THE UNIVERSITY
of EDINBURGH

Tissue-resident immune cells in health and disease: their heterogeneity and associated gene signatures

Anirudh Patir

Student ID: S1569052

A Thesis submitted for the degree of

Doctor of Philosophy

The University of Edinburgh

2021

Declaration

The thesis presented is the work of the author except where stated otherwise by reference and/or acknowledgment. Any work presented, which has been conducted by (or in collaboration with) others is explicitly acknowledged. No part of this work has been submitted in candidature for any other degree or qualification.

(ANIRUDH PATIR) 04/09/2020

Name Date

Anirudh Patir, S1569052
Ph.D. Genetics and Genomics,
Supervisors: Tom Freeman, Barry McColl

Abstract

The immune system is comprised of numerous cell types, molecules and pathways whose primary purpose is to regulate tissue homeostasis and protect an individual from disease. Researchers have tried to identify and characterise components of the immune system and their interactions as modulating aspects of this system is a major goal of the pharmaceutical industry. This work has described the significant heterogeneity of immune cell types which are defined by their microenvironment, extending from their lineage commitment in specialised tissues, e.g. bone marrow and thymus, to their activation states in disease. A facet of this heterogeneity are tissues-resident immune cells (TRICs) which have tissue specific homeostatic functions, and differ from their tissue naïve counterparts at the transcriptomic and epigenetic level. Furthermore, these cells display unique activation states in disease, making them a target for tissue-specific therapies. Hence, in this thesis, I have sought to expand on the current knowledge of TRICs in health and disease, investigating their heterogeneity and how to define them using various computational approaches.

Initially in chapter two a single well defined TRIC population, microglia, the tissue-resident macrophage of the brain is investigated. Microglia are the dominant immune cell type of the brain and are strongly implicated in neurodegenerative disease. These cells exhibit great heterogeneity depending on the brain region they reside in, also influencing their activation states. Fifteen studies had previously sought to define the functional profile of microglia in humans and mice, but these ‘gene signatures’ showed poor agreement overall. To address this issue a core human microglia signature conserved across brain regions was derived. Accordingly, data derived from intact brain tissue and pooled cells derived transcriptomic data from various brain regions was collated. This included nine datasets across three resources, the Genotype-Tissue Expression (GTEx) project, the Allen Brain Atlas (ABA) and a study of central nervous system (CNS) cells from Zhang et al. From each dataset, a microglial signature was derived using gene coexpression network (GCN) analysis to capture genes sharing a common expression profile across samples and which likely represented the same biology. The final human microglia signature comprised of 249 genes which were present in three or more of the dataset-derived microglial signatures. This gene set was validated using various sources of evidence. The average expression of signature genes correlated with microglial numbers and was significantly higher in myeloid populations relative to other immune and CNS cell types. Furthermore, the proteins encoded by signature genes positively

stained for microglia in different brain regions. The signature provides a means to understand the homeostatic state of these cells and a baseline against which their divergence in disease may be measured. Accordingly, the signature was used to analyse microglia in a transcriptomic dataset generated from post-mortem brain tissue of individuals of different ages and from Alzheimer's patients in four regions of human brain. This helped untangle the qualitative (activation states) and quantitative (cell proportions) differences in microglia between conditions and brain regions. Microglial cell numbers correlated with neuroinflammation and tau pathology in a region-dependent manner. The activation state of these cells was characterised by the downregulation of homeostatic genes (*CX3CR1* and *P2RY12*) and upregulation of TREM2-TYROBP pathway genes which have been implicated in Alzheimer's disease through genome-wide association studies (GWAS).

In chapter three, the analysis of TRICs was expanded upon from microglia to other TRIC populations. Currently, several immune cell types have been defined by selected markers and cytokine/chemokine profiles in the context of different diseases and tissues. However, a comprehensive unbiased analysis of these phenotypes in the context of other immune cells is required to appreciate the breadth of cellular heterogeneity, revealing commonalities and further subdivisions of known cell types. Given the availability of transcriptomic atlases of immune- and tissue-derived cells, there is a considerable scope for further analyses of TRIC biology. Hence, publicly available transcriptomic datasets from mouse including that derived from pooled-cells of marker-defined cell types and that from unbiased single cell RNA-Seq (scRNA-Seq) were considered. The former was taken from the Immunological genome project (ImmGen) resource, which comprised of 128 combinations of marker-defined cell types from 26 tissues. The relationship between these cells was studied, as were the gene signatures associated with them. Comparing cell types based on their transcriptome showed the relative similarity between lymphoid cell types relative to the heterogeneous myeloid cell populations. Using GCN analysis, 157 gene modules associated with either cell lineages, cell types, cell subsets or TRICs were identified. Interestingly, it was difficult to distinguish certain marker-defined cell types from others, either suggesting that cell types could be defined by a few genes or current markers may encompass overlapping cell populations. As a complimentary unbiased approach, we also analysed immune subsets defined by the Tabula Muris Atlas, which included scRNA-Seq data derived from twelve tissues. Forty-three cell clusters were identified which were associated with forty-four gene

clusters. The analyses highlighted gene clusters associated with the different cell lineages, cell types and TRICs, many of which significantly overlapped with those from the ImmGen GCN analysis. Some gene signatures were unique to TRICs or common across them, indicative of a tissue-dependent/independent biology. As expected, the greatest number, eleven signatures were associated with macrophages, eight of which agreed with cell types identified in the literature based on certain associated genes. To aid these analyses, novel approaches including annotating cells from a reference set of known cell types based on their transcriptomic profile; and capturing gene coexpression patterns using GCNs were developed, both of which are ongoing challenges unique to scRNA-Seq due to particular technical and biological variations.

Building upon the work described in chapter three, chapter four involved extending similar deconvolution analyses to human TRIC gene signatures from bulk tissue transcriptomic data taken from the GTEx resource. First, a set of reference immune signatures (RIS) was derived from a downsampled collection of all the 28 tissues of the GTEx, thus representing nine classifications of immune cell types. In the case of macrophages, monocytes and dendritic cells, combined gene signatures were identified, thus highlighting the resolution of bulk-transcriptomics for signature derivation. Finally, the RIS aided in deriving TRIC signatures individually from the 21 tissues of the GTEx considered for downstream analyses. As expected, signatures of macrophages-monocyte-dendritic cells were found in every tissue and across tissues 1,012 genes were associated with these cells, the highest number relative to other cell classifications. Interestingly, genes that were most commonly associated with a given cell type across tissues included many known markers for them, as found for macrophage-monocyte-dendritic cells, neutrophils, T cells, NK cells and B cells. Subsequently, each TRIC signature was compared with those derived from mouse in chapter three. Thirty-nine gene clusters overlapped between species and were associated with 12 TRICs. Seven TRIC populations and their associated genes were supported through literature.

In conclusion, this work has sought to examine the heterogeneity of TRICs, the transcriptomic signatures associated with them and the computational approaches to best derive them from tissue and cell level data. The work also shows the potential of using these TRIC signatures to explore disease states and the associated response of immune cells in this context.

Lay abstract

Humans possess a defence system called the immune system which helps in protecting them from disease. This system comprises of specialised cell types which are spread across different tissues of our body. These cells are constantly sensing for signs of infection or injury and respond accordingly to counteract them. Indeed, cells of the immune system are associated with the pathology of almost all diseases, from cancers to the infection of COVID-19. Understanding and modulating the immune system is a cornerstone of the biomedical research and a major target of pharmaceuticals. Modern research has studied this natural defence system and uses it to treat various diseases. Hence, there has been a great interest in understanding the different parts of the immune system. In order to respond appropriately to the numerous challenges associated with the spectrum of disease, the immune system consists of a diverse set of cell types each with unique characteristics. Though communicating with one another they can launch an array of responses. In addition to the commonly known “white blood cells” which are measured in blood tests, immune cell types display a great diversity as they also reside in various tissues. Within the tissues they help maintain proper functioning of the tissues and have unique responses to counteract infections or injuries. One of the ways we can understand these cells is to see which genes they utilise or express. Genes encode the instructions to make specific proteins, and when expressed by a cell allow it to carry out specialised functions associated with these proteins. Each cell type expresses a different combination of genes which define their identity and activity. In many cases, a group of genes work together to carry out a particular function and we try identifying those that are expressed by a particular cell type. In this thesis GCN analysis was used to identify these genes. The method constructs a network of genes, where genes related to the same biology, sharing a similar expression across samples are highly connected and hence can be captured. Using this approach this thesis examines the diversity of immune cells across tissues and the genes that are associated with them.

In chapter 2, we begin exploring these immune cells in the context of tissues by examining the major immune cell type of the brain, known as the microglia. These cells are implicated in various diseases such as Alzheimer’s disease, Parkinson’s and multiple sclerosis, all of which are associated with dying neurons, and together are termed as neurodegenerative diseases. Microglia are the most abundant immune cell type of the brain responsible for its proper function, e.g. redefining connections between neurons. These cells differ depending on the region of the brain, hence we tried to find a group of

genes or a gene signature that defines microglia across brain regions by using GCNs. To identify these genes nine publicly available datasets of gene expression data derived from the human brain were considered. From each dataset we used GCN analysis to identify a group of genes that represented microglia, and genes that were found in three or more of these networks were considered to represent the final core human microglia. 249 such genes were identified. Knowing this signature of genes in healthy brains we could then observe how these genes changed in disease. By examining the expression of these genes together, the number of microglial cells that changed in young, ageing and individual with Alzheimer's disease was examined. In ageing, the proportion of microglia increased the most in the hippocampus, a region of the brain known to be significantly affected by age and which is associated with learning and enforcing memories. In Alzheimer's, microglial numbers were higher in brain regions with inflammation and those known to have a greater accumulation of the protein tau, associated with Alzheimer's. Furthermore, changes were also observed within microglia as they reduced the expression of genes associated with their identity in healthy individuals and began to express genes with known associations to Alzheimer's.

In chapter 3, the analysis from microglia was expanded upon to the various immune cell types across different tissues in the mouse. This study examined how these cell types are currently defined, i.e. by a set of genes or markers which can uniquely separate them from other cell types based on their expression. For this, a public resource called the ImmGen was used which is an international effort to understand the biology of the immune system. This resource included data from cells isolated from 26 different tissues and using 128 combinations of markers, each representing a unique cell type. Similar to how we studied the similarity between genes using GCN, from this data we were able to find the relationships between these cell types. Cell types of similar origin were found connected to one another while other heterogeneous cells formed separate groups e.g. macrophages, a versatile immune cell type ubiquitously present across tissues. Subsequently, the Tabula Muris dataset was analysed, which has the gene expression from single cells also called scRNA-Seq from 12 tissues; however, in this case, we did not know the cell types. This enabled the exploration of these cell types unbiased by predefined markers. From the data, 44 signatures for cell types were identified either specific to a tissue or those found in multiple tissues. These signatures overlapped between the analyses from both resources. In the process, methods to identify cell types based on known cell types were developed. Additionally, a method to identify genes with

a similar expression in scRNA-Seq data was also developed, which has been challenging in this research area due to the immense variation found in this data.

In chapter 4, the analysis was extended to humans and the results were compared with the analysis in mouse. This comparison is essential as several drugs are based on studies in mouse but are found to be ineffective in human. For this the GTEx resource was considered, which provides gene expression data from 33 tissues. From each tissue, genes signatures were derived for the different immune cell types present using GCNs. Certain genes were repeatedly found to be associated with a particular immune cell type across tissues. Interestingly several of these genes included known markers which defined the immune cell type. Macrophages had the most number of genes in its signature and was identified in every tissues thus supporting its diversity. Lastly, the signatures were compared with those from mouse. Literature supported the signatures for seven immune cell types that overlapped across species. Still, the overlap across mouse and human immune signatures was minimal likely because when we examine the mixture of cells within a tissue it is difficult to find genes expressed by a certain cell type as they may be shared across several cell types. These findings highlighted the power of studying individual cells rather than a mixture of them like in tissues.

In summary, this thesis examines 1) how the number and states of microglia differed in ageing and Alzheimer's depending on the brain region; 2) The association between 128 immune cell types (as defined by markers) from 26 different tissues. Here, cell types, were similar to those of the same origin based on their gene expression, although heterogenous cell types like macrophages were very different, evident from them forming different groups within the sample network; 3) An unbiased derivation of 44 immune signatures across 12 tissues using scRNA-Seq data, revealing the diversity of immune cells and how much of it is dependent on tissue; 4) A methodology to identify the types of cell in scRNA-Seq data based on the gene expression of known cell types; 5) developing and comparing different methods to construct GCN from scRNA-Seq data; 6) human immune signatures from bulk RNA-Seq data spanning 21 tissues; and 7) A comparison of the immune signatures from human (bulk tissue) and mouse (cells), highlighting the lack of resolution in bulk RNA-Seq to identify tissue-specific genes of immune cell types.

Acknowledgments

This PhD has been one of the most enjoyable times of my life, as my curiosity has been welcomed and supported by all my associates to whom I owe a wealth of gratitude. First and foremost, I would like to thank my parents without whom I could not have journeyed so far. They have given me their endless love and support for allowing me to pursue my interests and as beautiful human beings.

I would like to thank my supervisor Tom Freeman, with whom I have had the pleasure to share some of the most though provoking conversations. He has allowed me to grow while entertaining several of my ideas and always keeping me grounded, adhering to the big picture, i.e. the biology. My secondary supervisor Barry McColl has aided me in the same way, and I have always been inspired by his clarity and amazement in the field. Gratefully, in addition to their intellect they have become close friends, making this journey effortless and enjoyable.

I have been fortunate to be able to collaborate with Megan Davey, Musa Hassan, Jayne Hope, Adam Balic and David Hume. They have introduced me to various new fields and methodologies, while providing great insight into these areas. Specially, members of Megan Davey's Lab, Lynn Mcteir, Amy Fraser and Joe Rainger, whose expertise have been quintessential to my progress. In addition, I would like to thank Barbara Shih and Anna Raper, who have been most helpful in my projects and been happy to share their professional opinion. From my lab I would like to thank Ajit Johnson, Mark Barnett, Josh Harling-Lee and Bruno Giotti, with whom I have spent much time discussing work and worldly matters. Furthermore, I am thankful for the other friends I have made along the way who have enriched this experience, especially Prasun and Shweta.

