xlsx). Box 1 provides an overview of the study design and sample. Details on the two populations used to generate the DNAm signatures for circulating proteins (the KORA cohort and Lothian Birth Cohorts) have been reported previously Barbu et al. (2021). The STRADL population sample comprises broadly healthy individuals free from neurodegenerative disease. Data on whether the participants were current smokers at the time of interview were missing for 70 participants; only 6% ($n = 38$) of the sample were self-reported current smokers. The rest of the sample had full lifestyle data recorded, with an average BMI of 27.8 kg/m$^2$ ($\pm$ 5.17), 33% ($n = 233$) had a self-reported diagnosis of hypertension, and 88% ($n = 623$) of participants self-reported that they were current drinkers at the time of participation. There was a fairly even distribution of sex (58% female).

Ethical approval was formally obtained from the NHS Tayside committee on research (reference 14/SS/0039), and all participants provided their written informed consent.

Proteomic Measurement in STRADL

Blood was collected by venepuncture, and plasma proteins were measured using the SOMAscan assay platform (SomaLogic Inc) in 1065 individuals in STRADL. SOMAScan is an aptamer-based assay that allows for the simultaneous measurement and quantification of proteins. After raw processing and initial quality control (QC), details of which have been outlined previously Gadd et al. (2022b), 4235 proteins were measured, and 793 epitopes matched between
KORA and STRADL population samples and were included for analysis. The abundance of each protein was transformed by rank-based inverse normalization and regressed onto age, sex, and 20 genetic principal components. There were 778 individuals with proteomics data and DNAm data in STRADL, and 709 who additionally had full neuroimaging data.

**DNAm Measurement in STRADL**

Measurements of blood DNAm in STRADL were processed in two sets on the Illumina EPIC array using the same methodology as those collected in the wider Generation Scotland cohort, as described previously (Gadd et al., 2022b), with QC details provided in the supplement of Gadd et al. (2022b). Briefly, samples were removed based on sex mismatches (DNAm-predicted and genotype-based), low detection p-values for CpGs and saliva samples, and genetic outliers. After quality control, 793,706 and 773,860 CpGs were available in sets 1 and 2, respectively.

**Elastic Net Protein DNAm Signatures**

DNAm signatures were created for 109 different circulating plasma proteins, many of which were classed as inflammatory or immune-related (see Supplementary Data 4). The detailed protocol of these DNAm signatures has been outlined fully by Gadd et al. (2022b). Penalized regression models were used to generate DNAm signatures from two separate population samples. An elastic net penalty was specified ($\alpha = 0.5$), and cross-validation was applied. DNAm signatures were then projected into STRADL for the present analyses.

In the first population, the Lothian Birth Cohort 1936 (LBC1936), 160 plasma proteins were analyzed by Olink assays (Olink neurology 92-plex or the Olink inflammation 92-plex proximity extension assays; Olink Bioscience, Uppsala Sweden). For LBC1936 trained signatures, as test-set comparisons were not available between population cohorts (due to different proteomic assays used in STRADL vs LBC1936), instead a holdout sample was defined with 10-fold cross-validation carried out on the remaining data. The outcome of the models was protein levels, with 428,489 potential CpG features per model in the LBC1936 training. This returned 36 DNAm signatures with $\geq 1$ CpG features that passed the testing threshold ($r > 0.1$ and $p < 0.05$). Following this, new predictors were trained for these 36 proteins and generated using 12-fold cross-validation and tested externally in STRADL ($n = 778$) and internally in LBC1921 ($n = 162$). Of these, 21 DNAm signatures passed the threshold ($r > 0.1$ and $p < 0.05$) in at least one of the external testing sets, and 4 signatures which did not have external comparisons were included based on holdout performance, resulting in a total of 25 DNAm signatures from the Olink trained set.

In the second population sample, the German-based KORA cohort, 793 proteins were mea-
sured from plasma by the SOMAscan platform (Version 3.2). As with the LBC1936 trained DNAm signatures, the outcome of the model was circulating protein levels, with 397,630 potential CpG features in the KORA model. For KORA-trained signatures, a 10-fold cross-validation was applied, and signatures were tested in STRADL \( (n = 778) \). Of the 480 signatures that generated more than one CpG site to encode for a protein, 84 passed the threshold for inclusion (Pearson \( r > 0.1 \) and \( p < 0.05 \)) after testing in STRADL.

These 109 DNAm signatures were then considered in the STRADL dataset alongside aspects of neuroimaging and cognitive data. DNAm at each CpG site was scaled to have a mean of zero and variance of one, with scaling performed separately for GS sets. As there was overlap for 6 signatures which corresponded to the same protein of interest (GZMA, MMP.1, CXCL10, NTRK3, CXCL11, EN.RAGE | S100A) between panels, in the manuscript text, a ‘x’ or ‘_olink’ demarcates that this score was derived from the Olink panel. Additionally, as there were 11 DNAm signatures derived from the Olink panel where no matched protein level could be compared (FcRL2, G.CSF, GDF.8, N.CDase, NEP, SIGLEC1, SKR3, CD6, EN.RAGE, FGF.21, TGF.alpha) because of assay differences between Olink vs SomaScan, these are not reported in paired-associations (Supplementary Data 11), but are reported in the wider analyses of DNAm signatures with brain health outcomes (Supplementary Data 12).

Distributions of both the 109 DNAm signatures and equivalent proteins are presented in Supplementary Fig.1-2. The weights used to derive these signatures are provided in Supplementary Data 2. SomaScan signatures were built from an average of 96 CpG sites, whereas Olink signatures were built from an average of 76 sites; the number of CpG sites per DNAm signature is reported in Supplementary Data 3, and CpG counts alongside mapped phenotypic traits from the MRC-IEU catalog (http://www.ewascatalog.org) are reported in Supplementary Data 4.

**Immune-cell counts**

For a sensitivity analysis, we adjusted for white blood cell (WBC) proportions; the meffil (Min et al., 2018) implementation of the Houseman method was used to calculate these. The ‘blood gse35069 complete’ panel (taken from (Reinius et al., 2012) was used to impute measures for Monocytes, Natural Killer cells, B cells, Granulocytes, CD4+ T cells, and CD8+ T cells. Full reporting of cell count distributions is provided in Supplementary Data 1.

**Clinical health and lifestyle data**

All lifestyle and health factors variables entered into models as covariates are summarized in Supplementary Data 1. Participants from the STRADL cohort had health and lifestyle data collected, either by self-report or, in the case of BMI (calculated using height and weight), mea-
sured by clinical staff during baseline recruitment, the details of which have been reported previously (Habota et al., 2019; Barbu et al., 2021). Alcohol intake was self-reported as part of a pre-clinical questionnaire where participants were asked whether they were “never,” “for-
mer,” or “current” drinkers. Similarly, smoking status was obtained by asking participants if they had smoked in the past 12 months. In analyses, both current smoking status and drinking status were coded as binary variables (current = 1), and former drinking/former smoking were not considered. For the population sample used in analyses (n = 709) with complete proteomic, DNAm, and brain phenotypic data, 88% (n = 682) of participants self-reported that they were current drinkers at the time of participation. Data on whether the participants were current smokers at the time of the interview was missing for 83 participants, leaving only 6% (n = 42) of the sample who were current smokers. APOE e4 status was coded as a dose dependent on how many alleles present - 0, 1, or 2 - to generate a continuous numeric variable (APOE haplotypes were coded as follows: e2e2 = 0, e2e3 = 0, e3e3 = 1, e3e4 = 2, e4e4 = 2).

MRI acquisition and analyses

Brain structural MRI scans were performed on STRADL participants (T1 imaging of N = 1070 participants scanned between 2015-2019) at two sites in Scotland (N = 544 from Aberdeen and N = 526 from Dundee). Full reporting of MRI acquisition and QC is supplied in Stolycyn et al. (2020) and in the Supplementary Methods. For Aberdeen MRI, data was acquired at the Aberdeen Royal Infirmary with a Philips Achieva 3T TX-series scanner (Philips Healthcare, Best, Netherlands). At the Dundee site (Ninewells Hospital), a Siemens 3T Prisma-FIT (Siemens Healthineers, Erlangen, Germany) scanner was used. The following data was obtained: 3D T1-weighted fast gradient echo with magnetization preparation; 3D T2-weighted fast spin echo; 3D Fluid Attenuation Inversion Recovery (FLAIR); Diffusion Tensor Imaging (DTI); and Susceptibility Weighted Imaging (SWI) or T2*-weighted gradient echo (Habota et al., 2019).

From T1 images, structural neuroimaging measures were obtained using FreeSurfer version 5.3 (Dale et al., 1999; Fischl et al., 2004). For 34 bilateral cortical regions, mean cortical thickness, cortical surface area, and cortical volume were derived (cortical regions defined by the Desikan-Killiany atlas (Desikan et al., 2006). Alongside this, the volumes of 14 bilateral subcortical structures - including accumbens, amygdala, caudate nucleus, hippocampus, pallidum, putamen, thalamus, and four cerebellar regions - were also extracted with FreeSurfer. The number of QC edits made per individual was recorded and used as a covariate in relevant analyses.

In total, N = 980 participants remained after QC - removing participants with any missing values, as well as participants whose intracranial volume (ICV) measure and global cortical
measures, i.e., overall cortical volume (sum of regional cortical volumes), overall surface area (sum of regional surface areas), were more than three standard deviations away from the sample mean (Stolicyn et al., 2020). Details of MRI acquisition and quality control process are outlined in Stolicyn et al. (2020) in Supplementary A.1.2 and A.1.3. Participants whose demographic information was missing were also removed. There were 225 FreeSurfer-derived features available for each participant (204 cortical and 21 subcortical features). Visually inspected Fluid-attenuated inversion recovery (FLAIR) scans were available for 940 subjects to obtain Fazekas score, an index of WMHs.

Diffusion Tensor Imaging (DTI) data was also collected to examine white matter microstructure. DTI data underwent pre-processing and QC using standard tools from FSL (https://fsl.fmrib.ox.ac.uk/fsl/fslwiki). Tract-Based Spatial Statistics (TBSS) was then performed based on the DTI protocol outlined by The Enhancing NeuroImaging Genetics through Meta-Analysis (ENIGMA) Consortium (http://enigma.ini.usc.edu/protocols/dti-protocols/). Additionally, we performed region of interest (ROI) extraction analyses using ENIGMA protocols to extract measures of fractional anisotropy (FA) and mean diffusivity (MD) (http://enigma.ini.usc.edu/protocols/dti-protocols/). To categorize white matter tracts, we used the Johns-Hopkins University DTI-based white matter atlas (Mori and Van Zijl, 2007), resulting in 5 unilateral tracts and 19 bilateral tracts. These tracts included ten association fibers, three commissural fibers, eight projection fibers, and four thalamic radiations. Abbreviations are used throughout this manuscript and can be found in Supplementary Data 16.

In total, we looked at 112 brain health phenotypes, consisting of cognitive test scores measures (including both global measures of cognitive function as well as scores pertaining to different cognitive domains, e.g., processing speed, n = 7), global neuroimaging measures (such as total brain volume, global grey matter volume, n = 8), regional neuroimaging measures (cortical and subcortical volume measures, n = 45), diffusion-tensor imaging (DTI) derived white matter tract microstructure measures for individual white matter tracts (fractional anisotropy and mean diffusivity; FA and MD, n = 50), as well as APOE haplotype and estimated brain age. The ways these individual metrics were derived are outlined below.

**Neuroimaging phenotypes**

Distributions of all global neuroimaging metrics (pertaining to results presented in Association of DNAm signatures with global aspects of brain and cognitive health) are presented in Supplementary Fig. 7. Six main global neuroimaging metrics were derived: global grey matter, global white matter, whole brain volume as well as global measures of cortical thickness, volume and surface area. Global grey matter was calculated by summing the volume of cor-
tical, subcortical and cerebellar grey matter within both hemispheres of the brain. Global white matter was calculated in the same fashion, excluding anything that was not white matter, and included white matter hyperintensities. Total brain volume summed both grey and white matter volume but excluded the brainstem, ventricles, cerebrospinal fluid and choroid plexus. Global cortical measures were derived by summing metrics from both left and right hemispheres. In addition to this, global measures of white matter integrity (gFA and gMD), a measure of white matter hyperintensity burden and estimated brain age were considered as part of the 'global' neuroimaging phenotypes (distinct from regional brain metrics). The final analysis sample for the STRADL sample with DNAm data was 709 (for most models, this was N = 702, i.e. including complete ICV data) included individuals with either T1 (N=709) or DTI (N=686) data, with N=680 having both. Mean values for all neuroimaging metrics used in analyses are reported in Supplementary Data 1.

White matter hyperintensity burden – Fazekas scores

FLAIR images were examined for the presence of WMH, which were defined by punctuate, focal or diffuse lesions in the deep or periventricular white matter, basal ganglia or brainstem, visible as areas of hyperintensity in respect to normal appearing white or grey matter. The amount of WMH was graded according to the Fazekas scale (Fazekas et al., 1987), distinguishing between periventricular and deep WMHs and grading them from 0-3 separately; these grades were summed to provide a total Fazekas grade (range 0-6), which is treated as continuous measure of global WMH burden within our statistical analyses. This approach has been adopted previously (Shi et al., 2021) to capture WMH burden within this cohort, and where full details of individual gradings can be found in the supplementary information.

White matter tract microstructure

White matter tract integrity was measured by looking at average FA and MD for 24 white matter tracts of interest (Supplementary Data 16). Alongside examining white matter microstructure for individual tracts, global white matter integrity was determined by applying principal component analysis (PCA) on the 24 tracts to extract a latent measure. Scores of the first unrotated principal component of FA/MD were extracted and set as the dependent variable (proportion of variance explained by gFA and gMD was 45.2% and 61.9% respectively). Distributions of gFA and gMD are presented in Supplementary Fig.8. Full details are supplied in Supplementary Results.
**Brain age measurements**

Brain age was derived using the software package brainageR (Version 2.1; available at [https://github.com/james-cole/brainageR](https://github.com/james-cole/brainageR)). This estimates brain age by applying machine learning to voxel-wise grey, white matter and CSF volumetric data derived from structural T1-weighted MRI (Cole et al., 2018). This estimate was then regressed on chronological age and the residuals from this model were used to index *relative brain age*. See Supplementary Fig. 8 for relative distributions and sample sizes for all global brain ageing metrics.

**Cognitive data**

Of the 778 individuals who had DNAm data available, 771-775 participants had both cognitive and DNAm data. Included in analyses are those that also had neuroimaging data collected. We examined seven measures of cognitive function: general cognitive ability (g), general fluid cognitive ability (g_f) and individual domains of matrix reasoning, processing speed, executive function, vocabulary, and verbal declarative memory as per previous studies in this cohort (Habota et al., 2019; Whalley et al., 2016). Full details are provided in *Supplementary Methods*.

**Statistical analyses**

All statistical analyses were performed in R (version 4.0.5) (R Core Team, 2020).

**Correlations between DNAm signatures**

Pairwise correlations between DNAm signatures (*Supplementary Data 9*) and DNAm with white blood cell proportions (*Supplementary Fig. 6*) were determined by calculating Pearson’s correlation coefficients (r) using the `cor()` function in R. Correlation (r) and p-values (p) are provided for the association between projected DNAm signatures and STRADL in relation to protein levels for both protein epitopes common to SomaScan and Olink in (*Fig. 1a*).

**Pathway analysis**

To annotate CpG sites comprising the DNAm signatures, we used the Infinium MethylationEPIC BeadChip database, which provides information concerning genes, chromosome location, start and end site, and other characteristics ([https://emea.support.illumina.com/array/array_kits/infinium-methylationepic-beadchip-kit/downloads.html](https://emea.support.illumina.com/array/array_kits/infinium-methylationepic-beadchip-kit/downloads.html)). We then used the Functional Mapping and Annotation of Genome-wide association studies (FUMA) (Watanabe et al., 2017) to identify biological pathways relevant to our proxied signatures of interest, a database that examines whether genes of interest are overrepresented in any pre-defined gene sets across a number of databases. Further gene-set enrichment was carried out to conduct both Gene
Ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) analyses to define biological pathways linked to proteins of interest using the ShinyGO tool (Ge et al., 2020). Functional annotations of the 109 trained signatures have been detailed previously (Gadd et al., 2022b) and were sourced from the STRING database (Jensen et al., 2009).

**Regression models**

DNAm signatures or measured protein levels were used as predictors in models and various brain phenotypes were set as outcomes. All betas were standardized. False Discovery Rate (FDR) multiple comparison correction was applied (using the `p.adjust` function in R) and all associations termed significant in this report are defined by pFDR < 0.05. Across all supplementary data, we additionally report basic (nominal p < 0.05) as well as adjusted (pFDR < 0.05) results. Three main models were used as illustrated in Fig.1c: a baseline model controlling for age and sex, a sensitivity model that adjusted for age and sex plus immune cell proportions (monocyte, granulocyte, B-cell, CD4T, CD8T, and natural killer cells), alongside a fully-adjusted model that further controlled for various lifestyle factors that are considered to influence inflammation (hypertension, smoking, alcohol consumption, and BMI). In cases where MRI metrics were used as outcome measures participant age, sex, batch, number of image edits per individual, imaging site, and estimated intracranial volume (ICV) were set as covariates in mixed-effect linear models for both proteomic and DNAm signature associations. Where cognitive metrics or APOE were set as outcome measures, age and sex were the baseline covariates. See Supplementary Fig.2-3. $R^2$ estimates for all models are reported. Throughout the results section, incremental $R^2$ is reported, with the change in $R^2$, presented as % difference, used to quantify the individual and combined explanatory power of immune and fully-adjusted models in terms of improvement in model fit relative to the null model. The results of all 12,208 associations (109 DNAm signatures x 112 brain health outcomes) and paired-proteome associations can be sourced from: [https://github.com/EleanorSC/Inflammatory-DNAm_STRADL/tree/main/Supplementary_Results](https://github.com/EleanorSC/Inflammatory-DNAm_STRADL/tree/main/Supplementary_Results), alongside code to run analyses.

**Mediation models**

We used the R package `lavaan` to investigate how the relationship between an exposure variable and an outcome variable are driven by a third intermediate variable, the mediator. Here, we wanted to test if aspects of brain structure (as mediator, M) mediate the relationship between DNAm signatures (as exposure, X) and cognitive ability (as outcome, Y).

A mediator needs to meet the following three criteria: (1) A change in levels of the exposure variable significantly affects the changes in the outcome (i.e., total effect of X on Y is significant). (2) There is a significant relationship between the mediator and the outcome (i.e., Path from...
M to Y). (3) A change in levels of the exposure variable significantly affects the changes in the mediator (i.e., Path from X to M). Mediation analyses were therefore run on significant associations between DNAm and cognitive outcomes, with brain MRI metrics set as mediator variables. An example path diagram is provided in Supplementary Fig.10.

Informed by these analyses, a second set of mediation modeling was run to determine more localized brain structural changes that may drive specific aspects of cognitive functioning. For these, analyses where regional cortical volumes were set as mediators were performed to examine whether alterations in specific cortical regions were driving cognitive changes.

**Code and data availability**

Access to and use of GS and STRADL data must be approved by the GS Access Committee under the terms of consent. Full details of the application process can be found at [www.generationscotland.org](http://www.generationscotland.org).

The data collected in the STRADL study have been incorporated into the larger Generation Scotland dataset. Nonidentifiable information from the Generation Scotland cohort is available to researchers in the United Kingdom and to international collaborators through application to the Generation Scotland Access Committee (access@generationscotland.org) and through the Edinburgh Data Vault ([https://doi.org/10.7488/8f68f1ae-0329-4b73-b189-c7288ea844d7](https://doi.org/10.7488/8f68f1ae-0329-4b73-b189-c7288ea844d7)).

Generation Scotland operates a managed data access process including an online application form, and proposals are reviewed by the Generation Scotland Access Committee. All code used for the analyses used in this study is available at [https://github.com/EleanorSC/Inflammatory-DNAm_STRADL](https://github.com/EleanorSC/Inflammatory-DNAm_STRADL), with Supplementary Tables accessed at [https://github.com/EleanorSC/Inflammatory-DNAm_STRADL/blob/main/Supplementary_Data](https://github.com/EleanorSC/Inflammatory-DNAm_STRADL/blob/main/Supplementary_Data).

**Acknowledgments**

We are grateful to the GS:SFHS participants who have taken part in the STRADL study. We also thank the families of the participants, and the wider GS:SFHS and STRADL project teams.
References


8.3 Conclusion

This study conducts an integrated examination of DNAm signatures generated from the circulating proteome in relation to brain health outcomes. The rationale for this study came from the success of investigating DNAm-based predictors of exposures applied to adjacent health outcomes, such as DNAm proxies of smoking predicting survival in cancer (Zhao et al., 2021) and risk of cardiovascular disease (Chamberlain et al., 2022) and DNAm signatures of circulating proteins in relation to age-related disease outcomes (Lu et al., 2019, p. 2; Hillary et al., 2020; Gadd et al., 2022). In Chapters 6 and 7, and in work by (Green, Shen, et al., 2021; Green, Squillace, et al., 2021) we demonstrate that this same approach can be extended to examine the impact of chronic inflammation on brain and cognitive health outcomes. In this study, we report 470 associations between DNAm signatures of the circulating proteome (where each DNAm signature is a composite metric of ~99 CpG sites) and brain and cognitive ageing phenotypes.

Of the 73 unique DNAm signatures that associated with poor brain and cognitive ageing metrics, DNAm MMP12, CRP and PIGR had the highest number of significant ($p_{FDR} < 0.05$) associations, each associating with eight separate brain health phenotypes (such as global grey and white matter volume, gFA, gMD and brain age) not including regional associations with cortical and subcortical volumes and individual white matter tract microstructure. These, alongside the other DNAm signatures that associated across multiple makers of brain and cognitive ageing (CRP, MMP12, PIGR, IGFBP4, FGF.21, SKR3, PRSS2, SERPIND1, THBS2, VEGFA, ICAM5, MMP.1, RARRES2, TGF.alpha, ACY1, AFM, CCL17, CCL18, HGF, MMP1, MMP9, SIGLEC1) govern the expression of proteins involved in the regulation of the innate and adaptative immune response, inflammatory signalling pathways (complement, JAK-STAT, PI3K-Akt, MAPK) as well as wider epiphenomena of inflammation, such as angiogenesis, extracellular matrix reorganisation, and metabolic pathways. DNAm signatures
outperformed circulating protein levels in explaining variance in brain health outcomes, indicating that DNAm may offer an augmented means to proximal associations of inflammaging. By contrast, the 10 top protective DNAm signatures for brain and cognitive health outcomes (SEMA3E, NTRK3, NCAM1, NOTCH1, CNTN4, OMD, GDF.8, WFIKKN2, SLITRK5, INSR) were those proxying proteins involved in neurological pathways, anti-inflammatory adipokine signalling, and synaptogenesis.

When assessing the degree to which alterations in brain structure mediated the associated between DNAm signatures and aspects of cognitive function, global grey matter emerged as the most significant mediator for associations with \( g \). Regional cortical volumes in the superior temporal and fusiform cortex emerged as the unifying significant mediators of the association between inflammation and processing speed (attenuation 6.9-13%), where inflammatory-related DNAm signatures that were associated with poorer test scores (DNAm MMP12, PIGR, SKR3, RARRES2, THBS2) appeared to mediate this effect via atrophy in these regions. These findings are in line with previous work which has established a link between temporal lobe atrophy and declines in processing speed both in non-clinical populations and studies investigating cognitive impairment in dementia (Zhao et al., 2018; Pelkmans et al., 2021).

In a previous study, several of these DNAm signatures were linked to three or more incident disease outcomes, and 78 DNAm signatures were found to associate with various disease phenotypes associated with inflammaging such as diabetes, rheumatoid arthritis and COPD. This and previous work linking chronic inflammation, the epigenome and chronic inflammatory diseases (Zaghlool et al., 2020; Wielscher et al., 2022) demonstrates that DNAm signatures could have value in anticipating diseases of inflammatory origin. Alongside direct links to immune and inflammatory responses, there were several DNAm signatures associated with poor brain health outcomes which corresponded to genes involved in cell-adhesion, extracellular matrix reorganisation and vascular endothelial functioning. These include the matrix
metalloproteinases, MMP12, MMP2, MMP1 and MMP9 which have been implicated in blood brain barrier disruption. This paints the picture of a complex network of inflammatory, and inflammatory-adjacent activity that may ultimately leads to glial, neuronal, axonal, and dendritic damage, driving our observations of higher WMH burden, lower white matter integrity and general atrophy of the brain at global and regional levels.

This study has some unique advantages, such as combining multimodal neuroimaging measures, cognitive assessments, and epigenetic data in a well-characterised cohort. Limitations of this work include the fact that peripheral blood was used to obtain DNAm, which may not be the best way to achieve mechanistic insights into cognitive dysfunction due to the limited overlap with brain profiles. Equally, while we included a variety of covariates in the regression analyses, we cannot exclude the possibility that the observed associations between the identified methylation markers and brain health outcomes might be explained to some extent by confounding factors not accounted for here.

In summary we establish that DNAm offers additive value over proteomic measurements in associating with relative brain age, global neuroimaging metrics, and aspects of cognitive functioning. We additionally find that multiple aspects of brain structure mediate the association between inflammatory DNAm signatures and cognitive functioning, with regional atrophy volume in the superior temporal cortex accounting for poorer processing speed ability associated with inflamming. Associations between multiple brain health outcomes and inflammatory-related DNAm signatures suggests that DNAm changes in blood could be a useful peripheral biomarker for indexing cumulative inflammatory exposure and potentially anticipating risk of age-related cognitive dysfunction and associated structural brain changes.
In this thesis, I explored the integration of neuroimaging, epigenomics and proteomics to understand the influence of chronic inflammation on brain ageing. The three empirical chapters of this thesis each included a discussion section which provided specific commentary on the analyses and findings of the respective studies. Therefore, the aim of this chapter is to provide a broader discussion of the findings and how they interconnect in the investigation of chronic inflammation in relation to brain ageing across the lifecourse. This chapter begins by giving an overview of the work before discussing the headline findings and points of discussion. I conclude by considering the limitations of the work presented before outlining future research directions.

9.1 Overview

This thesis aimed to characterise inflammatory-related DNAm signatures and their relation to brain and cognitive outcomes across the life course by leveraging data from multiple UK-based cohorts. The introductory chapters of this thesis reviewed the evidence for chronic inflammation’s role in the initiation and progression of various age-related diseases, including those concerning the brain. The first chapter outlines how, despite the complex molecular architecture of chronic inflammation, most studies conflate acute vs chronic inflammation and rely on singular samples of highly variable inflammatory proteins in the blood as their primary means of assessment. I raise the point that this, in part, could account for why studies investigating the association between chronic inflammation and its effects on the brain have yielded such conflicting results. Through Chapters 1-2 I highlight these inconsistencies – particularly those concerning the exact regional impact of chronic inflammation on brain structure and the related degree, type, and extent of cognitive impairment. I also point out that how these associations
present across the life-course is not well characterised, particularly in early-life and midlife, with the majority of studies focusing on elderly populations, or in patient-populations with specific neurodegenerative or neuropsychiatric diagnoses (see Chapter 2.4.3).

Given the complex nature of inflammatory signalling cascades, and the pitfalls of using singular serum measures for indexing a ‘baseline’ level of inflammatory exposure, Chapter 3 introduces the concept of leveraging DNAm to build surrogate markers of cumulative exposures. It has been suggested that the stability of DNAm patterns in cells provides a molecular record of prior experience and can potentially serve as a long-term biomarker of cumulative exposure (Ladd-Acosta, 2015; Ladd-Acosta and Fallin, 2019), and be adapted to profile chronic inflammation (Stevenson et al., 2020; Wielscher et al., 2022; Verschoor et al., 2023). This can be particularly useful for studying the link between chronic inflammation and age-related diseases such as cognitive decline (where it is thought that exposure to inflammation over many years contributes to disease onset and progression) as well as examining inflammation in very early life (where exposure during gestation is difficult to account for). Equally, in aggregating information across multiple loci, immune-related DNAm signatures can potentially offer an augmented signal capturing downstream effects of environment on gene expression and cellular function across many different sites in the genome. As such, DNAm signatures may provide a more complete representation of the pathways linking chronic inflammation to disease phenotypes beyond the original protein levels they were trained on.

With this background in mind, the overarching research objective of Chapters 6, 7 and 8 was to generate DNAm signatures of inflammatory burden and examine their associations with brain structure and function across the lifecourse – a process outlined in Figure 9.1, with Chapter-specific aims and hypothesis detailed in Chapter 4.
Figure 9.1 Overview of thesis aims

In Chapter 6, a DNAm signature of CRP was analysed alongside blood CRP levels in a cohort of older age adults (72.5 ± 0.7 years of age). In this study DNAm CRP, but not serum CRP, displayed significant associations with total grey matter volume, white matter volume and increased burden of white matter hyperintensities. Higher inflammation as indexed by DNAm CRP also associated with lower cortical volume and thickness in the superior temporal
gyrus, supramarginal gyrus and cuneus regions. Additionally, the relationship between DNAm CRP and a general factor of fluid intelligence (\(g_f\)) and processing speed were both mediated via lower white matter volume.

**Chapter 7** then adapted this approach to a neonatal cohort – this time looking at DNAm taken from saliva samples – and found similar differences in global brain structural metrics, and additionally displayed surprising similarities between which regional white matter tracts (such as the corpus callosum, anterior thalamic radiata, uncinate fasciculi) displayed the largest associations with DNAm CRP. This work also demonstrates that a DNAm signature of CRP associates with gestational age, with most brain structural changes found in the preterm subgroup, and suggests that developmental changes following preterm birth and sustained inflammation result in both global changes as well as complicated patterns of cortical and subcortical alterations and particular disruption of white matter tract development. These findings are in alignment with the wider literature on inflammation and neonatal neuroimaging outcomes covered in **Chapter 2.7.3**. This study also indicated that DNAm CRP reflected an allostatic load of inflammatory exposure, in being both higher in preterm vs term infants, and displaying a dose-response relationship with inflammatory-related morbidities.

Having looked at these two opposite ends of the lifespan, the first year of life and 7th decade of life, the validity of inflammatory DNAm signatures was then examined across a wider age range in adulthood in the STRADL cohort. Following the replication of white matter alterations and global brain structural differences with raised DNAm CRP levels (Green et al., 2021), in **Chapter 8**, this approach was extended to profile multiple inflammatory-related biomarkers. In this study I found 470 associations between 73 unique DNAm signatures and brain and cognitive ageing phenotypes, and outlined which immune-related signatures associate across multiple metrics and are particularly pertinent to brain health. I also demonstrated that (i) DNAm offers advantages over protein levels in explaining variance in relative brain age, global neuroimaging metrics, and aspects of cognitive functioning (ii)
characterise regional patterning of chronic inflammation as indexed by DNAm signatures on brain volumes and white matter integrity (iii) highlight that these associations are broadly independent of immune cell proportions and health and lifestyle factors (iv) examine the interrelationships between DNAm, brain and cognition.

The work presented here demonstrates that various inflammatory-related DNAm signatures display promising reproducibility, specificity to exposure, cross-tissue applicability, and relevance to brain health outcomes in different population samples and age ranges. The findings suggest that DNAm signatures can capture the cumulative effects of chronic inflammation on the brain and serve as a useful tool for investigating inflammation and brain health. The novelty of this work is that there were no inflammatory-related DNAm signatures (based on circulating blood proteins) that have been examined in relation to neuroimaging phenotypes prior to this PhD. DNAm-based biomarkers of inflammation are in nascent stages of investigation (see Figure 9.2) and more widely there are vanishingly rare studies that examine neuroimaging, cognitive, epigenetic and lifestyle data concurrently to explore these relationships. Application of DNAm signatures of inflammation in neonatal cohorts had also never been attempted in any context, despite the potential of this metric to index shared maternal-fetal inflammatory exposures.

Future studies may benefit from integrating multi-omics approaches to develop reliable biomarkers of inflammation and track inflammation levels to mitigate adverse brain health outcomes at different stages of the lifespan. Ultimately, the work presented in this thesis illustrates that DNAm may provide a new molecular dimension to stratify those at risk of inflammation-associated brain ageing, and improve the prediction of related adverse brain health outcomes in both neurodevelopment and later-life cognitive decline.
9.2 The case for DNAm signatures as indices of chronic inflammation

The overarching aims of this thesis were:

(1) to examine how DNAm signatures of inflammation compare with circulating inflammatory protein levels in relation to variance in brain structure and cognitive function across the lifecourse

(2) to examine the interrelationships between inflammation, brain structure and cognitive ability

(3) to investigate the impact of various aspects of health and lifestyle on these relationships

With respect to aim 1, Chapters 6 and 8 compared the associations of both proteomic and DNAm signatures of inflammation with various cognitive and neuroimaging outcomes. The rationale for this work was due to the high variability and phasic concentrations of inflammatory biomarkers in the blood, rendering their ability to reliably index baseline inflammation levels questionable. CRP, in particular, is a highly phasic protein that fluctuates on a day-to-day basis (Bower et al., 2012; DeGoma et al., 2012; Wu et al., 2012; Bogaty et al., 2013; Nash et al., 2013). As outlined in the first chapter (specifically, in Figure 1.4), CRP also circulates in the blood in two distinct isoforms that have different biological properties. These isoforms are not measured by standard assays\(^\text{37}\) and do not always display high correlation with one another.

As discussed in Chapter 1 of this thesis, composite DNAm measures are increasingly used as biomarkers of environmental exposures, either for those

\(^{37}\) The latest high sensitivity assays for CRP measure the level of pentameric CRP (pCRP) – very few assays have measured the monomeric form of CRP (which is considered to be more reactive)
that are complex and difficult to measure, or those that cannot be reliably obtained from clinical or self-reported data. Previous work has demonstrated the proof of concept that DNAm can act as a molecular readout of cumulative exposure in the case of smoking (Joehanes et al., 2016; Langdon et al., 2021), heavy metal exposures (Cardenas et al., 2017; Colicino et al., 2021) and alcohol consumption (Yousefi et al., 2019). Many of these same signatures have since been used to examine associated long-term health outcomes (Corley et al., 2019; Chamberlain et al., 2022; Gadd et al., 2022; Cheng et al., 2023), demonstrating the utility of epigenetic signatures of exposures to assess clinically relevant phenotypes. Building a biomarker of chronic inflammation was based on this same premise: that in taking a collection of CpG sites that are associated with inflammatory protein levels and aggregating them into a single index of inflammation, DNAm signatures might potentially capture the complex interplay between multiple inflammatory pathways, offering a more comprehensive view of baseline inflammation less prone to misclassification bias caused with fluctuating blood biomarkers.

Despite this, there are few studies that use DNAm to proxy inflammation (and assemble a composite DNAm biomarker) published to date. Similarly, the field of neuroimaging-epigenetics, where neuroimaging metrics are set as primary outcomes of interest, is much smaller than that of other complex traits (Lancaster, Morris and Connelly, 2018; Wheater et al., 2020; Walton et al., 2023). The relative proportion of published studies for both inflammation and neuroimaging is illustrated in Figure 9.2.
Figure 9.2 DNAm biomarkers of exposure

Select statistical summary of DNAm methylation studies to date (a) graph displaying number of epigenome-wide association studies (EWAS) publications (y-axis) per-year (b) publications that use DNAm to proxy or predict various exposures (c) platforms used across studies (d) publications broken down by peripheral tissue type (e) number of studies with neuroimaging outcomes (f) possible directional links between peripheral DNAm and brain structure. All data sourced from EWAS Atlas database, on 2 April 2023 (https://ngdc.cncb.ac.cn/ewas/downloads).
9.3 Multi-omic signatures reveal inflammation-brain ageing relationships

In **Chapter 8**, multiple immune-related DNAm signatures were found to associate with neuroimaging phenotypes and cognitive outcomes. One of the main takeaway messages from this work was the relevance of profiling multiple inflammatory markers when examining associations of chronic inflammation with health outcomes.

As discussed in the discussion of **Chapter 8** and in **section 1.5.4**, multiple DNAm signatures derived in this work proxy for circulating proteins in the blood that have obvious and direct roles in inflammatory signalling cascades. For example, I outline that several signatures (CRP included) are part of classical inflammatory signalling cascades such as the complement system discussed in **Chapter 1.5.2**. Others are part of various immune signalling cascades that are considered canonical pathways of inflammation such as JAK-STAT and PI3k-AKT cascades. It is important to recognise that many of these are pleiotropic proteins that have multiple roles, appear across multiple pathways, and in turn exert various cellular effects depending on context – the promotion or attenuation of inflammation being the most unifying and conserved function between them.

Equally, as outlined in **Chapter 1.5**, not all of these proteins are direct instigators of inflammation or ‘classical’ inflammatory biomarkers. What we consider to be an inflammatory biomarker is mostly historical: originally, we looked at the coagulation properties of blood (these markers are still used clinically today - ESR, fibrinogen, plasma viscosity) to make inferences about

---

38 a key arm of the innate immune response: its name comes from the concept of it triggering a cascade that enhances – *complements* – the ability of immune cells to deal with damaged cells or pathogens and initiate inflammatory responses.
inflammatory responses. The classical ‘inflammatory mediators’ are the acute phase proteins (like CRP), cytokines (TNF-α, IL1, IL6) and chemokines (CXC and CC motifs, e.g. CXCL9, CXCL10, CCL11 etc.), but there are also many proteins that work adjacent to these signalling cascades. For example, some signatures here proxy for binding proteins that facilitate the triggering of pro-inflammatory intracellular signalling pathways, such as EN-RAGE (S100A9) which activates the NF-κB signalling pathway (Loza et al., 2007), or adhesion molecules – such as SIGLEC1, which is expressed on the surfaces of macrophages, with a key role in recruiting other immune cells as part of the adaptive immune response (O’Neill, van den Berg and Mullen, 2013). Others, such as the proteases (PRSS2, MMP1, MMP2, MMP9, MMP12) are enzymes that remodel the extracellular matrix (ECM), and can weaken blood-brain barrier (BBB) integrity and have distinct roles in immune-vascular endothelial functioning (Liu et al., 2013; Rempe, Hartz and Bauer, 2016). Equally, the microvascular damage caused by upstream inflammatory activity is also captured by our array of tissue markers and growth factors. VEGFA, for example, is a mediator of angiogenesis - a response induced by platelets activated at the surface of inflammation-microvascular damage sites (Zhang et al., 2000), which has been found to be a key player in astrocyte-mediated BBB integrity (Argaw et al., 2012; Lan et al., 2022).

Many of the 73 signatures found to be relevant to brain health in Chapter 8 converge these various immune signalling or immune-endothelial / vascular interactions. The vascular-inflammation link is particularly interesting from a brain-health perspective, given the concordance between regional differences in brain structure between the inflammatory markers we describe here and previous accounts of vascular risk factors on brain morphology, a finding common in both Chapter 6 and 8 which were set in older-age adult cohort groups. This paints the picture of a complex network of pro-inflammatory and inflammatory adjacent activity that may ultimately lead to glial, axonal, and dendritic damage (Anwar, Özkan and Gürsoy-Özdemir, 2013).

Specifically, MMPs cleave tight junction proteins (such as claudins and occludins) that constitute the BBB, which makes it leakier; see Rempe et al. (2016).
driving our observations of higher WMH burden, lower white matter integrity and general atrophy of the brain at global and regional scales. In contrast, many of the DNAm signatures that were identified as conferring degrees of more favourable neuroimaging outcomes (e.g. lower relative brain age, higher grey matter volume) are involved in attenuating or resolving inflammation. Some, such as IL19 (Leigh, Scalia and Autieri, 2020, p. 19) and IL27, are anti-inflammatory cytokines – others are receptors on white blood cells involved in the immune response, such as CD209 (Lugo-Villarino et al., 2018) – and others have roles upstream of immune cells, and act in immune-regulatory capacities. The findings of Chapter 8 demonstrated that, generally, DNAm signatures that coded for proteins involved in the promotion of inflammation were those that showed significant associations with negative hallmarks of brain structure and function, whereas signatures that associated with proteins involved in attenuating inflammation or resolving innate immune cell activity associated with more favourable outcome measures.

There is clear benefit in more granular resolution of the molecular architecture of inflammation as shown here, and increasing studies call for this more comprehensive profiling of inflammatory responses when examining incident health outcomes (Alpert et al., 2019; Furman et al., 2019; Sayed et al., 2021; Al-Nesf et al., 2022; Walker et al., 2022). This is clearly relevant for the identification of factors involved in the initiation, progression of chronic inflammation and how it relates to differences in brain and cognitive ageing.

40 or, as in the case of those derived from the Olink Neurology panel, were otherwise involved in pathways that may confer brain resilience to neuroinflammatory insult, such as NCAM1(a component of perineuronal nets, which are extracellular matrix structures that support neurons) which was found here to associate with multiple measures of favourable brain health, a finding that aligns with previous studies that examine this signature’s association with higher brain volume and cognitive ability (Harris et al., 2020).
9.4 Links between chronic inflammation, brain structure and cognition

Chronic inflammation has long been associated with cognitive decline and is thought to contribute to the pathological processes underlying differences in the rates of brain ageing. The empirical work of this thesis adds to this literature, with three different population studies demonstrating an association with DNA methylation (DNAm) signatures of inflammation and brain structural measures, and in the cases of Chapters 6 and 8, that some brain structural alterations appear to mediate the association between inflammation and cognition. These findings address the second overarching aim of this thesis: examining interrelationships between inflammation, brain structure and cognitive ability.

In both LBC1936 and STRADL population groups, DNA methylation (DNAm) CRP was found to associate with poorer measures of global cognitive functioning ($g$, $g_I$); of individual cognitive abilities, processing speed showed the strongest association with inflammatory-related DNA methylation (DNAm) signatures in both study samples. Processing speed is a multifaceted cognitive measure that requires integration of several cognitive domains and distributed cortical networks – it has been identified as one of the first domains to decline in normative cognitive ageing (Salthouse, 2019, p. 20) and also in preclinical AD (Kaskikallio et al., 2020). This link between diffuse, global brain structural alteration mediating differences in processing speed ability has previously been associated with low-level chronic inflammation (Eckert, 2011; Heringa et al., 2014), and more generally elevated inflammatory markers have been linked to both declines in global cognitive function and processing speed (Lin et al., 2018). If the neural underpinnings of processing speed involves the coordination of different cortical regions, this might make processing speed particularly sensitive to the early stages of cognitive decline or prodromal cerebral pathology. However, as addressed in the discussion of Chapter 6.2,
our ability to make strong claims around the association of chronic inflammation with individual cognitive domains is moderated by the fact that variance across cognitive domains is shared at the general level (Tucker-Drob et al., 2014), so these relationships are likely reflecting a largely shared general process.

In LBC1936, DNAm CRP and processing speed in multiple SEM models was largely mediated via normal appearing white matter volume (see Appendix 11.1, Fig e4), supporting the concept that poorer processing speed may be partly due to systemic inflammation’s impact on global brain white matter. In addition to this, in the GS neuroimaging subset STRADL sample, gFA mediated the association between various inflammatory-DNAm signatures and processing speed. There were also regional reductions in the temporal regions of the brain which significantly mediated the association of DNAm CRP and several other inflammatory-related DNAm signatures (MMP12, PIGR, SKR3, FGF.21, RARRES2, THBS2) with poorer processing speed ability. Some of the regions that mediated the inflammation \( \rightarrow \) processing speed association identified here (lower cortical volume in the medial temporal, superior temporal and fusiform regions) are in keeping with recent work in UKBiobank, where causal associations have been found between these regions and higher inflammatory protein levels (Williams et al., 2022). Other studies on regional brain changes with inflammation also corroborate these findings (Bettcher et al., 2012; Marsland et al., 2015; Corlier et al., 2018). More generally, finding mediation via these structures makes sense in the larger context of cognition: the regions have varied roles in memory consolidation, language, semantic processing and visual processing. Multiple studies have noted changes in these regions associating with poorer processing speed (Zhao et al., 2018, 2019; Pelkmans et al., 2021). Several DNAm signatures were also mediated by volume of the precuneus, which has been suggested as a site of early atrophy in AD (Migliaccio et al., 2015).

---

41 Each of the associations between higher scores of DNAm PIGR, MMP12, CRP, SKR3, FGF.21 and AFM and poorer processing speed was significantly and partially mediated by lower gFA; Appendix 11.3, Supplementary Figure 9.
Given that both LBC1936 and STRADL samples consisted of individuals without neurodegenerative diagnoses, it remains unclear as to whether these associations are reflective of general trends with inflammation and age or are a consequence of subclinical pathology. It is worth noting that AD patients and cognitively healthy individuals of similar age exhibit considerable overlap in the spatial patterns of brain atrophy, although the rate of regional atrophy may be more pronounced in patients with AD (Fjell et al., 2014).

These findings suggest that while systemic inflammation may exert broad impacts on cognitive functioning via diffuse white matter damage and global reductions in grey matter volume\textsuperscript{42}, certain brain regions may be more susceptible to inflammation-related damage which confer specific cognitive deficits. Moreover, complex molecular interactions, receptor expression and inflammatory signalling pathways may contribute to this regional vulnerability, which highlights the need for further investigations into the causal pathways underpinning inflammation, brain structure, and cognitive functioning relationships.

Overall, the findings in Chapters 6 and 8 demonstrate that the relationship between peripheral inflammation and cognitive ability in adulthood may be largely driven by inflammation’s impact on the white matter of the brain, which brings us to the next discussion point.

\textsuperscript{42} Of the relationships between inflammation (DNAm CRP) and measures of general cognitive ability ($g$ or $g_i$), global grey matter volume proved to be a significant mediator across both studies in LBC1936 and STRADL.
9.5 Inflammation-related DNAm and white matter microstructure

One of the main findings of this thesis was that, across all three population cohorts, inflammatory-DNAm signatures displayed significant associations with white matter microstructure at both global and regional levels. These findings are consistent with previous reports that have linked inflammation to white matter damage in both diffuse and focal regions of the brain.

The various dMRI methods we used to examine white matter microstructure may be particularly well placed to identify some of the more subtle associations of chronic inflammation on the brain that reflect neuroinflammatory processes. Measures such as FA and MD have been shown previously to be sensitive to anatomical changes in white matter cytoarchitecture linked to inflammation (Wersching et al., 2010; Bettcher et al., 2015; Jiang et al., 2015; Walker et al., 2017; Dubner et al., 2019; O'Donovan et al., 2021), but determining the reasons for these alterations remains difficult (Jones, Knösche and Turner, 2013). In mind of this, ascribing direct relevance of these findings to the functionality or health of individual white matter tracts or aspects of brain connectivity should be tempered by the consideration of other factors (of less biological relevance) that can influence these measures. Are inflammation-associated alterations in FA or MD reflective of changes to white matter integrity – e.g. demyelination – or instead reflective of changes in the organisation of fibres, partial volume effects and membrane permeability? A significant strength of the empirical chapters used in this thesis are the array of multi-modal neuroimaging measures used. We integrate DTI with complementary imaging modalities to obtain a more comprehensive understanding of inflammation’s relationship to white matter. For example, in addition to FA and MD (examined across all three studies) we examine normal-appearing white matter and white matter hyperintensities in Chapter 6, neurite orientation dispersion and density imaging metrics in Chapter 7, and global white matter...
in Chapter 8. However, each technique still relies on assumptions and simplifications that may not fully capture the complexity of white matter microstructure, a topic further discussed in the limitation section (9.7.2).

With this caveat in mind, all three studies found that global FA and MD were altered with inflammatory-related DNAm signatures, illustrating a widespread effect of chronic inflammation and white matter microstructure at different stages of brain ageing. At the regional level in Chapter 6 we found DNAm CRP-related alterations in FA and MD in the genu, but not splenium, of the corpus callosum, as well as bilateral associations in the arcuate fasciculi, anterior thalamic radiata and uncinate fasciculi; we found similar findings within a preterm cohort in Chapter 7. The corpus callosum has been identified as a particular hotspot for inflammatory activation, where dysregulation in oligodendrocytes and enhanced activity of glia is thought to drive microstructural alterations (Allen et al., 2023). In Chapter 8, various inflammatory-related DNAm signatures were associated with significant alterations across several association fibre tracts, projection fibre tracts, and thalamic radiata. Though it is difficult to directly compare the regional white matter microstructure findings across these three studies – as different segmentation and tractography methods were used to segment the fibre tracts, and tracts were mapped according to different white matter atlases (a limitation further discussed in section 9.7.2.) – the degree of cross-study replication is notable in spite of these methodological discrepancies.

The findings of Chapter 7, set in a neonatal cohort, are of particular interest when reflecting on these associations of inflammatory-DNAm and white matter. Evidence that early development displays epigenetic plasticity and that changes to DNAm during this time have lifecourse implications (Krol et al., 2019) is key point of reflection for when brain-ageing trajectories are set.

---

43 e.g., in Chapter 6, 24 tracts ROIs were segmented using TractoR (https://www.tractor-mri.org.uk); in Chapter 7, subjects were registered to the Edinburgh Neonatal Atlas (ENA50) and tracts were delineated by drawing ROIs manually on the FA image; in Chapter 8 tracts were categorised according to the Johns-Hopkins University DTI-based white matter atlas, resulting in 24 tract ROIs (5 unilateral tracts and 19 bilateral tracts).
and whether sustained inflammation causes consequences that persist into adulthood. This ties into the developmental origins of health and disease (DOHaD) hypothesis, which posits that adverse exposures in early life are associated with later-life health outcomes (Felix and Cecil, 2019). Given that some of the findings we found here on the association with DNAm CRP and white matter tract microstructure are comparable to later life associations (for example, altered FA and MD was observed in the genu, but not splenium, of the corpus callosum, similar to what we reported in LBC1936 in Chapter 6), questioning what age chronic inflammation manifests a meaningful and detrimental impact on white matter is important. Future studies that examine the degree to which these observed alterations persist or resolve with age are therefore needed. Previous studies have drawn a parallel between alterations in white matter in preterm infants and later life cognitive function, arguing that changes to white matter in early life may persist well into adolescence and adulthood (Nosarti et al., 2008, 2014). As we present only cross-sectional findings here, we cannot draw conclusions on the wider links to later-life cognitive ability or brain structure, though TEBC is well placed to map such trends when data collection is complete.

Placing these findings in the wider context of spatiotemporal white matter tract development is of interest given that this population sample contained two groups of distinct age ranges (term vs preterm infants). As discussed in the introductory Chapter 4, a recent effort to profile global changes in brain structure across the lifespan examined over 100,000 participants (Bethlehem et al., 2022). One of their key findings was in the differential velocities of grey and white matter volume in early development: while grey matter increases non-linear fashion in perinatal development, white matter volume displays a lagged development, increasing linearly throughout childhood and early adolescence.
Figure 9.3 Grey and white matter exhibit distinct trajectories starting in the perinatal period

Figure adapted from Bethlehem et al., (2022) study described in Chapter 2. This early developmental period (filled black segment) demarcates the point of intersection between the trajectories of grey and white matter volume (shaded square; 298 post-conception days) until the point of maximum absolute difference between grey matter and white matter volume (shaded rectangle; 1395 post-conception days). X-axis denotes age, calculated as log-scaled post-conception days. Abbreviations: grey matter volume (GMV), white matter volume (WMV)

These distinct maturational profiles have important implications when it comes to considering how chronic inflammation may influence aspects of brain structure at different ages and how these relate to neurodevelopment. The first is that white matter development displays a delayed trajectory relative to grey matter; this is relevant as the maturational stage relates to the vulnerability of certain brain structures (their location to perforating arterioles and their respective proportions of immune cells, glia and neurons) to inflammatory processes. In line with this, while we saw significantly poorer global white matter volume and tract microstructure among preterms, no significant relationship between DNA methylation (DNAm) CRP and cortical grey matter was observed – although deep grey matter structures did display associations with DNAm CRP, in line with previous reporting on EoP characteristics linked to inflammation (Boardman et al., 2006).
Chapter 2.7.3 discussed how systemic inflammation can directly disrupt the maturation of pre-myelinating oligodendrocytes into mature ones (Favrais et al., 2011; Volpe et al., 2011). This could explain why we saw distinct trends in poorer white matter microstructure relative to gestational age; while preterm infants showed lower FA and higher MD across multiple tracts with elevated DNAm CRP, term infants displayed no significant white matter tract alterations (Appendix 11.2, Supplementary Figure 10).

The strong correlation between gestational age at birth and DNAm CRP, alongside the finding that DNAm CRP associations with brain structural alterations were reserved for the preterms in this cohort, further validate the proof of concept that DNAm signatures of circulating inflammatory proteins may capture a useful marker of sustained inflammation. Multiple studies have demonstrated that inflammation is associated with gestational age, partly owing to the fact that the earlier the onset of spontaneous preterm labour, the more likely an inflammatory trigger such as infection has occurred (Lamont and Sawant, 2005). Moreover, Infants born extremely preterm (< 28 weeks) are also exposed to the NICU environment – confronted with hospital-specific microbes, physical intervention, drugs, ventilation, and hypoxia or hyperoxia which have been associated (independent from other inflammatory risk factors) as drivers of higher inflammation (Bose et al., 2013). Having not benefited from later stage third-trimester maternal antigens, extremely preterm infants rely heavily on non-specific innate immunity, and often display significant imbalanced ratios of pro- vs anti-inflammatory mediators, as outlined in Figure 1.2 (Blencowe et al., 2013; Bennet et al., 2018).

However a curious aspect to this association is that none of the CpG sites which constitute the DNAm CRP were identified as top hits within a previous EWAS examining methylation associated with gestational birth in the sample (Wheater et al., 2022). This could demonstrate that inflammation may be driving differences in brain dysmaturation above and beyond preterm status itself. In support of this, the CpG sites which constitute the score are involved in vascular and immune function and differentially expressed across tissues (see Appendix 11.2). The gene mapped to cg18181703 (SOCS3) is involved
in pro-inflammatory cytokine pathways (Boyraz et al., 2016), its expression typically correlating with severity of inflammation, with upregulation of SOCS3 correlating with increased levels of the proinflammatory cytokines IL-1β, IL-6, and TNF-α (Rottenberg and Carow, 2014, p. 3). SOCS3 is additionally implicated in demyelination via oligodendrocyte disruption (Emery et al., 2006), has been previously implicated in a murine-model of preterm-related WM injury (Boccazzi et al., 2021) and has been found to be upregulated in the microglia of AD patients (Walker, Whetzel and Lue, 2015).

This evidence suggests that aberrant methylation of cg18181703 may play an important intermediary role linking peripheral inflammatory processes with white matter maturation in the preterm neonate. This locus may also be particularly susceptible and responsive to differences in health and lifestyle as this site has also been linked to metabolic differences across the lifecourse such as BMI (Ali et al., 2016) and incident type 2 diabetes (Chambers et al., 2015), conditions themselves which have been linked to elevated CRP levels (as discussed in Chapter 1.5.2). In previous studies of older age adults, the CpG site cg18181703 was also significantly associated with a general factor of cognitive ability (gf), as well as processing speed (Heringa et al., 2014), further supporting the importance of this locus in driving brain changes. How this maps to neurodevelopment and cognitive changes earlier in life is however unclear, and scope to study this association may be challenging owing to the variance in neurodevelopmental testing scores in preterm infants in early infancy (van Beek et al., 2022), with some evidence for standard neurodevelopmental testing paradigms underestimating impairment and developmental delay following preterm birth (Anderson et al., 2010; Vohr et al., 2012).

In the older cohorts, the broad findings of reduced FA and increased MD with higher inflammatory burden is consistent with previous cross-sectional (Wersching et al., 2010) and longitudinal (Bettcher et al., 2015; Walker, 2019; Walker et al., 2022) studies set in older age adults without dementia. Our results add to these findings in being, alongside the work by Green et al.,
(2021) the only examples of DTI associations with epigenetic indices of inflammatory burden, which may better capture cumulative exposure to inflammation. In these older-age cohort examples, it is important to keep in mind that any association with global white matter microstructure with inflammation may be obscured by ageing-specific effects differentially affecting individual white matter tracts (Cox et al., 2016). Our regional findings in Chapter 6 align with those reported in recent studies, where lower FA was observed in the genu of the corpus callosum, anterior thalamic radiata, and uncinate fasciculi (Prasad et al., 2015; Walker et al., 2017).

In both studies that were set in older age cohorts where measures of white matter hyperintensities were available, in neither instance were individuals with high white matter hyperintensity burden removed as a part of a sensitivity analysis. This has implications for our white matter microstructure findings, owing to the penumbra effect of white matter lesions on otherwise ‘normal’ appearing white matter (Maniega et al., 2015). Briefly, this effect describes how existing WMH may affect the microstructural properties nearby white matter. WMHs have been found to intersect or overlap some white matter tracts more than others – in particular the anterior thalamic radiation, inferior longitudinal fasciculus, forceps, posterior thalamic radiation, inferior fronto-occipital fasciculus and parietal and temporal superior longitudinal fasciculi – and in cases where there is overlap, sites further away from the WMH display less signs of ‘poorer’ white matter integrity (i.e increasing FA/ lower MD the further away from the lesion site) (Muñoz Maniega et al., 2019). Given that WMH burden has been closely linked to inflammation-associated microvascular damage (Wardlaw et al., 2017), it’s possible that the underlying mechanisms responsible for WMH may lead to subtle changes in the structure of the surrounding white matter that are not easily detected by conventional MRI techniques (Fernando et al., 2006; Wardlaw et al., 2009). This means that while this white matter may still be classed as ‘normal-appearing’, the values of FA and MD surrounding sites of focal damage may be altered, potentially leading to overstated effects of inflammation in certain regions of the brain.
This effect in part could contribute to the observed anterior-posterior gradient (aka the ‘frontal ageing’ hypothesis as discussed in Chapter 2.6.1) in diffusion parameter changes that we see in our findings. This caveat has been noted by several researchers in the DTI literature, as when analyses are repeated after accounting for specific white matter lesions, only small regional effects persist (Vernooij et al., 2008). The use of other techniques such as NODDI (as utilised in Chapter 7) could provide enhanced insight into the underlying neurobiological pathways via which inflammation might relate to altered FA or MD – owing to metrics such as NDI and ODI having enhanced resolution to pick up on changes to axonal density / fibre dispersion (Tariq et al., 2016). Interestingly in Chapter 8, 10 inflammatory-related DNAm signatures displayed lower FA in the fornix, an area of regional degeneration that has previously been linked to incipient cognitive decline in normative cognitive ageing (Fletcher et al., 2013) and is considered to potentially precede cognitive changes associated with AD (Nowrangi and Rosenberg, 2015). Given that no significant associations between global white matter volume were found in STRADL, but various global (gFA and gMD) and regional microstructural associations were found, inflammation may mediate subtle microstructural changes that precede overt white matter atrophy in mid-to late adulthood.

Peripheral inflammation via several mechanisms such as blood-brain barrier dysfunction (refer to mechanisms of immune crosstalk, Chapter 1.6.3) may promote neuroinflammation and subsequent structural alterations that underlie these observations. Increased infiltration of inflammatory mediators into the CNS has been shown to drive microglia, astrocytes and other glia into ‘activated’ phenotypes, leading to sustained cycles of elevated inflammation. At the microstructural level, this may result in demyelination, axonal loss, and glial scarring, all of which can result in changes in axonal density and fibre dispersion, ultimately affecting white matter microstructure as measured by DTI. The relationship between chronic inflammation and white matter is also related to age, where opposite sides of the lifespan may
represent windows of specific vulnerability: in the first year of life, when the immune system is still developing, and in old-age, where it enters functional decline. For example, at the cellular level, studies have shown that populations of microglia and astrocytes are diverse and display distinct age-related phenotypes (Grabert et al., 2016). Of relevance to later life, and in line with the ‘inflammaging’ hypothesis, aged glia display significantly increased inflammatory gene expression (Soreq et al., 2017; Boisvert et al., 2018; Rissanen et al., 2018), rendering the aged brain more prone to neuroinflammation and subsequent damage. Early-life presents a different window of vulnerability, where immature glia may be more liable to adopting reactive phenotypes (Volpe, 2019; Komada and Nishimura, 2022) and oligodendrocytes less efficient at myelinating axons (Back et al., 2001; Motavaf and Piao, 2021).

In summary, these studies provide further evidence that inflammation is associated with alterations in white matter microstructure at multiple stages of brain ageing. The diffuse nature of the observed effects suggests that elevated peripheral inflammation may exert a global influence on white matter microstructure, potentially through mechanisms such as endothelial dysregulation and related BBB weakening and subsequent neuroinflammatory propagation. These findings also suggest that subtle dysmaturation in early life and alterations in microstructure in mid to late adulthood may precede overt white matter atrophy, and underscore the overall relevance of chronic inflammation’s impact on brain structure.

9.6 The influence of health and lifestyle on inflammation-brain health relationships

With respect to the third aim of this thesis, across all three studies various aspects of health and lifestyle were controlled for in sensitivity analyses. These analyses took various forms according to the population cohorts used.
For example, in the oldest age cohort examined (LBC1936), an array of clinical health variables and lifestyle data was added into multivariate models given the evidence on how vascular risk factors influence brain and cognitive ageing differences (Cox et al., 2019). At the opposite end of the life-course, postnatal clinical variables pertaining to cases of acute infection (such as sepsis, bronchopulmonary dysplasia, and necrotising enterocolitis) were examined, given their association with both inflammation and EoP phenotypes. Across all three studies, common lifestyle factors such as smoking (maternal smoking in the case of the newborn population) were controlled for.

Across the inflammation-brain structural associations found in LBC1936 in Chapter 6, the addition of most clinical disease factors (cardiovascular disease history, diabetes, etc) did not attenuate associations, and of the lifestyle factors examined, smoking and BMI appeared to have the biggest impact (rendering some regional tract FA and MD associations non-significant; see Appendix 11.1, etables 8-9). This fits into the broader literature of lifestyle risk factors that influence brain and cognitive ageing. Given that many (correlated) risk factors are likely to influence cognitive trajectories but via small individual effects (Corley, Cox and Deary, 2018), modelling these factors simultaneously is encouraged to tease apart relative contributions to brain health phenotypes. As examined in section 1.5.2.4, chronic inflammation is considered a key feature of many determinants linked to poor brain health (interlinking many lifestyle factors) and may act as a central mechanism via which these risk factors exert their effects (Newcombe et al., 2018; Bieri, Schroer and Villeda, 2023).

One possible explanation as to why BMI attenuates some associations is the complex network tying systemic inflammation to adiposity. BMI serves as a proxy measure for adiposity and can reflect the overall level of body fat. When BMI is included as a covariate in these regression models, it accounts for the influence of adiposity on the association between inflammation and grey matter volume. Adipose tissue is a key producer of many pro-
inflammatory mediators (including CRP), and many adipokines produced by fat tissue can cross the BBB via bidirectional crosstalk routes (Gómez-Apo et al., 2021). Therefore, it is possible that the relationship between systemic inflammation and elements of brain structure is at least partially explained by the effects of adipose tissue on systemic inflammation which then drive neuroinflammatory and neurodegenerative consequences. This finding highlights the importance of considering the role of adiposity in studies examining the relationship between inflammation and brain structure or function, and examining BMI → inflammation → brain structure relationships is therefore a point of further enquiry.

The principal finding in Chapter 8 was that pro-inflammatory-related DNAm signatures in mid to late adulthood were associated various aspects of poor brain health. These associations were broadly independent of other vascular risk factors, such as BMI, smoking and hypertension (e.g. like Chapter 6) but also of interest was that several DNAm signatures that associated with alterations in brain structure and function had previously been linked to inflammatory-related diseases in the wider Generation Scotland cohort (Gadd et al., 2022). Full reporting of the CpG sites used to generate various DNAm signatures is reported in supplementary information attached to this study. As this pertains to over 9,000 different CpG sites, I will discuss only the top sites of interest that were present across multiple DNAm signatures. However full EWAS reporting assimilated from the EWAS catalogue is provided in Appendix, Supplementary Data 8.

For example, cg06690548 (SLC7A11) was a site that was included in both the DNAm CRP score used in Chapter 6-7 and across several immune-related DNAm signatures in Chapter 8 (n = 2444) represents a key area in which lifestyle could exert effects on various brain endophenotypes via inflammatory pathways – specifically, the lifestyle trait of alcohol

44 This site was included in the following DNAm proxies: CCL21, MIA, IGFBP1, IGFBP4, LGALS4, C9, ADAMTS13, PIGR, CNTN4, ACY1, SELE, ADIPOQ, ENPP7, NCAM1, SELL, SHBG, PRSS2, NOTCH1, OMD, NEP, NTRK3, CXCL10, CXCL9, FGF.21
consumption. In the largest EWAS (n = 8,161) of alcohol consumption to date, this site was identified as the top hit in a healthy population of adults and then replicated in an independent population sample of patients with alcoholism (n = 615). The study then examined both rat liver mRNA expression and human post-mortem brain mRNA expression, finding high expression of SLC7A11 of in both instances. This underscores the relevance of DNAm at the liver–brain-axis, and how alcohol directly impacts brain structure via a potential inflammatory mechanism. Other studies that examine methylation at cg06690548 with clinical endpoints suggest a regulatory role for this site in the body’s response to environmental stress. For example, a study that examined multi-omics phenotypes with DNAm found cg06690548 to relate to obesity (Sayols-Baixeras et al., 2017) and increased SLC7A11 expression has been linked to hepatic inflammation (Lee et al., 2019).

Equally, the significance of this probe in relation to brain health outcomes has also been remarked on previously, particularly in relation environmental mediated risk of Parkinson’s disease (Vallerga et al., 2020).

A further 25 of the DNAm signatures in this study contained the canonical AHRR-smoking loci (cg05575921). The AHRR gene plays a role in the metabolism of xenobiotics, including the toxicants present in cigarette smoke, and is known to be involved in the regulation of immune and inflammatory responses. Interestingly, this CpG site has also been identified as a surrogate for systemic inflammation, with hypomethylation at this site consistently associated with higher levels of circulating inflammatory biomarkers such as CRP and IL-6, even after adjusting for smoking status. Methylation at this site also appears to be influenced by diet, where greater adherence to the Mediterranean diet appears to attenuate the effects of smoking at this site (Fernández-Carrión et al., 2023).

9.7 Limitations

I have discussed limitations specific to each study in Chapters 5-8. In the following section, I outline the general limitations that apply to the cohorts,
methods and analytical approaches used in this thesis, and how these restrict the conclusions we can draw from this work.

9.7.1 Cohorts and generalisability

The empirical work presented in this thesis was based across three population groups: the LBC1936, TEBC and STRADL cohorts. While there are some significant strengths to these datasets – all of these cohorts are extremely well characterised, with extensive phenotypic data available, including DNAm profiling, neuroimaging data and clinical health linkage (see Chapter 4 for details) – there are some common limitations across all three populations used.

All three cohorts are based in Scotland and primarily consist of white European individuals, limiting the degree to which these findings can be generalised to the wider population. Several studies have highlighted the importance for diverse cohort studies in epigenetic epidemiological research, and given the ethnic variation in predisposition to chronic inflammation discussed in Chapter 1.5.2.4, this underscores the importance of replication in more diverse populations. Related to this is the relative health and homogeneity of the participants which may not be reflective of the wider population. Generally, these population groups were healthier than the average population with LBC1936 and STRADL in particular oversampled for individuals with higher educational attainment and SES\(^{45}\) than the wider population (though of the three, LBC1936 is the most homogenous in terms of SES and education, whereas TEBC and STRADL sample had a wider-range of SES strata). Sample selection and attrition bias also adds to this effect: in the cases of STRADL and LBC1936 cohorts, self-selected participants who remain for later waves of participation are more likely to be healthier, with attrition selectively affecting more cognitively impaired

\(^{45}\) SES was not adjusted for in these studies which is a limitation of this work, though work is currently ongoing to assess the influence of SES on DNAm CRP and DNAm IL6 in TEBC and has previously been examined in STRADL.
participants. Though this effect is more relevant for longitudinal study designs (Taylor, Pattie and Deary, 2018), this limits the degree to which these findings can be directly compared against those set in earlier waves.

Another point is that across all three populations, where data pertaining to lifestyle was concerned (such as current smoking status in Chapters 6 and 8, or maternal smoking in pregnancy in the case of Chapter 7), self-reported measures were used – a measure prone to recall bias, or, in the case of pregnancy, underreporting (Shipton et al., 2009). Similar bias is found in the collection of past medical history used to inform the covariates that formed part of the sensitivity analyses across the three empirical chapters – e.g. self report of hypertension, past CVD history, diabetes and so forth.