Lastly, I would like to thank the University of Edinburgh, for providing the cutting edge in facilities and resources, while also surrounding me with some of the greatest minds and academics at the forefronts of their fields.

Publications

Related work

1. *Patir, A.,* Shih, B., McColl, B. W., & Freeman, T. C. (2019). A core transcriptional signature of human microglia: Derivation and utility in describing region-dependent alterations associated with Alzheimer's disease. Glia, 67(7), 1240-1253.
2. *Patir A.,* Shih B., & Freeman, T.C. “Tissue resident immune signatures” (In preparation).

Unrelated work

1. *Patir, A.,* Fraser, A. M., Barnett, M. W., McTeir, L., Rainger, J., Davey, M. G., & Freeman, T. C. (2020). The transcriptional signature associated with human motile cilia. Scientific Reports, 10(1), 1-12.
2. Davies, C. L., *Patir, A.,* & McColl, B. W. (2019). Myeloid cell and transcriptome signatures associated with inflammation resolution in a model of self-limiting acute brain inflammation. Frontiers in immunology, 10, 1048.
3. Smith, C., McColl, B. W., *Patir, A.,* Barrington, J., Armishaw, J., Clarke, A., ... & Rice, G. I. (2020). Biallelic mutations in NRROS cause an early onset lethal microgliopathy. Acta Neuropathologica, 1-5.
4. *Patir, A.*, Gossner, A., Ramachandran, P., Alves, J., Freeman, T., Henderson, N. C., ... & Hassan, M. A. (2019). Single-cell RNA-Seq reveals CD16-monocytes as key regulators of human monocyte transcriptional response to Toxoplasma. bioRxiv, 863274.
5. Padhi, A. K., Narain, P., Dave, U., Satija, R., *Patir, A*., & Gomes, J. (2019). Insights into the role of ribonuclease 4 polymorphisms in amyotrophic lateral sclerosis. Journal of Biomolecular Structure and Dynamics, 37(1), 116-130.
6. *Patir, A.,* Barnett, M.W., Raper, A., Fleming, R., Henderson, B., Murphy L., Henderson, N. & Freeman T.C. Cellular heterogeneity of the developing worker honeybee pupa: a single cell transcriptomic analysis (in preparation).
7. *Patir, A.,* & McColl B. The myeloid landscape of a mouse stroke model using single cell transcriptomics (In preparation).

8. Freeman, T.C., Horsewell, S., *Patir, A.*, Harling-Lee, J., Regan, T., Shih, B., P. James, Hume, D.A. & Angus, T. Tim, (2020), Graphia: A platform for the graph-based visualisation and analysis of complex data. bioRxiv

Presentation at conferences and symposiums

1. Presented a talk at the "**The Genotype Tissue Expression (GTEx) Project Community Meeting**" held in Barcelona from the 20th to 21st April 2017, titled "Derivation of a microglial transcriptional signature by network deconvolution of human central nervous system expression data".
2. Presented a poster at the conference of "**XIII European Meeting on Glial Cells in Health and Disease**" held from the 8th to 11th July 2017, titled "Derivation of a microglial transcriptional signature by network deconvolution of human central nervous system expression data" in Edinburgh.
3. Attended the "**Open data science conference (ODSC)"** from the 12th to 14th October 2017 held in London.
4. Presented a talk at the conference of "**European conferences on computational biology (ECCB)**" held from the 8th to 12th September 2018, titled "Evaluation of network methods for the analysis scRNA-seq data and development of a new KNN-based method based on identified caveats" in Athens.
5. Attended the "**PacBio Symposium 2019**" on the 7th February 2019 held in Edinburgh.
6. Attended the "**Mass Spectrometry Symposium 2019**" on the 18th March 2019 held in Edinburgh.
7. Attended the "**Senescence Symposium 2019**" on the 10th May 2019 held in Edinburgh.

Workshops and courses

1. Undertook theory and practicals at the workshop of "**Analysis of single cell RNA-Seq data**" from the 5th to 9th February 2018, conducted by Physalia Courses and Welcome Trust Sanger Institute in Berlin.
2. Undertook theory for the workshop on "**Nature Masterclass in Scientific Writing and Publishing**" held on the 2nd October 2018 by the Nature editorial in Edinburgh.
3. Undertook theory and practicals at the workshop of "**ECDF 2: HPC Shared-Memory Programming with OpenMP**" on the 23rd October 2018 conducted by the University of Edinburgh in Edinburgh.

4. Undertook theory and practical at the workshop of "**UNIX 2: intermediate**" on the 1st March 2019 conducted by the University of Edinburgh in Edinburgh.

Teaching & activities

1. Held a talk at Merchiston Castle School (Edinburgh, United Kingdom) titled "**Wearing the many hats of a PhD to examine Alzheimer's"** on January 2018
2. Assisted in informing the public about honeybees on the campus **open day** at the Roslin institute, October 2018
3. Held a tutorial for the undergraduate "**Cell biology**" course at the University of Edinburgh, November 2018
4. Held tutorials for the undergraduate "**Cell signalling**" course at the University of Edinburgh, November 2018

Table of contents

List of abbreviations

ABA	Allen Brain Atlas
AD	Alzheimer's Disease
APC	Antigen Presenting Cell
Ab	Amyloid Beta
BCR	B Cell Receptor
BiTE	Bispecific T Cell Engagers
CAR	Chimeric Antigen Receptors
cDC	Conventional Dendritic Cells
CLP	Common Lymphoid Progenitor
CLR	C-Type Lectin Receptors
cMoP	Common Monocyte Progenitor
CMP	Common Myeloid Progenitor
CNS	Central Nervous System
CyTOF	Cytometry by Time-Of-Flight Mass Spectrometry
DAM	Disease Associated Microglia
DAMP	Damage-Associated Molecular Pattern
DC	Dendritic Cells
dsRNA	Double-Strand RNA
EBI	European Bioinformatics Institute
EMP	Erythroid-Myeloid Progenitors
eQTL	Expression Quantitative Trait Loci
FACS	Fluorescence Activated Cell Sorting
GC	Germinal Centre
GCN	Gene Coexpression Network
GEO	Gene Expression Omnibus
GMP	Granulocyte-Monocyte Progenitors
GO	Gene Ontology
GSEA	Gene Set Enrichment Analysis
GTEx	Genotype-Tissue Expression
GTExGC	GTEx Gene Cluster
GWAS	Genome-Wide Association Studies
HPA	Human Protein Atlas
HSC	Hematopoietic Stem Cell
IHC	Immunohistochemistry
ImmCC	ImmGen Cell Cluster
ImmGC	ImmGen Gene Cluster
ImmGen	Immunological Genome Project
ISH	*In Situ* Hybridization

kNN	K-Nearest Neighbour
LMP	Lymphoid-Myeloid Progenitors
Mac-Mono-Den	Macrophage-Monocytes-Dendritic Cells
MCL	Markov Cluster Algorithm
MEP	Megakaryocyte–Erythroid Progenitor
MLP	Multipotent Lymphoid Progenitor
MPP	Multi-Potential Progenitors
MZ	Marginal Zone
NCBI	National Centre of Biotechnology and Information
NLR	Nucleotide-Binding Oligomerization Domain Leucin Rich Repeats Containing Receptors
PAGA	Partition-Based Graph Abstraction
PAMP	Pathogen-Associated Molecular Pattern
PBMC	Peripheral Blood Mononuclear Cells
PCA	Principal Component Analysis
pDC	Plasmacytoid Dendritic Cell
PPI	Protein-Protein Interactions
PRR	Pathogen Recognition Receptors
RIS	Reference Immune Signature
RLR	Retinoic Acid-Inducible Gene 1-Like Receptors
SAM	Self-Assembling Manifolds
scRNA-Seq	Single Cell RNA-Seq
SINCERITIES	Single Cell Regularized Inference Using Time-Stamped Expression Profiles
snRNA-Seq	Single Nuclear RNA-Seq
ssRNA	Single-Strand RNA
STREAM	Single-Cell Trajectories Reconstruction, Exploration and Mapping
STRING	Search Tool for The Retrieval of Interacting Genes
TAA	Tumour Associated Antigen
TAM	Tumour Associated Macrophages
Tc	Cytotoxic T Cell
TCR	T Cell Receptor
Th	Helper T Cell
TLR	Toll-Like Receptors
TMCC	Tabula Muris Cell Clusters
TMGC	Tabula Muris Gene Clusters
TOM	Topological Overlap Matrix
TRIC	Tissue-Resident Immune Cell
TRICs	Tissue-Resident Immune Cells

TSLP	Thymic Stromal Lymphopoietin
tSNE	t-Distributed Stochastic Neighbour Embedding
UMAP	Uniform Manifold Approximation and Projection
WEP	Weighted Edge Pruning
WGCNA	Weighted Gene Co-Expression Network Analysis
YS	Yolk Sac

List of tables

Supplementary tables

- TABLE S2.1_Comparison of the current and published human microglia signatures
- TABLE S2.2_Dataset information
- TABLE S2.3_Dataset derived microglial clusters and annotated core microglia signature genes
- TABLE S2.4_ToppGene enrichment analysis of core microglial genes
- TABLE S2.5_Clusters from Figure 4
- TABLE S2.6_Fold change in core and MAG for aging and Alzheimer's across regions
- TABLE S2.7_ToppGene enrichment analysis of MAG
- TABLE S2.8_ToppGene enrichment analysis of genes differentially expressed only in Alzheimer's
- TABLE S3.1_mmGen, Sample clusters and annotation
- TABLE S3.2_ImmGen, gene clusters and annotation
- TABLE S3.3_Tabula Muris, sample meta
- TABLE S3.4_Tabula Muris, cell clusters and annotation
- TABLE S3.5_Tabula Muris, gene clusters from different methods
- TABLE S3.6_Unique and conserved gene clusters across GCN construction methods
- TABLE S3.7_Tabula Muris, gene cluster annotation and enrichment
- TABLE S4.1 GTEx, Enrichment analysis of combined dataset for all genes
- TABLE S4.2 GTEx, reference immune signature
- TABLE S4.3 GTEx, tissue resident immune signatures
- TABLE S4.4 Comparison of TRICs between human and mouse

List of figures

Supplementary figures

- FIGURE S2.2_IHC of published signature genes specifically coexpressed in the cerebellum
- FIGURE S2.3_Coexpression of human microglial signatures in CNS tissue transcriptomics

Supplementary data availability

The supplementary data (tables and figure) is available through the University of Edinburgh DataShare using the following DOI:

https://doi.org/10.7488/ds/2917

1. Chapter 1: Introduction

1.1 Components of the immune system

1.1.1 The innate and adaptive immune system

The immune system is responsible for an individual's defence response to pathogens and in maintaining tissue homeostasis. It consists of specialised cell types and molecules which orchestrate different responses. By convention they are classified into the innate and adaptive immune response. These responses are characterised based on the type of cells, humoral components, receptors, response times and their capacity to memorise previously encountered antigens. The innate immune system is considered the first line of defence and is defined by its quick response to infection/injury. One of the primary components of innate immunity are the physical (mucosal layers and ciliated epithelial layers of the lung which facilitate pathogen clearance) and chemical (anti-microbial peptides like defensins and ribonuclease) barriers which prevent pathogens from entering the body (Turvey and Broide, 2010). Part of the humoral innate immunity are molecules like LPS-binding proteins, C-reactive protein and complement proteins which possess antimicrobial activity and facilitate innate immune responses, e.g. the complement system disrupts a pathogen's cell membrane via a membrane attack complex and opsonizes them for subsequent phagocytosis (Dunkelberger and Song, 2010). The immediate response of the innate immune system is attributed to the germline encoded pathogen recognition receptors (PRR) which recognise large groups of pathogens based on evolutionarily conserved pathogen-associated molecular patterns (PAMPs). In addition, they also recognise damage-associated molecular patterns (DAMPs) of molecules introduced as a result of infection. To account for these different patterns PRRs comprise of several subfamilies including, C-type lectin receptors (CLR), nucleotide-binding oligomerization domain leucin rich repeats containing receptors (NLR), retinoic acid-inducible gene 1-like receptors (RLR) and Toll-like receptors (TLR) (Takeuchi and Akira, 2010). Depending on the type of receptor they are localised to different cellular compartments and recognise different molecular patterns. For example, TLR3, TLR7, TLR8 and TLR9 that reside in the endosomal compartment are specific to DNA and RNA molecules found intracellularly (Kawasaki and Kawai, 2014). Furthermore, some of these interactions can result in different signalling cascades such as TLR3 which uses a TRIF-dependant pathway as opposed to the MyD88-dependent pathway used by the majority of the other TLRs. On recognising pathogens, innate immune cells can mediate various types of responses based on their specialised

functions. For example, macrophages, dendritic cells (DCs) and neutrophils are considered professional phagocytes which can phagocytose pathogens and debris at the sight of infection or injury. Innate immune cells such as granulocytes and NK cells release cytotoxic compounds on activation. Furthermore, in combination with cells of the adaptive immune system, innate immune cells secrete cytokines and chemokines to steer the inflammatory response and recruit further immune cells, e.g. a conventional proinflammatory response termed as type 1 is defined by cytokines like IL2, IL12 and IFNγ, while that for a type 2 anti-inflammatory/wound healing response is characterised by IL4, IL5 and IL13, both of which are associated with cell types with corresponding activation states (Gause et al., 2013).

Following activation of an innate immune response, the adaptive immune response is initiated. This system is characterised by different cell types and organs which unlike the innate immune system generate antigen-specific molecules through gene rearrangement and somatic hypermutations (Chaplin, 2010). Cells of the adaptive immune system, e.g. T cells are triggered by engaging antigens via the T cell receptor (TCR) through antigen presentation cells (APCs) such as DCs which capture and process antigens for presentation. Subsequently some of these cells exhibit effector T cells functions (e.g. secretion of cytotoxic compounds) while other helper T cells (Th) facilitate B cell maturation. B cells are lymphocytes specialised for producing vast amounts of antibodies on activation through a series of phases ultimately resulting in a stable interaction between B and T cells. Much like the mediators of the innate immune system, antibodies also facilitate the recognition of pathogens thereby facilitating immune responses, however, these antibodies are dynamic and epitope specific. Furthermore, depending on the context of activation, some cells of the adaptive immune system can convert to memory cells, which retain receptors specific to the PAMP from an earlier infection. These T and B memory cells are therefore able to initiate an immediate response on reencountering the same antigen. Thus, the innate and adaptive immune system present two arms of the immune system carrying out different functions with the ultimate goal of pathogen clearance.