9.7.2 Neuroimaging metrics

All three cohorts used in this thesis (excepting a subsection of TEBC participants who were preterm and unwell) used healthy humans who were free of significant brain impairment (in the case of TEBC, participants with CP, for example, were excluded; in LBC1936 and STRADL, participants with evidence of neurodegenerative disease were excluded). However, within healthy individuals there is a significant natural variation in brain morphology, which can affect the sensitivity of detecting morphological differences across groups. Longitudinal analyses of within-participant age-related changes in brain structure are less affected by this inter-individual morphological variation and are therefore ideal for examining inflammation-brain structure relationships. Though we corrected for ICV across all three studies, this does not entirely eradicate the issue, as different subcortical tissue volume fractions may occur between individuals and sexes (Voevodskaya et al., 2014), and the choice of ICV adjustment can itself influence study results (Nordenskjöld et al., 2013).
Some of the neuroimaging metrics used across these chapters are more validated and widely reported than others, enabling for more meaningful comparison with the current evidence base. For example, while global grey matter volume is widely considered a strong neuroimaging substrate of brain ageing, white matter hyperintensity (WMH) volume is less validated, though there is good evidence to suggest this metric’s relevance to differences in ‘healthy’ or normative brain ageing. However, different age ranges in the population samples used in this thesis should be considered when examining associations of inflammation and this metric. For example, in Chapter 8 of this thesis, 9 inflammatory-related DNAm signatures were found to be significantly associated with increased WMH burden – with the strongest seen for DNAm MMP12 and WMH ($\beta = 0.185, p = 2.92 \times 10^{-4}$). However, the GS neuroimaging sample contains younger participants than those typically found in populations of high WMH burden – and as increased WMH burden has been shown to be strongly related to later-life, we may not be capturing the full impact of various inflammatory-related signatures in contributing to WMH burden. The second factor to consider is the overall accuracy and potential of the Fazekas score grading as an estimate of WMH burden in the first place (to see distributions of these among the GS neuroimaging participants, see Supplementary Figure 8 provided in the appendices attached to this chapter). As a contrast, in Chapter 6 of this thesis, WMH burden was also estimated but via a quantitative approach where WMH volume measures were estimated using a method that fuses T2*W and FLAIR images to get a volumetric estimate of WMH (Hernández et al., 2012). This quantitative approach is arguably superior to using Fazekas score gradings as (a) it provides a continuous measure of WMH burden, which captures the variation in WMH volume within and between individuals, and may better reflect the underlying pathological changes than the ordinal scale of the Fazekas score, and (b) it is less subject to grading bias (Ballerini et al., 2020). Had we compared quantitative WMH measures in the GS subset for the older age participants in this group who were matched to the ages of

---

46 This is reflected in how only 3 (0.4%) participants had severe Fazekas score gradings (>6)
participants in the LBC1936 sample, it is expected that we would observe similar findings.

In addition, different tractography-based methods were used to examine white matter tract microstructure across the three chapters of this thesis. In Chapter 6, probabilistic tractography approach was used to derive average FA and MD across each of the WM tract ROIs for each participant, whereas in Chapters 7-8, TBSS was used. As outlined in Chapter 2.5.3 and the methodology Chapter 5 subsections, both of these are approaches are used throughout the neuroimaging literature but use different approaches to construct fibre tracts. As briefly discussed in Chapter 2.5.3 the appropriateness of these methods pertains to the age of population sample. In LBC1936, an examination of older age adults, PNT method may be more appropriate than TBSS (despite a wealth of prior literature examining chronic inflammation and white matter integrity in old-age with TBSS). PNT uses a pool of candidate tracts to select the closest match (that resembles a predefined reference tract in length and shape), providing a native space representation of individual tract anatomy with preserved individual variance (Muñoz Maniega et al., 2018). This method also allows for visual inspection and correction or removal of failed QC, and is considered to be more optimised for older subject's brains which may exhibit certain hallmarks of pathology or structural aberrations (e.g. SVD markers).

Equally, when it comes to segmentation methods used in DTI analyses, the different chapters use distinct approaches including atlas-based and ROI methods. Atlas-based methods use pre-defined white matter atlases to segment white matter tracts, ROI-based methods manually or automatically define ROIs within the white matter. This means that even if similar tracts are reported across the studies, the way that they have been defined is not necessarily equivalent; atlas-based methods such as the Johns-Hopkins University DTI-based white matter atlas (Mori and Van Zijl, 2007) used in Chapter 8 are generally more conservative than manual ROI methods, such as the approach adopted in Chapter 7. This is because atlas-based methods
use a pre-defined set of regions that are based on a population average or a template, whereas manual ROI methods allow for more individual variation and flexibility in defining the ROIs. Atlas-based methods can be limited by the accuracy and resolution of the atlas used, and may miss small, but clinically important, areas of interest. On the other hand, manual ROI methods can be more prone to inter- and intra-rater variability, and may be influenced by researcher bias. Overall, the choice of segmentation method was largely down to the age context of the cohorts and the availability of appropriate atlases.

The regional cortical volume measures chosen also differed between studies. One metric, the ventral diencephalon, was only examined in STRADL (where lower volumes were associated with DNAm SERPIND1, IGFBP4 and FGF.21). This region is a FreeSurfer derived metric that includes the hypothalamus, subthalamic nuclei, substantia nigra, geniculate nuclei and surrounding white matter. Use of this region is contested given that its relatively vaguely-defined boundaries makes it more variable than other FreeSurfer derived regions (Frodl and Amico, 2014). We included this structure because of the relevance of the hypothalamus and sub-thalamic nuclei to the Hypothalamic-Pituitary-Adrenal (HPA) axis. Despite being a more variable measure than some of the other regional metrics used, given that changes in these structures may reflect dysregulation of the HPA axis and inflammation-related damage (DiSabato, Quan and Godbout, 2016), we elected to include it. The neuroimaging metrics used in this thesis are therefore differentially hampered by their resolution power and regional specificity; findings should be considered in the context of differences in the granularity / boundaries of the parcellation procedures.

A further caveat to raise is that in Chapter 8, mediation was tested only for significant three-way associations between inflammation/imaging/cognitive variables at \( p_{FDR} < 0.05 \), adhering to the assumptions outlined in Chapter 8 Methods. However, a hypothesis-free approach may be preferable when exploring the neuroanatomical substrates of different signatures of chronic
inflammation and their contribution to cognitive ability. As demonstrated in the multiple SEM models of Chapter 6, it is likely that various brain imaging variables, including white matter microstructural integrity, partly mediate the relationship between chronic inflammation and cognitive ageing.

Finally, while sample sizes in this thesis are large compared to that of the wider neuroimaging-epigenetics literature (Wheater et al., 2020), larger sample sizes are of particular importance to draw meaningful conclusions from neuroimaging-associations. The exception to this is the work presented in Chapter 7, which was set in a cohort of 258 infants. TEBC’s sample size is based on realistic assessment of recruitment and follow-up and power from computational modelling and previous data (Boardman et al., 2020). Compared to the other cohorts used in this thesis, a sample size of 258 infants seems relatively small. However TEBC is unusual in its depth of phenotypic data and multi-modal neuroimaging data compared to other early-life cohorts, and for a neonatal cohort this is still substantial given that the bulk of neonatal brain imaging research has been conducted in sample sizes of less than 100 participants (Volpe, 2019). Similarly, using cross-sectional methods heavily hampers our ability to draw conclusions. This is particularly true for the analyses presented in Chapter 8 which spans a large age range (28 – 81 years). This has been reinforced by the recent work mapping brain charts both cross-sectionally and longitudinally (Di Biase et al., 2023), which finds that age-related change in brain structure is underestimated by cross-sectional data. Equally as we did not adjust for age 11Q in these models, no conclusions pertaining to inflammatory-DNAm’s relationship to cognitive decline can be drawn. Longitudinal studies that aim to collect repeated measures of both neuroimaging and cognitive data alongside DNAm are needed to delineate these relationships.

9.7.3 Proteomic assays
In **Chapter 6**, a high-sensitivity assay was used to measure serum CRP levels\(^\text{47}\). In **Chapter 8**, three different proteomics panels (SOMAscan platform, Olink Neurology panel, Olink Inflammatory panel) were used to measure CRP and other inflammatory-related blood proteins.

There are strengths and limitations to the aptamer-based arrays (SOMAlogic) and antibody-based multiplex technology (Olink). SOMAscan offers broad coverage (greater coverage of circulating proteome than Olink) and high sensitivity, but provides *relative* quantification rather than *absolute* concentrations of proteins. Specifically, SomaScan has issues of aptamers binding to non-target proteins (around 7% of aptamers appear to bind off-target), reducing the certainty that these measures reflect the circulating proteome (Sun *et al.*, 2018). Olink technology allows for multiplexed immunoassays, but similarly the extent to which these capture *in vivo* circulating proteins is unclear. Also, the degree to which these two panels correlate with one another is mixed – in studies that have compared the performance of these panels (Raffield *et al.*, 2020, p. 202) some proteins (fittingly, CRP is one of them, with \(r > 0.7\)) displaying very strong correlations, and but others poorer correlations (\(r < 0.3\)). For example in a recent study that compared 817 proteins common to both platforms, Spearman’s correlations were high (\(r > 0.75\)) for 15% of proteins and poor (\(r < 0.40\)) for 45% (Haslam *et al.*, 2022). These sources of variation include technical differences, certain biological attributes of target proteins, binding affinity and detection limits (Pietzner *et al.*, 2021).

### 9.7.4 DNAm arrays

\(^{47}\) A low-sensitivity assay was also used in the original investigation, alongside a genetic risk score of CRP, and the relative performance of the ls-CRP, hs-CRP, prs-CRP and DNAm CRP was compared. After rounds of review by *Neurology*, these analyses were eventually removed from the manuscript upon reviewer’s suggestions, as hs-CRP was deemed a more appropriate measure for indexing the type of low-level inflammation of interest and the PRS-CRP score did not display any significant associations with brain health outcome measures.
Across all three chapters, microarrays were used to measure DNAm levels at specific CpG sites. Even the most comprehensive of these, the Illumina HumanMethylationEPIC BeadChip, is still limited in its capture of the genome (providing coverage up to 3%) and may miss key sites of crucial relevance to chronic inflammatory responses. This is particularly relevant when trying to build accurate and robust biomarkers of circulating protein levels, where CpG sites of importance could be undetected. Furthermore, DNAm levels are often correlated across the genome, and focusing on a single snapshot can result in the exclusion of relevant information. DNAm arrays are not as accurate at measuring DNAm in repetitive regions of the genome and distal regulatory regions (Pidsley et al., 2016).

To overcome these limitations, more advanced methods, such as single-cell sequencing may offer higher resolution and the ability to capture cell-specific changes in DNAm patterns, enabling a more precise characterisation of the inflammatory response at the cellular level (Mogilenko, Shchukina and Artyomov, 2022).

9.7.5 DNAm tissue and cell specificity

The choice of surrogate tissue is an important facet to study design in epigenetic epidemiological research. Ideally, tissue-type from which DNAm is sampled should be appropriate for both the study’s predictor and outcome measures.

For each of the three population studies in this thesis, DNAm was sampled from peripheral tissues that contain a mixture of different cell types: blood and saliva. Given that neuroimaging and cognition were the key outcome measures across these three studies, brain-tissue derived DNAm is an obviously ideal candidate. However, there clear drawbacks to using brain-tissue derived DNAm.
The first is that to obtain brain tissue from living subjects, a biopsy is required – an impractical and ethically dubious method of sample collection. An alternative approach is sampling DNA from CSF, though this too involves an invasive procedure (lumbar puncture), and may be particularly difficult to obtain in particular population samples (e.g., early-life neonatal cohorts as examined in Chapter 7). Because of these limitations, studies on brain-tissue derived DNA are generally restricted to post-mortem samples. Post-mortem samples, by nature, represent the end point of a disease process, providing limited potential to examine the extent to which changes are reflective of active disease processes or death itself. More pressingly, the stability of DNA in post-mortem tissue has been questioned, with global levels of DNA varying according to post-mortem sampling interval (Sjöholm et al., 2018).

Blood is an obvious choice for examining epigenetic variation in systemic inflammation. However, because we are examining the influence of chronic inflammation on brain-health relates traits, examining the degree to which variation at various CpG sites in blood is reflective of site-specific changes in the brain is important. Several studies have examined such blood-brain concordance and other interindividual DNA variation across peripheral tissue samples (Hannon et al., 2015; Braun et al., 2019; Perkeybile et al., 2019). There is divided opinion on this topic, with some researchers finding poor correspondence between peripheral tissues and brain (Walton et al., 2016) and others finding consistency in inter-individual variation across tissue types (Sommershof et al., 2009; Davies et al., 2012; Provençal et al., 2012; Slieker et al., 2013; Smith et al., 2015), and high concordance between blood and brain (Gregory et al., 2009; Braun et al., 2019; Perkeybile et al., 2019).

Of particular relevance to this thesis, DNA signatures of inflammation (DNA CRP and DNA IL6) were compared across peripheral blood and five post-mortem brain regions obtained from a small subgroup of participants in LBC1936 cohort, where inflammatory-related DNA displayed subtle variation across brain regions (Stevenson et al., 2022). Specifically,
while DNAm CRP displayed no significant variation across brain regions, DNAm IL6 was higher in the ventral anterior cingulate cortex relative to other regions – this indicates that inflammatory-DNAm signatures originally derived in blood may proxy methylation in some areas of the brain better than others. It should be caveated that the DNAm signatures used throughout this thesis have not been formally validated in brain tissue samples, and the study above by Stevenson et al (2022) was in itself a preliminary investigation that was underpowered (n = 14) owing to the difficulty of obtaining blood, methylation and post-mortem samples from the same individuals. However, the degree of concordance is less of a concern when causality is not the main aim of the study. In this instance, reliably and robustly proxying chronic inflammation is the main goal, and significant differences in DNAm have been robustly associated with a range of inflammatory-related lifestyle exposures and disease phenotypes—from smoking (Sugden et al., 2019), inflammatory-related diseases (Ligthart et al., 2016; McDermott et al., 2016; Ventham et al., 2016; Wielscher et al., 2022) and psychiatric disorders linked to inflammation (Barker et al., 2018; Crawford et al., 2018).

Many studies that investigate perinatal exposures, or neonatal outcomes opt for cord blood as the tissue of choice, which offers the advantage of a shared tissue between parent and offspring, and is also obtained at time of birth. Our approach in Chapter 7 of this thesis differed from this in two respects: we used saliva to obtain DNAm, and sampled at term-equivalent age (the same day as neuroimaging assessment) instead of the day of birth. There were two main reasons for this approach: the first being that saliva was the most accessible tissue source appropriate for this age-group, where unnecessary venepuncture (e.g. blood sampling for research purposes) is to be avoided. The second is that obtaining DNAm at term-equivalent age instead of birth means capturing an epigenetic profile of both maturational stage and also the allostatic load of early post-natal exposure and experience. Given this study was conducted on a cohort of preterm infants, this is particularly important for capturing aspects of an infant experiencing expedited arrival to the world and foreign NICU environment, which consists of unique stressors that may
influence susceptibility to sustained inflammation (e.g. medical treatments, stressful or painful interventions such as intubation, altered nutrition, such as formula and exposure to hospital-specific microbes). It should also be noted that while this study was cross-sectional in nature, saliva samples are also better suited to longitudinal sampling (repeated blood sampling is less ethical, and cord blood can only be obtained at time of birth so isn’t appropriate for longitudinal analyses).

Saliva sampling contains DNA from both leukocytes (of mesodermal origin) and squamous epithelial cells (of ectodermal origin). The latter of these shares the same embryonic origin of brain tissue, which some have argued make it a better surrogate tissue than blood for investigating brain-health related traits (Smith et al., 2015).

9.7.5.1 DNAm and cell-specificity

Generally, in instances where samples used for DNAm analysis involve multiple cell types (such as in the case of whole blood, or saliva), it is considered important to account for cell type cell in the analysis (Houseman et al., 2015). Because cell types have unique DNAm profiles, not accounting for these differences runs the risk of misattributing differences between measures as to do with the particular trait you are interested in, when it may in fact be a byproduct of more general methylation changes associated between cell types. In other words, when you measure DNAm from a bulk tissue type, you are getting a representation of a proportion of DNA sites at a given location that are methylated across all of the cells you are measuring – so you run the risk of getting a measure that mostly reflects the cell-type distributions in the bulk sample, rather than an informative signature of a disease or exposure process you’re interested in. An alternative perspective is that alterations in cell composition may represent biologically relevant states reflective of dysregulated disease processes (i.e. chronic inflammation has been shown to influence the relative proportion adaptive and innate
immune cells and overall composition of blood) and therefore adjustment for these differences should be forgone as they may provide key insights into the cellular disruption associated with disease pathogenesis (Shanthikumar et al., 2020).

So far, research has indicated that cell subtype-specific methylation patterns have been associated with particular disease states. For example, in a post-mortem study of 50 patients with schizophrenia and 45 controls cell type-specific methylomes were generated from both neurons and oligodendrocytes. Differential methylation between the two cell types was robust, whereas DNAm differences associated with schizophrenia were more subtle and relegated mainly to differential DNAm within oligodendrocytes (Mendizabal et al., 2019). This demonstrates that epigenetic dysregulation associated with neuropsychiatric phenotype is specific to brain cell types rather than being universal across brain tissue, which was underpowered to detect this difference. A limitation of this study is the lack of controlling for lifestyle-associated factors known to influence DNAm such as smoking (though of note, the CpGs found to associate with schizophrenia in this study were not those that had previously been linked to smoke exposure). In these instances, by purely examining heterogeneous tissue samples such as saliva or whole blood, differences in DNAm levels that are highly relevant to disease pathogenesis could be entirely missed. In another example, microglial cells in the brain (which are of particular relevance to neuroinflammatory responses) have been shown to different methylation profiles to neurons and glia (de Witte et al., 2022). While these cell-type differences may be of little relevance when taking a macro perspective on the brain, (e.g., examining the association of a DNAm-based blood biomarker with overall grey matter atrophy) their distinct methylation profiles may underscore why some of the regional observations in these studies.

In another example, a longitudinal study aimed to examine how DNAm varies in rheumatoid arthritis (RA) patients, some of whom received immunotherapy treatment (Julià et al., 2022). RA patients had distinct methylation profiles
compared to controls (an effect that was ultimately attributed to the downstream consequences of systemic inflammatory responses over and above other RA pathophysiology). Using cell-type deconvolution analysis, these authors were able to tease apart cell-specific changes in methylation profiles in response to immunotherapy treatment, finding that cells of the innate and adaptive system, particularly monocytes, demonstrated the largest epigenetic differences and may ultimately underlie heterogeneity in treatment response. Other measures of inflammation, such as serum CRP levels, showed no such difference. This illustrates the wider point that chronic inflammatory processes can lead to alterations in the DNAm pattern in immune cells. These methylation changes can persist even after the initial inflammatory signal has subsided, and thus may serve as a more stable biomarker of inflammation. In contrast, circulating protein levels may fluctuate more readily over time and subside quickly after the initiating acute inflammatory insult has subsided. Therefore, DNAm changes associated with inflammation might be a more reliable and specific biomarker for tracking the severity of disease, as well as for monitoring the efficacy of treatments. While cell-type specific changes might inform us about disease mechanisms, on the other hand, if a core part of disease pathogenesis results in DNAm changes that occur across cell types (as is likely the case for systemic inflammatory responses), bulk-tissue derived DNAm differences could prove useful as indices of disease risk or exposure.

Because of these two sides to the equation, it is generally considered that epigenome studies that are run on heterogeneous tissue types limit our ability to infer exact mechanistic insights into disease pathogenesis, but are potentially still useful as biomarkers of exposure, disease severity and progression (Relton, Hartwig and Davey Smith, 2015; Ladd-Acosta and Fallin, 2019; Cecil, Neumann and Walton, 2022; Yousefi et al., 2022; Nabais et al., 2023). In the context of DNAm signatures, their accuracy is likely related to cellular diversity of the training sets they were derived from, given that this seems to be the case for the more robust epigenetic clock measures (Yousefi et al., 2022).
9.7.6 Sources of confounding

Inter-individual variation in DNAm may be arbitrary, genetic, environmental, or governed by time (Lancaster, Morris and Connelly, 2018). Many DNAm differences could be attributed to a disease or trait of interest or could alternatively be driven by underlying genetics, environmental exposures, age, medications, corollaries of a disease (i.e., reverse causation) or even just shifts in populations of cells. Because of this, many of these factors are considered as important cofounders to account for.

The first confounder to consider is genetic variation. Genetic variants could influence DNAm profiles in both the technical sense and biological sense: technically, SNPs in the DNA sequence can overlap with CpG sites targeted by DNAm arrays, which can lead to variation in DNAm levels that are unrelated to the phenotype of interest but instead reflect genetic variation. This confounding can be mitigated through various strategies, such removing CpG sites that have known genetic variation, using SNP-aware normalisation techniques or imputing SNP genotypes (Zhou, Laird and Shen, 2017).

Biologically, genetic variants can modify the association between DNAm and the phenotype of interest by a). directly affecting the function of the CpG site via methylation quantitative trait loci (mQTLs) or b). by influencing the downstream biological processes that are impacted by DNAm. This can lead to false positive or false negative associations between DNAm and the phenotype of interest, and may require consideration of gene-environment interactions and mediation analysis to properly evaluate the relationship – in this sense, examining mQTLs is therefore important to delineate such sequences. Efforts to collect longitudinal sampling are needed to examine the direction of effect of DNAm in gene expression and complex traits like brain health outcomes. Statistical efforts to this effect – mendelian randomisation (MR) and causal inference analysis – that combine epigenetic
and genetic data should also be considered (a topic discussed further in section 9.7.9).

The second cofounder to consider has been touched on previously: intrinsic (cell heterogeneity-independent) and extrinsic signals (cell heterogeneity-dependent), and the degree to which they’re a confounding factor or reflective of the variable we’re interested in measuring. If DNA is sampled from peripheral tissue such as blood, which contains mixtures of cells with unique DNAm patterns, this can lead to inter-individual differences in methylation levels – teasing apart whether these DNAm differences are driven by intrinsic or extrinsic signals has been of considerable debate (Nabais et al., 2023)\(^{48}\). In Chapter 8, we adjusted for cell-type composition in models. While we achieved this via reference-based methods (where DNAm data from purified cell types is used to estimate cell-type proportions in mixed samples – in our case by using the Houseman method), another approach is deconvolution-based methods that use DNAm patterns to estimate cell-type proportions. Deconvolution\(^{49}\)-based methods have the advantage of having higher granularity, are more flexible, and do not require a reference dataset, but they can be more sensitive to noise in the DNAm data and thus may not perform as well in highly heterogenous samples. It's therefore important to caveat that we cannot completely eliminate the potential for confounding due

\(^{48}\) In the context of DNA methylation changes associated with chronic inflammation in peripheral whole blood, intrinsic signals refer to DNA methylation changes that are directly associated with the inflammatory process, regardless of the heterogeneity of the cell population. For example, certain CpG sites may become hypermethylated in response to chronic inflammation, and this hypermethylation may occur consistently across multiple cell types in the blood. These changes would be *intrinsic* signals because they are directly associated with the inflammatory process and are predominantly driven by differences in cell type proportions. By contrast, *extrinsic* signals refer to DNA methylation changes that are driven by differences in cell type proportions (monocytes, eosinophils, NK T-cells etc) between individuals with and without chronic inflammation. So if certain cell types are more abundant in individuals with chronic inflammation, but lower in people with low baseline inflammation levels, this could lead to differences in DNA methylation levels at certain CpG sites that are specific to those cell types.

\(^{49}\) Methods for measuring cell-types include estimation by Houseman et al.’s method vs more modern deconvolution methods such as IDOL method and the very recent EPIC IDOL-Ext by Salas et al. (2022). The latest methods display an enhanced ability to account for cellular heterogeneity – EPIC IDOL-Et in particular characterises sub-types of CD4+ and CD8+ memory T cells, which are known to be strongly influenced by sustained inflammatory responses.
to intrinsic or extrinsic signals in these findings. Study designs that combine genotyping with DNAm to attempt to distinguish between differential DNAm arising as a result of genetic variation and differences arising as a result of chronic inflammation are advised going forward. Other studies have performed transcriptomics investigations on the same cells tested for DNAm changes (Liu et al., 2012). Such additional steps allow a number of drivers and consequences of DNAm-changes to be investigated while in part accounting for meta-epigenomic effects, but run the risk of decreasing interpretability of the analyses (Birney, Smith and Greally, 2016).

9.7.7 Temporal stability of DNAm

Identifying which aspects of the methylome are stable vs dynamic is of particular relevance to predicting health and disease states, with CpG sites that are particularly prone to perturbation being potential mediators or moderators of environmental effects. Moreover, while DNAm is considered to be one of the more stable forms of epigenetic modification, some locations in the genome are more stable than others (Lancaster, Morris and Connelly, 2018).

Regarding DNAm's potential to support lifetime or long-term risk prediction, there are still a lot of unanswered questions. For instance, in Thompson et al., (2022)'s recent study that demonstrated the potential applicability of DNAm signatures to clinical health outcomes (where aggregate DNAm signatures were linked to electronic health record (EHR) data – medications, lab panel values, and diagnosis codes – showing enhanced predictive ability compared to baseline models and PRS models; for discussion of this study refer back to Chapter 3.8.2) DNAm data was obtained from whole blood following clinical diagnosis. As a result, it is unclear whether methylation modifications occurred before the diagnosis (thereby useful as a screening tool, or pre-diagnostic measure) or as a result of the progression of the disease (lending more applicability to profiling health trajectories). Though
the studies in this thesis obtained DNAm at the same time as neuroimaging assessment, in being cross-sectional, these studies do not track changes in brain health status over time.

It may be that there are key windows of plasticity in the aetiology of certain disease outcomes where altered methylation is informative for disease risk. Of relevance to the findings in Chapter 7 of this thesis, this is particularly relevant when considering time-sensitive periods in brain development in early-life. If methylation changes related to disease occur at specific time points, the relevance of deriving DNAm signatures outside of these windows could diminish. Relevant to this, a recent collaborative effort to characterise the human methylome across the lifecourse harmonising DNAm data across eight cohort groups (spanning human samples from birth to 100 years of age) found that there are widespread changes to mean levels of DNAm and DNAm variability with ageing (Walton et al., 2021). Longitudinal studies are essential for assessing intra-individual stability and determining whether DNAm changes induced by exposures persist or fluctuate over time, and whether fluctuations are stochastic or in response to modifiers of that exposure (e.g. anti-inflammatory treatment).

9.7.8 Inconsistencies across the chapters in terms of study design

One of the main distinctions outlined in Chapter 1.3 was the difference between acute inflammation vs low-level chronic inflammation. In Chapter 6, participants with raised serum CRP levels (> 10mg/L) were excluded from analyses (n = 32, Appendix 6), in line with other study designs that attempt to control for cases of acute inflammation indicative of infection of illness. However, this approach was not applied to the study presented in Chapter 7 for two main reasons (1) the variable response of CRP in preterm

50 In the literature, low-grade chronic inflammation (the type that is potentially undetected by your standard low-sensitivity CRP assay) is sometimes referred to as ‘metainflammation’
populations renders it a less useful indicator of inflammation (2) serum CRP was collected at different time point to DNAm sampling. This means that, along with not being able to directly compare the performance of DNAm CRP and serum CRP, the findings in Chapter 7 may reflect acute rather than sustained or chronic inflammation as we did not screen out cases of acute-inflammation. Given that a proportion of the preterm infants within this population went on to develop an inflammatory-related condition (78%, n = 112; see Figure 9.4), it is highly likely there are cases of acute as well as sustained inflammation within this sample. Having said this, the DNAm CRP score does appear to index related risk of these conditions better than previous approaches in TEBC, such using serum IL6 or IL8 (Sullivan et al., 2021), suggesting its utility of capturing a sustained profile of inflammation that has incident health consequences.

Can inflammation-related DNAm give insight into brain dysmaturation in preterm birth?

**Background:**
Inflammation has been linked to neuroimaging hallmarks of preterm birth (PTB) termed the encephalopathy of prematurity (EoP).

Looking at epigenetic modifications, such as DNA methylation, offer a new way to assess the cumulative impact of inflammation at the maternal-fetal interface

**Research objective:**
To examine how a DNAm signature of C-Reactive Protein (DNAm CRP) relates to PTB, perinatal exposures and brain structure & connectivity

**Figure 9.4 Chronic inflammation vs acute inflammation in preterm infants**
Graphical abstract for (Conole et al., 2023) displaying that many preterm infants had inflammatory comorbidities suggestive of acute inflammation compared to term infants (none of whom contracted postnatal clinical inflammatory phenotypes / infections).
Additionally, we did not examine serum levels of CRP from mothers, and instead examined indirect inflammatory risk factors that were grouped according to whether they were of maternal, fetal, or postnatal origin. As CRP does not cross the placenta (Nielsen et al., 1990), it is hypothesised maternal contributions to fetal inflammation are a result of downstream inflammatory activation resulting from raised maternal CRP rather than direct transference of CRP across the placental barrier. However a recent longitudinal study using the same DNAm CRP signature (same 7 CpG sites) found that cord blood DNAm CRP corresponds well with cord blood serum protein levels, but not maternal CRP levels (Yeung et al., 2020). This indicates that this signature is more likely capturing the perinatal influences of inflammation rather than maternal inflammation. The work in our study corroborates these findings, as maternal inflammatory risk factors (such as smoking and maternal age) were shown to have little influence on DNAm CRP levels.

Similarly, while the approach to profiling this DNAm CRP signature in the STRADL cohort was conducted on a sub-population of participants with serum CRP levels < 10mg/L (Green, Shen, et al., 2021), in the analysis of 709 individuals with 109 inflammatory-related signatures, serum CRP levels as measured by the hsCRP assay were not examined as this significantly reduced the sampled size. Instead, DNAm signatures were compared against their proteomic equivalents as measured by the SomaScan platform, which, due to being aptamer based (as discussed in Chapter 8), may not reflect exact circulating protein levels. This also meant that of the 109 DNAm signatures analysed, matched-DNAm-proteomic associations were only available for 98 signatures (as some of the DNAm signatures had been generated out-of-sample from the Olink panels).

Another inconsistency between the studies is that unlike Chapter 8, a limitation of the work presented in Chapter 6 is that we did not adjust for immune-cell proportions in our models (discussion of the of adjusting for potential sources of confounding has been outlined in 1.5.2) and also only
compared DNAm CRP to a singular inflammatory protein measure. The primary reason we did not adjust for these factors in our analyses in Chapter 6 is that concurrently measured cell-count and olink proteomic data were not available for the neuroimaging wave 2. However, other work suggests that DNAm CRP score correlates well with pro-inflammatory biomarkers and immune-cell counts in the LBC1936 cohort. In a study looking at 823 people from wave 1 of LBC1936 (Nabais et al., 2021), DNAm CRP was examined alongside 92 blood inflammatory protein markers measured with the Olink inflammation panel, as well as immune-cell including such as monocytes and granulocytes. DNAm CRP associated positively with pro-inflammatory coded Olink markers and also immune cell counts – particularly granulocytes counts ($r = 0.50$-$0.77$).

Additionally, Chapter 6 is the only set of analyses that attempted to adjust for medication usage that might specifically influence chronic inflammation, where any participant taking a drug with known anti-inflammatory action were coded as a binary variable (0/1) and included as a sensitivity analyses (See Appendix 11.1, eTable4). Chapter 7 attempted a similar medication adjustment in controlling for the administration of steroids or MgSO4 during the late stages of pregnancy, but other anti-inflammatory medications during the course of pregnancy were not recoded. Similarly, we did not attempt to adjust for anti-inflammatory medication or other medication in Chapter 8.

**Neuroimaging approaches:**

Aside from the limitations previously covered in section 9.7.2, there were some broad differences in the neuroimaging data that should be stated. A technical consideration is that different FreeSurfer versions were used between the cohorts, which is likely to contribute some degree of variance to

51 In this study, it should be cautioned that the weights used to build the DNAm CRP score were taken from the same population sample (LBC1936), which means that these scores likely inflated. Equally, DNAm CRP scores were highly associated with granulocytes which are the most common WBC types, which may mask biologically relevant actions from less abundant lineages
estimations. Moreover, different approaches were used across the three population studies to parcellate cortical measurements: in Chapter 6, a vertex-wise approach was employed, and in Chapters 7-8, an ROI-based approach. The ROI-based method\(^{52}\) involves examining subcortical and cortical volumes based on predefined atlases (ENA50 in Chapter 7 and Desikan-Killany in Chapter 8). In contrast, the vertex-wise approach fits a general linear model at each surface vertex to compare values of cortical ROIs (e.g. thickness, surface area, volume). The resulting statistical maps are overlaid on a template brain to represent contrast estimates in different hues (see Figure 3, Chapter 6 for cortical volume associations with DNAm CRP; Appendix 11.1, eFigure 1 for cortical thickness). Using pre-defined parcellations may increase statistical power if the regions of interest follow gyral borders. However, this method may miss effects in smaller or more specific regions.

Though there were attempts to make these study designs consistent, particularly in terms of the outcome measures assessed and the types of statistical analyses run, these listed differences limit the degree to which comparison of chronic inflammation on brain structure across different age ranges can be assessed.

9.7.9 Causality

Is inflammation a consequence of disease states or processes such as neurodegeneration which then results in differential methylation at certain sites? Or do aberrant epigenetic processes drive inflammatory states and subsequent susceptibility of brain health outcomes? Or are these mechanisms not mutually distinct, with co-occurring inflammation and epigenetic dysregulation exacerbating each other? To what degree do other exposures, or aspects of lifestyle, associate with these relationships? The

\(^{52}\) ROI-based analysis allows for the segmentation of both subcortical and parcellated cortical metrics, whereas the vertex-wise analysis is applicable to deriving cortical ROIs: volume, thickness and surface area.
empirical chapters of this thesis attempt to address some of these questions by examining epigenetic, inflammation, neuroimaging and lifestyle data concurrently, but do not perform causative modelling. As a result, the findings presented here do demonstrate robust evidence for causality in any instance, and further work is therefore needed to determine the directionality between inflammation, DNAm and brain relationships. This is key to determining whether chronic inflammation plays an instigating, perpetuating or double-hit role in the development of cognitive impairment (such as those seen in neurodevelopmental delay in early-life) or cognitive decline, or instead is a corollary or auxiliary product of brain-related changes. While in this thesis I did not examine the causal role of DNAm other studies have made some progress in this area using two-step mendelian randomisation (MR) which I will discuss here.