During an infection or injury immune cells migrate to the affected site as part of the immune response. There are also various tissues which are associated with the immune system where lymphocytes reside, develop and undergo activation. Lymphoid organs are categorised into primary, secondary and tertiary lymphoid organs. Primary lymphoid organs include the thymus and bone marrow. These tissues are involved in lymphocyte

production and development. The bone marrow is where both lymphoid and myeloid cells originate through the process of haematopoiesis. Subsequently, T cells are 'educated' by engaging MHC and antigen complexes in the thymus through various checkpoints filtering out self-reactive T cells. Secondary lymphoid organs such as lymph nodes, Peyer's patch, and spleen, have an assortment of cell types from both the innate and adaptive immune system and are involved in immune cell activation and pathogen clearance. For example, in the spleen APCs migrate to activate cells of the adaptive immune response, additionally phagocytic cells, such as macrophages await antibody bound pathogens. Whilst similar in function, tertiary lymphoid organs develop in tissues during adulthood in response to inflammation as observed in cancers, autoimmune diseases and infections (Buettner and Lochner, 2016).

There is an increasing interest in immune cell types that reside in non-lymphoid tissues and the homeostatic functions they carry out. For example, adipose macrophages have increased β-oxidation which aids in lipid catabolism thereby minimising lipotoxicity. This alternatively activated phenotype is maintained by IL4 and IL13 released by tissue resident Th2 cells, ILC2s and iNKT cells (Odegaard and Chawla, 2015, Zhang et al., 2019). Such cells or tissue-resident immune cells (TRIC) contribute significantly to immune cell diversity. Differences between TRICs and their tissue-naïve counterparts can be observed at the proteomic, transcriptomic and epigenetic level. One of the earliest exhaustive studies in the field described the transcriptomic profile of seven tissue-resident macrophage populations, as well as the transcription factors and enhancers that define them (Lavin et al., 2014a). Distinctions are also observed in TRIC ontogeny and how they are sustained through adulthood, e.g. microglia, the tissue macrophages of the nervous system, originate from primitive haematopoiesis in the yolk sac (Mass, 2018), in contrast to majority of the other macrophage populations which are derived from definitive haematopoiesis in the bone marrow. Furthermore, microglia are sustained by a local pool of cells in the brain (Ajami et al., 2007). In summary, the immune system has numerous components from molecules to complex systems of cells, all which work in concert to protect and maintain homeostasis.

1.1.2 Production of immune cell types from haematopoiesis

The immune system consists of cell types originating from two lineages, the myeloid and lymphoid lineage. Whilst, innate immune cell types are derived from both lineages, those of the adaptive immune system originate from only the lymphoid lineage. Both lineages originate from a common multipotent cell, the hematopoietic stem cell (HSC). Cell

production and differentiation in the immune system is referred to as haematopoiesis and in adult life this process primarily takes place in the bone marrow. The factors regulating the differentiation towards lymphoid or myeloid commitment are still under investigation, e.g. aged HCSs have shown a preference towards committing to the myeloid lineage, suggesting a correlation between lineage commitment and the age of progenitors as defined by the number of cell divisions (Young et al., 2016). Alternately, certain mediators like TGFβ1 can direct HSC differentiation to myeloid cell types (Challen et al., 2010). Recent studies also suggest a precommitment towards cellular states, as cells which clonally expand from the same progenitor tend to share the same lineage (Naik et al., 2013). Supporting this, cells in the bone marrow were found to be predominately precommitted and unipotent (Notta et al., 2016). In contrast, HSCs presenting multi-lineage gene expression profiles have been identified suggesting a plastic phenotype (Olsson et al., 2016). These results could be indicative of the plasticity of HSCs and their ability to precommit to certain states. The question then remains, what factors determine cell type commitment.

Definitive hemopoieses define many of the immune cell types in human, beginning from HSCs which sequentially differentiate through various states to become multi-potential progenitors (MPP) (**Figure 1.1**). These cells then diverge giving rise to lymphoid and myeloid lineages by differentiating into the common lymphoid progenitor (CLP) and common myeloid progenitor (CMP), respectively. CMPs can differentiate into either megakaryocyte–erythroid progenitor (MEP) which give rise to erythrocyte and megakaryocytes (platelets), or granulocyte-monocyte progenitors (GMP) which generate granulocytes (basophils, neutrophils, eosinophils and mast cells), monocytes, macrophages and DCs. CLPs differentiate into cell of the lymphoid lineage including B cell, T cells, NK cells, ILCs, NKT and DCs with certain cell types being developed in the thymus, excluding innate lymphocytes.

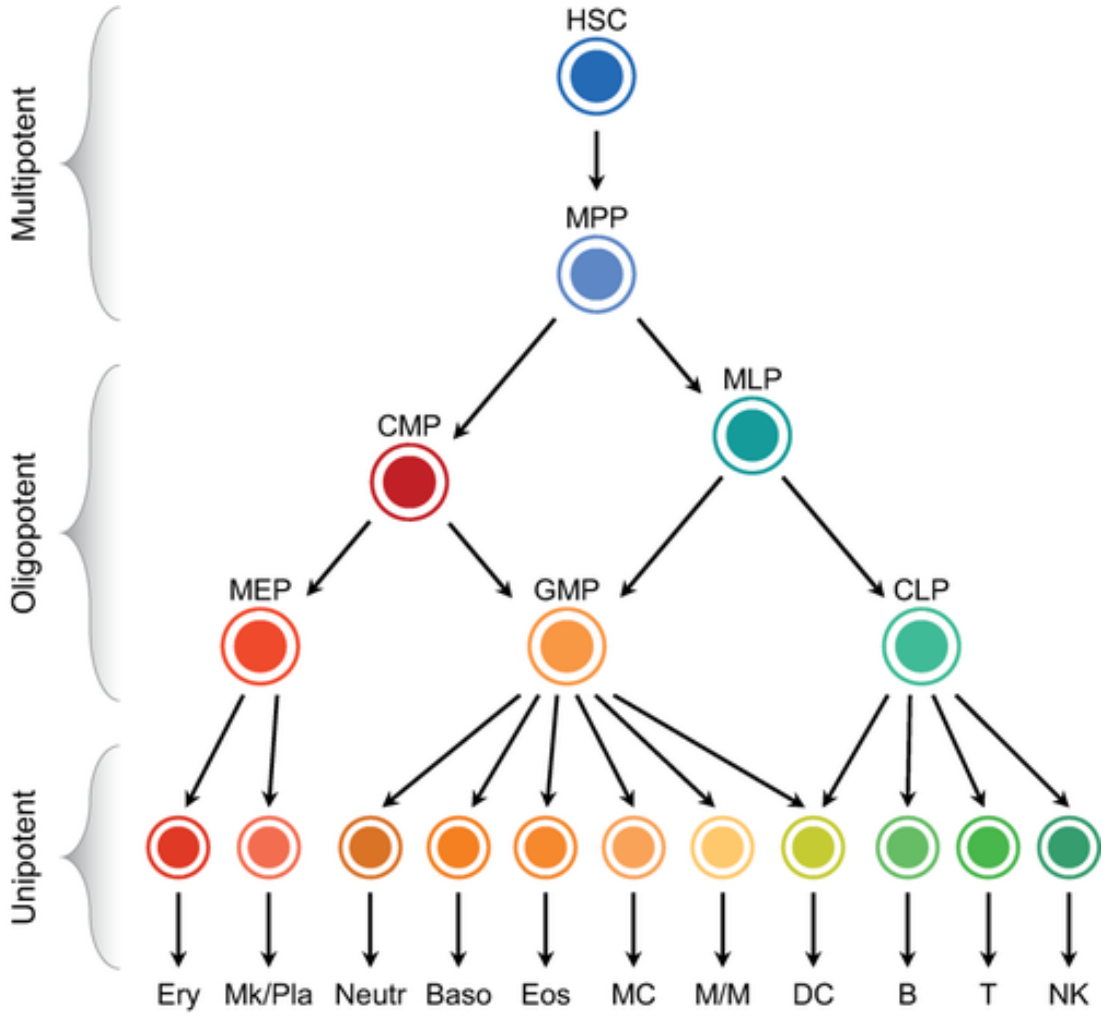


Figure 1.1. Human haematopoiesis.

Taken from (Antoniani et al., 2017). The simplified schematic of human haematopoiesis from multipotent HSC (top) to unipotent cell types (bottom). HSC: hematopoietic stem cells, MPP: Multipotent progenitors, CMP: Common myeloid progenitor, MLP: multipotent lymphoid progenitor, MEP: megakaryocyte–erythroid progenitor, GMP: granulocyte-monocyte progenitors, CLP: Common lymphoid progenitor, Ery: Erythrocyte, Mk/Pla: Megakaryocyte/Platelet, Neutr: Neutrophils, Baso: Basophils, Eos: Eosinophils, MC: Mast cells, M/M: Macrophage/monocyte, DC: Dendritic cell, B: B cells, T: T cells, NK: Natural killer cells.

1.2 Disease as a disorder of the immune system

1.2.1 Types of diseases and therapies

The innate immune system facilitates an immediate response to infection or injury, inhibiting pathogen spread and recruiting neighbouring cells including those of the

adaptive immune system. A complex set of interactions allows the immune system to display a range of response types. Alternatively, disorders of the immune system can lead to: 1) inflammatory diseases where inflammation is induced by either infection or injury continues without resolution resulting in adverse effects such as the permeabilization of the blood-brain barrier during neuroinflammation (Erickson et al., 2012); 2) allergies, where the immune system produces an inappropriate response to molecules, e.g. asthmatic patients exposed to aeroallergens like pollen and dust mite trigger a type 2 immune response (Caminati et al., 2018); 3) autoimmune diseases, where the immune system is unable to distinguish self from non-self, e.g. rheumatoid arthritis where antibodies are generated towards molecules like fibrin and collagen resulting in inflammation of the synovial joints (Guo et al., 2018); and 4) cancers, where the immune system produces an inefficient response towards abnormal and highly proliferative cells. Therefore, to develop therapies for such diseases it is important to understand the components of the immune system and their role, e.g. immune-checkpoint blockade therapy targets key regulators which mediate the immune response of which anti-PD1 therapy has been found effective in treating various cancers and a recent scRNA-Seq study attributes it to a subset of tissue-resident memory T cells and classical monocytes (Krieg et al., 2018). Various types of immunotherapies have been developed based on our understanding of the immune system, including vaccines, cytokine therapies, antibody/ligand-based therapies, and engineering T cells which can e.g. regulate immune checkpoints (**Figure 1.2**) (Naran et al., 2018).

Further studies in TRIC biology have led to a great interest in tissue-specific immunotherapies which are largely focused on treating cancers as TRICs attain alternate activation states relative to their tissue naïve counterparts under pathological conditions. For example, tumour-associated macrophages (TAMs) are pro-tumorigenic (mediate angiogenesis and metastasis) with clinical studies showing the significance in depleting these cells (through mAb induced apoptosis or inhibiting their CCL2 mediated recruitment) and modulation of their phenotype, e.g. reprograming macrophages from M2 to M1 state (Li et al., 2019). Other TRIC based immunotherapies unrelated to cancer include microglia-based therapies being developed for neurological diseases. Microglia are implicated in several neurological diseases and as a result numerous *in vitro* and *in vivo* models have been generated to describe their activation states (Masuda et al., 2020). In health, these cells phagocytose apoptotic cells, neuronal synapsis and unfolded proteins such as amyloid beta (Aβ). However, in Alzheimer's disease (AD), a

neurodegenerative disease characterised by aggerates of Aβ plaques, microglia exhibit certain pro-inflammatory features but can become anti-inflammatory on using metformin (Wang and Colonna, 2019). Hence, given the essential role of TRICs in tissue homeostasis it is crucial to understand these cell types and their activation states in disease.

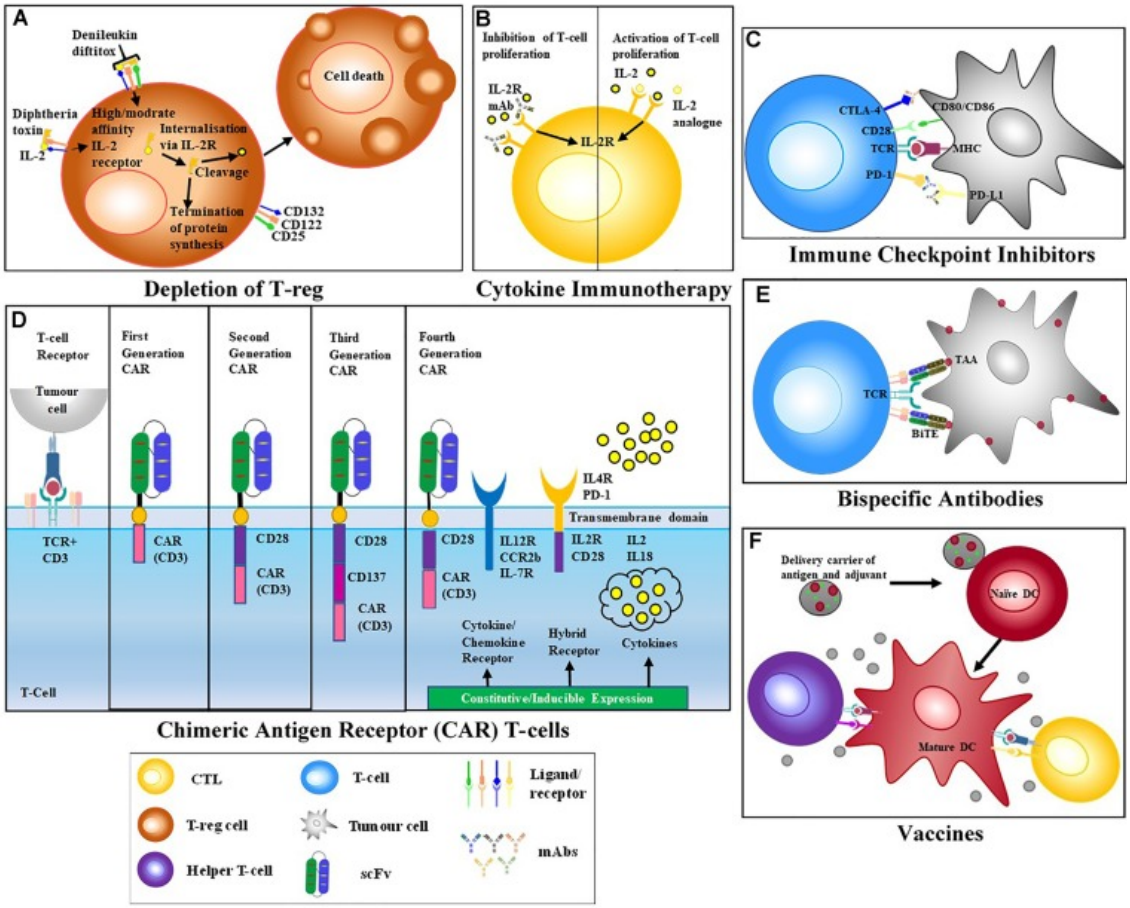


Figure 1.2. Types of immunotherapies.