In two step MR, in step 1, the causal impact of the exposure (e.g, inflammatory proteins) on the epigenetic signature (e.g. DNAm CRP, or individual CpG sites) is assessed using an SNP as a proxy for the exposure. In step 2, the causal nature of the epigenetic signature on the outcome measure is then evaluated using a genetic surrogate for the epigenetic signature. Some caution is advised here: alongside the requirement for large sample size and satisfying conventional MR assumptions, the epigenetic MR strategy faces unique challenges of tissue/cell specificity as discussed in sections 9.7.2.4-5, as well as availability of robust genetic variants that can be used as a genetic instrument for the epigenetic signature of interest. As first discussed in Chapter 1.5.2.3, in the case of inflammation (circulating serum CRP specifically), the largest GWAS to date by Said et al., (2022) identified 266 genetic loci that associate with CRP – 57 of which were replications of the prior largest GWAS which found 58 loci to associate with

53 With an N = 575,531, this GWAS was almost double the size of its ruling predecessor by Lightart et al. (2018), which was conducted in 204,402 individuals. In the 2018 study, 58 loci were identified, the top hits of which explained ~ 7% of variance in serum CRP; in the 2022 study, percentage variance explained was raised to 16.3%, though for any individual SNP the top variance explained was only 1.4%
serum CRP (Ligthart et al., 2018), the top hits of which explained ~ 7% of variance.

The coefficients from this study were used in the recent study by Wielscher et al (2022), described in section 3.8.3, where MR was deployed to interrogate two hypotheses (1.) DNAm is causal for changes in circulating CRP and (2.) DNAm is consequential of changes in circulating CRP. To test (1.) the authors regressed the DNAm of each of the 1,511 CRP-relevant CpG sites identified against all SNPs present in the cis of the concordant CpG. To test (2.), the authors generated a PRS of CRP based off 52 SNPs (i.e. \( \text{CpG} \sim \text{CRP}_{\text{PRS}} + \text{age} + \text{sex} + \text{technical covariates} \)). For hypothesis (1.) the authors found 709 genetic instruments for CRP-CpGs – ultimately finding that 8 of these loci (cg04111102, cg14099685, cg26470501, cg14702231, cg02039839, cg17580616, cg00138407) went on to be causally linked to increased blood CRP levels. However, a further triangulation analysis indicated limited causality. By contrast for hypothesis (2.) they found that their triangulation analysis demonstrated a causal effect of serum CRP on the majority of CpGs. So overall, this study’s MR approach confirmed that observed changes in DNAm were most likely a consequence of low-grade, chronic inflammation, with altered methylation at CpG sites linked to blood CRP levels rather than a driver of it. With mediation analysis, they highlighted additionally that that lifestyle factors (such as obesity and smoking) were key factors driving alterations in CpG methylation levels. They also found that increased DNAm CRP significantly increased risk of incident disease outcomes such as cardiometabolic diseases and COPD. An earlier study by Walton et a., (2018) only found evidence for hypothesis (1.), with DNAm at certain sites (mapped to genes \text{TMEM49, BCL3 and MIR21, regulators of immune responses and apoptosis} appearing to influence circulating CRP levels. The hypothesis proposed by Walton et al. (2018) is that DNAm changes drive an upregulation in CRP (i.e. the causal effect) but simultaneously result in a cascade of immune signalling that might influence DNAm in whole blood (the observed effect).
Aside from these two direct examinations of DNAm in relation to CRP, one of the more interesting studies that went about disentangling cause and effect in relation to DNAm, chronic inflammation and disease is the longitudinal study by Somineni et al. (2019) (discussed in Chapter 3.7.2) which examined blood-DNAm in Chron’s disease patients. This study’s headline findings were that differential DNAm in Chron’s disease patients were attributed to inflammation beyond other aspects of disease aetiology, and with treatment, these DNAm profiles returned to that of patients without intestinal inflammation. MR analyses demonstrated that the vast majority of dysregulated DNAm was a result of inflammation, an effect that was seemingly reversible following targeted treatment. This study brings into sharp relief the potential of DNAm to track disease severity and response to treatment, a topic further discussed in section 9.8.

However, unlike the examples above this issue of causality is particularly challenging in the context of brain health outcomes, where longitudinal sampling of the primary tissue of interest (brain) are not feasible. Even samples that contain longitudinal neuroimaging, cognition and concurrent DNAm data are scarce (Wheater et al., 2020; Walton et al., 2023). In the case of traditional protein-based inflammation measures (and no examination of DNAm) and brain health outcomes examinations are on the rise. Two recent MR studies found strong evidence for a casual association with CRP with Parkinson’s via phenome-wide MR in the UK Biobank (Si et al., 2021). Another study, using pooled European population cohorts (totalling an n of 13,955 - 204,402 participants), found that elevated circulating IL6 protein levels were causally associated with age of onset of Parkinson’s disease (Bottigliengo et al., 2022). These causal associations are also much more numerous in the psychoimmunology literature, where MR analyses in UKBiobank have demonstrated that both CRP and IL6 were causally associated with depression (Khandaker et al., 2019) and schizophrenia (Williams et al., 2022). However there is still wide dissent on causality of inflammation and other brain health outcomes, such as cognition, as shown in a recent two-step MR of inflammatory proteins and measures of cognitive
ability across a broad, but notably young\textsuperscript{54}, age range (Slaney et al., 2023), in which multiple MR methods were adopted. This study also used the SNPs identified by Ligthart et al. (2018) in their analyses, finding no strong causal role for inflammatory proteins (CRP, IL6 and GlycA\textsuperscript{55}) on aspects of cognitive ability. Here, the authors suggest that other biological pathways linked to inflammation and cognitive functioning may drive the observed effect in other studies, although the younger age range (24 years) in this study may also indicate that this is too early to detect divergencies in cognitive ageing trajectories as a result of ‘inflammaging’. As limited studies have looked at DNA methylation signatures of inflammation in relation to brain health outcomes (excepting those presented in this thesis), these are a key next step in determining causal associations and delineating pathways via which inflammation relates to brain health.

In Chapter 8, we carried out pathway analyses to investigate whether the most associated CpG probes cluster within distinct biological functions using FUMA (http://fuma.ctglab.nl) which tests the relationships between tissue-specific gene expression and disease-gene associations, using gene expression data from a variety of sources such as GTEx and the BrainSpan project. Genes closest to CpG probes linked to circulating inflammatory proteins were largely not preferentially expressed in brain tissue (Appendix 11.3, Supplementary Figures 12-13). While this corroborates previous findings that that systemic inflammation can nonetheless associate and potentially exert deleterious effects on the brain (reviewed in sections 1.6, 2.4.3, 2.6.1, and 2.7.3), it also indicates that the relationship between systemic inflammation and brain structure is complex and multifactorial. It is possible that systemic inflammation triggers a cascade of events that ultimately lead to neuroinflammation, neurodegeneration, and cognitive decline via the pathways outlined in Chapter 1.6. This cascade may involve

\textsuperscript{54} In contrast to most studies examining associations between inflammatory proteins and cognition, this study looked at measures of ‘hot’ and ‘cold’ cognitive functioning at age 24 in 3,305 individuals from the ALSPAC cohort.

\textsuperscript{55} Glycoprotein acetylation (GlycA) can, in its own way, be classed as a composite biomarker of inflammation. It’s a nuclear magnetic resonance imaging derived measure of the integrated concentration (‘glycosylation’) of a range of pro-inflammatory acute phase proteins.
interactions between immune cells, blood-brain barrier dysfunction, neurotransmitter dysregulation, and other inflammatory milieu that DNAm-based proxies are well placed to capture (in covering a range of immune and vascular loci). Given that many of the DNAm signatures described in Chapter 8 were also found to have CpG sites mapped to metabolic pathways and vascular-endothelial interactions, it is also likely that systemic inflammation influences brain structure through effects on cardiovascular health, which in turn can affect cerebral blood flow. While we used various methods to try and determine the functional relevance of these signatures, genotyping could help to further refine our understanding of the causal relationships between these factors.
9.8 Recommendations for future research

In addition to the implications for this work discussed in Chapters 6-8, it is also important to consider the future directions for this work within the wider research community.

This thesis has indicated that epigenetic signatures are a valuable tool for profiling the type of chronic, low-level inflammation that may not be well characterised by single serum sampling. The empirical work of this thesis demonstrated that these approaches explain variance in brain structural substrates of cognitive ageing. Owing to DNAm’s stability over-time, and durability in archived tissue samples, this also offers a means to impute inflammation (or other exposures) in studies that do not have phenotypic data (e.g. serum inflammatory protein levels) originally collected. However, larger sample sizes, replication across populations of different ages, longitudinal study designs and integrated multi-omic methods are necessary to elucidate and appraise legitimate potential pathways between risk factors, chronic inflammation and brain health outcomes.

Equally, while this thesis aimed to examine concurrently measured DNAm and neuroimaging /cognitive measures, outcome measures that are assessed after DNAm acquisition could help establish temporal patterning associations of DNAm, and longitudinal follow-up will clarify how robust these associations are. In general, repeated DNAm measurement is both optimal for demonstrating the robustness of associations and essential to help further our understanding of DNAm changes over time. Investigating the temporal stability of DNAm signatures is important to establish how these indices of chronic inflammation respond in cases of treatment, lifestyle intervention, disease progression, and age. This is also relevant for examining how DNAm signatures present during pregnancy, whether they signpost any risk of deviations in healthy gestation trajectories, and the degree to which
signatures persist postnatally, as previously discussed in Chapter 7 of this thesis.

Future studies should focus on building more DNAm-based predictors of inflammation and their applicability to various phenotypes. Of particular translational value to assessing the relationship between inflammation and clinical phenotypes, inflammatory-related DNAm signatures should be investigated in brain-health conditions suspected to have a significant inflammatory component – such as brain-fog associated long COVID, cognitive decline in autoimmune populations and MS patients. To date, no studies have looked at inflammation-associated DNAm and specific neuropathology, although studies have extended this paradigm to examine neuropsychiatric risk pertaining to inflammation (as indexed by DNAm signatures) (Green, Shen, et al., 2021; Green, Squillace, et al., 2021; Edmondson-Stait et al., 2022). There is good evidence for inflammation-related differences in DNAm profiles in MS patients in particular, making this a promising area of further study (Celarain and Tomas-Roig, 2020).

Moreover, future studies could focus on identifying new CpG sites associated with chronic inflammation to improve the accuracy of DNAm risk scores with larger EWAS or training samples. This is particularly relevant considering the number of CpGs that might confer minor effects individually but significant effects in aggregate (Yousefi et al., 2022). The number of epigenetic modifications linked to chronic inflammation is expanding and will likely continue to do so (Gonzalez-Jaramillo et al., 2019; Ramos-Lopez et al., 2021). While the work of this thesis focuses on DNAm, the variety of epigenetic modifications (histone, ncRNAs and chromatin modifications) linked to inflammation is also anticipated to increase.

Furthermore, integration of other -omic data types could provide a more comprehensive understanding of the molecular mechanisms underlying chronic inflammation and its relationship with brain health outcomes. For example, examining the correlation between DNAm levels and gene
expression levels provides insights into the functional relevance of specific epigenetic modifications, which could help cover regulatory networks and immune pathways that are dysregulated in cases of accelerated brain ageing. **Figure 9.5** illustrates the potential applications of DNAm in relation to understanding the mechanisms that underlie disease states as well as efforts towards disease prediction, characterisation and prognostic / treatment outcomes. This is particularly relevant to the growing interest in multi-omic integration in line with AI for precision medicine (MacEachern and Forkert, 2021). Since DNAm at different sites displays different degrees of plasticity (Wu and Zhang, 2014), chronic inflammation associated DNAm signatures such as those explored in this thesis could present as a tool from which to monitor how intervention, treatment and lifestyle effectively decrease the risk of brain health outcomes – as has been demonstrated in longitudinal profiling of DNAm in chronic inflammatory diseases and autoimmune conditions (Somineni et al., 2019; Nair et al., 2020; Julià et al., 2022).

More broadly, the potential of measuring DNAm as a way to monitor the efficacy of personalised interventions is of particular interest to the growing market of personalised health tracking. In areas other than brain health, DNAm could serve as a marker for the transition between metabolically healthy and unhealthy states, or a signpost risk of flare ups in disease conditions – relevant to conditions such as MS, arthritis, inflammatory bowel disease, ulcerative colitis and Chron’s disease.
Figure 9.5 Future of DNAm biomarkers: potential, hurdles, solutions

Schematic illustrating some of the technical and non-technical roadblocks in the translational potential of DNAm biomarkers, as well as potential solutions, next steps and general advances that could help facilitate progress in the field. Note that technical setbacks (e.g. batch effects, cell-type heterogeneity, confounders and collider bias) are listed once for clarity, but are problems across all of the main research avenues (using DNAm for uncovering disease mechanisms, prediction, prognosis, treatment).

While there is great potential in leveraging epigenomic data to (a) understand disease mechanisms (b) identify disease susceptibility (c) characterise stage or severity of disease (d) monitor or track progression and ultimately (e)
assess treatment efficacy, each of these goals faces unique and overlapping hurdles and roadblocks (illustrated in Figure 9.5). DNAm is unlikely to tell the full story: different data types may reveal different aspects of disease, and the interactions between these might be more informative than single -omic type alone. While integrated multi-omic approaches are likely to provide a high dimensional capture of inflammation and immune dynamics, and in doing so show particular promise for overcoming the issue of mischaracterisation of inflammation, efforts in pursuit of this aim are setback by technological caveats in the integration of different data types - a key focus of several reviews (Chater-Diehl et al., 2021; Sudhakar et al., 2022; Yousefi et al., 2022; Feldner-Busztin et al., 2023). Technical contributions of noise and myriad confounding factors are important to define and account for, and there is a need for harmonisation and collaborative effort to establish clear guidelines surrounding the statistical handling, access, clinical feasibility and translational ethics of multi-omic data. A unifying feature is the need for standardised data analysis, scalable technologies, and sharing to facilitate progress in the field. Improvements in data management infrastructure and the development of user-friendly software and tools to streamline approaches to handle these complex data sets is therefore a key area of focus (Babu and Snyder, 2023).
9.9 Final summary

In conclusion, the work presented in this thesis is the first time DNAm signatures of inflammation have been investigated in relation to multi-modal neuroimaging outcomes across different age ranges. This doctoral thesis uncovers the global and regional repercussions of chronic inflammation on brain structure across various life stages, shedding new light on the mechanisms that contribute to individual variations in cognitive abilities. It emphasises the potential of the burgeoning field of epigenetics to overcome the conventional constraints associated with evaluating inflammatory burden in human cohorts. Overall, the successful performance of a range of inflammatory-related DNAm signatures across three different populations and two different surrogate tissue types provides rationale for future exploration of DNAm biomarkers for indexing chronic inflammation and examining associated differences in complex traits. By harnessing this approach, a deeper comprehension of the intricate interplay between inflammation, aberrant neurodevelopment, and cognitive decline can be attained. These findings not only offer novel insights into the far-reaching impact of inflammation on brain health but also underscore the significance of DNAm as a valuable tool for risk prediction and stratification concerning cognitive outcomes. Moving forward, these results lay the foundation for future work using DNAm signatures to track inflammation levels and potentially aid in the development of biomarkers or interventions to mitigate adverse brain health outcomes at different stages of the lifespan.


Bettcher, B.M. et al. (2021) 'Peripheral and central immune system crosstalk in Alzheimer disease — a research prospectus', *Nature Reviews Neurology*, 17(11), pp. 689–701. Available at: https://doi.org/10.1038/s41582-021-00549-x.


Bojesen, S.E. et al. (2017) 'AHRR (cg05575921) hypomethylation marks smoking behaviour, morbidity and mortality', *Thorax*, 72(7), p. 646. Available at: https://doi.org/10.1136/thoraxjnl-2016-208789.


Cheng, Y. *et al.* (2023) 'Development and validation of DNA methylation scores in two European cohorts augment 10-year risk prediction of type 2 diabetes', *Nature Aging* [Preprint]. Available at: https://doi.org/10.1038/s43587-023-00391-4.


Chilungu, F.P. et al. (2021) 'Genome-wide DNA methylation analysis on C-reactive protein among Ghanaians suggests molecular links to the emerging risk of cardiovascular diseases', npj Genomic Medicine, 6(1), p. 46. Available at: https://doi.org/10.1038/s41525-021-00213-9.


Davies, G. et al. (2015) 'Genetic contributions to variation in general cognitive function: a meta-analysis of genome-wide association studies in the CHARGE consortium (N=53 949)’, *Molecular Psychiatry*, 20(2), pp. 183–192. Available at: https://doi.org/10.1038/mp.2014.188.


Dik, M.G. et al. (2005) ‘Serum inflammatory proteins and cognitive decline in older persons’, Neurology, 64(8), p. 1371. Available at: https://doi.org/10.1212/01.WNL.0000158281.08946.68.


Galea, I. (2021) 'The blood–brain barrier in systemic infection and inflammation', *Cellular & Molecular Immunology*, 18(11), pp. 2489–2501. Available at: https://doi.org/10.1038/s41424-021-00757-x.


Heringa, S.M. et al. (2014) 'Markers of low-grade inflammation and endothelial dysfunction are related to reduced information processing speed and executive functioning in an older population–the Hoorn study', *Psychoneuroendocrinology*, 40, pp. 108–118.


309


Johnson, W., Nijenhuis, J. te and Bouchard, T.J. (2008) ‘Still just 1 g: Consistent results from five test batteries’, Intelligence, 36(1), pp. 81–95. Available at: https://doi.org/10.1016/j.intell.2007.06.001.


Melnikov, I. et al. (2022) 'The monomeric C-reactive protein level is associated with the increase in carotid plaque number in patients with subclinical carotid atherosclerosis', *Frontiers in Cardiovascular Medicine*, p. 1940.


Ohsawa, M. et al. (2005) ‘CRP levels are elevated in smokers but are unrelated to the number of cigarettes and are decreased by long-term smoking cessation in male smokers’, *Preventive Medicine*, 41(2), pp. 651–656. Available at: https://doi.org/10.1016/j.ypmed.2005.02.002.


Randomization Study’, PLOS Medicine, 13(6), p. e1001976. Available at: https://doi.org/10.1371/journal.pmed.1001976.


Ridker, P.M. et al. (2003) ‘C-Reactive Protein, the Metabolic Syndrome, and Risk of Incident Cardiovascular Events’, Circulation, 107(3), pp. 391–397. Available at: https://doi.org/10.1161/01.CIR.0000055014.62083.05.


Singh-Manoux, A. *et al*. (2014) 'Interleukin-6 and C-reactive protein as predictors of cognitive decline in late midlife', *Neurology*, 83. Available at: https://doi.org/10.1212/WNL.0000000000002866.


Sluiman, A.J. *et al*. (2022) 'Higher baseline inflammatory marker levels predict greater cognitive decline in older people with type 2 diabetes: year 10 follow-up of the Edinburgh Type 2 Diabetes Study', *Diabetologia*, 65(3), pp. 467–476. Available at: https://doi.org/10.1007/s00125-021-05634-w.


Smith, R. *et al*. (2020) 'Quantitative streamlines tractography: methods and inter-subject normalisation'.


Sudhakar, P. et al. (2022) ‘Tailoring Multi-omics to Inflammatory Bowel Diseases: All for One and One for All’, Journal of Crohn’s and Colitis, 16(8), pp. 1306–1320. Available at: https://doi.org/10.1093/ecco-jcc/jiac027.


Tegeler, C. et al. (2016) 'The inflammatory markers CRP, IL-6, and IL-10 are associated with cognitive function—data from the Berlin Aging Study II', Neurobiology of aging, 38, pp. 112–117.


Thompson, M. et al. (2022) 'Methylation risk scores are associated with a collection of phenotypes within electronic health record systems', npj Genomic Medicine, 7(1), p. 50. Available at: https://doi.org/10.1038/s41525-022-00320-1.


Tillett, W.S. and Francis, T., Jr. (1930) 'SEROLOGICAL REACTIONS IN PNEUMONIA WITH A NON-PROTEIN SOMATIC FRACTION OF PNEUMOCOCCUS', Journal of Experimental Medicine, 52(4), pp. 561–571. Available at: https://doi.org/10.1084/jem.52.4.561.


Van Steenwinckel, J. et al. (2014) 'Brain damage of the preterm infant: new insights into the role of inflammation', Biochemical Society Transactions, 42(2), pp. 557–563. Available at: https://doi.org/10.1042/BST20130284.

Ventham, N.T. et al. (2016) 'Integrative epigenome-wide analysis demonstrates that DNA methylation may mediate genetic risk in inflammatory bowel disease', Nature communications, 7, pp. 13507–13507. Available at: https://doi.org/10.1038/ncomms13507.


Walton, E. et al. (2023) ‘A systematic review of neuroimaging epigenetic research: calling for an increased focus on development’, Molecular Psychiatry [Preprint]. Available at: https://doi.org/10.1038/s41380-023-02067-2.


Whalley, H.C. et al. (2016) 'Dissection of major depressive disorder using polygenic risk scores for schizophrenia in two independent cohorts', Translational Psychiatry, 6(11), pp. e938–e938. Available at: https://doi.org/10.1038/tp.2016.207.


Wheater, E.N.W. et al. (2020) 'DNA methylation and brain structure and function across the life course: A systematic review', Neuroscience and biobehavioral reviews. 2020/03/06, 113, pp. 133–156. Available at: https://doi.org/10.1016/j.neubiorev.2020.03.007.


de Witte, L.D. et al. (2022) 'Contribution of Age, Brain Region, Mood Disorder Pathology, and Interindividual Factors on the Methylome of Human Microglia', Predicting the Natural History and Treatment Response in Mood Disorders, 91(6), pp. 572–581. Available at: https://doi.org/10.1016/j.biopsych.2021.10.020.


11 Appendix – Supplementary data

11.1 Chapter 6 Supplementary data

An online version of this can be accessed at:

https://cdn-links.lww.com/permalink/wnl/b/wnl_2021_10_19_conole_1_sdc1.pdf
Supplementary Methods

DNA Methylation and Protein Markers of Chronic Inflammation and their Associations with Brain and Cognitive Ageing

Authors: Eleanor L.S. Conole¹,²,³, Anna J. Stevenson²,⁴, Sarah E. Harris¹, Claire Green⁵, Susana Muñoz Maniega¹, María del C Valdés-Hernández¹, Mathew A. Harris¹,⁵, Mark E. Bastin¹,³, Joanna M. Wardlaw¹,³, Ian J. Deary¹, Veronique E. Miron⁴,⁶, Heather C. Whalley³,⁵, Riccardo E. Marioni¹,², Simon R. Cox¹*

¹ Lothian Birth Cohorts group, Department of Psychology, University of Edinburgh, Edinburgh EH8 9JZ, UK.

² Centre for Genomic and Experimental Medicine, Institute of Genetics and Molecular Medicine, University of Edinburgh, Edinburgh EH4 2XU, UK

³ Centre for Clinical Brain Sciences, University of Edinburgh, Edinburgh EH16 4SB, UK

⁴ UK Dementia Research Institute, Edinburgh Medical School, University of Edinburgh, Edinburgh

⁵ Division of Psychiatry, Royal Edinburgh Hospital, University of Edinburgh, Edinburgh, EH10 5HF, UK

⁶ The Queen’s Medical Research Institute, Edinburgh BioQuarter, University of Edinburgh, Edinburgh, EH16 4TJ

*Correspondence to: Simon. R. Cox

Email: Eleanor. L. S. Conole

Email:
Supplementary Methods

Neuroimaging

LBC1936 Image Acquisition and MRI parameters

Participants underwent whole brain structural and high angular resolution 2mm isotropic voxel diffusion tensor MRI (DT-MRI) using a 1.5T GE Sigma Horizon HDxt clinical scanner (General Electric, Milwaukee, WI, USA) operating in research mode using a shelf-shielding gradient set with maximum gradient of 33 mT/m, and an 8 channel phased-array head coil. Mean age at scanning was 72.7 (SD 0.7) years. The DT-MRI protocol comprised single-shot spin-echo echo-planar (EP) diffusion weighted volumes (b = 1000 s mm\(^{-2}\)) acquired in 64 noncollinear directions, alongside 7 T2-weighted EP volumes (b = 0 s mm\(^{-2}\)). This protocol was run with 72 contiguous axial slices with a field of view of 256 × 256 mm, an acquisition matrix of 128 × 128 and 2mm isotropic voxels. Repetition and echo times were 16.5 s and 95.5 ms, respectively. The full imaging protocol for the LBC1936 has been published previously\(^1\).

Freesurfer Quality Control

For cortical volume and cortical thickness measurements, participants with a T1-weighted scan and complete inflammation data were selected (n=666). Acquired volumes were then processed in FreeSurfer v5.1\(^{2,3}\). This involved segmentation of each volume, identifying brain tissue types, followed by parcellation of cortical grey matter into 34 regions per hemisphere, according to the Desikan- Killiany atlas\(^4\). Quality control involved visually assessing each image output for segmentation and parcellation errors, which were then corrected manually; segmentations with errors that could not be corrected were excluded (n = 37), bringing the total eligible to 629 participants. Participants were then excluded if they had self-reported history of neurodegeneration or signs of cognitive impairment (n=69); additionally, participants with high inflammation levels indicative of infection or illness were excluded (n=32). After exclusions, a total of 521 participants had complete inflammation, cognitive neuroimaging and relevant health data; see flow diagram of attrition, figure 2.

Structural MRI volumetric analysis

Intracranial volume (ICV) was semi-automatically extracted on T2*W images using the Object Extraction Tool in Analyze 9.0 (Mayo Clinic, Analyze 9.0. AnalyzeDirect, Inc. Mayo Clinic) followed by manual editing to remove erroneous structures as described in (Valdés Hernández et al, 2012). In order to estimate total brain tissue volume, cerebrospinal fluid (CSF), venous sinuses and meninges were subtracted from ICV. The CSF and these non-brain tissue structures were extracted
using the combination of T2*W and FLAIR sequences and the MCMxxxVI method as described in\(^5\). Quantitative estimates of WMH were estimated using the same MCMxxxVI method\(^5\) fusing T2*W and FLAIR images. False positives and old infarcts were visually identified and removed manually. Total brain (TB), grey matter (GM), normal-appearing white matter (NAWM) volumes were also segmented using this semi-automated multi-spectral technique\(^5\). WMH was log-transformed, and to control for overall size of the head, we present all MRI measures (TB, GM, NAWM, WMH) as a percentage of ICV in subsequent analyses.

**White matter tract analysis**

White matter connectivity data – measures of fractional anisotropy (FA) and mean diffusivity (MD) – were estimated using the BEDPOSTX/ProbTrackX algorithm in FSL (https://fsl.fmrib.ox.ac.uk) and 12 major tracts of interest were segmented using TractoR (https://www.tractor-mri.org.uk): the genu and splenium of the corpus callosum; arcuate fasciculi; anterior thalamic radiation; rostral cingulum; uncinate fasciculi; inferior longitudinal fasciculi. Tract-average white matter FA and MD were derived as the average of all voxels contained within the resultant tract maps, as described previously\(^6\). A general factor of FA (g\(_{FA}\)) and MD (g\(_{MD}\)) was derived for each participant from the first un-rotated principal component of a principal components analysis (PCA) on twelve of the white matter tracts FA and MD values; these general factors reflect common microstructural properties across main white matter pathways and capture the common variance in white matter integrity\(^7\). Participants with up to 2 missing values from specific tracts had data replaced with the mean value for that tract. Details of individual test loadings are provided in the supplementary document (eTable 14).

**Cognitive variables**

All participants in the LBC1936 underwent a detailed battery of standardised cognitive tests; from these, participant scores for three distinct cognitive domains (visuospatial ability, processing speed and verbal memory) were created, based upon well-fitting, hierarchical structural equation models tested in our previously published work\(^8\)–\(^10\).

Visuospatial ability was assessed by performance on Matrix Reasoning and Block Design from the Wechsler Adult Intelligence Scale III\(^{UK}\) (WAIS III\(^{UK}\))\(^11\) and the sum of Spatial Span Forward and Backward from the Wechsler Memory Scale III\(^{UK}\) (WMS III\(^{UK}\))\(^12\). Processing Speed was measured with Symbol Search and Digit Symbol Substitution from the WAIS III\(^{UK}\), Visual Inspection Time\(^13\) and Four-Choice Reaction Time\(^14\). Verbal Memory ability was measured using Logical Memory (sum of immediate and delayed) and Verbal Paired Associates (sum of immediate and delayed) from the WMS III\(^{UK}\), and Digit Span Backwards from the WAIS III\(^{UK}\).
We also derived a single general fluid-type cognitive ability score ($g_f$) for each participant from the first un-rotated principal component of a principal components analysis on six of the Wechsler Adult Intelligence Scale-III tests – an approach that has been used both in this cohort$^{15}$ and in others$^{16}$. This fluid-type cognitive ability measure ($g_f$) represents the capacity to perform basic information processing and extemporary thinking tasks, rather than those involving acquired knowledge or experience. Scores on different tests of fluid-type cognitive ability are typically highly correlated$^{17}$, indicative of a latent general cognitive ability factor, and $g_f$ has shown to be particularly sensitive to the effects of ageing$^8$. Relevant cognitive tests and individual weightings of $g_f$ and the latent variables of processing speed, visuospatial ability and verbal memory can be found in supplementary document (eTable 3)

**Lifestyle variables**

All subjects were interviewed and tested individually by a trained psychologist and a research nurse during a visit to the Wellcome Trust Clinical Research Facility (http://www.wtcrf.ed.ac.uk), Western General Hospital, Edinburgh. Trained research nurses measured height and weight as part of a physical examination using a standardized protocol. Height (in centimeters) was measured with a SECA stadiometer on individuals not wearing shoes. The research nurses measured weight (in kilograms) for individuals without outer clothing or shoes using electronic SECA scales with a digital readout. BMI was calculated as weight (in kilograms) divided by height squared (in square meters). The clinical interview then assessed the participant’s self-reported health and lifestyle: cardiovascular disease history (CVD); hypertension; diabetes; smoking status (coded as current smoker [1] versus ex/non-smoker [0]) and current alcohol use (alcohol units per week).
### Table 1. Participant characteristics

<table>
<thead>
<tr>
<th>Class</th>
<th>Variable</th>
<th>Units</th>
<th>mean</th>
<th>SD</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>Demographics</strong></td>
<td>sex</td>
<td>M: F</td>
<td>358:322</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>Age</td>
<td>years</td>
<td>72.494</td>
<td>0.716</td>
</tr>
<tr>
<td><strong>Inflammation</strong></td>
<td>DNAm CRP</td>
<td>Std units</td>
<td>-0.018</td>
<td>0.001</td>
</tr>
<tr>
<td></td>
<td>serum CRP (high sensitivity)</td>
<td>mg/L</td>
<td>2.016</td>
<td>1.92</td>
</tr>
<tr>
<td></td>
<td>TB</td>
<td>cm³</td>
<td>992.0</td>
<td>89.8</td>
</tr>
<tr>
<td></td>
<td>GM</td>
<td>cm³</td>
<td>473.5</td>
<td>45.1</td>
</tr>
<tr>
<td></td>
<td>NAWM</td>
<td>cm³</td>
<td>477.0</td>
<td>50.67</td>
</tr>
<tr>
<td></td>
<td>WMH</td>
<td>cm³</td>
<td>8.800</td>
<td>1.31</td>
</tr>
<tr>
<td><strong>Neuroimaging</strong></td>
<td>Cortical thickness</td>
<td>mm</td>
<td>3.130</td>
<td>0.15</td>
</tr>
<tr>
<td></td>
<td>gFA</td>
<td>Std units</td>
<td>0.040</td>
<td>0.97</td>
</tr>
<tr>
<td></td>
<td>gMD</td>
<td>Std units</td>
<td>-0.023</td>
<td>0.97</td>
</tr>
<tr>
<td></td>
<td>Genu of corpus callosum</td>
<td>Std units</td>
<td>0.376</td>
<td>0.046</td>
</tr>
<tr>
<td></td>
<td>Splenium of corpus callosum</td>
<td>Std units</td>
<td>0.511</td>
<td>0.066</td>
</tr>
<tr>
<td></td>
<td>Left Arcuate Fasciculus</td>
<td>Std units</td>
<td>0.436</td>
<td>0.039</td>
</tr>
<tr>
<td></td>
<td>Right Arcuate Fasciculus</td>
<td>Std units</td>
<td>0.415</td>
<td>0.039</td>
</tr>
<tr>
<td></td>
<td>Left Anterior Thalamic Radiation</td>
<td>Std units</td>
<td>0.324</td>
<td>0.032</td>
</tr>
<tr>
<td></td>
<td>Right Anterior Thalamic Radiation</td>
<td>Std units</td>
<td>0.335</td>
<td>0.035</td>
</tr>
<tr>
<td></td>
<td>Left Rostral Cingulum</td>
<td>Std units</td>
<td>0.437</td>
<td>0.058</td>
</tr>
<tr>
<td></td>
<td>Right Rostral Cingulum</td>
<td>Std units</td>
<td>0.410</td>
<td>0.047</td>
</tr>
<tr>
<td></td>
<td>Left Uncinate fasciculus</td>
<td>Std units</td>
<td>0.325</td>
<td>0.034</td>
</tr>
<tr>
<td></td>
<td>Right Uncinate fasciculus</td>
<td>Std units</td>
<td>0.320</td>
<td>0.031</td>
</tr>
<tr>
<td></td>
<td>Left Inf. Longitudinal Fasciculus</td>
<td>Std units</td>
<td>0.386</td>
<td>0.051</td>
</tr>
<tr>
<td></td>
<td>Right Inf. Longitudinal Fasciculus</td>
<td>Std units</td>
<td>0.377</td>
<td>0.048</td>
</tr>
<tr>
<td><strong>Fractional</strong></td>
<td><strong>anisotropy</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>Genu of corpus callosum</td>
<td>Std units</td>
<td>798.7</td>
<td>78.3</td>
</tr>
<tr>
<td></td>
<td>Splenium of corpus callosum</td>
<td>Std units</td>
<td>809.0</td>
<td>123.6</td>
</tr>
<tr>
<td></td>
<td>Left Arcuate Fasciculus</td>
<td>Std units</td>
<td>658.5</td>
<td>50.4</td>
</tr>
<tr>
<td></td>
<td>Right Arcuate Fasciculus</td>
<td>Std units</td>
<td>644.8</td>
<td>50.6</td>
</tr>
<tr>
<td></td>
<td>Left Anterior Thalamic Radiation</td>
<td>Std units</td>
<td>756.5</td>
<td>63.9</td>
</tr>
<tr>
<td></td>
<td>Right Anterior Thalamic Radiation</td>
<td>Std units</td>
<td>739.1</td>
<td>76.2</td>
</tr>
<tr>
<td></td>
<td>Left Rostral Cingulum</td>
<td>Std units</td>
<td>636.1</td>
<td>45.3</td>
</tr>
<tr>
<td></td>
<td>Right Rostral Cingulum</td>
<td>Std units</td>
<td>622.4</td>
<td>44.1</td>
</tr>
<tr>
<td></td>
<td>Left Uncinate fasciculus</td>
<td>Std units</td>
<td>764.6</td>
<td>57.7</td>
</tr>
<tr>
<td></td>
<td>Right Uncinate fasciculus</td>
<td>Std units</td>
<td>760.6</td>
<td>49.0</td>
</tr>
<tr>
<td></td>
<td>Left Inf. Longitudinal Fasciculus</td>
<td>Std units</td>
<td>772.7</td>
<td>105.4</td>
</tr>
<tr>
<td></td>
<td>Right Inf. Longitudinal Fasciculus</td>
<td>Std units</td>
<td>757.9</td>
<td>97.1</td>
</tr>
<tr>
<td><strong>Mean</strong></td>
<td><strong>diffusivity</strong></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td></td>
<td>gf</td>
<td>Std units</td>
<td>0.052</td>
<td>0.975</td>
</tr>
<tr>
<td></td>
<td>processing speed</td>
<td>Std units</td>
<td>0.058</td>
<td>0.947</td>
</tr>
<tr>
<td></td>
<td>visuospatial ability</td>
<td>Std units</td>
<td>0.048</td>
<td>0.997</td>
</tr>
<tr>
<td></td>
<td>verbal memory</td>
<td>Std units</td>
<td>0.04</td>
<td>0.972</td>
</tr>
<tr>
<td><strong>Cognitive</strong></td>
<td>hypertension</td>
<td>% Yes</td>
<td>53%</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>CVD</td>
<td>% Yes</td>
<td>29%</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>diabetes</td>
<td>% Yes</td>
<td>10%</td>
<td>-</td>
</tr>
<tr>
<td><strong>Lifestyle</strong></td>
<td>alcohol units per week</td>
<td>Std units</td>
<td>14.2</td>
<td>14.9</td>
</tr>
<tr>
<td></td>
<td>smokers</td>
<td>% Yes</td>
<td>8%</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>ex smokers</td>
<td>% Yes</td>
<td>44%</td>
<td>-</td>
</tr>
<tr>
<td></td>
<td>BMI</td>
<td>Kg/m²</td>
<td>27.8</td>
<td>4.25</td>
</tr>
<tr>
<td></td>
<td>anti-inflammatory medication</td>
<td>% Yes</td>
<td>7.2%</td>
<td>-</td>
</tr>
</tbody>
</table>

TB: total brain, GM: grey matter, NAWM: normal-appearing white matter, WMH: white matter hyperintensity, gf, general cognitive ability; gFA: general fractional anisotropy, gMD: general mean diffusivity
**eTable 2. CpG sites and relative weights** (from Lighthart et al. 2016) used to generate DNAm CRP score

<table>
<thead>
<tr>
<th>CpG</th>
<th>Gene</th>
<th>Beta (discovery sample)</th>
</tr>
</thead>
<tbody>
<tr>
<td>cg06126421</td>
<td>TUBB</td>
<td>-0.0052</td>
</tr>
<tr>
<td>cg06690548</td>
<td>SLC7A11</td>
<td>-0.0048</td>
</tr>
<tr>
<td>cg10636246</td>
<td>AIM 2 &amp; IF116</td>
<td>-0.0069</td>
</tr>
<tr>
<td>cg18181703</td>
<td>SOCS3</td>
<td>-0.0053</td>
</tr>
<tr>
<td>cg19821297</td>
<td>DNASE2</td>
<td>-0.0051</td>
</tr>
<tr>
<td>cg25325512</td>
<td>FGD2</td>
<td>-0.0051</td>
</tr>
<tr>
<td>cg27023587</td>
<td>HEATR6</td>
<td>-0.005</td>
</tr>
</tbody>
</table>
### eTable 3: Test loadings for general factors of cognitive ability measures

<table>
<thead>
<tr>
<th>Cognitive test</th>
<th>gf*</th>
<th>processing speed</th>
<th>visuospatial ability</th>
<th>verbal memory</th>
</tr>
</thead>
<tbody>
<tr>
<td>digit-span backwards</td>
<td>0.634</td>
<td>-</td>
<td>-</td>
<td>0.608</td>
</tr>
<tr>
<td>symbol search</td>
<td>0.746</td>
<td>0.816</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>digit-symbol coding</td>
<td>0.739</td>
<td>0.834</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>matrix reasoning</td>
<td>0.691</td>
<td>-</td>
<td>0.796</td>
<td>-</td>
</tr>
<tr>
<td>letter-number sequencing</td>
<td>0.708</td>
<td>-</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>block design</td>
<td>0.71</td>
<td>-</td>
<td>0.839</td>
<td>-</td>
</tr>
<tr>
<td>Inspection Time</td>
<td>-</td>
<td>0.628</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Four choice reaction</td>
<td>-</td>
<td>-0.773</td>
<td>-</td>
<td>-</td>
</tr>
<tr>
<td>Verbal Paired Associates</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.809</td>
</tr>
<tr>
<td>Logical Memory</td>
<td>-</td>
<td>-</td>
<td>-</td>
<td>0.827</td>
</tr>
<tr>
<td>Spatial Span</td>
<td>-</td>
<td>-</td>
<td>0.739</td>
<td>-</td>
</tr>
</tbody>
</table>

| Proportion of variance          | 0.498| 0.588            | 0.628                | 0.569        |

*Note* *the tests used to generate a general score of cognitive ability were used to facilitate comparison with previous work by Stevenson et al. (2019)*
**Table 4.** Sensitivity analysis: ancillary effect models; interaction effects between age*inflammation, sex*inflammation and anti-inflammatory drug status*inflammation, for DNAm CRP brain-health associations

<table>
<thead>
<tr>
<th>Brain health phenotype</th>
<th>DNAm CRP*sex</th>
<th></th>
<th></th>
<th>DNAm CRP*age</th>
<th></th>
<th></th>
<th>DNAm CRP*anti-inflammatory drug use</th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>( \beta )</td>
<td>SE</td>
<td>pFDR</td>
<td>( \beta )</td>
<td>SE</td>
<td>pFDR</td>
<td>( \beta )</td>
<td>SE</td>
<td>pFDR</td>
</tr>
<tr>
<td>TB</td>
<td>0.073</td>
<td>0.082</td>
<td>0.684</td>
<td>-0.070</td>
<td>0.042</td>
<td>0.276</td>
<td>-0.051</td>
<td>0.119</td>
<td>0.834</td>
</tr>
<tr>
<td>GM</td>
<td>0.003</td>
<td>0.086</td>
<td>0.972</td>
<td>-0.038</td>
<td>0.044</td>
<td>0.688</td>
<td>-0.013</td>
<td>0.126</td>
<td>0.972</td>
</tr>
<tr>
<td>NAWM</td>
<td>0.080</td>
<td>0.086</td>
<td>0.675</td>
<td>-0.013</td>
<td>0.044</td>
<td>0.932</td>
<td>0.127</td>
<td>0.124</td>
<td>0.652</td>
</tr>
<tr>
<td>WMH</td>
<td>0.045</td>
<td>0.087</td>
<td>0.828</td>
<td>-0.049</td>
<td>0.045</td>
<td>0.631</td>
<td>-0.138</td>
<td>0.128</td>
<td>0.631</td>
</tr>
<tr>
<td>gFA</td>
<td>0.023</td>
<td>0.089</td>
<td>0.834</td>
<td>0.032</td>
<td>0.044</td>
<td>0.774</td>
<td>0.255</td>
<td>0.133</td>
<td>0.168</td>
</tr>
<tr>
<td>gMD</td>
<td>-0.049</td>
<td>0.089</td>
<td>0.828</td>
<td>-0.011</td>
<td>0.044</td>
<td>0.932</td>
<td>-0.122</td>
<td>0.132</td>
<td>0.675</td>
</tr>
<tr>
<td>g</td>
<td>0.176</td>
<td>0.074</td>
<td>0.060</td>
<td>0.001</td>
<td>0.037</td>
<td>0.972</td>
<td>-0.018</td>
<td>0.118</td>
<td>0.972</td>
</tr>
<tr>
<td>visuospatial_ability</td>
<td>0.119</td>
<td>0.074</td>
<td>0.286</td>
<td>-0.009</td>
<td>0.037</td>
<td>0.932</td>
<td>-0.004</td>
<td>0.118</td>
<td>0.972</td>
</tr>
<tr>
<td>processing_speed</td>
<td>0.033</td>
<td>0.077</td>
<td>0.834</td>
<td>0.023</td>
<td>0.038</td>
<td>0.828</td>
<td>-0.053</td>
<td>0.115</td>
<td>0.834</td>
</tr>
<tr>
<td>verbal_memory</td>
<td>0.047</td>
<td>0.076</td>
<td>0.828</td>
<td>-0.020</td>
<td>0.038</td>
<td>0.828</td>
<td>-0.152</td>
<td>0.119</td>
<td>0.501</td>
</tr>
</tbody>
</table>
### eTable 5. Sensitivity analysis: health and lifestyle factors added as covariates into inflammation-brain health regressions.

<table>
<thead>
<tr>
<th>Brain health metric</th>
<th>(H0)</th>
<th>(H1)</th>
<th>(H2)</th>
<th>(H3)</th>
<th>(H4)</th>
<th>(H5)</th>
<th>(H6)</th>
<th>(H7)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>β</td>
<td>p(FDR)</td>
<td>β</td>
<td>p(FDR)</td>
<td>β</td>
<td>p(FDR)</td>
<td>β</td>
<td>p(FDR)</td>
</tr>
<tr>
<td><strong>Model 1</strong></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>TB</td>
<td>-0.197</td>
<td>8.42E-06</td>
<td>-0.137</td>
<td>0.002</td>
<td>-0.196</td>
<td>1.19E-04</td>
<td>-0.194</td>
<td>1.19E-04</td>
</tr>
<tr>
<td>GM</td>
<td>-0.200</td>
<td>1.66E-05</td>
<td>-0.153</td>
<td>0.001</td>
<td>-0.192</td>
<td>1.92E-04</td>
<td>-0.199</td>
<td>1.91E-04</td>
</tr>
<tr>
<td>NAWM</td>
<td>-0.150</td>
<td>5.47E-04</td>
<td>-0.089</td>
<td>0.052</td>
<td>-0.147</td>
<td>0.004</td>
<td>-0.147</td>
<td>0.003</td>
</tr>
<tr>
<td>WMH</td>
<td>0.108</td>
<td>0.017</td>
<td>0.071</td>
<td>0.118</td>
<td>0.096</td>
<td>0.052</td>
<td>0.109</td>
<td>0.030</td>
</tr>
<tr>
<td>gFA</td>
<td>-0.162</td>
<td>6.97E-04</td>
<td>-0.108</td>
<td>0.022</td>
<td>-0.151</td>
<td>0.005</td>
<td>-0.155</td>
<td>0.003</td>
</tr>
<tr>
<td>gMD</td>
<td>0.124</td>
<td>0.010</td>
<td>0.130</td>
<td>0.007</td>
<td>0.121</td>
<td>0.022</td>
<td>0.123</td>
<td>0.020</td>
</tr>
<tr>
<td>g</td>
<td>-0.158</td>
<td>6.55E-05</td>
<td>-0.101</td>
<td>0.009</td>
<td>-0.154</td>
<td>4.16E-04</td>
<td>-0.154</td>
<td>4.26E-04</td>
</tr>
<tr>
<td>visuospatial ability</td>
<td>-0.097</td>
<td>0.014</td>
<td>-0.051</td>
<td>0.189</td>
<td>-0.093</td>
<td>0.030</td>
<td>-0.090</td>
<td>0.034</td>
</tr>
<tr>
<td>processing speed</td>
<td>-0.144</td>
<td>4.64E-04</td>
<td>-0.067</td>
<td>0.091</td>
<td>-0.138</td>
<td>2.00E-03</td>
<td>-0.139</td>
<td>2.00E-03</td>
</tr>
<tr>
<td>verbal memory</td>
<td>-0.095</td>
<td>0.017</td>
<td>-0.076</td>
<td>0.061</td>
<td>-0.096</td>
<td>0.030</td>
<td>-0.099</td>
<td>0.027</td>
</tr>
</tbody>
</table>

**Note:** Regression models with added health covariates entered individually. Standardised betas (p values) reported. Bold typeface denotes p < 0.05 (FDR corrected).

The following regressions are reported:

Brain health phenotype ~ DNAm CRP + age + sex + BMI + smoking + alcohol + diabetes + hypertension + CVD history
### eTable 6: FA white matter tract associations with inflammation measures

<table>
<thead>
<tr>
<th>White matter tract</th>
<th>Side of hemisphere</th>
<th>Serum CRP (model 1)</th>
<th>Epigenetic CRP (model 2)</th>
<th>Δ association magnitudes</th>
<th>p</th>
<th>p</th>
<th>p</th>
</tr>
</thead>
<tbody>
<tr>
<td>gFA</td>
<td></td>
<td>β</td>
<td>SE</td>
<td>p</td>
<td>β</td>
<td>SE</td>
<td>p</td>
</tr>
<tr>
<td>Corpus callosum</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Genu</td>
<td></td>
<td>-0.016</td>
<td>0.044</td>
<td>0.999</td>
<td>-0.119</td>
<td>0.045</td>
<td>0.008</td>
</tr>
<tr>
<td>Splenium</td>
<td></td>
<td>-0.090</td>
<td>0.044</td>
<td>0.302</td>
<td>-0.078</td>
<td>0.044</td>
<td>0.077</td>
</tr>
<tr>
<td>Arcuate fasciculus</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Left</td>
<td></td>
<td>0.002</td>
<td>0.043</td>
<td>0.999</td>
<td>-0.108</td>
<td>0.044</td>
<td>0.014</td>
</tr>
<tr>
<td>Right</td>
<td></td>
<td>-0.053</td>
<td>0.045</td>
<td>0.713</td>
<td>-0.150</td>
<td>0.046</td>
<td>0.001</td>
</tr>
<tr>
<td>Anterior thalamic radiation</td>
<td></td>
<td>-0.036</td>
<td>0.045</td>
<td>0.830</td>
<td>-0.120</td>
<td>0.044</td>
<td>0.007</td>
</tr>
<tr>
<td>Left</td>
<td></td>
<td>-0.039</td>
<td>0.043</td>
<td>0.830</td>
<td>-0.126</td>
<td>0.044</td>
<td>0.004</td>
</tr>
<tr>
<td>Right</td>
<td></td>
<td>-0.006</td>
<td>0.044</td>
<td>0.999</td>
<td>-0.050</td>
<td>0.044</td>
<td>0.253</td>
</tr>
<tr>
<td>Rostral cingulum</td>
<td></td>
<td>-0.124</td>
<td>0.045</td>
<td>0.089</td>
<td>-0.140</td>
<td>0.045</td>
<td>0.002</td>
</tr>
<tr>
<td>Left</td>
<td></td>
<td>0.001</td>
<td>0.044</td>
<td>0.999</td>
<td>-0.106</td>
<td>0.045</td>
<td>0.020</td>
</tr>
<tr>
<td>Right</td>
<td></td>
<td>-0.034</td>
<td>0.044</td>
<td>0.830</td>
<td>-0.099</td>
<td>0.044</td>
<td>0.025</td>
</tr>
<tr>
<td>Uncinate fasciculus</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Left</td>
<td></td>
<td>-3.31 x 10^{-5}</td>
<td>0.043</td>
<td>0.999</td>
<td>-0.032</td>
<td>0.044</td>
<td>0.466</td>
</tr>
<tr>
<td>Right</td>
<td></td>
<td>-0.073</td>
<td>0.044</td>
<td>0.496</td>
<td>-0.084</td>
<td>0.043</td>
<td>0.054</td>
</tr>
</tbody>
</table>

**Note** Supplementary eTable 6. Results of linear regression analyses of epigenetic CRP score and serum CRP score with white matter tract microstructural metrics (FA). Furthest right column displays difference between association magnitudes of both models, as assessed by Williams’ test.
**eTable 7: MD white matter tract associations with inflammation measures**

<table>
<thead>
<tr>
<th>White matter tract</th>
<th>Side of hemisphere</th>
<th>Serum CRP (model 1)</th>
<th>Epigenetic CRP (model 2)</th>
<th>Δ association magnitudes</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>β</td>
<td>SE</td>
<td>p</td>
</tr>
<tr>
<td>gMD</td>
<td></td>
<td>-0.025</td>
<td>0.045</td>
<td>0.758</td>
</tr>
<tr>
<td>Corpus callosum</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Genu</td>
<td></td>
<td>-0.034</td>
<td>0.044</td>
<td>0.758</td>
</tr>
<tr>
<td>Splenium</td>
<td></td>
<td>0.056</td>
<td>0.043</td>
<td>0.730</td>
</tr>
<tr>
<td>Arcuate fasciculus</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Left</td>
<td></td>
<td>-0.028</td>
<td>0.043</td>
<td>0.758</td>
</tr>
<tr>
<td>Right</td>
<td></td>
<td>0.018</td>
<td>0.044</td>
<td>0.758</td>
</tr>
<tr>
<td>Anterior thalamic radiation</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Left</td>
<td></td>
<td>-0.029</td>
<td>0.045</td>
<td>0.758</td>
</tr>
<tr>
<td>Right</td>
<td></td>
<td>-0.022</td>
<td>0.043</td>
<td>0.758</td>
</tr>
<tr>
<td>Rostral cingulum</td>
<td></td>
<td>0.074</td>
<td>0.044</td>
<td>0.610</td>
</tr>
<tr>
<td>Uncinate fasciculus</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Left</td>
<td></td>
<td>-0.015</td>
<td>0.046</td>
<td>0.758</td>
</tr>
<tr>
<td>Right</td>
<td></td>
<td>0.037</td>
<td>0.043</td>
<td>0.758</td>
</tr>
<tr>
<td>Inferior longitudinal fasciculus</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Left</td>
<td></td>
<td>-0.066</td>
<td>0.043</td>
<td>0.610</td>
</tr>
<tr>
<td>Right</td>
<td></td>
<td>0.022</td>
<td>0.045</td>
<td>0.758</td>
</tr>
</tbody>
</table>

*Note Supplementary eTable 7. Results of linear regression analyses of epigenetic CRP score and serum CRP score with white matter tract microstructural metrics (MD). Furthest right column displays difference between association magnitudes of both models, as assessed by Williams’ test.*
**eTable 8.** Sensitivity analysis: health and lifestyle factors added as covariates into inflammation-white matter tract regressions (FA).

<table>
<thead>
<tr>
<th>White matter tract FA</th>
<th>Side</th>
<th>Model 1</th>
<th>Fully-adjusted model</th>
<th>Hypertension</th>
<th>CVD history</th>
<th>Diabetes</th>
<th>BMI</th>
<th>Alcohol</th>
<th>Smoking</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>(H0)</td>
<td>(H1)</td>
<td>(H2)</td>
<td>(H3)</td>
<td>(H4)</td>
<td>(H5)</td>
<td>(H6)</td>
<td>(H7)</td>
</tr>
<tr>
<td></td>
<td></td>
<td><strong>β</strong></td>
<td><strong>p(FDR)</strong></td>
<td><strong>β</strong></td>
<td><strong>p(FDR)</strong></td>
<td><strong>β</strong></td>
<td><strong>p(FDR)</strong></td>
<td><strong>β</strong></td>
<td><strong>p(FDR)</strong></td>
</tr>
<tr>
<td>Corpus callosum</td>
<td>Genu</td>
<td>-0.119</td>
<td>0.008</td>
<td>-0.090</td>
<td>0.058</td>
<td>-0.117</td>
<td>0.010</td>
<td>-0.116</td>
<td>0.011</td>
</tr>
<tr>
<td></td>
<td>Splenum</td>
<td>-0.078</td>
<td>0.077</td>
<td>-0.014</td>
<td>0.763</td>
<td>-0.072</td>
<td>0.101</td>
<td>-0.074</td>
<td>0.095</td>
</tr>
<tr>
<td></td>
<td>Left</td>
<td>-0.108</td>
<td>0.014</td>
<td>-0.097</td>
<td>0.040</td>
<td>-0.100</td>
<td>0.023</td>
<td>-0.102</td>
<td>0.021</td>
</tr>
<tr>
<td></td>
<td>Right</td>
<td>-0.150</td>
<td>0.001</td>
<td>-0.129</td>
<td>0.008</td>
<td>-0.143</td>
<td>0.002</td>
<td>-0.144</td>
<td>0.002</td>
</tr>
<tr>
<td>Arcuate fasciculus</td>
<td>Left</td>
<td>-0.120</td>
<td>0.007</td>
<td>-0.101</td>
<td>0.033</td>
<td>-0.110</td>
<td>0.012</td>
<td>-0.117</td>
<td>0.008</td>
</tr>
<tr>
<td></td>
<td>Right</td>
<td>-0.126</td>
<td>0.004</td>
<td>-0.066</td>
<td>0.153</td>
<td>-0.120</td>
<td>0.006</td>
<td>-0.121</td>
<td>0.006</td>
</tr>
<tr>
<td>Anterior thalamic radiation</td>
<td>Left</td>
<td>-0.050</td>
<td>0.253</td>
<td>-0.029</td>
<td>0.536</td>
<td>-0.051</td>
<td>0.246</td>
<td>-0.048</td>
<td>0.278</td>
</tr>
<tr>
<td></td>
<td>Right</td>
<td>-0.140</td>
<td>0.002</td>
<td>-0.108</td>
<td>0.023</td>
<td>-0.134</td>
<td>0.003</td>
<td>-0.142</td>
<td>0.002</td>
</tr>
<tr>
<td>Rostral cingulum</td>
<td>Left</td>
<td>-0.106</td>
<td>0.020</td>
<td>-0.068</td>
<td>0.161</td>
<td>-0.101</td>
<td>0.026</td>
<td>-0.101</td>
<td>0.027</td>
</tr>
<tr>
<td></td>
<td>Right</td>
<td>-0.099</td>
<td>0.025</td>
<td>-0.058</td>
<td>0.222</td>
<td>-0.094</td>
<td>0.035</td>
<td>-0.091</td>
<td>0.040</td>
</tr>
<tr>
<td>Uncinate fasciculus</td>
<td>Left</td>
<td>-0.032</td>
<td>0.466</td>
<td>-0.010</td>
<td>0.837</td>
<td>-0.029</td>
<td>0.514</td>
<td>-0.027</td>
<td>0.541</td>
</tr>
<tr>
<td></td>
<td>Right</td>
<td>-0.084</td>
<td>0.054</td>
<td>-0.048</td>
<td>0.296</td>
<td>-0.076</td>
<td>0.079</td>
<td>-0.078</td>
<td>0.075</td>
</tr>
</tbody>
</table>

**Note:** Inflammation-FA regression models with added health covariates entered individually. Standardised betas (p values) reported. **bold** typeface denotes p < 0.05 (FDR corrected). The following regressions are reported:

[H0] Brain health phenotype ~ DNAm CRP + age + sex
[H1] Brain health phenotype ~ DNAm CRP + age + sex + hypertension + CVD history + diabetes + BMI + alcohol + smoking
<table>
<thead>
<tr>
<th>White matter tract MD</th>
<th>Side</th>
<th>(H0) Model 1 Fully-adjusted model</th>
<th>(H1) hypertension</th>
<th>(H2) CVD history</th>
<th>(H3) diabetes</th>
<th>(H4) BMI</th>
<th>(H5) alcohol</th>
<th>(H6) smoking</th>
<th>(H7)</th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>β</td>
<td>p(FDR)</td>
<td>β</td>
<td>p(FDR)</td>
<td>β</td>
<td>p(FDR)</td>
<td>β</td>
<td>p(FDR)</td>
<td>β</td>
</tr>
<tr>
<td>Corpus callosum</td>
<td>Genu</td>
<td>0.060</td>
<td>0.180</td>
<td>0.077</td>
<td>0.105</td>
<td>0.064</td>
<td>0.154</td>
<td>0.064</td>
<td>0.157</td>
<td>0.062</td>
</tr>
<tr>
<td></td>
<td>Splenium</td>
<td>0.079</td>
<td>0.074</td>
<td>0.058</td>
<td>0.217</td>
<td>0.076</td>
<td>0.088</td>
<td>0.085</td>
<td>0.056</td>
<td>0.080</td>
</tr>
<tr>
<td>Arcuate fasciculus</td>
<td>Left</td>
<td>0.105</td>
<td><strong>0.016</strong></td>
<td>0.104</td>
<td><strong>0.027</strong></td>
<td>0.098</td>
<td><strong>0.025</strong></td>
<td>0.101</td>
<td><strong>0.022</strong></td>
<td>0.102</td>
</tr>
<tr>
<td></td>
<td>Right</td>
<td>0.111</td>
<td><strong>0.015</strong></td>
<td>0.106</td>
<td><strong>0.029</strong></td>
<td>0.107</td>
<td><strong>0.018</strong></td>
<td>0.112</td>
<td><strong>0.014</strong></td>
<td>0.111</td>
</tr>
<tr>
<td>Anterior thalamic radiation</td>
<td>Left</td>
<td>0.083</td>
<td>0.063</td>
<td>0.108</td>
<td><strong>0.024</strong></td>
<td>0.080</td>
<td>0.072</td>
<td>0.082</td>
<td>0.066</td>
<td>0.077</td>
</tr>
<tr>
<td></td>
<td>Right</td>
<td>0.016</td>
<td>0.717</td>
<td>0.016</td>
<td>0.737</td>
<td>0.015</td>
<td>0.735</td>
<td>0.011</td>
<td>0.806</td>
<td>0.014</td>
</tr>
<tr>
<td>Rostral cingulum</td>
<td>Left</td>
<td>0.078</td>
<td>0.073</td>
<td>0.081</td>
<td>0.081</td>
<td>0.081</td>
<td>0.062</td>
<td>0.075</td>
<td>0.084</td>
<td>0.083</td>
</tr>
<tr>
<td></td>
<td>Right</td>
<td>0.093</td>
<td><strong>0.033</strong></td>
<td>0.091</td>
<td>0.050</td>
<td>0.093</td>
<td><strong>0.034</strong></td>
<td>0.098</td>
<td><strong>0.025</strong></td>
<td>0.088</td>
</tr>
<tr>
<td>Uncinate fasciculus</td>
<td>Left</td>
<td>0.046</td>
<td>0.318</td>
<td>0.061</td>
<td>0.212</td>
<td>0.043</td>
<td>0.347</td>
<td>0.043</td>
<td>0.349</td>
<td>0.041</td>
</tr>
<tr>
<td></td>
<td>Right</td>
<td>0.139</td>
<td><strong>0.002</strong></td>
<td>0.140</td>
<td><strong>0.003</strong></td>
<td>0.137</td>
<td><strong>0.002</strong></td>
<td>0.136</td>
<td><strong>0.002</strong></td>
<td>0.136</td>
</tr>
<tr>
<td>Inferior longitudinal fasciculus</td>
<td>Left</td>
<td>0.054</td>
<td>0.220</td>
<td>0.090</td>
<td>0.052</td>
<td>0.055</td>
<td>0.212</td>
<td>0.059</td>
<td>0.179</td>
<td>0.059</td>
</tr>
<tr>
<td></td>
<td>Right</td>
<td>0.013</td>
<td>0.763</td>
<td>0.019</td>
<td>0.680</td>
<td>0.010</td>
<td><strong>0.813</strong></td>
<td>0.011</td>
<td><strong>0.797</strong></td>
<td>0.019</td>
</tr>
</tbody>
</table>

Note: Inflammation-MD regression models with added health covariates entered individually. Standardised betas (p values) reported. bold typeface denotes $p < 0.05$ (FDR corrected). The following regressions are reported:
[H0] Brain health phenotype ~ DNAm CRP + age + sex
[H1] Brain health phenotype ~ DNAm CRP + age + sex + hypertension + CVD history + diabetes + BMI + alcohol + smoking
eTable 10: Bivariate associations among study variables

<table>
<thead>
<tr>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
<th></th>
</tr>
</thead>
<tbody>
<tr>
<td>(1) DNAm CRP</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(2) serum CRP</td>
<td>0.288</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(3) BMI</td>
<td>0.189</td>
<td>0.292</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(4) hypertension</td>
<td>0.059</td>
<td>0.047</td>
<td>0.208</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(5) diabetes</td>
<td>0.121</td>
<td>0.027</td>
<td>0.162</td>
<td>0.123</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(6) CVD history</td>
<td>0.129</td>
<td>0.002</td>
<td>0.079</td>
<td>0.194</td>
<td>0.080</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(7) smoking</td>
<td>0.221</td>
<td>0.138</td>
<td>-0.066</td>
<td>0.023</td>
<td>-0.054</td>
<td>-0.028</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(8) alcohol consumption</td>
<td>0.241</td>
<td>0.060</td>
<td>0.014</td>
<td>0.048</td>
<td>-0.013</td>
<td>0.106</td>
<td>0.047</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(9) gf</td>
<td>-0.162</td>
<td>-0.092</td>
<td>-0.080</td>
<td>-0.082</td>
<td>-0.088</td>
<td>-0.062</td>
<td>-0.153</td>
<td>0.004</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(10) processing speed</td>
<td>-0.147</td>
<td>-0.083</td>
<td>-0.127</td>
<td>-0.090</td>
<td>-0.102</td>
<td>-0.065</td>
<td>-0.181</td>
<td>0.045</td>
<td>0.789</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(11) visuospatial ability</td>
<td>-0.072</td>
<td>-0.079</td>
<td>0.011</td>
<td>-0.074</td>
<td>-0.061</td>
<td>-0.056</td>
<td>-0.148</td>
<td>0.057</td>
<td>0.810</td>
<td>0.546</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(12) verbal memory</td>
<td>-0.116</td>
<td>-0.040</td>
<td>-0.025</td>
<td>0.005</td>
<td>-0.070</td>
<td>0.006</td>
<td>-0.085</td>
<td>0.013</td>
<td>0.586</td>
<td>0.382</td>
<td>0.418</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(13) TB</td>
<td>-0.249</td>
<td>-0.026</td>
<td>-0.117</td>
<td>-0.038</td>
<td>-0.104</td>
<td>-0.110</td>
<td>-0.164</td>
<td>-0.213</td>
<td>0.172</td>
<td>0.205</td>
<td>0.093</td>
<td>0.133</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(14) GM</td>
<td>-0.213</td>
<td>-0.023</td>
<td>-0.124</td>
<td>-0.100</td>
<td>-0.081</td>
<td>-0.049</td>
<td>-0.124</td>
<td>-0.122</td>
<td>0.134</td>
<td>0.147</td>
<td>0.099</td>
<td>0.099</td>
<td>0.741</td>
<td>1.000</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>(15) NAWM</td>
<td>-0.182</td>
<td>-0.023</td>
<td>-0.092</td>
<td>-0.053</td>
<td>-0.070</td>
<td>-0.071</td>
<td>-0.169</td>
<td>-0.150</td>
<td>0.241</td>
<td>0.312</td>
<td>0.155</td>
<td>0.161</td>
<td>0.530</td>
<td>0.188</td>
<td>1.000</td>
<td></td>
<td></td>
</tr>
<tr>
<td>(16) WMH</td>
<td>0.103</td>
<td>0.018</td>
<td>0.076</td>
<td>0.137</td>
<td>0.023</td>
<td>-0.005</td>
<td>0.099</td>
<td>0.021</td>
<td>-0.176</td>
<td>-0.230</td>
<td>-0.141</td>
<td>-0.100</td>
<td>0.006</td>
<td>-0.219</td>
<td>-0.585</td>
<td>1.000</td>
<td></td>
</tr>
<tr>
<td>(17) gFA</td>
<td>-0.160</td>
<td>-0.054</td>
<td>-0.159</td>
<td>-0.146</td>
<td>0.002</td>
<td>-0.086</td>
<td>-0.094</td>
<td>-0.108</td>
<td>0.118</td>
<td>0.159</td>
<td>0.056</td>
<td>0.060</td>
<td>0.251</td>
<td>0.210</td>
<td>0.386</td>
<td>-0.370</td>
<td>1.000</td>
</tr>
<tr>
<td>(18) gMD</td>
<td>0.126</td>
<td>0.024</td>
<td>-0.047</td>
<td>0.049</td>
<td>0.012</td>
<td>0.019</td>
<td>0.063</td>
<td>-0.009</td>
<td>-0.093</td>
<td>-0.148</td>
<td>-0.092</td>
<td>-0.124</td>
<td>-0.049</td>
<td>-0.105</td>
<td>-0.313</td>
<td>0.428</td>
<td>-0.480</td>
</tr>
</tbody>
</table>

*Note.* Pearson’s r reported. TB: total brain volume, WMH: white matter hyperintensity volume, GM: grey matter volume; NAWM: normal appearing white matter volume, gf, general cognitive ability; gFA: general fractional anisotropy, gMD: general mean diffusivity; CVD: cardiovascular disease history
### eTable 11. Results of single SEM mediation models assessing the relationship of brain structure with cognitive ability.