The figure is taken from (Naran et al., 2018). Shown are the different types of immunotherapies for cancers and infectious diseases. A) Drugs like denileukin diftitox target receptors of immunosuppressing Tregs to initiate apoptosis thereby skewing the immune response towards proinflammation. B) Depending on the response required different cytokines or associated antibodies are used to propagate either an inflammatory response, like IL2 mediated T cell activation or anti-inflammatory/inflammation suppression e.g. mAbs for blocking IL2R. C) mAbs can also block immune checkpoints which are essential steps in immune response, e.g. T cell activation activated by tumour cells can be inhibited by engaging PD1 and CTLA4, however these receptors can be blocked using mAbs. D) Recombinant receptors comprising of scFv

can redirect the specificity of T cells by mediating MHC-independent activation. Such T cells are called chimeric antigen receptors (CAR) T cells. Shown are the four generations of CAR-T cells which attempt to strengthen signalling (co-stimulatory receptors CD28 and CD137), while the fourth generation also introduces separate constitutive/inducible expression of coreceptors and cytokines. E) Certain antibodies, like bispecific T cell engagers (BiTE) can induce T cell activation which is MHC independent as the antibody brings T cells proximal to tumours cells due to their specific to both via CD3 and TAA, respectively. F) Vaccines introduce non-pathogenic antigens which can be presented by APC such as DCs, thus initiating a T cell response which subsequently result in memory immune cells. mAb: monoclonal antibody, TAA: tumour associated antigen, BiTE: bispecific T cell engagers, CTL: cytotoxic T cell, APC: antigen presentation cell.

1.2.2 A challenge in drug development: modelling the human immune system from mouse

One of the challenges in drug development is the lack of translation from animal models to human. Currently, the chances of a drug successfully clearing all clinical trial phases is less than 10% (Dowden and Munro, 2019). Several factors determine the success of a drug including its efficacy, toxicity and risks. Furthermore, one of the requirements is to conduct trials in mouse. Unfortunately, this is also one of the factors that can mislead drug development. Indeed, mouse models have advanced our understanding of diseases as they allow testing which would be unethical/impossible to conduct in human, e.g. engrafting tumours subcutaneously for testing drug treatments (Tentler et al., 2012), and making genetically engineered mouse models to replicate human disease pathologies. A recent success story for drug development using mouse models is for sickle cell disease, where a gene therapy developed in mouse successfully translated to humans (Hoban et al., 2016). Conversely, few drugs have been approved for neurodegenerative diseases due to a lack of translation (Mehta et al., 2017).

Whilst there are similarities between the mouse and human immune system such as the structure and function of immune tissues like the thymus, lymph node and spleen (Haley, 2003), there are also many differences which could explain the lack of translation. There is different gene usage between similar cell types across species, e.g. a study found two thirds of genes are differently expressed across species amongst similar neuronal cell types (Hodge et al., 2019). Certain genes like *APP* (a precursor to Aβ), where mutations of the gene in humans are associated with early onset AD, in mouse the protein is less

likely to aggregate as it differs in three residues (Xu et al., 2015). Pathways/signalling also vary between species, e.g. whilst, IFN and LPS induce iNOS production in mouse macrophages through the arginine metabolism pathway (Arg1, Gch1 and Ass1) the response and signalling is not observed in human macrophages, which express low levels of *ARG1* and *ARG2* (Young et al., 2018). The diversity of immune cells can also vary, as microglia in humans are more diverse relative to mouse strains and highly express genes enriched in AD-associated genes and ferroptosis also linked with AD (Geirsdottir et al., 2019, Li et al., 2020). The distribution of immune cells across certain tissues are also different, e.g. in humans there are a larger proportion of neutrophils in blood (30-50% lymphocyte and 50-70% neutrophils) in contrast to the mouse (75-90% lymphocytes and 10-25% neutrophils) (Doeing et al., 2003). Lastly, mouse models may not perfectly replicate all symptoms of the disease observed in humans, even the use of stable mouse strain has been questioned due to their lack of genetic diversity or exposure to the environment (Beura et al., 2016).

1.3 Myeloid immune cell heterogeneity

Immune cell types have been described in context of their ontogeny, the phenotypes they develop in disease or following stimulation, and the tissues they reside in. Only in the past decade has research on tissue-resident immune cells (TRICs) been expanded upon. This is likely due to initial studies being limited to cell cultures as the complexities of reconstructing the tissue environment are still being explored today, e.g. through organoids and tissue culture techniques (Pampaloni et al., 2007, Walsh et al., 2005).

1.3.1 Monocytes

Monocytes primarily reside in blood and are a direct product of haematopoiesis. Three states have been characterised in human monocytes, the classical (CD14+ CD16-), non-classical (CD14+ CD16+) and intermediate ($CD14^{int}$ $CD16^{int}$). The corresponding states in mouse include the Ly6C+ classical monocytes and Ly6C- non-classical monocytes. Additionally, a recent study examining the transcriptomic profile of monocytes at the single cell level has proposed two other subsets (Villani et al., 2017). As part of the innate response and as first responders to tissue injury, these cells secrete proinflammatory mediators, instigate the adaptive immune response by differentiating into DCs and macrophages which present antigen while also regulating wound healing. For GMPs to become monocytes, they first differentiate into common monocyte progenitor (cMoP) after which they require a combination of signals including the activation of certain transcription factors, PU.1, IRF8 and KLF4, as well as the inhibition of GATA1, GATA2

and CEBPα. The molecules determining the commitment towards the different monocyte subsets are still being determined. At least for non-classical Ly6C- monocytes it is known that one of the routes is through classical Ly6C+ monocytes on activation of CEBPβ and Nr4a1 (Mildner et al., 2017).

Classical monocytes are scavenger cells. They comprise 80-95% of circulating monocytes and are highly phagocytic. These cells are considered more proinflammatory, as they secret the highest amounts of TNFα, IL6, and IL1β in response to TLR agonists (Boyette et al., 2017). Non-classical monocytes represent 2-11% of circulating monocytes and are characterise by pathways involving the complement system, Fc gamma-mediated phagocytosis and wound healing. These cells express higher levels of CX3CR1 which enables them to migrate towards damaged tissues more than other subsets (Ancuta et al., 2003). Intermediate monocytes are defined by their antigen presentation capability, expressing genes like *CD74*, *HLA-DR*, *IFI30* and *CTSB* (Gren et al., 2015). Of the two recent monocyte subsets identified through single cell sequencing, Mono-3 displayed a high expression of genes (*MXD1*, *CXCR1*, *CXCR2* and *VNN2*) thought to influence cell cycle differentiation and trafficking, while Mono-4 possessed a cytotoxic signature (*PRF1*, *GNLY* and *CTSW*) (Villani et al., 2017). Upon external inflammatory signals, classical monocytes can differentiate into macrophages and DCs, while also being one of the sources for tissues resident myeloid cells (Bain et al., 2013). For example, listeria infection of monocytes showed a subset of FcγRIII+CD209a−MHCII−Ly6C+ (non-classical) monocytes preferentially differentiated into iNOS+ inflammatory macrophage (Menezes et al., 2016).

1.3.2 Macrophages

It is well known in the field of immunology that macrophages represent one of the most diverse and heterogeneous cell types as supported by their ubiquitous presence across tissues. These cells are highly phagocytic and as part of the innate immune response are responsible for removing cellular debris, apoptotic cells and pathogens. Once they ingest microbes, they process associated proteins or antigens and present them to Th cells which carry out the adaptive immune response. Studies are continuing to explore the plasticity of macrophages to find novel subsets. Accordingly, macrophages have traditionally been categorised into eight classifications, M1, the four subsets of M2 (M2a, M2b, M2c, and M2d), tumour associated macrophages (TAM), TCR+ macrophage and CD169+ macrophage. However, several studies have challenged these classifications and rather proposed a spectrum of states for macrophages, e.g. on activating monocyte

derived macrophages with various stimuli including those associated with M1 and M2 macrophages, cells attained multidirectional activation states as opposed to the dual M1 and M2 based on their transcriptomic profile (Mosser and Edwards, 2008, Xue et al., 2014). Indeed, we must begin to comprehend these activation states as a combination of different programs, however, as a reference of the functional diversity of these cells, the eight macrophage subsets which have been well characterised are described.

Conventionally, M1 and M2 macrophages were characterised based on their reaction to two subsets of Th cells in mouse, Th1 and Th2, respectively (Mills et al., 2000). Macrophage differentiated into M1 or classically activated macrophages on exposure to IFNγ secreted by Th1 cells which can also be artificially induced by LPS, IFNγ and GM-CSF. The resultant phenotype is considered to be the 'classical' proinflammatory macrophage and is characterised by the production of NO, proinflammatory cytokines (IL1B, IL12, IL18, and TNF) and chemoattractants (CCL9 and CXCL10) for cells like Th1 (Arango Duque and Descoteaux, 2014). This state is governed by various transcription factors (STAT1, IRF5, and SOCS3) which in turn regulate pathways like IRF/STAT, IFN/TLR, NFKB/PI3 pathways (Raza et al., 2014) all of which result in proinflammatory responses and downregulation of immunosuppressive molecules like *IL10*. In contrast,

M2 macrophages or 'alternately-activated' macrophages are considered anti-inflammatory and secrete high levels of IL10. M2 activation states are associated with Th2 secretory molecules and further diversify (M2a - d) (**Figure 1.3**). M2a (wound healing macrophage) is associated with tissue repair and is induced by IL4 and IL13 (Martinez et al., 2008); M2b (regulatory macrophages) expresses higher levels of IL10 than M2a, however, they also produce various proinflammatory cytokines (TNFα, IL1β and IL6) and are induced by immune complexes and LPS (Wang et al., 2019a); M2c macrophages are referred to as deactivated macrophages as they downregulate proinflammatory cytokines and redirect their functions to scavenging, angiogenesis and wound healing (Lurier et al., 2017). These cells are induced by IL10, TGFβ and glucocorticoids; and M2d macrophages are induced by IL6 and adenosine, thereby secreting growth factors (VEGF) and anti-inflammatory cytokines to promote tissue repair, tumour progression and angiogenesis (Rőszer, 2015, Yao et al., 2019). Certain classifications are further divided giving a spectrum of activation states as previously proposed e.g. M2c is divided into M(GC), M(IL10) and M(GC+TGFβ).

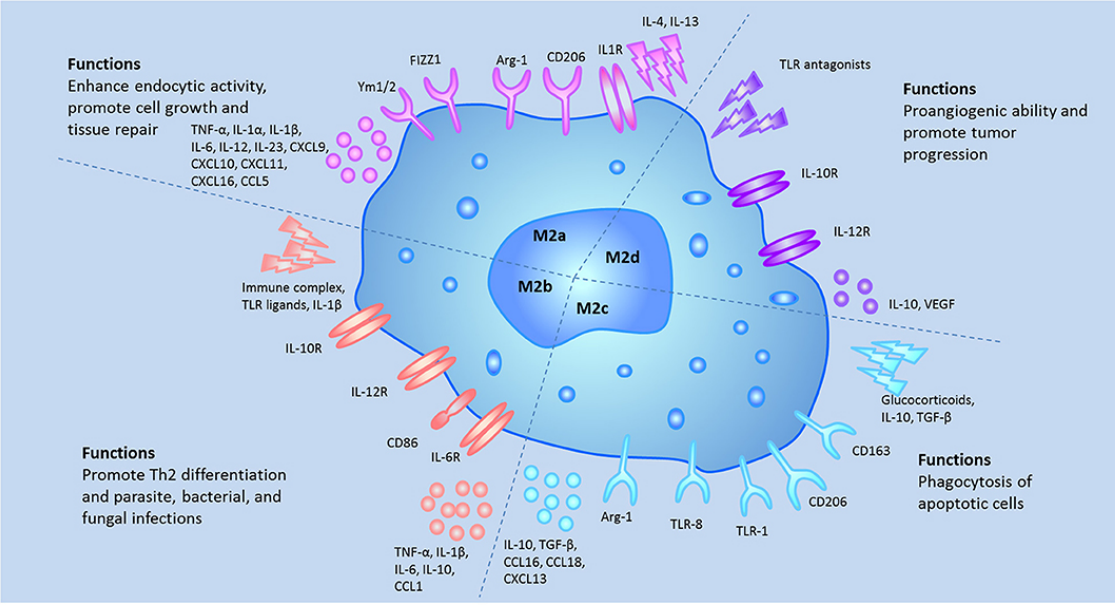


Figure 1.3. The subsets of M2 macrophage.