<table>
<thead>
<tr>
<th>Mediator variable</th>
<th>Independent variable</th>
<th>Model</th>
<th>mediation (ab)</th>
<th>Total effect (c)</th>
<th>Direct effect (c')</th>
<th>Attenuation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td>β</td>
<td>SE</td>
<td>p</td>
<td>β</td>
</tr>
<tr>
<td>TB volume</td>
<td>serum CRP</td>
<td>model 1</td>
<td>-0.007</td>
<td>0.007</td>
<td>0.345</td>
<td>-0.099</td>
</tr>
<tr>
<td></td>
<td></td>
<td>model 2</td>
<td>0.001</td>
<td>0.006</td>
<td>0.874</td>
<td>-0.061</td>
</tr>
<tr>
<td></td>
<td>DNAm CRP</td>
<td>model 1</td>
<td>-0.031</td>
<td>0.011</td>
<td><strong>0.005</strong></td>
<td>-0.158</td>
</tr>
<tr>
<td></td>
<td></td>
<td>model 2</td>
<td>-0.018</td>
<td>0.008</td>
<td>0.051</td>
<td>-0.102</td>
</tr>
<tr>
<td>GM volume</td>
<td>serum CRP</td>
<td>model 1</td>
<td>-0.004</td>
<td>0.005</td>
<td>0.492</td>
<td>-0.099</td>
</tr>
<tr>
<td></td>
<td></td>
<td>model 2</td>
<td>0.002</td>
<td>0.004</td>
<td>0.641</td>
<td>-0.060</td>
</tr>
<tr>
<td></td>
<td>DNAm CRP</td>
<td>model 1</td>
<td>-0.022</td>
<td>0.010</td>
<td><strong>0.025</strong></td>
<td>-0.158</td>
</tr>
<tr>
<td></td>
<td></td>
<td>model 2</td>
<td>-0.014</td>
<td>0.008</td>
<td>0.077</td>
<td>-0.101</td>
</tr>
<tr>
<td>NAWM volume</td>
<td>serum CRP</td>
<td>model 1</td>
<td>-0.009</td>
<td>0.010</td>
<td>0.377</td>
<td>-0.098</td>
</tr>
<tr>
<td></td>
<td></td>
<td>model 2</td>
<td>0.001</td>
<td>0.009</td>
<td>0.886</td>
<td>-0.060</td>
</tr>
<tr>
<td></td>
<td>DNAm CRP</td>
<td>model 1</td>
<td>-0.033</td>
<td>0.011</td>
<td><strong>0.003</strong></td>
<td>-0.159</td>
</tr>
<tr>
<td></td>
<td></td>
<td>model 2</td>
<td>-0.018</td>
<td>0.010</td>
<td>0.065</td>
<td>-0.102</td>
</tr>
<tr>
<td>WMH volume</td>
<td>serum CRP</td>
<td>model 1</td>
<td>-0.007</td>
<td>0.011</td>
<td>0.510</td>
<td>-0.098</td>
</tr>
<tr>
<td></td>
<td></td>
<td>model 2</td>
<td>0.002</td>
<td>0.011</td>
<td>0.852</td>
<td>-0.060</td>
</tr>
<tr>
<td></td>
<td>DNAm CRP</td>
<td>model 1</td>
<td>-0.021</td>
<td>0.011</td>
<td>0.060</td>
<td>-0.158</td>
</tr>
<tr>
<td></td>
<td></td>
<td>model 2</td>
<td>-0.011</td>
<td>0.011</td>
<td>0.330</td>
<td>-0.101</td>
</tr>
<tr>
<td>gFA volume</td>
<td>serum CRP</td>
<td>model 1</td>
<td>-0.006</td>
<td>0.006</td>
<td>0.258</td>
<td>-0.099</td>
</tr>
<tr>
<td></td>
<td></td>
<td>model 2</td>
<td>0.000</td>
<td>0.004</td>
<td>0.954</td>
<td>-0.061</td>
</tr>
<tr>
<td></td>
<td>DNAm CRP</td>
<td>model 1</td>
<td>-0.015</td>
<td>0.008</td>
<td>0.079</td>
<td>-0.158</td>
</tr>
<tr>
<td></td>
<td></td>
<td>model 2</td>
<td>-0.008</td>
<td>0.006</td>
<td>0.189</td>
<td>-0.101</td>
</tr>
<tr>
<td>gMD volume</td>
<td>serum CRP</td>
<td>model 1</td>
<td>-0.002</td>
<td>0.004</td>
<td>0.571</td>
<td>-0.098</td>
</tr>
<tr>
<td></td>
<td></td>
<td>model 2</td>
<td>-0.002</td>
<td>0.003</td>
<td>0.551</td>
<td>-0.060</td>
</tr>
<tr>
<td></td>
<td>DNAm CRP</td>
<td>model 1</td>
<td>-0.007</td>
<td>0.006</td>
<td>0.229</td>
<td>-0.158</td>
</tr>
<tr>
<td></td>
<td></td>
<td>model 2</td>
<td>-0.007</td>
<td>0.006</td>
<td>0.265</td>
<td>-0.101</td>
</tr>
</tbody>
</table>

**Note:** Model 1 = Brain health variable ~ age + sex + inflammation; Model 2 = Brain health variable ~ age + sex + inflammation + BMI + hypertension + smoking status + alcohol use + CVD history + diabetes.
Table 12. Results of multiple mediation models (TB, GM, NAWM, WMH, gFA, gMD) assessing the relationship of brain structure with cognitive ability (gf) and cognitive domains of visuospatial ability, processing speed and verbal memory

<table>
<thead>
<tr>
<th>Outcome variable</th>
<th>Mediator variable</th>
<th>Independent variable</th>
<th>Model</th>
<th>mediation (ab)</th>
<th>Total effect (c)</th>
<th>Direct effect (c’)</th>
<th>Attenuation</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td></td>
<td></td>
<td>β</td>
<td>SE</td>
<td>p</td>
<td>β</td>
</tr>
<tr>
<td>g</td>
<td>Brain structure</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>serum CRP</td>
<td></td>
<td></td>
<td>model 1</td>
<td>-0.008</td>
<td>0.012</td>
<td>0.650</td>
<td>-0.099</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>model 2</td>
<td>0.007</td>
<td>0.011</td>
<td>0.715</td>
<td>-0.060</td>
</tr>
<tr>
<td>DNAm CRP</td>
<td></td>
<td></td>
<td>model 1</td>
<td>-0.047</td>
<td>0.015</td>
<td>0.004</td>
<td>-0.159</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>model 2</td>
<td>-0.023</td>
<td>0.013</td>
<td>0.167</td>
<td>-0.102</td>
</tr>
<tr>
<td>Visuospatial ability</td>
<td>Brain structure</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>serum CRP</td>
<td></td>
<td></td>
<td>model 1</td>
<td>-0.005</td>
<td>0.010</td>
<td>0.676</td>
<td>-0.081</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>model 2</td>
<td>0.004</td>
<td>0.009</td>
<td>0.800</td>
<td>-0.071</td>
</tr>
<tr>
<td>DNAm CRP</td>
<td></td>
<td></td>
<td>model 1</td>
<td>-0.036</td>
<td>0.013</td>
<td>0.013</td>
<td>-0.097</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>model 2</td>
<td>-0.020</td>
<td>0.012</td>
<td>0.170</td>
<td>-0.051</td>
</tr>
<tr>
<td>Processing speed</td>
<td>Brain structure</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>serum CRP</td>
<td></td>
<td></td>
<td>model 1</td>
<td>-0.011</td>
<td>0.014</td>
<td>0.631</td>
<td>-0.086</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>model 2</td>
<td>0.005</td>
<td>0.013</td>
<td>0.852</td>
<td>-0.029</td>
</tr>
<tr>
<td>DNAm CRP</td>
<td></td>
<td></td>
<td>model 1</td>
<td>-0.058</td>
<td>0.016</td>
<td>0.001</td>
<td>-0.143</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>model 2</td>
<td>-0.031</td>
<td>0.015</td>
<td>0.105</td>
<td>-0.066</td>
</tr>
<tr>
<td>Verbal memory</td>
<td>Brain structure</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>serum CRP</td>
<td></td>
<td></td>
<td>model 1</td>
<td>-0.002</td>
<td>0.009</td>
<td>0.830</td>
<td>-0.046</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>model 2</td>
<td>0.003</td>
<td>0.009</td>
<td>0.873</td>
<td>-0.029</td>
</tr>
<tr>
<td>DNAm CRP</td>
<td></td>
<td></td>
<td>model 1</td>
<td>-0.032</td>
<td>0.013</td>
<td>0.026</td>
<td>-0.096</td>
</tr>
<tr>
<td></td>
<td></td>
<td></td>
<td>model 2</td>
<td>-0.022</td>
<td>0.012</td>
<td>0.141</td>
<td>-0.076</td>
</tr>
</tbody>
</table>

Results of multiple SEM mediation models (where TB, GM, NAWM, WMH, gFA, gMD are entered simultaneously) assessing the inter-relationship between inflammation, brain structure and cognitive measures. The final column ‘attenuation’ indicates the β-ratio (the proportion of inflammation’s effect on cognition that is explained by brain structural variable). For example, all brain structural variables accounted for 29.7% of the association between DNAm CRP and general cognitive ability, whereas the same brain structural variables accounted for 41% of the association between DNAm CRP and processing speed. Model 1 = Brain structure variable ~ age + sex + inflammation; Model 2 = Brain structure variable ~ age + sex + inflammation + BMI + hypertension + smoking status + alcohol use + CVD history + diabetes.
<table>
<thead>
<tr>
<th>Mediator variable</th>
<th>Independent variable</th>
<th>Model</th>
<th>( g_f ) mediation (ab)</th>
<th>Visuospatial ability mediation (ab)</th>
<th>Processing speed mediation (ab)</th>
<th>Verbal memory mediation (ab)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td></td>
<td>( \beta )</td>
<td>SE</td>
<td>( p )</td>
<td>( \beta )</td>
<td>SE</td>
</tr>
<tr>
<td>( \Sigma [GM, NAWM, WMH, gFA, gMD] )</td>
<td>serum CRP</td>
<td>model 1</td>
<td>-0.008</td>
<td>0.012</td>
<td>0.650</td>
<td>-0.005</td>
</tr>
<tr>
<td></td>
<td>DNAm CRP</td>
<td>model 2</td>
<td>0.007</td>
<td>0.011</td>
<td>0.715</td>
<td>0.004</td>
</tr>
<tr>
<td></td>
<td></td>
<td>model 1</td>
<td>-0.047</td>
<td>0.015</td>
<td>0.004</td>
<td>-0.036</td>
</tr>
<tr>
<td></td>
<td>DNAm CRP</td>
<td>model 2</td>
<td>0.023</td>
<td>0.013</td>
<td>0.167</td>
<td>-0.020</td>
</tr>
<tr>
<td>GM volume</td>
<td>serum CRP</td>
<td>model 1</td>
<td>-0.002</td>
<td>0.003</td>
<td>0.512</td>
<td>-0.002</td>
</tr>
<tr>
<td></td>
<td>DNAm CRP</td>
<td>model 2</td>
<td>0.001</td>
<td>0.004</td>
<td>0.659</td>
<td>0.002</td>
</tr>
<tr>
<td></td>
<td></td>
<td>model 1</td>
<td>-0.016</td>
<td>0.009</td>
<td>0.091</td>
<td>-0.019</td>
</tr>
<tr>
<td></td>
<td>DNAm CRP</td>
<td>model 2</td>
<td>0.009</td>
<td>0.007</td>
<td>0.232</td>
<td>-0.012</td>
</tr>
<tr>
<td>NAWM volume</td>
<td>serum CRP</td>
<td>model 1</td>
<td>-0.008</td>
<td>0.009</td>
<td>0.391</td>
<td>-0.005</td>
</tr>
<tr>
<td></td>
<td>DNAm CRP</td>
<td>model 2</td>
<td>0.001</td>
<td>0.008</td>
<td>0.881</td>
<td>0.002</td>
</tr>
<tr>
<td></td>
<td></td>
<td>model 1</td>
<td>-0.030</td>
<td>0.012</td>
<td>0.012</td>
<td>-0.024</td>
</tr>
<tr>
<td></td>
<td>DNAm CRP</td>
<td>model 2</td>
<td>0.002</td>
<td>0.007</td>
<td>0.232</td>
<td>-0.012</td>
</tr>
<tr>
<td>WMH volume</td>
<td>serum CRP</td>
<td>model 1</td>
<td>0.002</td>
<td>0.003</td>
<td>0.580</td>
<td>0.001</td>
</tr>
<tr>
<td></td>
<td>DNAm CRP</td>
<td>model 2</td>
<td>0.004</td>
<td>0.004</td>
<td>0.350</td>
<td>0.002</td>
</tr>
<tr>
<td></td>
<td></td>
<td>model 1</td>
<td>-0.002</td>
<td>0.006</td>
<td>0.795</td>
<td>4.68E-04</td>
</tr>
<tr>
<td></td>
<td>DNAm CRP</td>
<td>model 2</td>
<td>0.003</td>
<td>0.004</td>
<td>0.448</td>
<td>-0.001</td>
</tr>
<tr>
<td>gFA volume</td>
<td>serum CRP</td>
<td>model 1</td>
<td>-4.41E-04</td>
<td>0.003</td>
<td>0.883</td>
<td>0.003</td>
</tr>
<tr>
<td></td>
<td>DNAm CRP</td>
<td>model 2</td>
<td>1.58E-06</td>
<td>1.39E-04</td>
<td>0.991</td>
<td>1.38E-04</td>
</tr>
<tr>
<td></td>
<td></td>
<td>model 1</td>
<td>-0.001</td>
<td>0.009</td>
<td>0.866</td>
<td>0.010</td>
</tr>
<tr>
<td></td>
<td>DNAm CRP</td>
<td>model 2</td>
<td>-2.68E-04</td>
<td>0.006</td>
<td>0.965</td>
<td>0.007</td>
</tr>
<tr>
<td>gMD volume</td>
<td>serum CRP</td>
<td>model 1</td>
<td>0.001</td>
<td>0.002</td>
<td>0.712</td>
<td>-0.001</td>
</tr>
<tr>
<td></td>
<td>DNAm CRP</td>
<td>model 2</td>
<td>0.002</td>
<td>0.002</td>
<td>0.720</td>
<td>-0.001</td>
</tr>
<tr>
<td></td>
<td></td>
<td>model 1</td>
<td>0.002</td>
<td>0.006</td>
<td>0.789</td>
<td>-0.004</td>
</tr>
<tr>
<td></td>
<td>DNAm CRP</td>
<td>model 2</td>
<td>0.007</td>
<td>0.007</td>
<td>0.668</td>
<td>-0.003</td>
</tr>
</tbody>
</table>

Note: Model 1 = Brain health variable ~ age + sex + inflammation;
Model 2 = Brain health variable ~ age + sex + inflammation + BMI + hypertension + smoking status + alcohol use + CVD history + diabetes
**Table 14: Tract loadings for general factors of white matter fractional anisotropy and mean diffusivity**

<table>
<thead>
<tr>
<th>White matter tract</th>
<th>PC1 (FA)</th>
<th>PC1 (MD)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Genu of corpus callosum</td>
<td>0.647</td>
<td>0.646</td>
</tr>
<tr>
<td>Splenium of corpus callosum</td>
<td>0.490</td>
<td>0.308</td>
</tr>
<tr>
<td>Left arcuate fasciculus</td>
<td>0.722</td>
<td>0.735</td>
</tr>
<tr>
<td>Right arcuate fasciculus</td>
<td>0.69</td>
<td>0.767</td>
</tr>
<tr>
<td>Left anterior thalamic radiation</td>
<td>0.65</td>
<td>0.72</td>
</tr>
<tr>
<td>Right anterior thalamic radiation</td>
<td>0.634</td>
<td>0.645</td>
</tr>
<tr>
<td>Left rostral cingulum</td>
<td>0.55</td>
<td>0.648</td>
</tr>
<tr>
<td>Right rostral cingulum</td>
<td>0.573</td>
<td>0.752</td>
</tr>
<tr>
<td>Left uncinate fasciculus</td>
<td>0.637</td>
<td>0.665</td>
</tr>
<tr>
<td>Right uncinate fasciculus</td>
<td>0.669</td>
<td>0.744</td>
</tr>
<tr>
<td>Left Inferior longitudinal fasciculus</td>
<td>0.494</td>
<td>0.41</td>
</tr>
<tr>
<td>Right Inferior longitudinal fasciculus</td>
<td>0.477</td>
<td>0.371</td>
</tr>
</tbody>
</table>

**Proportion of variance**

|                      | 0.370    | 0.405    |
eFigure 1. DNAm CRP shows stronger and more widespread associations with regional brain cortical thickness than serum CRP. Regional cortical thickness regressed against serum CRP (i-ii) and DNAm CRP (iii-iv). Colours denote the magnitude (T-maps; top) and significance (Q values; bottom) of the negative associations between inflammation and brain cortical thickness. Panel (v) shows the percentage attenuation for the significant associations between DNAm-CRP and cortical thickness when also controlling for serum CRP. Conjunction plot (vi) shows the spatial extent of independent contributions and overlap (red) in cortical loci that exhibit FDR-corrected unique associations with simultaneously-modelled serum (pink) and epigenetic (blue) inflammation measures; results are corrected for sex, age and ICV.
Supplementary eFigure 2.

Fig. e-2. Associations between serum and epigenetic CRP measures and brain structure and the impact of lifestyle covariates on associations; shapes show standardised regression coefficients for different models (corresponding with table e-6), error bars show standard errors. TB: total brain, GM: grey matter, NAWM: normal-appearing white matter, WMH: white matter hyperintensity, gf, general cognitive ability; gFA: general fractional anisotropy, gMD: general mean diffusivity.
Supplementary eFigure 3.

Fig. e-3. Associations between serum and epigenetic CRP measures and WM tract FA and MD and the impact of lifestyle covariates on associations; shapes show standardised regression coefficients for different models (corresponding with eTable 9 and eTable 10), error bars show standard errors.
Supplementary eFigure 4.

**A**

Brain structure: GM, gMD, WMH, NAWM, DNAm CRP

Chronic inflammation

**B**

Brain structure: GM, gMD, WMH, NAWM, DNAm CRP

Chronic inflammation

**C**

Brain structure: GM, gMD, WMH, NAWM, DNAm CRP

Chronic inflammation

**Fig e.4. Multiple mediator models of the association of inflammation with individual cognitive domains**

**Left** displays structural equation model path diagrams; right displays mediation of individual MRI metrics (indirect effect size on y axis) and standard error bars. Light bars show model 1 (includes covariates age and sex), dark bars show model 2 which contains additional health covariates (age + sex + BMI + hypertension + smoking status + alcohol use + CVD history + diabetes). Asterisks denotes FDR p <0.05. TB: total brain, GM: grey matter, NAWM: normal-appearing white matter, WMH: white matter hyperintensity, gf, general cognitive ability; gFA: general fractional anisotropy, gMD: general mean diffusivity; n = 521
Fig e.5. Mechanisms of neurodegeneration via increased systemic chronic inflammation

Suggested mechanisms by which the causes of inflammaging (immunosenescence, lifestyle, clinical health) and related consequences may drive brain health (structural and cognitive) outcomes.
e-references


11.2 Chapter 7 Supplementary data

An online version of this can be accessed at:
#s0150
Supplemental Document

Immuno-epigenetic signature derived in saliva associates with the encephalopathy of prematurity and perinatal inflammatory disorders

Authors: Eleanor L.S. Conole¹,²,³, Kadi Vaher⁴, Manuel Blesa Cabez⁴, Gemma Sullivan⁴, Anna J. Stevenson², Lee Murphy, Jill Hall⁴, Michael J. Thrippleton³, Mark E. Bastin¹,², Alan J. Quigley⁵, Veronique E. Miron⁴, Heather C. Whalley³, Riccardo E. Marioni², James P. Boardman⁴,³, Simon R. Cox¹*

¹ Lothian Birth Cohorts group, Department of Psychology, University of Edinburgh, Edinburgh EH8 9JZ, UK.

² Centre for Genomic and Experimental Medicine, Institute of Genetics and Cancer, University of Edinburgh, Edinburgh EH4 2XU, UK

³ Centre for Clinical Brain Sciences, University of Edinburgh, Edinburgh EH16 4SB, UK

⁴ MRC Centre for Reproductive Health, Queen’s Medical Research Institute, Edinburgh BioQuarter, University of Edinburgh, Edinburgh, EH16 4TJ

⁵ Imaging Department, Royal Hospital for Children and Young People, Edinburgh, EH16 4TJ

*Correspondence to: Simon. R. Cox
Email: Eleanor. L. S. Conole
Email:

This document includes:
- Supplementary Methods (pages 2-4)
- Supplementary Results (page 5)
- Supplementary Figures 1-10 (pages 6-15)
- Supplementary Tables 1-12 (pages 16-29)
Supplementary methods:

DNA methylation pre-processing

Raw intensity (.idat) files were read into R environment (version 3.4.4) using minfi. watermelon and minfi used for preprocessing, quality control and normalisation (1). The pfilter function in watermelon was used to exclude: samples with 1% of sites with a detection p-value greater than 0.05; sites with beadcount 0.05. Cross hybridising probes and probes targeting single nucleotide polymorphisms with overall minor allele frequency ≥0.05 were also removed. Control probes were also removed. Samples were removed if there was a mismatch between predicted sex (minfi) and recorded sex (n = 3). Probes located on sex chromosomes were removed prior to analysis. Data from one of each twin pair was removed randomly (n=20). Data was datanormalised which includes background correction and dye bias correction (1). Saliva contains different cells types, including buccal epithelial cells and leukocytes. Epithelial cell proportions were estimated with epigenetic dissection of intra-sample heterogeneity with the reduced partial correlation method implemented in the R package EpiDISH (2). Prior to implementation of statistical models, β-values for CpG sites were adjusted (regressed as dependent variables) to remove batch effects using ComBat (3), where each BeadChip was considered to be one batch, and effects of estimated epithelial cell proportions.

dMRI pre-processing

Phase 1 dMRI acquisition were denoised using a Marchenko-Pastur-PCA-based algorithm (4); eddy current distortion and head movement were corrected using outlier replacement (5,6); bias field inhomogeneity correction was performed by calculating the bias field of the mean b0 volume and applying the correction to all the volumes (7). The processing for phase 2 was very similar as the one for phase 1. The main differences are that because the phase 2 consists in two different acquisitions, the data was concatenated before starting the pre-processing, then due to the availability of reverse encoded data, we corrected for EP distortion and within volume movement (5,8)

Tract segmentation and extraction of tract-averaged dMRI metrics

Briefly, the ENA50 neonatal template space (brain atlas optimised for infants) was used to perform whole brain tractography (9) and the SingleTensorFT tool within DTI-TK (10) was used to parse out white matter tractography from the ENA50 atlas tensor volume.
From here, segmentation of white matter tracts was performed within the ENA50 atlas and tracts were delineated by drawing regions of interest (ROIs) manually on the FA image, using the protocols outlined in (11,12). Placement of ROIs is described in Supplementary Table 3 and these were drawn using the Paintbrush mode in ITK-SNAP (13). (http://www.itksnap.org/). The ROIs were used to filter whole brain tractography either to select or to exclude tracts crossing the ROIs using TractTool within DTI-TK. The resulting tract images were binarized and manually refined. T2w processed images were registered to the ENA50 T2w structural template using rigid, affine and symmetric normalization (SyN) implemented in Advanced Normalization Tools (ANTS) (14). The resulting transformation was concatenated with the previously computed transformation from B0 to T2w and used to bring the tract ROIs defined in the ENA50 space to each subject's native space in a single step. The average multi-tissue response function was calculated across the full population using the function dwi2response implemented in MRtrix3 (15), with a FA threshold of 0.1. Then, using the function dwi2fod in MRtrix3 the multi-tissue fibre orientation distribution (FOD) was calculated (16), with the average response function using a spherical harmonic order (Lmax) of 8. Only two (white matter and cerebrospinal fluid) response functions were used. Finally, a joint bias field correction and multi-tissue informed log-domain intensity normalisation on the FODs images was performed using the function mtnormalise in MRtrix3 (17).

The tracts in native space were created using the iFOD2 algorithm using the command tckgen in MRtrix3 (18). The propagated tract ROIs were dilated and the original tract ROIs were used as seed images for the tractography, while the dilated tract ROIs were used as masks to constrain the tracts. The length of the fibres was set with a minimum length of 20 mm and a maximum of 250 mm. Finally, for each tract, a track density image (TDI) map (number of tracts per voxel) was created and normalized between 0 and 1 (19) using the MRtrix3's command tckmap. For each tract, the TDI map was multiplied by each of the DTI and NODDI maps, summed and divided by the average of the TDI map to calculate the weighted tract-averages for FA and MD. We calculated TDI-weighted tract averages to better capture the core of the tracts and reduce bias arising from partial volume effects as highlighted by (20).

Selection criteria of covariates

Our choice of covariates in analyses was informed by variables that could either cause variance in inflammation, aka the DNAm CRP score (the exposure), variance in MRI metrics (the outcome), or both. These are summarised below:
### Covariate | Reasoning
--- | ---
Gestational age at birth (GA) | We controlled for gestational age at birth in all models as we are assessing the relationship between inflammation and brain dysmaturation within a sample of neonates, over half of which are born before 37 weeks gestation with a considerable range in GA (23.28 – 34.84 weeks). Within preterm infants, it has been shown that there is a dose-dependent relationship between GA and EoP features, so it is important to account for GA in our analyses; similarly, the allostatic load of inflammation is anticipated to be higher in gestationally younger infants within this subgroup.

Birthweight (BW_Z) | We control for birthweight via birthweight Z score, which provides an estimate measure of fetal growth restriction; this was calculated according to International Fetal and Newborn Growth Consortium for the 21st Century (INTERGROWTH-21st) standards (21). We control for this as preterm infants with a lower birthweight are more likely to experience white matter injury than those with a higher birthweight, so controlling for birthweight is necessary to prevent this from biasing the findings. We control for these factors in both preterms and the control group (> 37 weeks GA), given that infants might be growth restricted but gestationally older (e.g. term subgroup birthweight range: 2346 – 4670g) and the range of gestational age within both groups.

Gestational age at MRI scan; GA at scan (TEA for preterm infants) | Age at scan for both preterm infants and term infants included a range (full term infants: 39.84 – 47.14 weeks; preterm infants at TEA: 37.70 - 45.14 weeks), where MRI features are dynamic, meaning this is important to account for in both groups. This approach has been adopted previously in this cohort (22–26) and others, such as the ELGAN cohort (27–30).

Inflammatory risk factors (antenatal and postnatal) | We focused on antenatal and neonatal inflammatory risk factors or co-morbidities that exhibit well-established associations with brain growth and neurodevelopmental outcomes, including severe retinopathy of prematurity (ROP) (≥ Stage III), bronchopulmonary dysplasia (BPD), necrotising enterocolitis (NEC), defined as stages two or three according to the modified Bell’s staging for NEC (31) and fetal / postnatal infections e.g., sepsis and histological chorioamnionitis alongside maternal factors linked to increased sustained inflammation or EoP phenotypes (smoking during pregnancy, preeclampsia, maternal age). Inflammatory hits were included as covariates in the models examining brain outcomes as preterm birth has been linked to increased risk for various inflammatory morbidities that may relate to poor neurocognitive outcomes by pathways other than inflammation but which may nonetheless be accompanied by raised inflammation levels in neonates (e.g. for preeclampsia, placental insufficiency is considered one factor for the encephalopathy of prematurity (EoP) phenotype, whereby decreased oxygen and nutrient supply to the fetus could alter brain development; other studies suggest that some of the hormones associated with preeclampsia, such as anti-angiogenic factors, may also interfere with the development of the fetal brain. Similarly, for the postnatal complications, such as BPD and NEC, malnutrition may be a driving factor above and beyond inflammation). Ultimately, by controlling for these various factors, we can better isolate the effects of inflammation on brain health outcomes.

Controlling for these variables safeguards our hypothesis that any observed relationship between DNAm CRP and neuroimaging metrics is not wholly attributable to specific clinical characteristics of the preterms in this cohort, but rather driven by the effects of sustained inflammation within this group.
Supplementary results:

Bivariate analysis of potential covariates with DNAm CRP

Prior to assessing the relationships between DNAm CRP with brain structural outcomes or postnatal inflammatory morbidities, we explored the interrelationships between DNAm CRP and the other variables in our study (supplementary figure 2). Correlations were assessed with Pearson correlations between continuous variables (e.g. association between DNAm CRP scores and gestational age) and point-biserial correlations between continuous variables and binary variables (e.g. DNAm CRP scores with incidence of sepsis). Preterm status and DNAm CRP score were moderately correlated ($r$ (256) = .57, $p > 0.001$), and DNAm CRP scores were inversely associated with gestational age ($r$ (256) = -.62, $p > 0.001$). There were moderately strong correlations with risk of BPD (.5) moderate correlations with incidences of HCA and sepsis (.34-.37) and weak correlations with NEC, ROP and maternal smoking (.2-.28). There were no significant correlations between DNAm CRP and infant sex, birthweight Z score or preeclampsia but these were included in models due to biological significance (see DAG; supplementary figure 1). There were no significant correlations between DNAm CRP and maternal age or gestational diabetes, so these covariates were omitted from regression models.
Supplementary figures:

Supplemental Figure 1. Correlations between study variables (n = 258)

![Figure 1. Correlations between study variables (n = 258)](image-url)
Supplemental Figure 2: Directed acyclic graph (DAG) demonstrating the expected interrelationships between preterm birth, DNAm CRP and brain structure. Covariates are in bold and black font and include maternal and perinatal inflammatory risk factors: smoking in pregnancy, preeclampsia and histologic chorioamnionitis (HCA); postnatal neonatal inflammatory morbidities, including bronchopulmonary dysplasia (BPD), necrotising enterocolitis (NEC), sepsis, and retinopathy of prematurity (ROP) as well as administration of MgSO4 and corticosteroids in pregnancy. The predictor (DNAm CRP score, measured at same time as MRI scan), potential confounders (sex, gestational age at birth, birthweight, time of buccal swab collection), batch variables, cellular heterogeneity and outcome variables (brain structural metrics). Variables that were adjusted for within models have are in bold; maternal age, MgSO4 and corticosteroids were not adjusted for in all models, since they may result in over-adjustment.
Supplemental Figure 3. Flow chart of study sample and data acquisition

**TEBC cohort**

<table>
<thead>
<tr>
<th>Cpg</th>
<th>Beta</th>
</tr>
</thead>
<tbody>
<tr>
<td>cg06690548</td>
<td>-0.0048</td>
</tr>
<tr>
<td>cg10636246</td>
<td>-0.0069</td>
</tr>
<tr>
<td>cg18181703</td>
<td>-0.0053</td>
</tr>
<tr>
<td>cg19821297</td>
<td>-0.0051</td>
</tr>
<tr>
<td>cg25325512</td>
<td>-0.0051</td>
</tr>
</tbody>
</table>

**Theirworld Edinburgh Birth Cohort**

\[ b_1 \text{Cpg}_1 + b_2 \text{Cpg}_2 + \ldots + b_n \text{Cpg}_n \]

**DNAm CRP score**

\[ n = 311 \text{ neonates} \]

**DNAm data**

43 excluded
- 32 failed DNAm pre-processing QC criteria
- 1 congenital abnormality
- 10 twins (20 twins in total; random removal)

**Collected DNAm data:**

\[ n = 258 \text{ neonates} \]

103 Term
155 Preterm

**Complete DNAm data suitable for derivation of composite DNAm CRP score**

**Complete global dMRI data:**

(PsFA, PsMD, PsAD, PsRD)

**Complete MRI volume data:**

(deep GM, cortical GM, brainstem, CSF, cerebellum, hippocampi & amygdalae, WM)

**PHASE 1 (n=93) & PHASE 2 data (n=121)**

**dMRI data**

\[ n = 214 \text{ neonates} \]

87 Term
127 Preterm

**PHASE 2 data (n=121)**

**DTI & NODDI data**

**Global DTI tract data:**

(g factors for FA, MD derived from individual WM tracts and PsNDI)

**Individual DTI tract data:**

(FA and MD for left and right hemispheric tracts of AF, ATR, CCG CST, IFOF, ILF, UNC & CC genu, CC splenium)

\[ n = 121 \text{ neonates} \]

57 Term
64 Preterm
Supplemental Figure 4. Model performance metrics for the association of DNAm CRP with white matter volume

Figure 4. Model performance for white matter volume ~ DNAm CRP + gestational age at birth + gestational age at scan + infant sex + birthweight (Z score)
Supplemental Figure 5. Models controlling for inflammatory risk factors in DNAm-brain relationships

Regression models examining association of DNAm CRP with neuroimaging metrics, controlling for perinatal factors:

1. Baseline models (H_1, H_3) =
   brain volume ~ DNAm CRP + GA at birth + GA at scan + scanner variable + birthweight

2. Fully adjusted models (H_2, H_4) =
   brain volume ~ DNAm CRP + GA at birth + GA at scan + scanner variable + birthweight + maternal age + maternal smoking in pregnancy + preeclampsia + HCA + ROP + NEC + sepsis + BPD

3. Individual risk factors entered in models (in preterm subgroup only) H_5-H_12 =
   brain volume ~ DNAm CRP + GA at birth + GA at scan + scanner variable + birthweight + maternal age | maternal smoking in pregnancy | preeclampsia | HCA | ROP | NEC | sepsis | BPD

Significant associations:
- significance (pFDR < 0.05)
- non significance (pFDR > 0.05)

Supplemental Figure 6. Raw weighted FA and MD within sample
Figure 6. Raw weighted FA and MD values within subsample (n = 121)

Supplemental Figure 7. Derivation of general factors of FA and MD
Figure 7. Derivation of general factors of FA and MD. (A) PCA variable contribution plot; the colours represent the contribution of the dMRI metric to the components (B) Scree plot of the eigenvalues (C) visualisation of gFA and gMD between term (blue circles) and preterm (yellow triangles) on the multimodal principal component axes FA = fractional anisotropy, MD = mean diffusivity.
Figure 8. Breakdown of inflammatory exposures according to multiple-hit categories.
Supplemental info to figure 2, *Multiple inflammatory hits associate with raised DNAm CRP* in main text (A) displays distributions of DNAm CRP according to number of inflammatory episodes experienced by infants within the sample (B) breaks down the contributions to these categories in the form of venn diagrams which illustrate, in preterms, the overlap of postnatal inflammatory morbidities in study sample and in term infants, the antenatal exposures experienced by a small subset of infants (n = 14).
**Supplemental Figure 9.** Interaction effects of DNAm CRP with gestational age and neuroimaging outcomes

**Figure 9.** Interaction effects of DNAm CRP, gestational age and DTI outcomes (A) distributions of DNAm CRP in study sample (n = 258) (B-E) associations of DNAm with aspects of white matter microstructure - PSMD, PSDA, PSNDI and PSRD
Supplemental Figure 10. DTI-tract associations with DNAm CRP

Figure 9. DTI-tract associations with DNAm CRP. Standardized regression coefficients for DNAm CRP associations between tract fractional anisotropy (FA) [left] and mean diffusivity [right] for term (first row) preterm (second row) and preterm infants in models controlling for additionally inflammatory risk factors (third row). Filled circles are left hemispheric tracts and open shapes are right hemispheric tracts, except in the case of the CC where filled shapes are the splenium and open shapes are the genu of the corpus callosum. Points show standardized coefficients and 95% confidence intervals. All models are controlled for sex, gestational age at birth, gestational age at scan and birthweight Z score. Model H2 [second row] additionally controls for inflammatory risk factors.
Supplementary Tables