Shown are the four subsets of M2 macrophages taken from (Yao et al., 2019). The figure highlights the molecules which induce the phenotype, the resultant receptors and cytokine profiles, including the function these subsets carry out.

In addition to the M1 and M2 macrophages are the TAM, TCR+ and CD169+ macrophages. TAMs, as the name suggests, are tissue macrophages at the tumour site or circulating monocytes which on exposure to the tumour microenvironment differentiate into TAMs and have shown to express both M1 and M2 associated genes (Schmieder et al., 2012). Much like other macrophage subsets, TAMs can be subdivided further based on their cytokine profile. These cells are associated with tumour progression, e.g. they secrete various proteolytic enzymes such as serine proteases and metalloproteinase (MMP2, MMP7 and MMP9) which lead to the degradation of the extracellular matrix thereby promoting metastasis (Lin et al., 2019). Hence, therapeutic strategies aim to suppress their immune-suppressive phenotype by reducing infiltration or causing them to differentiate into an immune-active phenotype (Cassetta and Kitamura, 2018). TCR+ macrophages have been identified in mouse and human. These cells express CD3, TCRαβ, and TCRγδ, thought to be unique to T cells (Chávez-Galán et al., 2015). This subset is elicited in early differentiating macrophages with TNFα promoting the phenotype with IL4 or IFNγ redirecting them towards M1 types (Beham et al., 2011). These macrophages also express other genes associated with lymphocytes and have been detected in atherosclerosis, tuberculosis and bacterial challenges (Fuchs et al.,

2015, Fuchs et al., 2013). CD169+ macrophages are located in secondary lymphoid organs and migrate during immune activation. Hence, these cells are more involved in immunomodulation, e.g. they can present antigen to B cells and activate CD8+ T cells, evident from their expression of *MHCII* and *CD68* (Martinez-Pomares and Gordon, 2012). Interestingly, these cells are unable to phagocytose.

1.3.3 Dendritic cells

DCs were first identified as phagocytic cells of the spleen which elicited an antibody response (Steinman and Cohn, 1973), and even now are defined by their efficacy in taking up antigen and present them to T cells. Current studies have revealed the immense diversity of these cell types, including the multiple subsets residing in peripheral tissues. In this section we discuss the seven subsets that have been identified in blood, these include conventional DC1 (cDC1), cDC2 (CD1C_A and CD1C_B), CD16+ DC, Axl DC (CD123+ Axl+ DC and CD11c+ Axl+ DC) and plasmacytoid DC (pDC).

cDC1s constitute only 0.05% of peripheral blood mononuclear cells (PBMC) (Jongbloed et al., 2010). These cells are identifiable by markers such as CLEC9A, CADM1, CD141 and XCR1 (Villani et al., 2017), and their development is regulated by transcription factors such as IRD8, BATF3 and ID2 (Murphy et al., 2016). cDC1s are efficient in presenting antigen and under bacterial and viral infections they prime CD8+ T cells (Theisen and Murphy, 2017). Furthermore, these cells can detect intracellular double strand (dsRNA) and DNA using TLR receptors (TLR3, TLR9 and TLR10) thereby triggering the production of type 1 IFNs and IL12 in an IRF3-dependent manner (Liu et al., 2015). In contrast to cDC1, the cDC2 subset is the most abundant DC of PBMC, with a higher expression of CD11c, and are regulated by a different set of transcription factors (IRF4, NOTCH2 and KLF4) (Murphy et al., 2016). These cells express various lectins (CLEC12A, CLEC4A, and CLEC10A) and TLRs (TLR2, TLR4 and TLR5) (Rhodes et al., 2019). On TLR stimulation, they secrete chemokines (CCL3 and CCL4) and cytokines (TNF, IL6, and IL12) (Piccioli et al., 2007). cDC2 can be subdivided into CD1C_A and cDC_B. CD1C_A highly express FCGR2B, MHC class 2 genes, CD1C and CD11C. Relatively, CD1C_B are similar in their transcriptomic profile to classical monocytes and express innate inflammatory genes (Villani et al., 2017). Functionally, these subsets have shown to trigger different differentiation trajectories in Th cells. CD1C_A polarise Th cells to regulatory T cells (Treg), Th2, Th17 and Th22 phenotypes while CD1C_B induce Th1 (Yin et al., 2017). Villani et al. also described DC4 or CD16+ DCs, which lack expression of typical DC markers such as *CD141* and *CD1c*. Furthermore, these cells

are considered proinflammatory, as on engaging TLR agonists they secrete inflammatory cytokines. However, further studies are required to understand these cells as they show a greater similarity to non-classical monocytes based on their transcriptomic profile, and majority of studies into DC apply gating strategies to remove this CD16+ monocyte subset.

Unlike the cDC, the pDCs are characterised by their production of type 1 and type 3 interferon and cannot induce a T cell response. These cells are specialised towards antiviral responses, detecting sense single-strand RNA (ssRNA) and dsRNA via TLR7 and TLR9, respectively (Gilliet et al., 2008). The TLR pathway is regulated by IRF8 which also governs the production of chemokines (CCL3-5 and CXCL10-11) and induce the NFKB pathway to secrete cytokines (TNFα and IL6) (Honda and Taniguchi, 2006). Interestingly, pDCs were initially isolated using CD123 and BDCA2, however recent studies using scRNA-Seq reveal minor subpopulations of cells within them, termed as Axl+ DCs. Such analyses show the heterogeneity of cells and the power of approaches such as scRNA-Seq in deconvoluting this diversity. Clustering of these cells suggests they represent a continuum between pDC and cDC2 states. Further subdivisions have been observed in these Axl+ DCs, those that are $CD123^{low}CD1c^{high}$ resemble cDC2s and $CD123^{high}CD1c^{low}$ resemble pDCs. This is further supported by their expression of transcription factors *ID2* and *TCF4* required for the commitment towards cDC2s and pDCs, respectively.

1.3.4 Granulocytes

As part of the innate immune system neutrophils mediate immunoregulatory and anti-microbial immunity, e.g. degranulation, phagocytosis and a specialised process called neutrophil extracellular trap (NET). In human, four subsets (CD177+, OLFM4+, CD49d+CXCR4+VEGFR1+ and TCRαβ+) have been described in health with further segregation of aged neutrophils ($CXCR4^{high}CD62L^{low}$). Majority of circulating neutrophils in healthy individuals are CD177+, a glycoprotein which mediates their transmigration out of the vasculature via endothelial contact during inflammation (Sachs et al., 2007). This subset has been associated with various autoimmune diseases, e.g. systematic vasculitis (Deng et al., 2018) and inflammatory bowel disease (Zhou et al., 2018). OLFM4+ neutrophils comprise of 20-25% of circulating neutrophils. This subset of neutrophil presented NETs also stained for OLFM4s (Welin et al., 2013). In sepsis, an increase in OLFM4+ neutrophils correlated with organ failure while OLFM4-deficient models showed resistance to experimental sepsis (Alder et al., 2017, Liu et al., 2012).

CD49d+CXCR4+VEGFR1+ neutrophils are proangiogenic, have increased chemokinesis and their high levels of CXCR4 suggest CXCR4-mediated recruitment of these cells (Massena et al., 2015). The TCRαβ+ neutrophil represents 5-8% of circulating neutrophils. Triggering the TCR negatively regulates apoptosis and results in the secretion of IL8 which is observed in activated neutrophils and mediates further recruitment of neutrophils (Puellmann et al., 2006). The maturation of neutrophils is described by the conversion from young CXCR4-CD62L+ to CXCR4+CD62L- aged cells, the balance of which is governed by the circadian clock (Adrover et al., 2019). Neutrophils become CXCR4+ and CD62L- before they migrate to the bone marrow or spleen to be phagocytosed or undergo apoptosis. Furthermore, the abundance of the subset correlates with greater vascular damage as observed in a mouse model (Adrover et al., 2019). Similar to macrophages, neutrophils display numerous activation phenotypes under pathological conditions. However, a formal comparison of these states is still required as they are currently defined by the condition/disease. Some of these include anti-inflammatory neutrophils (CD11b+CD49d-IL10+), proinflammatory neutrophils (CD11b-CD49d+L12+) and those expressing receptors for antigen presentation ($CD11b^{high}CD62L^{low}CXCR2^{low}$) (Silvestre-Roig et al., 2016, Hampton et al., 2015).

Eosinophils are granulated cells and provide immunity for parasitic infections, asthma and allergies through phagocytosis, degranulation and antigen presentation. These cells are in a small proportions in the bone marrow (6%), however, can increase in number in response to IL3, IL5 and GM-CSF (Ramirez et al., 2018). Subsequently, they migrate to inflamed tissues and contribute to the proinflammatory response by producing reactive oxygen species, cytokines (TNFα, TGFβ, IL1α and type 2: IL4, and IL13) and cationic proteins via degranulation (Hogan et al., 2008). The diversity of eosinophils has been described in various tissues and much like neutrophils their activation states are identified based on the disease, rather than a comparison across states. In a tumour model, eosinophils were essential in the recruitment of T cells via the production of CCL22, suggesting a role in metastasis (Zaynagetdinov et al., 2015). In a model of chronic colitis, eosinophils were crucial in promoting inflammation in a GM-CSF dependent manner, through various cytokine IL13, TNFα, and EPX (Griseri et al., 2015).

Basophils are known for their role in allergies, barrier inflammation and anti-helminth immunity. These rare populations comprise <1% of PBMCs. Due to their similar morphology and function to mast cells, they were primarily analysed to understand mast

cell biology. However, subsequently their relevance in mediating anti-helminth immunity was realised (Schroeder, 2009). GATA2, STAT5, and IRF8 have shown to be key transcription factors in basophil and mast cell development. These cells have been proposed to originate from different progenitors depending on the tissue, these include the spleen (basophil-mast cell progenitor) and bone marrow (pre-basophil and mast cell progenitors) (Arinobu et al., 2005, Li et al., 2015). Studies describe two basophil subsets, those elicited by IL13 which have been extensively studied and the recently identified thymic stromal lymphopoietin (TSLP) stimulated subset (Siracusa et al., 2011). Both IL13-elicited basophils and TSLP-elicited basophils mediate a type 2 response producing cytokines like IL4 and IL13. A notable functional difference is the IgE dependent degranulation of IL3-elicited basophils resulting in the release of histamines. This response is relatively suppressed in TSLP-elicited basophils. Further differences are observed in terms of their receptors, cytokines and chemokines, e.g. TSLP-elicited basophils have higher levels of IL33, IL6, CCL3 and CCL4 relative to IL13-elicited basophils. Further studies are required to expand on their functional differences.

Similar to basophils are mast cells, both in terms of their origin and function. The heterogeneity of mast cells has been described by the protease composition of their secretory granules, and these subsets are differently distributed across tissues. Three subsets of mast cells have been described, those containing tryptases (lung and gut); tryptases and carboxypeptidase (lung); and tryptases, chymase and carboxypeptidase (skin and peritoneum) (Dougherty et al., 2010). A recent study identifies two additional subsets distinguished by the expression of CD25, although these still require a comprehensive comparison with other subsets (Deho et al., 2014). Nonetheless, these relatively proliferative CD25+ mast cells expressed greater levels of certain cytokines (IL6 and TNFα), proteases (*CTSG* and *PRSS34*) and preferentially expressed the transcription factor *KLF1* relative to CD25-.

1.4 Lymphoid immune cell heterogeneity

1.4.1 T cells

T cells are part of the adaptive immune system and are characterised by TCR expression which recognises antigen-MHC complexes presented by APCs. As a result, T cells can respond in various ways depending on the context of engagement, their ontogeny and/or history. T cell development involves the selection of the TCR chain loci which can either generate αβ- (95% of T cell) or γδ-TCR T cells based on the respective TCR chain loci.

From haematopoiesis in the bone marrow, T cell progenitors migrate to the thymus through sequential states. These double negative CD4-CD8- thymic cells ultimately undergo selection where those successfully rearranging their TCR chain locus survive. These cells follow very different fates thereby resulting in different cell types. On successful TCR rearrangement immature αβ thymocytes become double positive CD4+Cd8+ and go through positive and negative selection based on their response to self-antigen in the context of MHC. Consequently, the process generates either mature CD4+ T cells (Th cells) that are MHC class 2 specific or CD8+ T cells (cytotoxic T cells) that are MHC class 1 specific (Germain, 2002), which can be further subdivided. Upon antigen stimulation naïve CD4+ Th cells can differentiate into various subsets depending on the cytokines they are exposed to (**Figure 1.4**). These include Th1, Th2, Th9, Th17, Th22, follicular Th cells, Tregs. Th1 and Th2 correspond to type 1 proinflammatory cytokines, e.g. IL12, IFNγ and TNF and type 2 anti-inflammatory cytokines, e.g. IL4, IL5

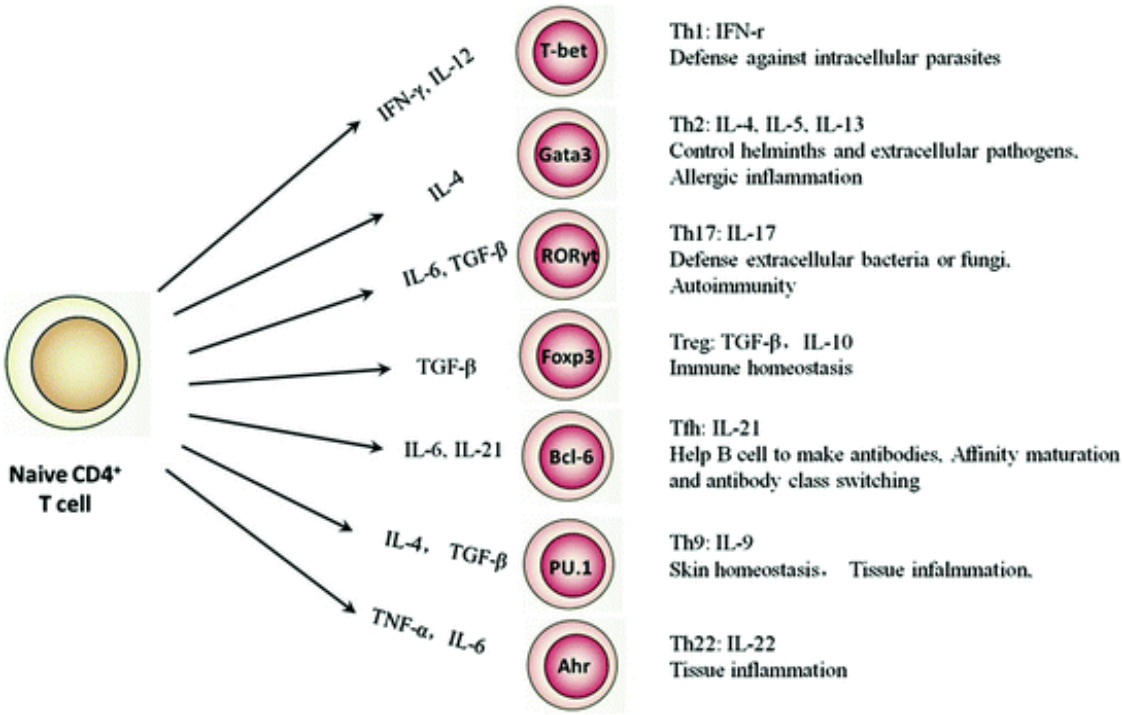


Figure 1.4. CD4+ T cell subsets

This figure was taken from (Sun and Zhang, 2014) which shows how Naïve CD4+ T cells can differentiate into various subsets based on the cytokines they are exposed to. The respective cytokines regulate transcription factors resulting in the production of certain cytokines and propagating specific immune responses.

and IL13, respectively. Other subsets like Th9, Th17 and Th22 correspond to the cytokines they produce, IL9, IL17 and IL22, respectively (Golubovskaya and Wu, 2016). Unlike CD4+ T cells, CD8+ T cells can detect antigen bound MHC-1 and release cytotoxic granules which result in apoptosis of infected cells. These cytotoxic T cells include two subsets Tc1 and Tc2 corresponding to the type 1 and type 2 responses. As in macrophages and CD4+ T cells, type 1 is associated with IL2, IL6 with the ability to produce high amounts of IFNγ, while type 2 are defined by their secretion of IL4 and IL5 (Kemp et al., 2005). On activation these T cell subsets differentiate into effector cells which are short-lived, as well as long-lived memory T cells which reside either in tissues (effector memory T cells) or lymphoid organs (central memory T cells). The memory T cells engage with antigen to produce a faster response in terms of proliferation and cytokine production than effector cells (Rosenblum et al., 2016).

Much like the $\alpha\beta$T cells, the $\gamma\delta$T cells can polarize into various activation states depending on the stimuli. $\gamma\delta$T cells have shown type 1 and type 2-like phenotypes on exposure to IL12 and IL4 respectively in the presence of isopentenyl pyrophosphate (Wesch et al., 2001). Other similar subsets include follicular helper $\gamma\delta$T cells (Crotty, 2011), IL17 $\gamma\delta$T cells (Caccamo et al., 2011) and regulatory $\gamma\delta$T cells. The latter has shown to be a stronger immunosuppressor than CD4+ $\alpha\beta$Treg in colorectal cancer (Wu et al., 2017). Interestingly, $\gamma\delta$T cells can also present antigen during microbial infections and stimulate CD4+ and CD8+ $\alpha\beta$T cells via MHC (Brandes et al., 2005).

1.4.2 B cells

B cells function as part of the humoral immunity as they are the main source of antibodies. In addition, these cells present antigen and can secrete various cytokines on activation. Much like T cells, the development of B cells goes through various rounds of selection for the B cell receptor (BCR). In the bone marrow these cells sequentially differentiate through pro-B cells, pre-B cells to immature cells, through which they are positively and negatively selected for self-antigens (LeBien and Tedder, 2008). Subsequently, the immature B cells migrate to the spleen as transitional B cells, consisting of T1 and T2 B cells. T1 B cells refer to those migrating to the spleen. Once within the spleen they transition to T2 which in turn can differentiate into mature marginal zone (MZ) B cell or follicular B cells depending on the context of BCR signalling in T2 B cells. (Chung et al., 2003). Follicular B cells reside in follicles of secondary lymphoid tissues and during infection can differentiate into plasmablasts or germinal centre (GC)

B cells. Plasmablasts have a high affinity towards antigens, secrete large numbers of antibodies, and are proliferative, however these cells are also short-lived. GC B cells undergo various changes (class switching and somatic hypermutations) to convert to plasma cells which produce specific reactive antibodies. In contrast to plasmablasts, which are formed early on in infections, plasma cells are formed later on and are long-lived (Vale et al., 2015). GC B cells also generate memory B cells which are in circulation and ready to respond to pre-exposed antigens to begin generating antigen-specific antibodies. As the name suggests, MZ B cells reside in the marginal zone of the spleen. These cells are more innate-like, as they readily respond to TLR stimuli during infection mediated by proximal APCs like marginal sinus associated macrophages. Like Tregs, there are also immunosuppressive regulatory B cells. These cells produce IL10 and TGBβ to supress proinflammatory responses while promoting differentiation of Tregs (Rosser and Mauri, 2015). Primarily arising from foetal and neonatal progenitors are B-1 B cells, which are further divided into CD5+ B1-a and CD5- B1-b. These cells migrate through the spleen homing in on the peritoneal cavity. Much like MZ B cells, B1-a cells are fast reacting innate-like cells involved in early protection and immediately produce natural antibodies (polyreactive with low affinity and are generated before antigen exposure). Alternately, B1-b cells produce antibodies after antigen exposure (Haas et al., 2005).

1.4.3 Innate lymphocytes and NKT cells

The nomenclature for innate lymphoid cells (ILCs) was introduced recently between 2009 and 2010, where studies reported NK cell-like lymphoid cells (Spits et al., 2013). Further research is being conducted into the ontogeny of these cells which suggest a non-MHC mechanism within the bone marrow mediated through transcription factor ID2 which counteracts B and T cell lymphopoiesis (Cherrier et al., 2012). The resultant immature ILCs are found to circulate in peripheral blood and in various tissues, where they can be primed based on environmental ques (Bal et al., 2020). Hence, ILC subsets are distributed differently across tissues (Yudanin et al., 2019). Studies have identified five subsets of ILCs including, IL1, ILC2, ILC3, lymphoid tissue inducer cells (LTi) and NK cells. ILC1, ILC2 and ILC3 bare resemblance to Th1, Th2 and Th17/Th22 cells, respectively. ILC1s are dependent on the transcription factor TBX21 (T-bet) and characterised by their production of IFNγ (Spits et al., 2016). ILC2s are defined by GATA3 and produces IL4, IL5, and IL13. ILC3s are dependent on RORγt and produce IL3 and IL22 with a proportion also expressing IL17 (Cupedo et al., 2009). LTi's are

required for organogenesis, including the development of lymph nodes and Peyer's patch (Veiga-Fernandes et al., 2007). NK cells are part of ILCs and also possess immunoregulatory roles. These cells produce various cytokines and cytolytic granules on detecting MHC1, thus generating stress and apoptotic signals. NK cells originate from CD34+ HSCs via growth factors (FLT3- and ckit-ligand), which generate CD34+CD122+CD56− immature NK cells (iNK cells) (McKenna et al., 2000). These cells then mature to CD3-CD56+ NK cells by responding to IL15 via CD122. Mature NK cells are further divided into $CD56^{dim}$ and $CD56^{bright}$ cells. $CD56^{dim}$ cells are the major circulating cell populations which express KIRs and are highly cytotoxic relative to $CD56^{bright}$ NK cells. The latter are localised to lymphoid organs and are characterised by the production of pro-inflammatory cytokines like IFNγ and TNFα (Pende et al., 2019).

NKT cells are innate-like T cells that develop in the thymus and accordingly have a TCR. However, these cells recognise antigens in the context CD1D which is a non-polymorphic MHC class 2 molecule (Bendelac et al., 1995). Like T cells, NKT cells can be subdivided based on their TCR chain usage into type 1 (invariant) which detects the TCR-αGalCer/CD1d complex and type 2 (diverse) which recognise the TCR-sulfatide/CD1d complex. Type I are further divided into NKT1, NKT2, and NKT17 cells which correspond to Th cell subsets (Watarai et al., 2012). Furthermore, similar to T cells they can respond to co-stimulatory signals from CD28 (Shissler et al., 2020) while also possessing various NK cells receptors and their cytotoxic capabilities (Makino et al., 1995). In contrast, type 2 NKT cells are immunoregulatory, as they supress proinflammatory response of type 1 NKT cells (Arrenberg et al., 2011).

1.5 Tissue resident immune cell heterogeneity

Depending on environmental cues cells can undergo changes at the transcriptional and epigenetic level thus developing tissue-specific homeostatic functions. Some TRICs include immune subsets described in the previous sections and are differently distributed across tissues. In the following section, the various TRIC populations are described, while highlighting certain aspects tissue-resident macrophages as these populations have been well studied, including their ontogeny, how they are sustained through life and the unique phenotypes that can develop relative to their tissue naïve counterparts.

1.5.1 The haematopoietic waves which generate tissue-resident macrophages

Macrophage subsets originate from haematopoiesis, which can be subdivided chronologically into three waves, primitive, transient definitive and definitive haematopoiesis. The first wave or primitive haematopoiesis begins in the yolk sac on embryonic day 7.5 (E7.5) to E8.5 in mouse. The early Csf1r^{high} c-Myb- erythroid-myeloid progenitors (EMP) originate from the posterior plate mesoderm in blood islands of the yolk sac (YS). In this phase, cells have a limited differentiation potential and give rise to primitive macrophages, erythrocytes and megakaryocytes (Palis et al., 1999, Tober et al., 2007). Even before infiltrating the brain, primitive macrophages have been shown to differentiate into microglial cells within the yolk sac (Ferrero et al., 2018). Of all the tissue-resident immune cells, microglia remain the only cell type originating from YS-derived macrophages which can self-renew independent of circulating monocytes once they have migrated to the brain. Supported by fate-mapping studies labelling RUNX1+, YS-derived cells are abundant in the mouse brain through to adulthood, whilst they are gradually replaced by unlabelled cells in other tissues (Ginhoux et al., 2010, Hoeffel et al., 2012).

The second wave, termed as transient definitive haematopoiesis, from E9.5, involves late Csf1r^{low} c-Myb+ EMPs and lymphoid-myeloid progenitors (LMPs) originating from the hemogenic endothelium of the YS (McGrath et al., 2015). These progenitors have a broad differentiation potential once they migrate to the fetal liver, as late EMPs can further differentiate into macrophages, monocytes and granulocytes, and LMPs into lymphoid cell lineages (Palis et al., 1999, Adolfsson et al., 2005). Further differences are observed between late EMPs and early EMPs. Runx1 and Csf1r fate mapping models suggest that late EMPs and their derived macrophages largely replace early EMPs and give rise to tissue-resident macrophage populations, e.g. the lung, skin, and kidney, but not microglia (Hoeffel et al., 2015). However, whether these two populations are distinct, or part of a continuum remains to be investigated. Studies have proposed that late EMPs can give rise to tissue-resident macrophage subsets through two intermediate progenitors, fetal liver EMP-derived monocytes and pre-macrophages (Mass et al., 2016). The recent study by Mass *et al.* shows pre-macrophages migrating into different tissues in a Cx3cr1-dependent manner and differentiating into respective tissue-resident macrophages based on environmental cues, e.g. Id3 is a key transcriptional factor for macrophage differentiation towards Kupffer cells, as indicated by Id3 deficient mice that

show a significant reduction in Kupffer cells while other tissue-resident macrophage populations are maintained. Further studies are required to identify the differences between late EMP monocytes and pre-macrophages, their differentiation trajectories and how they contribute to different tissue-resident macrophage populations.

The final wave of definitive haematopoiesis involves HSCs from the hemogenic endothelium of the aorta, gonads and mesonephros region at E10.5 (Bertrand et al., 2005). These cells have the potential to differentiate into myeloid and lymphoid lineages once they migrate to the fetal liver. Once the bone marrow cavity is formed around E17.5, definitive haematopoiesis begins to shift to the bone marrow where it continues throughout adulthood. HSCs are separated from EMPs by markers like Sca1 and unlike EMPs the development of macrophages from HSCs is strictly dependent on a monocyte intermediate (Oguro et al., 2013). With fetal liver monocytes arising from HSCs and EMPs, their respective contributions towards macrophage subsets and how that compares with the newly identified pre-macrophages are points which require further research. Interestingly, in addition to the brain, not all organs are entirely populated by bone marrow HSCs, these include the epidermis, liver and lung which are largely renewed from precursor which originate from the fetal liver and can be considered as "closed" systems (**Figure 1.5**) (Hashimoto et al., 2013, Merad et al., 2002). In certain tissues, macrophage contribution changes over time, e.g. through adulthood the contribution from bone marrow-derived monocytes increases in the gut, heart and pancreas (Ensan et al., 2016, Ginhoux and Guilliams, 2016, Bain et al., 2014, Epelman et al., 2014a). Furthermore, during inflammation even closed systems such as the brain are infiltrated by circulating monocytes, gaining a microglial-like phenotype (Leone et al., 2006) with similar findings in microglial depletion studies (Ajami et al., 2011, Sere et al., 2012, Lund et al., 2018). Hence, with the different waves of haematopoiesis the field is open to questions regarding the contribution of respective precursors towards the different tissue-resident macrophage, their differentiation pathways, the regulatory networks governing them and if their contribution changes over time and in inflammation.