Supplementary table 1. CpG sites and relative weights

<table>
<thead>
<tr>
<th>CpG</th>
<th>Gene</th>
<th>Beta (discovery sample)</th>
</tr>
</thead>
<tbody>
<tr>
<td>cg06690548</td>
<td>SLC7A11</td>
<td>-0.0048</td>
</tr>
<tr>
<td>cg10636246</td>
<td>AIM 2 &amp; IF116</td>
<td>-0.0069</td>
</tr>
<tr>
<td>cg18181703</td>
<td>SOCS3</td>
<td>-0.0053</td>
</tr>
<tr>
<td>cg19821297</td>
<td>DNASE2</td>
<td>-0.0051</td>
</tr>
<tr>
<td>cg25325512</td>
<td>FGD2</td>
<td>-0.0051</td>
</tr>
<tr>
<td>cg27023587</td>
<td>HEATR6</td>
<td>-0.005</td>
</tr>
<tr>
<td>cg06126421</td>
<td>TUBB</td>
<td>-0.0052</td>
</tr>
</tbody>
</table>

Supplementary Table 1. CpG sites and relative weights (from Lighthart et al. 2016) used to generate DNAm CRP score

Supplementary table 2. Cell-type expression of CpG sites used in composite methylation score

<table>
<thead>
<tr>
<th></th>
<th>macrophage</th>
<th>conventional dendritic cell</th>
<th>memory B cell</th>
<th>CD4-positive, alpha-beta T cell</th>
<th>CD8-positive, alpha-beta T cell</th>
<th>regulatory T cell</th>
<th>erythroblast</th>
<th>mature neutrophil</th>
<th>endothelial cell of umbilical vein (proliferating)</th>
<th>endothelial cell of umbilical vein (resting)</th>
</tr>
</thead>
<tbody>
<tr>
<td>AIM2</td>
<td>59</td>
<td>1</td>
<td>8</td>
<td>5</td>
<td>3</td>
<td>4</td>
<td>3</td>
<td>2</td>
<td>11</td>
<td>14</td>
</tr>
<tr>
<td>DNASE2</td>
<td>3</td>
<td>30</td>
<td>24</td>
<td></td>
<td>2</td>
<td>3</td>
<td>2</td>
<td>3</td>
<td>15</td>
<td>14</td>
</tr>
<tr>
<td>FGD2</td>
<td>5</td>
<td>10</td>
<td>8</td>
<td>4</td>
<td>2</td>
<td>3</td>
<td>4</td>
<td>5</td>
<td>5</td>
<td>5</td>
</tr>
<tr>
<td>HEATR6</td>
<td>16</td>
<td>80</td>
<td></td>
<td>2</td>
<td>5</td>
<td>0.6</td>
<td>21</td>
<td>152</td>
<td>30</td>
<td>41</td>
</tr>
<tr>
<td>IER3</td>
<td>22</td>
<td>20</td>
<td>40</td>
<td>21</td>
<td>18</td>
<td>38</td>
<td>9</td>
<td>16</td>
<td>44</td>
<td>51</td>
</tr>
<tr>
<td>SLC7A11</td>
<td>2</td>
<td>185</td>
<td></td>
<td>51</td>
<td></td>
<td>51</td>
<td>1</td>
<td>62</td>
<td>49</td>
<td>49</td>
</tr>
<tr>
<td>SOCS3</td>
<td>0.7</td>
<td>49</td>
<td>30</td>
<td>57</td>
<td>18</td>
<td>28</td>
<td>3</td>
<td>124</td>
<td>9</td>
<td>12</td>
</tr>
<tr>
<td>TUBB</td>
<td>128</td>
<td>145</td>
<td>34</td>
<td>40</td>
<td>40</td>
<td>49</td>
<td>366</td>
<td>33</td>
<td>311</td>
<td>316</td>
</tr>
</tbody>
</table>

Supplementary Table 2. Peripheral expression of CpG sites within DNAm CRP score in tissue types. Strand-specific RNA-Seq from different cell types from healthy individuals in the BLUEPRINT epigenome project. Light blue box: expression level is low (between 0.5 to 10 FPKM or 0.5 to 10 TPM); Medium blue box: expression level is medium (between 11 to 1000 FPKM or 11 to 1000 TPM); white: there is no data available. This study makes use of data generated by the Blueprint Consortium. A full list of the investigators who contributed to the generation of the data is available from www.blueprint-epigenome.eu.
<table>
<thead>
<tr>
<th>Tract</th>
<th>First ROI</th>
<th>Second ROI</th>
<th>Exclusion ROI</th>
</tr>
</thead>
<tbody>
<tr>
<td>CC genu</td>
<td>A coronal plane at the middle point between the anterior tip of frontal lobe and the anterior edge of the genu of the corpus callosum is selected using the mid-sagittal plane; the intense FA region in the left frontal lobe is drawn as the ROI</td>
<td>A mirror image of the first ROI is drawn in the right hemisphere on the same plane</td>
<td>To exclude fibres projecting posteriorly, an exclusion ROI is drawn encompassing the entire coronal slice at the middle of corpus callosum identified using the mid-sagittal plane</td>
</tr>
<tr>
<td>CC splenium</td>
<td>A coronal plane at the level of the posterior edge of the cerebellum and at the posterior edge of the parietooccipital sulcus is selected using the mid-sagittal plane; the intense FA region in the left occipital lobe is drawn as the ROI</td>
<td>A mirror image of the first ROI is drawn in the right hemisphere on the same plane</td>
<td>N/A</td>
</tr>
<tr>
<td>CST</td>
<td>An axial plane at the level of the decussation of the superior cerebellar peduncle is selected; a ROI is drawn to include the CST</td>
<td>An axial plane at the level of centrum semiovale is selected; a ROI is drawn to include the primary motor cortex</td>
<td>Exclusion ROIs are drawn to exclude all projections to the cerebellum and the other hemisphere</td>
</tr>
<tr>
<td>IFOF</td>
<td>A coronal plane at the middle of the posterior edge of the cingulum and the posterior edge of the parietooccipital sulcus is selected; the ROI is drawn to include the occipital lobe</td>
<td>A coronal plane at the level of the genu of corpus callosum is selected using the mid-sagittal plane; the entire hemisphere is drawn as the ROI</td>
<td>N/A</td>
</tr>
<tr>
<td>ILF</td>
<td>A coronal plane at the posterior edge of the cingulum is selected; the entire hemisphere is drawn as the ROI</td>
<td>The most posterior coronal plane in which the temporal lobe is not connected to the frontal lobe is selected; the ROI includes the entire temporal lobe</td>
<td>Exclusion ROIs are drawn to exclude fibres projecting to the frontal and temporal lobes or to the thalamus</td>
</tr>
<tr>
<td>AF</td>
<td>At the lowest axial level in which the fornix can be identified as a single intense structure, at the middle of the posterior limb of the internal capsule a coronal slice is selected; the ROI is drawn to include the entire frontal lobe</td>
<td>A coronal slice at the middle of the splenium of the corpus callosum is selected using the mid-sagittal plane; ROI is drawn to include the projections along the superior longitudinal fasciculus</td>
<td>N/A</td>
</tr>
<tr>
<td>UNC</td>
<td>The most posterior coronal plane in which the temporal lobe is not connected to the frontal lobe is selected; the ROI includes the entire temporal lobe</td>
<td>At the same coronal plane, the ROI is drawn to include projections to the frontal lobe</td>
<td>Exclusion ROIs are drawn to exclude fibres projecting posteriorly, laterally to the external capsule and to the other hemisphere</td>
</tr>
<tr>
<td>CCG</td>
<td>The coronal plane at the middle of splenium of the corpus callosum is selected using the mid-sagittal slice; the ROI is drawn to include the cingulum</td>
<td>A coronal plane at the middle of the genu of corpus callosum is selected using the mid-sagittal slice; the ROI is drawn to include the cingulum</td>
<td>Exclusion ROI is drawn to exclude all projections to the other hemisphere</td>
</tr>
<tr>
<td>ATR</td>
<td>A coronal slice at the middle of genu of corpus callosum is selected using the mid-sagittal slice; the ROI is drawn to include the anterior limb of the internal capsule</td>
<td>A coronal slice at the anterior edge of the pons is selected using the mid-sagittal slice; the ROI is drawn to include the thalamus</td>
<td>Exclusion ROIs are drawn to exclude projections to the spinal cord, temporal lobe, and posteriorly towards the back of the brain</td>
</tr>
</tbody>
</table>

**Supplementary Table 3. ROI placement for tract segmentation.** CC genu = corpus callosum genu/forceps minor, CC splenium = corpus callosum splenium/forceps major, CST = corticospinal tract, IFOF = inferior fronto-occipital fasciculus, ILF = inferior longitudinal fasciculus, AF = arcuate fasciculus, UNC = uncinate fasciculus, CCG = cingulum cingulate gyrus, ATR = anterior thalamic radiation, ROI = region of interest.
Supplementary Table 4. Association between DNA methylation CRP and inflammatory exposures.

<table>
<thead>
<tr>
<th>outcome</th>
<th>model</th>
<th>OR</th>
<th>Lower CI</th>
<th>Upper CI</th>
<th>p value</th>
</tr>
</thead>
<tbody>
<tr>
<td>preeclampsia</td>
<td>H1</td>
<td>0.62</td>
<td>0.38</td>
<td>0.97</td>
<td>0.042</td>
</tr>
<tr>
<td></td>
<td>H2</td>
<td>0.61</td>
<td>0.37</td>
<td>0.98</td>
<td>0.045</td>
</tr>
<tr>
<td></td>
<td>H3</td>
<td>0.65</td>
<td>0.38</td>
<td>1.08</td>
<td>0.098</td>
</tr>
<tr>
<td></td>
<td>H1</td>
<td>1.22</td>
<td>0.81</td>
<td>1.87</td>
<td>0.344</td>
</tr>
<tr>
<td>smoked in pregnancy</td>
<td>H2</td>
<td>1.87</td>
<td>1.29</td>
<td>2.78</td>
<td>1.29E-03</td>
</tr>
<tr>
<td></td>
<td>H3</td>
<td>2.67</td>
<td>1.72</td>
<td>4.72</td>
<td>1.26E-03</td>
</tr>
<tr>
<td></td>
<td>H1</td>
<td>1.87</td>
<td>1.29</td>
<td>2.78</td>
<td>1.29E-03</td>
</tr>
<tr>
<td>HCA</td>
<td>H2</td>
<td>1.95</td>
<td>1.32</td>
<td>2.97</td>
<td>1.21E-03</td>
</tr>
<tr>
<td></td>
<td>H3</td>
<td>2.00</td>
<td>1.32</td>
<td>3.16</td>
<td>1.83E-03</td>
</tr>
<tr>
<td></td>
<td>H1</td>
<td>1.95</td>
<td>1.32</td>
<td>2.97</td>
<td>1.21E-03</td>
</tr>
<tr>
<td>BPD</td>
<td>H2</td>
<td>3.77</td>
<td>2.36</td>
<td>6.44</td>
<td>1.97E-07</td>
</tr>
<tr>
<td></td>
<td>H3</td>
<td>4.71</td>
<td>2.74</td>
<td>8.87</td>
<td>1.90E-07</td>
</tr>
<tr>
<td></td>
<td>H1</td>
<td>2.10</td>
<td>1.38</td>
<td>3.34</td>
<td>9.16E-04</td>
</tr>
<tr>
<td>Sepsis</td>
<td>H2</td>
<td>2.33</td>
<td>1.50</td>
<td>3.79</td>
<td>3.20E-04</td>
</tr>
<tr>
<td></td>
<td>H3</td>
<td>2.30</td>
<td>1.46</td>
<td>3.81</td>
<td>6.36E-04</td>
</tr>
<tr>
<td></td>
<td>H1</td>
<td>2.33</td>
<td>1.50</td>
<td>3.79</td>
<td>3.20E-04</td>
</tr>
<tr>
<td>NEC</td>
<td>H2</td>
<td>3.35</td>
<td>1.48</td>
<td>9.57</td>
<td>9.67E-03</td>
</tr>
<tr>
<td></td>
<td>H3</td>
<td>3.86</td>
<td>1.58</td>
<td>12.03</td>
<td>8.05E-03</td>
</tr>
<tr>
<td></td>
<td>H1</td>
<td>0.64</td>
<td>0.35</td>
<td>1.20</td>
<td>0.135</td>
</tr>
<tr>
<td>Term control (0 inflammatory episodes)</td>
<td>H2</td>
<td>0.68</td>
<td>0.37</td>
<td>1.20</td>
<td>0.196</td>
</tr>
<tr>
<td></td>
<td>H3</td>
<td>0.67</td>
<td>0.37</td>
<td>1.20</td>
<td>0.199</td>
</tr>
<tr>
<td></td>
<td>H1</td>
<td>0.68</td>
<td>0.37</td>
<td>1.20</td>
<td>0.196</td>
</tr>
<tr>
<td>Preterms (0 inflammatory episode)</td>
<td>H2</td>
<td>0.42</td>
<td>0.27</td>
<td>0.63</td>
<td>4.90E-05</td>
</tr>
<tr>
<td></td>
<td>H3</td>
<td>0.38</td>
<td>0.24</td>
<td>0.59</td>
<td>3.05E-05</td>
</tr>
<tr>
<td></td>
<td>H1</td>
<td>0.42</td>
<td>0.27</td>
<td>0.63</td>
<td>4.90E-05</td>
</tr>
<tr>
<td>Preterms (1 inflammatory episode)</td>
<td>H2</td>
<td>0.91</td>
<td>0.66</td>
<td>1.27</td>
<td>0.586</td>
</tr>
<tr>
<td></td>
<td>H3</td>
<td>0.92</td>
<td>0.65</td>
<td>1.30</td>
<td>0.651</td>
</tr>
<tr>
<td></td>
<td>H1</td>
<td>0.91</td>
<td>0.66</td>
<td>1.27</td>
<td>0.586</td>
</tr>
<tr>
<td>Preterms (2 inflammatory episode)</td>
<td>H2</td>
<td>1.60</td>
<td>1.04</td>
<td>2.53</td>
<td>0.038</td>
</tr>
<tr>
<td></td>
<td>H3</td>
<td>1.73</td>
<td>1.08</td>
<td>2.86</td>
<td>0.026</td>
</tr>
<tr>
<td></td>
<td>H1</td>
<td>1.60</td>
<td>1.04</td>
<td>2.53</td>
<td>0.038</td>
</tr>
<tr>
<td>Preterms (3 inflammatory episode)</td>
<td>H2</td>
<td>2.83</td>
<td>1.65</td>
<td>5.33</td>
<td>4.58E-04</td>
</tr>
<tr>
<td></td>
<td>H3</td>
<td>3.06</td>
<td>1.72</td>
<td>6.05</td>
<td>4.31E-04</td>
</tr>
</tbody>
</table>

Supplementary Table 4. Association between DNA methylation CRP and inflammatory exposures. Results presented as mean (standard deviation) unless specified. CI = confidence interval, OR = odds ratio, model H1 = DNA methylation CRP only; model H2 = adjusted for gestation at sample collection, birthweight Z score, infant sex; model H3 = adjusted for gestation at sample collection, birthweight Z score, infant sex and corticosteroid and MgSO4 administration during pregnancy.
### Supplementary Table 5. Association between DNAm CRP and inflammatory exposure categories

Post-hoc testing (Tukey) of inflammatory exposure categories.

<table>
<thead>
<tr>
<th>group1</th>
<th>group2</th>
<th>estimate</th>
<th>Lower CI</th>
<th>Upper CI</th>
<th>p.adj</th>
<th>p.adj.signif</th>
</tr>
</thead>
<tbody>
<tr>
<td>[Term] 0 hits</td>
<td>[Term] 1 hit</td>
<td>0.000236</td>
<td>-3.50E-04</td>
<td>0.000822</td>
<td>8.56E-01</td>
<td>ns</td>
</tr>
<tr>
<td>[Term] 0 hits</td>
<td>[Preterm] 0 hits</td>
<td>0.000654</td>
<td>2.75E-04</td>
<td>0.001033</td>
<td>1.92E-05</td>
<td>****</td>
</tr>
<tr>
<td>[Term] 0 hits</td>
<td>[Preterm] 1 hit</td>
<td>0.001058</td>
<td>7.17E-04</td>
<td>0.001399</td>
<td>7.59E-14</td>
<td>****</td>
</tr>
<tr>
<td>[Term] 0 hits</td>
<td>[Preterm] 2 hits</td>
<td>0.001381</td>
<td>9.45E-04</td>
<td>0.001817</td>
<td>7.31E-14</td>
<td>****</td>
</tr>
<tr>
<td>[Term] 0 hits</td>
<td>[Preterm] 3 or more hits</td>
<td>0.001689</td>
<td>1.22E-03</td>
<td>0.002158</td>
<td>6.95E-14</td>
<td>****</td>
</tr>
<tr>
<td>[Term] 1 hit</td>
<td>[Preterm] 0 hits</td>
<td>0.000418</td>
<td>-2.09E-04</td>
<td>0.001044</td>
<td>3.95E-01</td>
<td>ns</td>
</tr>
<tr>
<td>[Term] 1 hit</td>
<td>[Preterm] 1 hit</td>
<td>0.000822</td>
<td>2.17E-04</td>
<td>0.001426</td>
<td>1.67E-03</td>
<td>**</td>
</tr>
<tr>
<td>[Term] 1 hit</td>
<td>[Preterm] 2 hits</td>
<td>0.001145</td>
<td>4.82E-04</td>
<td>0.001807</td>
<td>1.89E-05</td>
<td>****</td>
</tr>
<tr>
<td>[Term] 1 hit</td>
<td>[Preterm] 3 or more hits</td>
<td>0.001453</td>
<td>7.69E-04</td>
<td>0.002138</td>
<td>6.06E-08</td>
<td>****</td>
</tr>
<tr>
<td>[Preterm] 0 hits</td>
<td>[Preterm] 1 hit</td>
<td>0.000404</td>
<td>-2.88E-06</td>
<td>0.000811</td>
<td>5.29E-02</td>
<td>ns</td>
</tr>
<tr>
<td>[Preterm] 0 hits</td>
<td>[Preterm] 2 hits</td>
<td>0.000727</td>
<td>2.38E-04</td>
<td>0.001216</td>
<td>3.99E-04</td>
<td>***</td>
</tr>
<tr>
<td>[Preterm] 0 hits</td>
<td>[Preterm] 3 or more hits</td>
<td>0.001035</td>
<td>5.17E-04</td>
<td>0.001554</td>
<td>4.20E-07</td>
<td>****</td>
</tr>
<tr>
<td>[Preterm] 1 hit</td>
<td>[Preterm] 2 hits</td>
<td>0.000323</td>
<td>-1.37E-04</td>
<td>0.000783</td>
<td>3.37E-01</td>
<td>ns</td>
</tr>
<tr>
<td>[Preterm] 1 hit</td>
<td>[Preterm] 3 or more hits</td>
<td>0.000632</td>
<td>1.40E-04</td>
<td>0.001123</td>
<td>3.71E-03</td>
<td>**</td>
</tr>
<tr>
<td>[Preterm] 2 hits</td>
<td>[Preterm] 3 or more hits</td>
<td>0.000309</td>
<td>-2.53E-04</td>
<td>0.00087</td>
<td>6.14E-01</td>
<td>ns</td>
</tr>
</tbody>
</table>
Supplementary Table 6. Associations between DNAm CRP and brain structure. Standardized betas (β) and P-values are reported from regression models where DNAm CRP is regressed onto MRI measures for all infants (first column, n=214), preterm infants only (second column, n=127), and term infants only (first column, n=87), covarying for gestational age, sex, birthweight Z score, gestational age at scan, scanner variable (volumetric data are also corrected for head size). Additional R2 refers to the amount of variance in MRI measures accounted for DNAm CRP, beyond covariates. Bold text denotes FDR q-value <0.05.

<table>
<thead>
<tr>
<th>Term &amp; Preterm infants (n=214)</th>
<th>Preterm infants (n=127)</th>
<th>Term infants (n=87)</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>b</td>
<td>Upper CI</td>
</tr>
<tr>
<td>Cortical_grey_matter</td>
<td>-0.102</td>
<td>-0.047</td>
</tr>
<tr>
<td>Deep_grey_matter</td>
<td>-0.143</td>
<td>-0.088</td>
</tr>
<tr>
<td>White_matter</td>
<td>-0.219</td>
<td>-0.150</td>
</tr>
<tr>
<td>Hippocampi_and_Amygdala</td>
<td>-0.133</td>
<td>-0.064</td>
</tr>
<tr>
<td>Cerebellum</td>
<td>-0.100</td>
<td>-0.042</td>
</tr>
<tr>
<td>CSF</td>
<td>-0.018</td>
<td>0.043</td>
</tr>
<tr>
<td>Ventricles</td>
<td>0.127</td>
<td>0.211</td>
</tr>
<tr>
<td>Brainstem</td>
<td>0.001</td>
<td>0.068</td>
</tr>
<tr>
<td>PSFA</td>
<td>-0.236</td>
<td>-0.168</td>
</tr>
<tr>
<td>PSMD</td>
<td>0.130</td>
<td>0.193</td>
</tr>
<tr>
<td>PSAD</td>
<td>0.062</td>
<td>0.136</td>
</tr>
<tr>
<td>PSRD</td>
<td>0.094</td>
<td>0.161</td>
</tr>
</tbody>
</table>
### Supplementary Table 7. Interaction effects of gestational age on DNAm CRP in full cohort (n=214)

<table>
<thead>
<tr>
<th></th>
<th>DNAm CRP b</th>
<th>Upper CI</th>
<th>Lower CI</th>
<th>p</th>
<th>DNAm_CRP x gestational age b</th>
<th>Upper CI</th>
<th>Lower CI</th>
<th>p</th>
<th>r^2</th>
<th>add r^2</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cortical_grey_matter</td>
<td>-0.067</td>
<td>-0.006</td>
<td>-0.128</td>
<td>0.273</td>
<td>0.080</td>
<td>0.140</td>
<td>0.021</td>
<td>0.180</td>
<td>0.639</td>
<td>0.009</td>
</tr>
<tr>
<td>Deep_grey_matter</td>
<td>-0.071</td>
<td>-0.010</td>
<td>-0.131</td>
<td>0.242</td>
<td><strong>0.163</strong></td>
<td>0.222</td>
<td>0.104</td>
<td><strong>0.006</strong></td>
<td>0.645</td>
<td>0.025</td>
</tr>
<tr>
<td>White_matter</td>
<td><strong>-0.152</strong></td>
<td>-0.076</td>
<td>-0.227</td>
<td><strong>0.046</strong></td>
<td><strong>0.152</strong></td>
<td>0.226</td>
<td>0.078</td>
<td><strong>0.041</strong></td>
<td>0.444</td>
<td>0.039</td>
</tr>
<tr>
<td>Hippocampi_and_Amygdala</td>
<td>-0.082</td>
<td>-0.005</td>
<td>-0.158</td>
<td>0.285</td>
<td>0.115</td>
<td>0.190</td>
<td>0.041</td>
<td>0.123</td>
<td>0.435</td>
<td>0.017</td>
</tr>
<tr>
<td>Cerebellum</td>
<td>-0.054</td>
<td>0.010</td>
<td>-0.117</td>
<td>0.401</td>
<td>0.105</td>
<td>0.168</td>
<td>0.043</td>
<td>0.094</td>
<td>0.604</td>
<td>0.011</td>
</tr>
<tr>
<td>CSF</td>
<td>-0.026</td>
<td>0.042</td>
<td>-0.094</td>
<td>0.698</td>
<td>-0.019</td>
<td>0.047</td>
<td>-0.086</td>
<td>0.774</td>
<td>0.551</td>
<td>0.000</td>
</tr>
<tr>
<td>Ventricles</td>
<td>0.099</td>
<td>0.192</td>
<td>0.006</td>
<td>0.290</td>
<td>-0.064</td>
<td>0.028</td>
<td>-0.155</td>
<td>0.485</td>
<td>0.151</td>
<td>0.111</td>
</tr>
<tr>
<td>Brainstem</td>
<td>0.055</td>
<td>0.130</td>
<td>-0.019</td>
<td>0.457</td>
<td>0.123</td>
<td>0.196</td>
<td>0.050</td>
<td>0.093</td>
<td>0.460</td>
<td>0.007</td>
</tr>
<tr>
<td>PSFA</td>
<td>-0.111</td>
<td>-0.038</td>
<td>-0.183</td>
<td>0.127</td>
<td><strong>0.283</strong></td>
<td>0.354</td>
<td>0.212</td>
<td><strong>8.88E-05</strong></td>
<td>0.491</td>
<td>0.072</td>
</tr>
<tr>
<td>PSMD</td>
<td><strong>0.131</strong></td>
<td>0.201</td>
<td>0.061</td>
<td><strong>0.064</strong></td>
<td>0.002</td>
<td>0.070</td>
<td>-0.067</td>
<td>0.981</td>
<td>0.522</td>
<td>0.010</td>
</tr>
<tr>
<td>PSAD</td>
<td>-0.047</td>
<td>0.033</td>
<td>-0.128</td>
<td>0.555</td>
<td><strong>-0.248</strong></td>
<td>-0.170</td>
<td>-0.327</td>
<td><strong>0.002</strong></td>
<td>0.374</td>
<td>0.033</td>
</tr>
<tr>
<td>PSRD</td>
<td>0.144</td>
<td>0.218</td>
<td>0.070</td>
<td>0.053</td>
<td>0.113</td>
<td>0.185</td>
<td>0.040</td>
<td>0.122</td>
<td>0.464</td>
<td>0.011</td>
</tr>
</tbody>
</table>

**Supplementary Table 7. Interaction effects of gestational age and DNAm CRP with brain structure; associations between DNAm CRP on global brain MRI parameters.** Standardized betas (β) and P-values are reported from regression models where DNAm CRP is regressed onto MRI measures, covarying for gestational age, sex, birthweight Z score, gestational age at scan, scanner variable (volumetric data are also corrected for head size). Additional R2 refers to the amount of variance in MRI measures accounted for DNAm CRP, beyond covariates. Bold text denotes FDR q-value <0.05.
<table>
<thead>
<tr>
<th>Brain Structure</th>
<th>b</th>
<th>Lower CI</th>
<th>Upper CI</th>
<th>p</th>
<th>r²</th>
<th>add r²</th>
<th>Lower CI</th>
<th>Upper CI</th>
<th>p</th>
<th>r²</th>
<th>add r²</th>
</tr>
</thead>
<tbody>
<tr>
<td>Cortical_grey_matter</td>
<td>-0.121</td>
<td>-0.184</td>
<td>-0.058</td>
<td>0.058</td>
<td>0.641</td>
<td>0.007</td>
<td>0.055</td>
<td>-0.031</td>
<td>0.142</td>
<td>0.523</td>
<td>0.641</td>
</tr>
<tr>
<td>Deep_grey_matter</td>
<td>-0.208</td>
<td>-0.271</td>
<td>-0.144</td>
<td><strong>0.001</strong></td>
<td>0.637</td>
<td>0.019</td>
<td>0.127</td>
<td>0.041</td>
<td>0.214</td>
<td>0.144</td>
<td>0.637</td>
</tr>
<tr>
<td>White_matter</td>
<td>-0.268</td>
<td>-0.348</td>
<td>-0.188</td>
<td><strong>0.001</strong></td>
<td>0.428</td>
<td>0.031</td>
<td>0.166</td>
<td>0.057</td>
<td>0.275</td>
<td>0.129</td>
<td>0.428</td>
</tr>
<tr>
<td>Hippocampi_and_Amygdala</td>
<td>-0.206</td>
<td>-0.286</td>
<td>-0.125</td>
<td><strong>0.011</strong></td>
<td>0.421</td>
<td>0.018</td>
<td>0.164</td>
<td>0.054</td>
<td>0.273</td>
<td>0.137</td>
<td>0.421</td>
</tr>
<tr>
<td>Cerebellum</td>
<td>-0.177</td>
<td>-0.244</td>
<td>-0.111</td>
<td><strong>0.008</strong></td>
<td>0.605</td>
<td>0.014</td>
<td>0.126</td>
<td>0.035</td>
<td>0.216</td>
<td>0.166</td>
<td>0.605</td>
</tr>
<tr>
<td>CSF</td>
<td>-0.047</td>
<td>-0.117</td>
<td>0.024</td>
<td>0.508</td>
<td>0.554</td>
<td>0.001</td>
<td>0.029</td>
<td>-0.067</td>
<td>0.125</td>
<td>0.764</td>
<td>0.554</td>
</tr>
<tr>
<td>Ventricles</td>
<td>0.051</td>
<td>-0.046</td>
<td>0.148</td>
<td>0.599</td>
<td>0.153</td>
<td>0.145</td>
<td>0.008</td>
<td>0.107</td>
<td>-0.026</td>
<td>0.239</td>
<td>0.421</td>
</tr>
<tr>
<td>Brainstem</td>
<td>-0.071</td>
<td>-0.150</td>
<td>0.008</td>
<td>0.370</td>
<td>0.440</td>
<td>0.002</td>
<td>0.054</td>
<td>-0.054</td>
<td>0.162</td>
<td>0.616</td>
<td>0.440</td>
</tr>
<tr>
<td>PSFA</td>
<td>-0.166</td>
<td>-0.243</td>
<td>-0.088</td>
<td><strong>0.034</strong></td>
<td>0.458</td>
<td>0.041</td>
<td>-0.177</td>
<td>-0.283</td>
<td>-0.071</td>
<td>0.097</td>
<td>0.458</td>
</tr>
<tr>
<td>PSMD</td>
<td><strong>0.211</strong></td>
<td>0.139</td>
<td>0.283</td>
<td><strong>0.004</strong></td>
<td>0.535</td>
<td>0.020</td>
<td>-0.103</td>
<td>-0.201</td>
<td>-0.005</td>
<td>0.295</td>
<td>0.535</td>
</tr>
<tr>
<td>PSAD</td>
<td>0.120</td>
<td>0.034</td>
<td>0.205</td>
<td>0.163</td>
<td>0.346</td>
<td>0.339</td>
<td>0.007</td>
<td>-0.056</td>
<td>-0.173</td>
<td>0.060</td>
<td>0.629</td>
</tr>
<tr>
<td>PSRD</td>
<td><strong>0.189</strong></td>
<td>0.112</td>
<td>0.266</td>
<td><strong>0.015</strong></td>
<td>0.469</td>
<td>0.016</td>
<td>-0.148</td>
<td>-0.253</td>
<td>-0.043</td>
<td>0.161</td>
<td>0.469</td>
</tr>
</tbody>
</table>

Supplementary Table 8. Interaction effects of infant sex and DNAm CRP with brain structure (n= 214, term and preterm infants); associations between DNAm CRP on global brain MRI parameters. Standardized betas (β) and P-values are reported from regression models where DNAm CRP is regressed onto MRI measures, covarying for gestational age, sex, birthweight Z score, gestational age at scan, scanner variable (volumetric data are also corrected for head size). Additional R² refers to the amount of variance in MRI measures accounted for DNAm CRP, beyond covariates. Bold text denotes FDR q-value <0.05.
**Supplementary Table 9. Interaction effects of gestational age on DNAm CRP in preterm subgroup cohort (n=127)**

<table>
<thead>
<tr>
<th>Brain Structure</th>
<th>DNAm CRP</th>
<th>DNAm CRP x gestational age</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>b</td>
<td>Lower CI</td>
</tr>
<tr>
<td>Cortical_grey_matter</td>
<td>-0.134</td>
<td>-0.204</td>
</tr>
<tr>
<td>Deep_grey_matter</td>
<td>-0.217</td>
<td>-0.284</td>
</tr>
<tr>
<td>White_matter</td>
<td>-0.244</td>
<td>-0.330</td>
</tr>
<tr>
<td>Hippocampi_and_Amygdala</td>
<td>-0.206</td>
<td>-0.286</td>
</tr>
<tr>
<td>Cerebellum</td>
<td>-0.180</td>
<td>-0.250</td>
</tr>
<tr>
<td>CSF</td>
<td>-0.023</td>
<td>-0.092</td>
</tr>
<tr>
<td>Ventricles</td>
<td>0.073</td>
<td>-0.027</td>
</tr>
<tr>
<td>Brainstem</td>
<td>-0.097</td>
<td>-0.186</td>
</tr>
<tr>
<td>PSFA</td>
<td>-0.212</td>
<td>-0.281</td>
</tr>
<tr>
<td>PSMD</td>
<td>0.371</td>
<td>0.283</td>
</tr>
<tr>
<td>PSAD</td>
<td>0.232</td>
<td>0.147</td>
</tr>
<tr>
<td>PSRD</td>
<td>0.332</td>
<td>0.236</td>
</tr>
</tbody>
</table>

**Supplementary Table 9. Interaction effects of infant sex and DNAm CRP with brain structure (n = 127, preterm infants); associations between DNAm CRP on global brain MRI parameters.** Standardized betas (β) and P-values are reported from regression models where DNAm CRP is regressed onto MRI measures, covarying for gestational age, sex, birthweight Z score, gestational age at scan, scanner variable (volumetric data are also corrected for head size). Additional R2 refers to the amount of variance in MRI measures accounted for DNAm CRP, beyond covariates. Bold text denotes FDR q-value <0.05.
<table>
<thead>
<tr>
<th>Brain Structure</th>
<th>DNAm CRP</th>
<th>DNAm CRP x gestational age</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>b</td>
<td>Lower CI</td>
</tr>
<tr>
<td>Cortical grey matter</td>
<td>-0.033</td>
<td>-0.119</td>
</tr>
<tr>
<td>Deep grey matter</td>
<td>0.008</td>
<td>-0.088</td>
</tr>
<tr>
<td>White matter</td>
<td>-0.058</td>
<td>-0.159</td>
</tr>
<tr>
<td>Hippocampi and Amygdala</td>
<td>0.092</td>
<td>-0.011</td>
</tr>
<tr>
<td>Cerebellum</td>
<td>-0.031</td>
<td>-0.112</td>
</tr>
<tr>
<td>CSF</td>
<td>-0.125</td>
<td>-0.203</td>
</tr>
<tr>
<td>Ventrilces</td>
<td>-0.103</td>
<td>-0.215</td>
</tr>
<tr>
<td>Brainstem</td>
<td>-0.009</td>
<td>-0.106</td>
</tr>
<tr>
<td>PSFA</td>
<td>-0.093</td>
<td>-0.180</td>
</tr>
<tr>
<td>PSMD</td>
<td>-0.099</td>
<td>-0.212</td>
</tr>
<tr>
<td>PSAD</td>
<td>-0.097</td>
<td>-0.106</td>
</tr>
<tr>
<td>PSRD</td>
<td>-0.067</td>
<td>-0.179</td>
</tr>
</tbody>
</table>