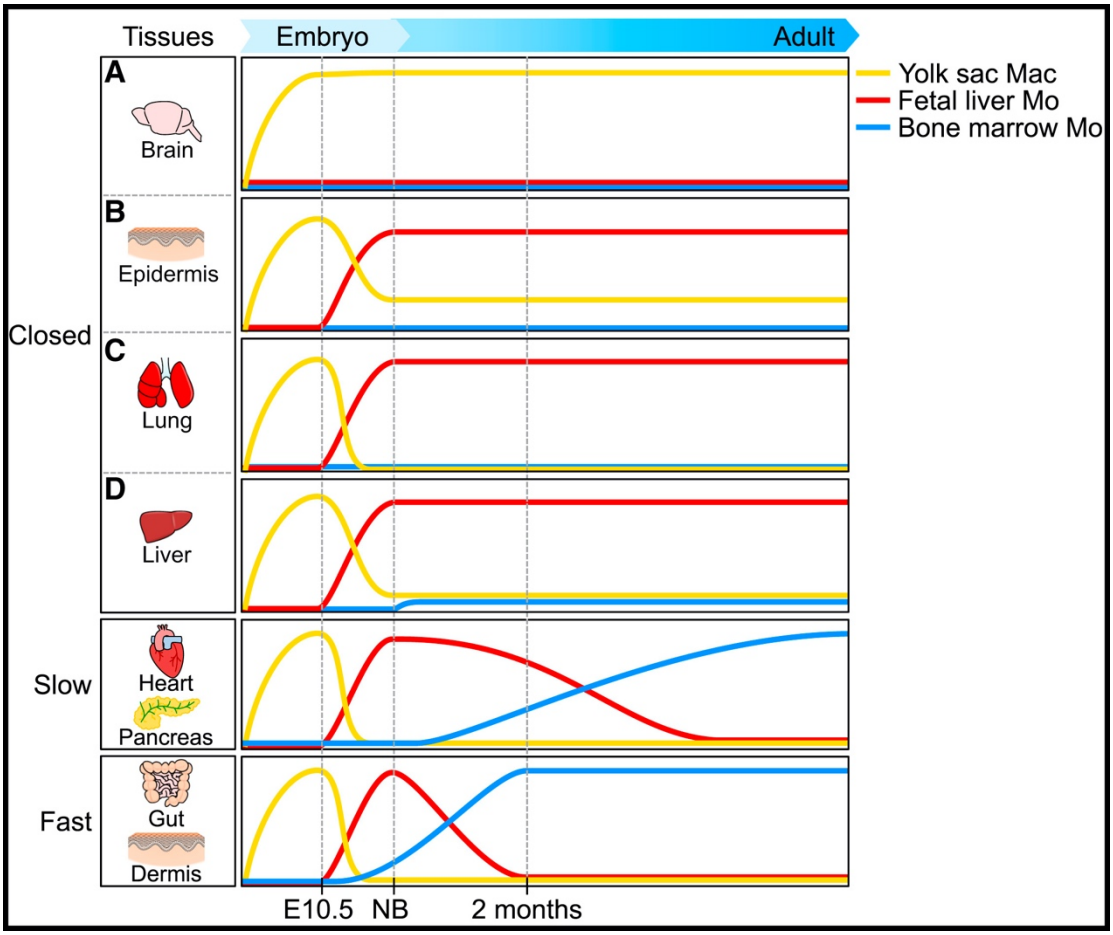


Figure 1.5. The contribution of progenitors from different compartments to tissue macrophages.

Taken from (Ginhoux and Guilliams, 2016) the figure shows the contribution of yolk sac macrophages, fetal liver macrophage and bone marrow macrophages towards the various tissues (left) from embryogenesis to adults. Certain tissues are closed systems like the brain epidermis, lung and liver which are derived from the yolk sac and foetal populations and are maintained through adulthood. Conversely tissues like the heart, pancreas, gut and dermis are derived from different precursor populations with time as these systems are open. Mac: Macrophage, Mo: Monocyte.

1.5.2 Precursors determine tissue colonisation and maintenance of macrophages

Recent studies have suggested that certain macrophage subsets are determined by their ontogeny. Relative to bone marrow-derived macrophages, embryonic-derived macrophages resist genotoxic stress giving them a survival advantage (Price et al.,

2015). These cells also show a self-renewal capacity which is imprinted as their proliferation in GM-CSF continues for weeks unlike bone marrow-derived monocytes, irrespective of the fact that both cell types present a similar GM-CSF receptor sensitivity (van de Laar et al., 2016). Similar observations have been observed in a depletion study of alveolar macrophages, where fetal liver monocyte-derived macrophages more efficiently occupied the lung relative to those from the bone marrow (van de Laar et al., 2016). Embryonic-derived macrophages also have longer life spans, e.g. the lifespan of intestinal and dermal bone marrow-derived macrophages is 4-6 weeks, while that of embryonic-derived cardiac macrophages is 8-12 weeks (Ginhoux and Guilliams, 2016). Interestingly, in certain cases bone marrow-derived macrophages have displayed a capacity to resemble embryonic-derived tissue resident macrophages as well. bone marrow-derived monocytes which infiltrate the lung and liver differentiate into alveolar and Kupffer cells, respectively, with few genes differently expressed relative to their corresponding embryonic-derived tissue-resident macrophages (Scott et al., 2016, Gibbings et al., 2015). Additionally, while bone marrow-derived monocytes are considered to have a poor self-renewable capacity as indicated by their progressive replenishment in tissues like the intestine and heart, depletion studies of Kupffer cells suggest the contrary, as bone marrow monocyte-derived Kupffer cells self-maintain for months (Scott et al., 2016). Therefore, precursors of macrophages determine their phenotypes through processes such as imprinting. To what extent these programs are transferred to bone marrow derived subsets based on conditions like cell depletion or inflammation, and why this differs between tissues are questions which require further research.

1.5.3 Tissue-resident macrophages

As infiltrating monocyte-derived macrophages are known to adopt tissue macrophage phenotypes, it is evident that the microenvironment has a strong impact in defining cell identity (Lavin et al., 2014b, Gosselin et al., 2014). Accordingly, macrophages are highly heterogeneous across tissues and adopt multifunctional roles in maintaining tissue homeostasis. For example, bone or osteal macrophages have a unique multinucleated morphology, proposed to increase bone resorption efficiency (Bar-Shavit, 2007). In the brain, microglia regulate neuronal synaptic formation and signalling, through the process of synaptic pruning and regulation of myelin sheath formation, while also removing plaques implicated in neurodegenerative diseases (Perry and Holmes, 2014). Alveolar macrophages are involved in recycling surfactant molecules, neutralizing pathogens and

phagocytosing allergens and apoptotic cells (Forbes and Haczku, 2010). As this tissue is in frequent contact with external stimuli, a balance is required between initiating immune responses towards agonists and opportunistic pathogens vs. tolerating innocuous stimuli. Supporting such functionality, depletion experiments for alveolar macrophages during bacterial infection show an increase in inflammation and pathogen burden (Archambaud et al., 2010). Interestingly, although the gut is frequently exposed to external stimuli, gut macrophages are derived from circulating Ly6^{high} monocytes and are short-lived relative to alveolar macrophages which are derived from foetal-monocytes (Faria et al., 2017, Varol et al., 2015). In contrast to alveolar macrophages, depletion of Kupffer cells reduces hepatic inflammation and apoptosis, while their activation can cause hepatic damage (Traeger et al., 2010, Dixon et al., 2013). Furthermore, Kupffer cells are able to clear red blood cells undergone oxidative damage using specialized scavenger receptors and also remove haemoglobin containing vesicles, features specific to the liver (Willekens et al., 2005). Apart from function, these populations vary in turnover rate, ontogeny, the source of renewal and longevity, as discussed previously

Diversity amongst tissue-resident macrophage populations has also been investigated through their transcriptomic and epigenetic programs. A study from the ImmGen, comparing mouse-derived macrophages from the liver, peritoneum, brain, splenic, and lung, shows great diversity in the genes expressed between macrophages from different regions (Gautier et al., 2012). Hundreds of genes presented a two-fold differential expression for macrophages specific to a given tissue, e.g. *Cx3cr1* in microglia, the associated histone of which has shown to be exclusively acetylated (Lavin et al., 2014a). Alternately, certain genes were uniform across cell types, such as *Cd14*, *Tlr7* and *Ctsd*. Research is still being conducted to describe the transcriptional networks which determine tissue-specificity, e.g. Spic (Spi-C) is required for differentiation towards red pulp macrophage found in the spleen, *Gata6* is a regulator for peritoneal macrophage and *Sall1* has been implicated in microglial development (**Figure 1.6)** (Kohyama et al., 2009, Lavin et al., 2014a, Mass et al., 2016). In the case of *Gata6*, tissue-specific localization and phenotype has shown to be reversibly induced by modulating *Gata6* expression. However, this did not prove so for development, suggesting different mechanisms for acquiring tissue-specific cell identity in development and for maintaining tissue populations (Okabe and Medzhitov 2014).

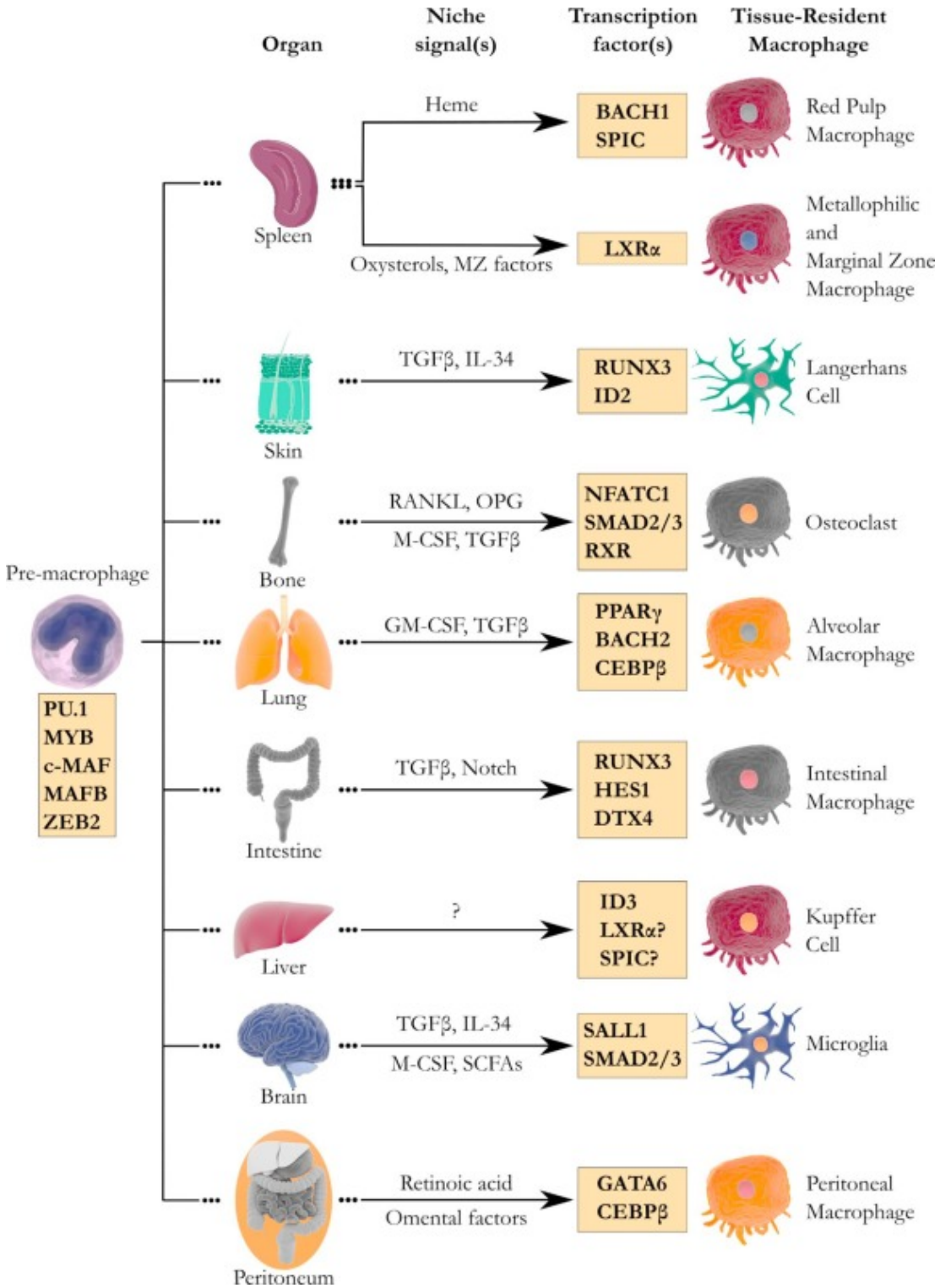


Figure 1.6. Molecular signals determining tissue-resident macrophages

Taken from (T'Jonck et al., 2018). The figure shows the macrophages precursor or pre-macrophage and their defining (orange box) transcription factors. Also shown are the tissues

they migrate to and the molecular signals involved in their differentiation towards the respective tissue-resident macrophages.

1.5.4 Tissue-resident dendritic cells

DC populations are highly diverse across tissues with further studies required for standardising the nomenclature of these subsets. Here we describe some of the main tissue-resident DC populations which largely function in acquiring antigen and migrating to present them to T cells. DCs have been identified in non-lymphoid tissues (liver, lung, gut, kidney and skin) and lymphoid tissues (spleen, thymus and lymph nodes). Furthermore, although extremely heterogenous tissue-resident DC have been mostly described in the context of inflammation as oppose to their tissue homeostatic functions. In lymphoid organs, two key populations of cDC exist CD4+ and CD8+ DCs which present MHC1 and MHC2, respectively. These cells are essential in priming T cell populations by presenting antigen (Henri et al., 2001). Of the gut DCs, CD103+CD11b+ DCs are well characterised. These cells secrete IL6 and TGFβ required for Th17 differentiation, which are essential for the mucosal barrier (Denning et al., 2011, Blaschitz and Raffatellu, 2010). The lung has an assortment of DCs such as cDCs, pDCs and moDCs. Studies also differentiate cDC populations in the lung based on the CD103 marker. CD103+ cDCs represent the dominant population in the lung which are efficient in cross-presenting antigens and in phagocytising dead cells to present to CD8+ T cells (Desch et al., 2011). The skin and associated dermis are home to several DC populations defined by CD207 and CD103 (Henri et al., 2010). The CD207+ DCs or Langerhans cells are essential in tissue injury as they are highly phagocytic and express chemokine receptors (CCR6 and CCR7) required to migrate to lymph nodes, eventually presenting antigens to T cells (Dieu et al., 1998). CD103+ CD207+ DCs are effective in cross-presenting antigen while CD103-CD207+ induce Treg differentiation (Guilliams et al., 2010).