**Supplementary Table 10.** Interaction effects of gestational age at birth and DNAm CRP with brain structure (n = 127, preterm infants); associations between DNAm CRP on global brain MRI parameters. Standardized betas (β) and P-values are reported from regression models where DNAm CRP is regressed onto MRI measures, covarying for gestational age, sex, birthweight Z score, gestational age at scan, scanner variable (volumetric data are also corrected for head size). Additional R2 refers to the amount of variance in MRI measures accounted for DNAm CRP, beyond covariates. Bold text denotes FDR q-value <0.05.
<table>
<thead>
<tr>
<th>model</th>
<th>neuroimaging metric</th>
<th>beta</th>
<th>SE</th>
<th>pFDR</th>
<th>r²</th>
</tr>
</thead>
<tbody>
<tr>
<td><strong>H₁ (term infants baseline model)</strong></td>
<td>Cortical grey matter</td>
<td>0.00618471</td>
<td>0.0863116</td>
<td>0.94305486</td>
<td>0.48887307</td>
</tr>
<tr>
<td></td>
<td>Deep grey matter</td>
<td>0.0321895</td>
<td>0.09439479</td>
<td>0.7339096</td>
<td>0.38865478</td>
</tr>
<tr>
<td></td>
<td>White matter</td>
<td>-0.0788723</td>
<td>0.0995926</td>
<td>0.4307631</td>
<td>0.31938319</td>
</tr>
<tr>
<td></td>
<td>Hippocampi and Amygdalae</td>
<td>0.13971501</td>
<td>0.10227963</td>
<td>0.17576367</td>
<td>0.2822574</td>
</tr>
<tr>
<td></td>
<td>Cerebellum</td>
<td>0.03497887</td>
<td>0.07957274</td>
<td>0.66142395</td>
<td>0.5565707</td>
</tr>
<tr>
<td></td>
<td>CSF</td>
<td>-0.0543354</td>
<td>0.07863558</td>
<td>0.49158016</td>
<td>0.57574343</td>
</tr>
<tr>
<td></td>
<td>Ventrices</td>
<td>-0.0937758</td>
<td>0.11201194</td>
<td>0.40497714</td>
<td>0.13916667</td>
</tr>
<tr>
<td></td>
<td>Brainstem</td>
<td>0.09886336</td>
<td>0.09618272</td>
<td>0.30710868</td>
<td>0.36527654</td>
</tr>
<tr>
<td><strong>H₂ (term infants fully adjusted model)</strong></td>
<td>Cortical grey matter</td>
<td>0.0144616</td>
<td>0.08980972</td>
<td>0.87249016</td>
<td>0.49315628</td>
</tr>
<tr>
<td></td>
<td>Deep grey matter</td>
<td>0.04421015</td>
<td>0.09847791</td>
<td>0.6547236</td>
<td>0.39059632</td>
</tr>
<tr>
<td></td>
<td>White matter</td>
<td>-0.0587154</td>
<td>0.10349359</td>
<td>0.57211703</td>
<td>0.32693919</td>
</tr>
<tr>
<td></td>
<td>Hippocampi and Amygdalae</td>
<td>0.16750002</td>
<td>0.10609542</td>
<td>0.1184365</td>
<td>0.2822574</td>
</tr>
<tr>
<td></td>
<td>Cerebellum</td>
<td>0.02489204</td>
<td>0.0830011</td>
<td>0.76505111</td>
<td>0.56709256</td>
</tr>
<tr>
<td></td>
<td>CSF</td>
<td>-0.0509749</td>
<td>0.07891948</td>
<td>0.52023311</td>
<td>0.60862252</td>
</tr>
<tr>
<td></td>
<td>Ventrices</td>
<td>-0.0826478</td>
<td>0.1135196</td>
<td>0.46735334</td>
<td>0.19356069</td>
</tr>
<tr>
<td></td>
<td>Brainstem</td>
<td>0.11039329</td>
<td>0.09925249</td>
<td>0.26944829</td>
<td>0.38097211</td>
</tr>
<tr>
<td><strong>H₃ (preterm infants baseline model)</strong></td>
<td>Cortical grey matter</td>
<td>-0.1386345</td>
<td>0.0683071</td>
<td>0.04461144</td>
<td>0.55556875</td>
</tr>
<tr>
<td></td>
<td>Deep grey matter</td>
<td>-0.1919936</td>
<td>0.06686441</td>
<td>0.00483272</td>
<td>0.57414387</td>
</tr>
<tr>
<td></td>
<td>White matter</td>
<td>-0.2472808</td>
<td>0.08513538</td>
<td>0.00437939</td>
<td>0.30961239</td>
</tr>
<tr>
<td></td>
<td>Hippocampi and Amygdalae</td>
<td>-0.1985864</td>
<td>0.07939704</td>
<td>0.10372603</td>
<td>0.39954359</td>
</tr>
<tr>
<td></td>
<td>Cerebellum</td>
<td>-0.1337108</td>
<td>0.07253394</td>
<td>0.06773452</td>
<td>0.49886425</td>
</tr>
<tr>
<td></td>
<td>CSF</td>
<td>0.00201049</td>
<td>0.06937298</td>
<td>0.87692797</td>
<td>0.54195047</td>
</tr>
<tr>
<td></td>
<td>Ventrices</td>
<td>0.1061455</td>
<td>0.09730149</td>
<td>0.27750564</td>
<td>0.09819657</td>
</tr>
<tr>
<td></td>
<td>Brainstem</td>
<td>-0.0480842</td>
<td>0.09041304</td>
<td>0.59582805</td>
<td>0.22136305</td>
</tr>
<tr>
<td><strong>H₄ (preterm infants fully adjusted model)</strong></td>
<td>Cortical grey matter</td>
<td>-0.1733511</td>
<td>0.07823228</td>
<td>0.00461144</td>
<td>0.55556875</td>
</tr>
<tr>
<td></td>
<td>Deep grey matter</td>
<td>-0.2056703</td>
<td>0.07653611</td>
<td>0.00483272</td>
<td>0.57414387</td>
</tr>
<tr>
<td></td>
<td>White matter</td>
<td>-0.3458994</td>
<td>0.09797648</td>
<td>0.00058828</td>
<td>0.3682006</td>
</tr>
<tr>
<td></td>
<td>Hippocampi and Amygdalae</td>
<td>-0.1591075</td>
<td>0.09400796</td>
<td>0.09330818</td>
<td>0.41620281</td>
</tr>
<tr>
<td></td>
<td>Cerebellum</td>
<td>-0.2006698</td>
<td>0.07927367</td>
<td>0.012737</td>
<td>0.58486369</td>
</tr>
<tr>
<td></td>
<td>CSF</td>
<td>0.01221838</td>
<td>0.07942535</td>
<td>0.87801164</td>
<td>0.58329272</td>
</tr>
<tr>
<td></td>
<td>Ventrices</td>
<td>0.01833798</td>
<td>0.11127664</td>
<td>0.86939872</td>
<td>0.18202366</td>
</tr>
<tr>
<td></td>
<td>Brainstem</td>
<td>-0.0844561</td>
<td>0.10451606</td>
<td>0.42074928</td>
<td>0.27839626</td>
</tr>
<tr>
<td><strong>H₅ (preterm infants maternal age model)</strong></td>
<td>Cortical grey matter</td>
<td>-0.1351839</td>
<td>0.06932893</td>
<td>0.00535407</td>
<td>0.55598835</td>
</tr>
<tr>
<td></td>
<td>Deep grey matter</td>
<td>-0.1964085</td>
<td>0.06781494</td>
<td>0.00451092</td>
<td>0.57483076</td>
</tr>
<tr>
<td></td>
<td>White matter</td>
<td>-0.2509687</td>
<td>0.08641975</td>
<td>0.004392</td>
<td>0.3109169</td>
</tr>
<tr>
<td></td>
<td>Hippocampi and Amygdalae</td>
<td>-0.203742</td>
<td>0.08055993</td>
<td>0.01274454</td>
<td>0.40048033</td>
</tr>
<tr>
<td></td>
<td>Cerebellum</td>
<td>-0.1379828</td>
<td>0.0736065</td>
<td>0.06329942</td>
<td>0.49950741</td>
</tr>
<tr>
<td></td>
<td>CSF</td>
<td>-0.0026771</td>
<td>0.0703845</td>
<td>0.96972358</td>
<td>0.54236845</td>
</tr>
<tr>
<td></td>
<td>Ventrices</td>
<td>0.13389897</td>
<td>0.09730531</td>
<td>0.17138397</td>
<td>0.12534167</td>
</tr>
<tr>
<td></td>
<td>Brainstem</td>
<td>-0.0619962</td>
<td>0.09140591</td>
<td>0.49892929</td>
<td>0.2281838</td>
</tr>
<tr>
<td></td>
<td>Cortical grey matter</td>
<td>-0.1386899</td>
<td>0.06859866</td>
<td>0.04544441</td>
<td>0.55558224</td>
</tr>
<tr>
<td>Deep grey matter</td>
<td>-0.1909589</td>
<td>0.06677905</td>
<td>0.00501113</td>
<td>0.5788463</td>
<td></td>
</tr>
<tr>
<td>------------------</td>
<td>------------</td>
<td>------------</td>
<td>------------</td>
<td>------------</td>
<td></td>
</tr>
<tr>
<td>White matter</td>
<td>-0.2471562</td>
<td>0.08549585</td>
<td>0.00456802</td>
<td>0.30968054</td>
<td></td>
</tr>
<tr>
<td>Hippocampi and Amygdalae</td>
<td>-0.197883</td>
<td>0.07959275</td>
<td>0.01429926</td>
<td>0.40171632</td>
<td></td>
</tr>
<tr>
<td>Cerebellum</td>
<td>-0.1335772</td>
<td>0.07283895</td>
<td>0.06917194</td>
<td>0.49894267</td>
<td></td>
</tr>
<tr>
<td>CSF</td>
<td>0.0026117</td>
<td>0.06954942</td>
<td>0.97010799</td>
<td>0.53417783</td>
<td></td>
</tr>
<tr>
<td>Ventricles</td>
<td>0.10613826</td>
<td>0.09771829</td>
<td>0.27960095</td>
<td>0.09819696</td>
<td></td>
</tr>
<tr>
<td>Brainstem</td>
<td>-0.0470805</td>
<td>0.09054199</td>
<td>0.60403995</td>
<td>0.22578535</td>
<td></td>
</tr>
<tr>
<td>Cortical grey matter</td>
<td>-0.1359852</td>
<td>0.06822376</td>
<td>0.04852478</td>
<td>0.56081374</td>
<td></td>
</tr>
<tr>
<td>White matter</td>
<td>-0.189846</td>
<td>0.06690801</td>
<td>0.00534808</td>
<td>0.57759047</td>
<td></td>
</tr>
<tr>
<td>Hippocampi and Amygdalae</td>
<td>-0.1972531</td>
<td>0.07968399</td>
<td>0.01471474</td>
<td>0.40087193</td>
<td></td>
</tr>
<tr>
<td>Cerebellum</td>
<td>-0.1312811</td>
<td>0.07255528</td>
<td>0.07291396</td>
<td>0.5032755</td>
<td></td>
</tr>
<tr>
<td>CSF</td>
<td>0.00379633</td>
<td>0.06951943</td>
<td>0.95654219</td>
<td>0.54397364</td>
<td></td>
</tr>
<tr>
<td>Ventricles</td>
<td>0.10245412</td>
<td>0.09720787</td>
<td>0.2940322</td>
<td>0.10837887</td>
<td></td>
</tr>
<tr>
<td>Brainstem</td>
<td>-0.0427015</td>
<td>0.08956847</td>
<td>0.6341676</td>
<td>0.24301401</td>
<td></td>
</tr>
<tr>
<td>Cortical grey matter</td>
<td>-0.1393088</td>
<td>0.06857403</td>
<td>0.04443077</td>
<td>0.55611458</td>
<td></td>
</tr>
<tr>
<td>Deep grey matter</td>
<td>-0.190779</td>
<td>0.06702712</td>
<td>0.00521059</td>
<td>0.57591521</td>
<td></td>
</tr>
<tr>
<td>White matter</td>
<td>-0.2480269</td>
<td>0.0854792</td>
<td>0.00442432</td>
<td>0.31028063</td>
<td></td>
</tr>
<tr>
<td>Hippocampi and Amygdalae</td>
<td>-0.1964021</td>
<td>0.07937499</td>
<td>0.01475721</td>
<td>0.40527132</td>
<td></td>
</tr>
<tr>
<td>Cerebellum</td>
<td>-0.1320997</td>
<td>0.07263524</td>
<td>0.07157625</td>
<td>0.5019805</td>
<td></td>
</tr>
<tr>
<td>CSF</td>
<td>0.00594989</td>
<td>0.06825603</td>
<td>0.93068276</td>
<td>0.5602219</td>
<td></td>
</tr>
<tr>
<td>Ventricles</td>
<td>0.10826447</td>
<td>0.09744919</td>
<td>0.26881504</td>
<td>0.10358719</td>
<td></td>
</tr>
<tr>
<td>Brainstem</td>
<td>-0.0434888</td>
<td>0.08933123</td>
<td>0.62777278</td>
<td>0.24671699</td>
<td></td>
</tr>
<tr>
<td>Cortical grey matter</td>
<td>-0.1171</td>
<td>0.07294198</td>
<td>0.11105947</td>
<td>0.55824282</td>
<td></td>
</tr>
<tr>
<td>Deep grey matter</td>
<td>-0.1484135</td>
<td>0.0706903</td>
<td>0.03788694</td>
<td>0.58509553</td>
<td></td>
</tr>
<tr>
<td>White matter</td>
<td>-0.2322611</td>
<td>0.09110091</td>
<td>0.01205888</td>
<td>0.31091323</td>
<td></td>
</tr>
<tr>
<td>Hippocampi and Amygdalae</td>
<td>-0.1640755</td>
<td>0.08455291</td>
<td>0.05468345</td>
<td>0.4061136</td>
<td></td>
</tr>
<tr>
<td>Cerebellum</td>
<td>-0.1463296</td>
<td>0.07761849</td>
<td>0.06183887</td>
<td>0.49978246</td>
<td></td>
</tr>
<tr>
<td>CSF</td>
<td>0.01208732</td>
<td>0.07425659</td>
<td>0.87096952</td>
<td>0.54217601</td>
<td></td>
</tr>
<tr>
<td>Ventricles</td>
<td>0.07770291</td>
<td>0.10394785</td>
<td>0.45622543</td>
<td>0.1026147</td>
<td></td>
</tr>
<tr>
<td>Brainstem</td>
<td>-0.0515676</td>
<td>0.09683531</td>
<td>0.59535291</td>
<td>0.22143302</td>
<td></td>
</tr>
<tr>
<td>Cortical grey matter</td>
<td>-0.2070385</td>
<td>0.07216831</td>
<td>0.00487599</td>
<td>0.57797935</td>
<td></td>
</tr>
<tr>
<td>Deep grey matter</td>
<td>-0.2394995</td>
<td>0.07156956</td>
<td>0.0010961</td>
<td>0.58495287</td>
<td></td>
</tr>
<tr>
<td>White matter</td>
<td>-0.3428695</td>
<td>0.08933005</td>
<td>0.00019975</td>
<td>0.35339979</td>
<td></td>
</tr>
<tr>
<td>Hippocampi and Amygdalae</td>
<td>-0.1997654</td>
<td>0.08608312</td>
<td>0.02201073</td>
<td>0.39955024</td>
<td></td>
</tr>
<tr>
<td>Cerebellum</td>
<td>-0.2141247</td>
<td>0.07617365</td>
<td>0.005777</td>
<td>0.52983511</td>
<td></td>
</tr>
<tr>
<td>CSF</td>
<td>-0.0318009</td>
<td>0.07476168</td>
<td>0.67135302</td>
<td>0.54706588</td>
<td></td>
</tr>
<tr>
<td>Ventricles</td>
<td>0.01826599</td>
<td>0.10330974</td>
<td>0.85995975</td>
<td>0.1358503</td>
<td></td>
</tr>
<tr>
<td>Brainstem</td>
<td>-0.1061019</td>
<td>0.09700719</td>
<td>0.27627235</td>
<td>0.23748479</td>
<td></td>
</tr>
<tr>
<td>Cortical grey matter</td>
<td>-0.1210103</td>
<td>0.07102402</td>
<td>0.09102965</td>
<td>0.55866454</td>
<td></td>
</tr>
<tr>
<td>Deep grey matter</td>
<td>-0.2097287</td>
<td>0.06951009</td>
<td>0.00312029</td>
<td>0.57728766</td>
<td></td>
</tr>
<tr>
<td>White matter</td>
<td>-0.274854</td>
<td>0.08834273</td>
<td>0.00233237</td>
<td>0.31718974</td>
<td></td>
</tr>
<tr>
<td>Hippocampi and Amygdalae</td>
<td>-0.1991712</td>
<td>0.08284386</td>
<td>0.01775157</td>
<td>0.399547</td>
<td></td>
</tr>
<tr>
<td>Cerebellum</td>
<td>-0.1004817</td>
<td>0.07484742</td>
<td>0.18199361</td>
<td>0.50986904</td>
<td></td>
</tr>
<tr>
<td>CSF</td>
<td>0.03403775</td>
<td>0.07157315</td>
<td>0.6352555</td>
<td>0.5518136</td>
<td></td>
</tr>
<tr>
<td>---------------------------</td>
<td>---------------------------------</td>
<td>---------------------------------</td>
<td>-------------------------------------</td>
<td>-------------------------------------</td>
<td>-------------------------------------</td>
</tr>
<tr>
<td>Ventricles</td>
<td>0.13843843 0.10093913 0.17279836 0.10859001</td>
<td>0.1405061 0.06832719 0.04193334 0.5933809</td>
<td>0.149390536 0.58498283</td>
<td>0.2493061 0.08525002 0.00413305 0.3142649</td>
<td>0.200235 0.07956486 0.1317947 0.40246834</td>
</tr>
<tr>
<td>Brainstem</td>
<td>-0.0371258 0.09426582 0.69440311 0.22255991</td>
<td>-0.0512681 0.09018717 0.57079192 0.23227113</td>
<td>-0.0512681 0.09018717 0.57079192 0.23227113</td>
<td>-0.0512681 0.09018717 0.57079192 0.23227113</td>
<td>-0.0512681 0.09018717 0.57079192 0.23227113</td>
</tr>
</tbody>
</table>

**Supplementary table 11. Associations between DNAm CRP on global brain MRI parameters, adjusting for inflammatory exposures.** Standardized betas (β) and P-values are reported from regression models where DNAm CRP is regressed onto MRI measures. Bold text denotes FDR q-value <0.05. Models are as follows:

- **Model H1**: MRI metric ~ DNAm CRP + GA birth, infant sex, birthweight, GA scan, scanner variable (term subgroup, n = 87)
- **Model H2**: MRI metric ~ DNAm CRP + GA birth, infant sex, birthweight, GA scan, all inflammatory risk factors (term subgroup, n = 87)
- **Model H3**: MRI metric ~ DNAm CRP + GA birth, infant sex, birthweight, GA scan, scanner variable (preterm subgroup, n = 127)
- **Model H4**: MRI metric ~ DNAm CRP + GA birth, infant sex, birthweight, GA scan, all inflammatory risk factors (preterm subgroup, n = 127)
- **Model H5**: MRI metric ~ DNAm CRP + GA birth, infant sex, birthweight, GA scan, scanner variable + maternal age (preterm subgroup, n = 127)
- **Model H6**: MRI metric ~ DNAm CRP + GA birth, infant sex, birthweight, GA scan, scanner variable + smoking in pregnancy (preterm subgroup, n = 127)
- **Model H7**: MRI metric ~ DNAm CRP + GA birth, infant sex, birthweight, GA scan, scanner variable + preeclampsia (preterm subgroup, n = 127)
- **Model H8**: MRI metric ~ DNAm CRP + GA birth, infant sex, birthweight, GA scan, scanner variable + HCA (preterm subgroup, n = 127)
- **Model H9**: MRI metric ~ DNAm CRP + GA birth, infant sex, birthweight, GA scan, scanner variable + ROP (preterm subgroup, n = 127)
- **Model H10**: MRI metric ~ DNAm CRP + GA birth, infant sex, birthweight, GA scan, scanner variable + NEC (preterm subgroup, n = 127)
- **Model H11**: MRI metric ~ DNAm CRP + GA birth, infant sex, birthweight, GA scan, scanner variable + Sepsis (preterm subgroup, n = 127)
- **Model H12**: MRI metric ~ DNAm CRP + GA birth, infant sex, birthweight, GA scan, scanner variable + BPD (preterm subgroup, n = 127)
Supplementary table 12. Associations between individual DTI tract metrics and DNAm CRP controlling for inflammatory exposures

<table>
<thead>
<tr>
<th>model</th>
<th>Tract</th>
<th>FA</th>
<th>MD</th>
</tr>
</thead>
<tbody>
<tr>
<td></td>
<td>beta</td>
<td>lower CI</td>
<td>upper CI</td>
</tr>
<tr>
<td>AF left</td>
<td>-0.433</td>
<td>-0.687</td>
<td>-0.179</td>
</tr>
<tr>
<td>AF right</td>
<td>-0.434</td>
<td>-0.677</td>
<td>-0.190</td>
</tr>
<tr>
<td>ATR left</td>
<td>-0.380</td>
<td>-0.631</td>
<td>-0.129</td>
</tr>
<tr>
<td>ATR right</td>
<td>-0.261</td>
<td>-0.517</td>
<td>-0.004</td>
</tr>
<tr>
<td>CC genu</td>
<td>-0.381</td>
<td>-0.596</td>
<td>-0.166</td>
</tr>
<tr>
<td>CC splenium</td>
<td>-0.119</td>
<td>-0.371</td>
<td>0.133</td>
</tr>
<tr>
<td>CCG left</td>
<td>-0.325</td>
<td>-0.588</td>
<td>-0.062</td>
</tr>
<tr>
<td>CCG right</td>
<td>-0.455</td>
<td>-0.718</td>
<td>-0.191</td>
</tr>
<tr>
<td>CST left</td>
<td>-0.411</td>
<td>-0.638</td>
<td>-0.184</td>
</tr>
<tr>
<td>CST right</td>
<td>-0.549</td>
<td>-0.773</td>
<td>-0.326</td>
</tr>
<tr>
<td>IFOF left</td>
<td>-0.391</td>
<td>-0.629</td>
<td>-0.153</td>
</tr>
<tr>
<td>IFOF right</td>
<td>-0.414</td>
<td>-0.642</td>
<td>-0.185</td>
</tr>
<tr>
<td>ILF left</td>
<td>-0.285</td>
<td>-0.515</td>
<td>-0.055</td>
</tr>
<tr>
<td>ILF right</td>
<td>-0.337</td>
<td>-0.581</td>
<td>-0.093</td>
</tr>
<tr>
<td>UNC left</td>
<td>-0.421</td>
<td>-0.659</td>
<td>-0.182</td>
</tr>
<tr>
<td>UNC right</td>
<td>-0.384</td>
<td>-0.625</td>
<td>-0.143</td>
</tr>
</tbody>
</table>

Supplementary Table 12. Associations between DNAm CRP and the dMRI metrics of FA and MD in each tract. The β coefficients are in the units of standard deviations. Reported p-values are adjusted for false discovery rate (FDR).
References


11.3 Chapter 8 Supplementary data

An online version of this can be accessed at:
https://github.com/EleanorSC/Inflammatory-DNAm_STRADL/tree/main/Supplementary_Data
Examining multi-omic signatures of inflammation in relation to brain structure, cognitive ability, and brain ageing

Authors: Eleanor L.S. Conole¹,²,³, Danni A Gadd², Aleks Stoilcyn⁴, Xueyi Shen⁴, Mark E. Bastin¹,³, Stephen M. Lawrie⁴, Andrew M. McIntosh²,⁴,⁵, Veronique E. Miron⁶, Heather C. Whalley³,⁴,⁵, Riccardo E. Marioni², Simon R. Cox¹

* Authors contributed equally

¹ Lothian Birth Cohorts, Department of Psychology, University of Edinburgh, Edinburgh EH8 9JZ, UK.
² Centre for Genomic and Experimental Medicine, Institute of Genetics and Cancer, University of Edinburgh, Edinburgh EH4 2XU, UK
³ Centre for Clinical Brain Sciences, University of Edinburgh, Edinburgh EH16 4SB, UK
⁴ Department of Psychiatry, University of Edinburgh, Edinburgh, UK
⁵ Scottish Imaging Network, A Platform for Scientific Excellence Collaboration (SINAPSE), Edinburgh, UK
⁶ MRC Centre for Reproductive Health, Queen’s Medical Research Institute, Edinburgh BioQuarter, University of Edinburgh, Edinburgh, EH16 4TJ

Correspondence to:
Simon. R. Cox
Email:

Eleanor. L. S. Conole
Email:

This document includes:
Supplementary Results
Supplementary Methods
Supplementary Figures 1-12
Supplementary Results

Firstly, we assessed paired protein and DNAm signature contributions to brain and cognitive phenotypes with the aim of highlighting instances where DNAm signatures statistically outperformed the measured protein equivalent in associations with brain health outcomes. Of the 109 trained DNAm signatures, 84 were derived from SomaScan and 25 from Olink panels (Supplementary Data 2), using elastic net penalised regression models as described previously (27). Both sets of DNAm signatures explained differential levels of variance in circulating proteins (Pearson’s r range .1 – .73), as illustrated in Fig.2a (Supplementary Data 3; distributions of signatures illustrated in supplementary figs 1-2), comparable to previously reported DNAm surrogates of circulating plasma proteins of CRP (r = .21-.29) (14) and IL6 ( r = .21) (15). In total, we looked at 112 brain health outcome measurements, consisting of both cognitive and brain-imaging metrics (see study design, Fig.1, for a breakdown of brain-health outcomes, supplementary methods for further details).

The DNAm signatures were composed of 9101 unique CpG sites, which can be found alongside respective weights in Supplementary Data 2; the number of CpG sites per unique DNAm signature is reported in Supplementary Data 7 and top contributing CpGs to signatures alongside mapped phenotypic traits from the MRC-IEU catalog are reported in Supplementary Data 8. For the six DNAm signatures common to both Olink and SomaScan panels, there was variable correlation strength: S100A9/EN.RAGE: r = 0.63; GZMA r = 0.83, MMP.1 r = 0.56, CXCL10 r = 0.27, NTRK3 r = 0.20, and CXCL11 r = 0.42 (a heatmap depicting the correlation matrix between signatures is presented in Supplementary Fig.4; full results reported in Supplementary Data 9). A principal components analysis (PCA) revealed that 31 principal components explained 80.1% cumulative variance in the projected 109 DNAm signatures. The top three components are plotted in Supplementary Fig.5 (PC1 = 18.8%, PC2 = 11.6%, PC3 = 7.3%) alongside plots of the DNAm signatures that contribute most strongly to the overall variation among the 109 signatures studied (table of eigenvalues presented in Supplementary Data 10).

Functional mapping and GeneSet enrichment (GO, STRING and KEGG databases; supplementary data 4-6, supplementary fig.3) highlighted that, of the original proteins used to train the 109 DNAm signatures, the majority (67%) were associated with inflammation and the immune response, and others with inflammation-adjacent roles in chemotaxis, angiogenesis, and extracellular matrix function. A subset of 13 DNAm signatures from the Olink-trained set were from the Neurology protein panel with roles in synaptogenesis, neural development, and cellular signalling, though many had also had immune-regulatory roles such as recruitment of adaptive immune cells (Supplementary Data 4).

Throughout analyses, DNAm signatures or measured protein levels were used as predictors in models and various brain phenotypes were set as outcomes. Three main models were used as illustrated in Fig.2c: a baseline model controlling for age and sex, a sensitivity model that adjusted for age and sex plus immune cell proportions (monocyte, granulocyte, B-cell, CD4T, CD8T, and natural
killer cells), alongside a fully-adjusted model that further controlled for various lifestyle factors that are considered to influence inflammation (hypertension, smoking, alcohol consumption and BMI). In cases where MRI metrics were used as outcome measures, there were additional adjustments for scanner site, batch, number of QC edits and intracranial volume (ICV) in all models as per a previous protocol (27) (see Methods for details). All supplementary data (tables) are provided in the accompanying file: https://github.com/EleanorSC/Inflammatory-DNAm_STRADL/blob/main/Supplementary_Data/Supplementary_document_Conole2023.xlsx.

**Supplementary Methods**

**Cognitive metrics**

The latent score of general cognitive ability ($g$) was derived by conducting principal component analysis (PCA) on the tests from the cold cognitive test battery and extracting the first un-rotated principal component. The tests from which the $g$-factor was derived were (1) the matrix reasoning test (84), (2) verbal fluency test (85), (3) Mill Hill vocabulary test (86), (4) logical memory (87) and (5) digit-symbol coding tests (87). The proportion of variance explained by $g$ was 44%. Another global measure, general fluid cognitive ability ($g_f$), was extracted using the same approach, but with the Mill Hill vocabulary test (a crystallised measure of intelligence) excluded from the composite (the proportion of variance explained was 55%).

For individual cognitive domains, we examined processing speed (The Wechsler Digit Symbol Coding test), executive function (letter-based phonemic verbal fluency test; letters C, F and L, for one minute each) and verbal declarative memory (summed measure of immediate and delayed scores from the recall section of one story of the Wechsler Logical Memory III UK test). Maximum possible scores have been reported previously (71); distributions of all cognitive outcomes are presented in supplementary fig.7.


Supplementary Figure 1. Histogram distribution plots for all matched protein signatures used in study sample. Mean scores are illustrated in vertical lines, SD presented as dotted lines, n = 709.
Supplementary Figure 2. Histogram distribution plots for all 109 DNAm signatures projected into STRADL population sample. Mean scores are illustrated in vertical lines, SD presented as dotted lines, n = 709.
Supplementary Figure 3. Biological pathways of protein-coding genes proxied by DNAm signatures. Number of DNAm signatures is presented in the y axis and pathways on the X axis. ShinyGO 0.77
Supplementary Figure 4. Global correlation map on the left displaying pairwise correlations between DNAm signatures, where correlation coefficients (r) are plotted (n=709), with an inset of the DNAm clusters on the right. Red patches in the correlation maps indicate positive and blue patches negative correlations. B highlights clusters of strongly correlated signatures, where optimal clusters were determined by the gap statistic method (54).
Supplementary Figure 5. Principal components analyses for 109 DNAm (a) displays a scree plot demonstrating the cumulative proportion of variance explained by the first 10 components and (c-d) outlines the DNAm signature contributions to the first 3 components. Cumulative variance and respective eigenvalues for components is provided in Supplementary Data 10.
**Supplementary Figure 6.** Global correlation map on the left displaying pairwise correlations between DNA methylation (DNAm) signatures and EDTA-derived full blood count (FBC) count measured at G5:SFHS baseline 2006-2011 assessment, where correlation coefficients ($r$) are plotted (n=709). Red patches in the correlation maps indicate positive and blue patches negative correlations. Abbreviations: neutrophil-to-lymphocyte ratio (NLR), platelet-to-lymphocyte ratio (PLR) systemic immune-inflammatory index (SII), neutrophils (NE), whole blood cells (WBC), monocytes (MO), red blood cells (RBC), haemoglobin (Hb), leucocytes (LUC), MCH, MCV, PLT, EO, HCT, BA, LY, LMR.
Supplementary Fig. 7. Histogram distribution plots for cognitive metrics from the STRADL dataset where participants had complete DNAm data. Metric means are illustrated in vertical lines, SD presented as dotted lines, n numbers reported alongside. Variables were trimmed to remove outliers (>3.5 SD from mean) before plotting.

Supplementary Fig. 8. Histogram distribution plots for global neuroimaging and brain health metrics from the STRADL dataset where participants had complete DNAm data. Metric means are illustrated in vertical lines, SD presented as dotted lines. Variables were trimmed to remove outliers (>3.5 SD from mean) before plotting. *note here that the estimated ICV was not standardised for each neuroimaging assessment centre site
Supplementary Figure 9. The relationship between inflammation and processing speed is mediated by brain structure. (A) Significant associations between DNAm signatures and processing speed are mediated by various neuroimaging metrics; points show indirect effect size, where diamonds represent significant mediation. (B) Example path diagram of relationship between DNAm PIGR, total brain volume and processing speed (controlling for age, sex, neuroimaging site, batch and edits).
Supplementary Figure 10. The relationship between DNA methylation (DNAm) CRP and cognitive ability is mediated by global grey matter volume. Path diagram of relationship between DNAm CRP, global grey matter volume and cognitive ability ($g$), controlling for age, sex, neuroimaging site, batch and edits.
Supplementary Figure 11. Tissue expression profiles for DNAm signatures that associate with positive brain and cognitive health outcomes. (A) RNA tissue expression for top 3 DNAm signatures that associated with multiple favourable brain health outcomes. Graphs are generated using the consensus dataset (consists of normalized expression (nTPM) levels for 55 tissue types, created by combining the HPA and GTEx transcriptomics datasets: https://www.proteinatlas.org/ENSG00000149294-NCAM1/tissue). (B) FUMA brain-tissue expression heatmap for all signatyres that were associated with favourable brain or cognitive outcomes. Average expression per tissue and per gene is provided in log2 transformed scale allowing comparison of gene expression across tissue types for each gene. Red rectangles indicate higher expression, whereas blue rectangles indicate lower expression. Genes and tissues are ordered by clusters. Normalized gene expression data (reads per kilo base per million) for each tissue type were obtained from GTEx v8.
Supplementary Figure 12. FUMA tissue expression heatmap for the 22 genes that were associated with favourable brain or cognitive outcomes. Note that GDF.8 = MST1 and N.CDase = ASHA2. Average expression per tissue and per gene is provided in log2 transformed scale allowing comparison of gene expression across tissue types for each gene. Red rectangles indicate higher expression, whereas blue rectangles indicate lower expression. Genes and tissues are ordered by clusters. Normalized gene expression data (reads per kilo base per million) for each tissue type were obtained from GTEx v8. (B) RNA tissue expression for top 3 DNAm signatures that associated with multiple brain health outcomes. Graphs are generated using the consensus dataset (consists of normalized expression (nTPM) levels for 55 tissue types, created by combining the HPA and GTEx transcriptomics datasets using the internal normalization pipeline.)

https://www.proteinatlas.org/ENSG00000149294-NCAM1/tissue
Supplementary Figure 13. FUMA tissue expression heatmap for the 31 genes that were associated with poor brain or cognitive outcomes. Note that MMP12 is mapped to MME, SKR3 = ACVRL and EN.RAGE = S100A12. Average expression per tissue and per gene is provided in log2 transformed scale allowing comparison of gene expression across tissue types for each gene. Red rectangles indicate higher expression, whereas blue rectangles indicate lower expression. Genes and tissues are ordered by clusters. Normalized gene expression data (reads per kilo base per million) for each tissue type were obtained from GTEx v8.