1.5.5 Tissue-resident neutrophils

Conventionally, neutrophils have been thought to exist in the bone marrow, blood, lung and spleen and during inflammation they migrate to the site of inflammation as part of the innate immune response. However, recent studies have shown a wider distribution of these cells across tissues (liver, thymus, brain and kidney) as these cells actively invade certain tissues (Becher et al., 2014, Casanova-Acebes et al., 2018). Only few of the tissue-resident neutrophils have been examined for their tissue homeostatic

functions. For example, splenic neutrophils produce cytokines which promote B cell differentiation and antibody production (Puga et al., 2011). Neutrophils, especially aged neutrophils which express CXCR4, preferentially migrate to the lungs which has relatively large numbers of neutrophils. Furthermore, neutrophils migrate into the lung in a circadian pattern, although the function of these are unknown (Zhang et al., 2014). Studies have also shown the phagocytosis of neutrophils by macrophages results in cytokine secretion, remotely regulated the bone marrow niche (Casanova-Acebes et al., 2018). .

1.5.6 Tissue-resident lymphocytes

Studies into conventional tissue resident $\alpha\beta$T cells mostly describe tissue-resident memory T cells derived from either CD4+ and CD8+ T cells. Of these subsets cytotoxic CD8+ T cells have been extensively studied mainly due to their protective role against tumour cells. Tissue-resident memory T cells have been identified in several tissues such as the brain, lung, gut, salivary glands and female reproductive tract. Much like other lymphocytes, effector T cells are recruited to the site of inflammation, where they can be identified by their expression of KLRG1 (Herndler-Brandstetter et al., 2018). Post-pathogen clearance, a subpopulation of these cells down-regulate KLRG1, these represent the precursors to tissue-resident memory T cells. These cells are also identified by markers CD69 and CD103, as well as transcription factors such as Blimp-1, Eomes, Hobit and T-bet required for their differentiation and tissue retention (Mackay et al., 2013, Mackay et al., 2016). In other tissues they may be controlled by additional regulators, e.g. in the CNS tissue-resident memory T cells are regulated by CD274 (B7-H1), required for its long-term maintenance (Pavelko et al., 2017). Tissue-resident memory T cells provide an innate-like protective role as they are one of the first responders to viral infections across tissues, providing a stronger immune response to circulating memory T cells as is seen in the skin (Gebhardt et al., 2009). They produce a combination of cytokines IFNγ and TNFα in a tissue-dependent manner and their accumulation is associated with a reduced pathology, as observed in the brain in *Toxoplasma gondii* infections (Landrith et al., 2017), viral infections of respiratory tract (Teijaro et al., 2011) and various cancers (Koh et al., 2017, Duhen et al., 2018). The longevity of tissue-resident T cells depends on the tissue, surviving in nasal tissues for up to three months and a year in skin after infection (Zaid et al., 2014, Pizzolla et al., 2017).

Unconventional T cells which are nonpolymorphic, have also been found to be tissue-resident such as invariant NKT cells (iNKT) and γδT cells. iNKT cells are found in the liver and adipose tissue, where in the latter they have shown to display markers of tissue-resident memory T cells, CD69, as well as regulate Treg and macrophage population through IL2 and IL10 secretion (Lynch et al., 2015). Tissue-resident γδT cells are derived from thymic precursors which migrate to tissues such as the skin, lung and reproductive tract, where they are most enriched. In the skin these cells secrete IL17 and IL22, while also displaying tissue-resident memory cell markers like CD69 and CD103 (Jiang et al., 2017). Tissue-resident memory γδT cells have shown to respond to infections with higher levels of IL17 in the lung and liver compared to naïve cells (Misiak et al., 2017, Romagnoli et al., 2016).

The subsets of ILCs have been mostly described in terms of infections and are differently distributed across tissues, e.g. ILC2s have a type 2 response during helminth infections which promotes tuft cell differentiation via IL13 which is enriched in mucosal sites (von Moltke et al., 2016). IL22 production by ILC3 is found to be essential in defending against gut bacterium (Rankin et al., 2016). Tissue-resident NK cells have been identified in adipose tissue, kidney, liver, lung, skin and uterus, with CD49a serving as a marker (Sojka et al., 2014a, Sojka et al., 2014b). Interestingly in mouse, certain subsets do not express Eomes and are defined by transcriptional factors such as Plzf, Hobit (ZNF683) and T-bet. However, this analysis needs to be revisited in human as studies in human liver have shown the opposite, where ITGA1- (CD49A) cells are the resident liver NK cells and express EOMES (Aw Yeang et al., 2017). In the mouse liver these cells also present memory potential for hapten specific responses (Zhang et al., 2016). Certain populations also show self-renewal capacity, like in the uterus, of the circulating and tissue-resident NK cells only the tissue-resident population was able to proliferate (Sojka et al., 2018).

1.6 Defining a cell types in the age of single cell multi-omics

What is a cell type and how do we define it? Such questions have been the basis of several discussions into the identity of the cell, the conclusion to which have constantly evolved over time, largely driven by advances in technology. Preliminary descriptions of cells using light microscopy were based on their morphology. Subsequently, various staining techniques were developed to highlight certain cell types, e.g. the neuron doctrine of Santiago Ramón y Cajal (Jones, 1999) revealed various types of neurons

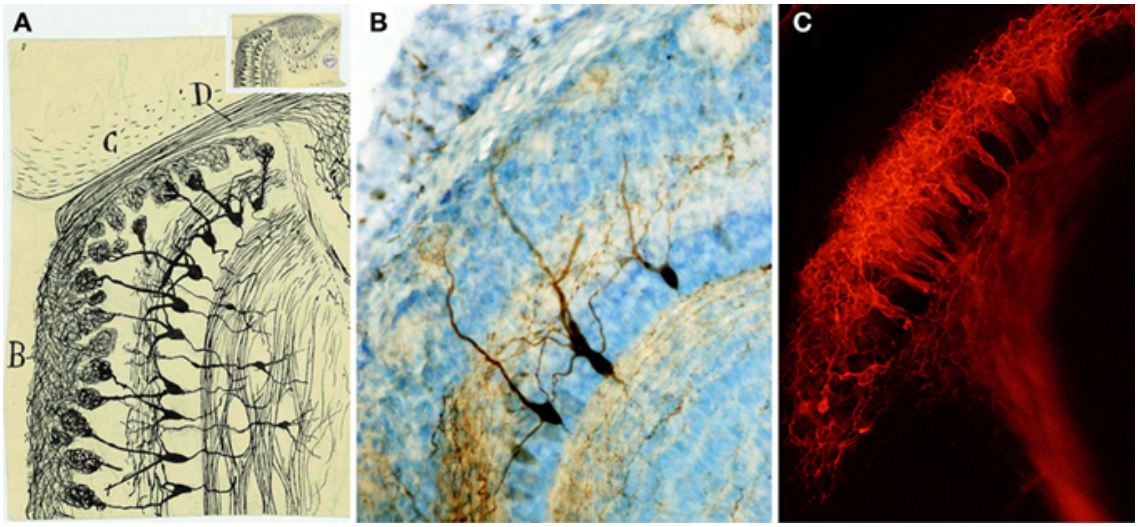


Figure 1.7. Describing the brain cytoarchitecture through the ages

Taken from the (Figueres-Oñate et al., 2014). A horizontal section of the mouse olfactory bulb as described through the ages where neurons were stained for, using A) Golgi stains subsequently drawn by Cajal, B) biotinylated dextran amine injection specific to neuronal cells, C) fluorescent staining through DiI injection.

based on their morphology which were drawn from stainings using the Golgi silver staining method, with certain structures, such as Cajal bodies in neurons, still being used as a reference in studies today (**Figure 1.7)**. Further development enabled staining of cells based on their expression of specific receptors using fluorescently tagged antibodies to track their localisation in living sections (Coons et al., 1941, Peel et al., 1997). Even at this stage there was a continues revaluation e.g. earlier markers for Th cells were found to stain myeloid lineage cell types under different conditions (Wood et al., 1983). In the current age, cells may be defined through multiple dimensions spanning various omic technologies.

New technologies, such as single cell omic platforms have enabled us to investigate the heterogeneity of cell types which can exist as discrete subsets, as well as a continuum of states. This has been shown for cells going through the different phases of cell cycle (Hsiao et al., 2020) and those going through haematopoiesis, gradually acquiring unipotency (Pellin et al., 2019). Certainly, integrating such technologies will help us gather the data needed to define cell types, e.g. a recent study in acute myeloid leukaemia used scRNA-Seq and genotyping to distinguish cell types and evaluate if they

were malignant (van Galen et al., 2019). Indeed, integrating such data will enable researchers to conduct cross study comparisons to evaluate cellular heterogeneity based on our current resources/databases. Projects generating datasets derived from a diverse set of cell types are already being initiated, like the Human Cell Atlas (Regev et al., 2017) and mouse cell atlases (Schaum et al., 2018, Han et al., 2018). However, studying cells in this high-dimensional space poses its own problems in terms of integrating these different omics data and standardising approaches to compare across studies, such that we have a common ontology to define cell types.

Researchers have also developed high-dimensional approaches such as cytometry by time-of-flight mass spectrometry (CyTOF) which tags antibodies with metal isotopes to reduce spectral overlap, hence, enabling the measurement of 40 proteins at a time from individual cells using mass spectrometry (Leipold, 2015). This is further expanded upon by other omic technologies used to measure molecular properties at the single cell level, including scTRrio-seq (CNV, CpG methylation, and transcriptomics)(Hou et al., 2016), scCOOL-seq (CNV, chromatin accessibility and CpG methylation)(Guo et al., 2017) and CITE-Seq (transcriptomics and proteomics) (Stoeckius et al., 2017) (**Figure 1.8**). Together these systems can capture thousands of variables with which to characterise cell types and provide a vast amount of data to aid in deconvoluting cellular heterogeneity, however the current challenge is in interpreting all this information and reaching a consensus.

With the shift in technological capabilities we must revaluate how we conceptually define cell types in the context of the cell features we can now quantify. Our interests and accordingly our knowledge of immune cell types diversity and biology have been determined by the prevalent cell types across diseases, e.g. as shown in previous sections numerous macrophages and T cells have been identified and referenced with one another to generate a comprehensive nomenclature, however, while such exhaustive studies of cell states across diseases and conditions are only beginning to take place for cells like neutrophils and other granulocytes (Silvestre-Roig et al., 2019). The latter are largely described in the context of few proteins presented or secreted, enzymes produced, diseases they are associated with and tissues they reside in (Gurish and Austen, 2012, Dwyer et al., 2016). Presently, much of how we define cell types has originated from making clear demarcations between them using selected markers. In certain cases, however we are beginning to see how this strategy masks cellular subsets, e.g. pDC isolated using CD123 and BDCA-2 were assumed to drive the IFN response

and T cell activation, however recent studies using scRNA-Seq reveal a minor subpopulations of cells within them termed as Axl+ DCs responsible for the potent activation of T cells (Villani et al., 2017). Such studies have helped revaluate our definition of cell types and their function.

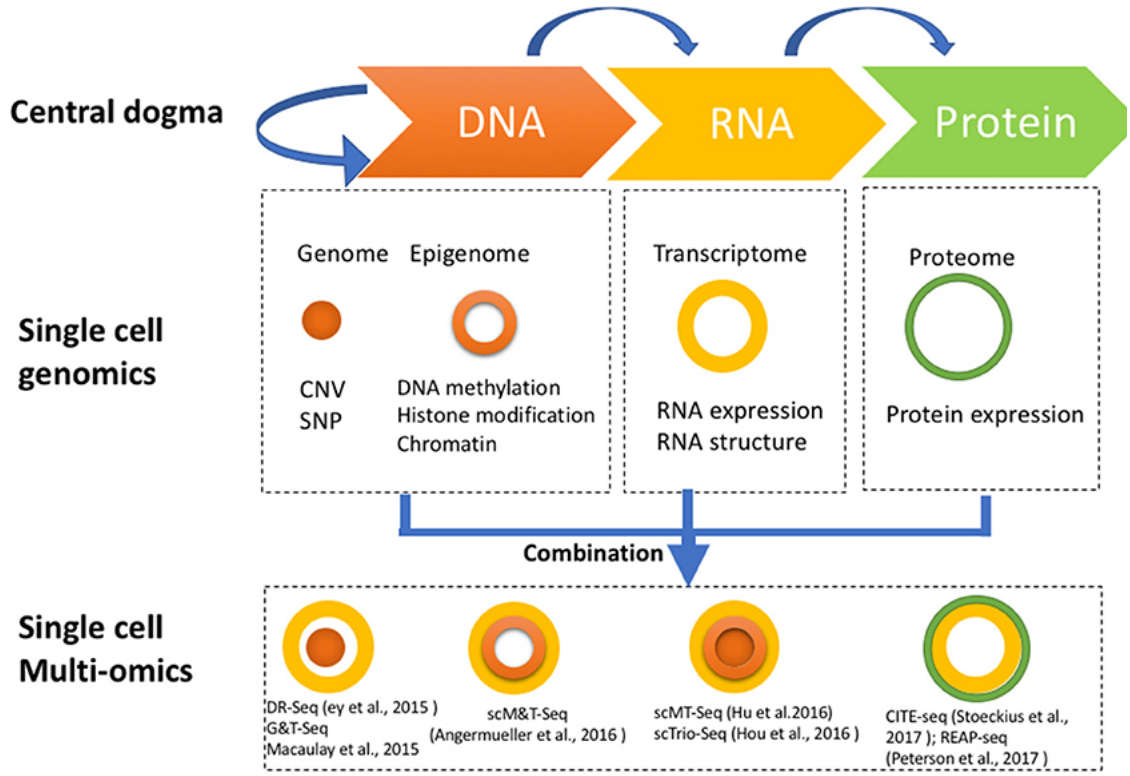


Figure 1.8. Single cell multiomics technologies

Taken from (Hu et al., 2018). Single cell omic technologies for studying the central dogma at various levels including the genome, epigenome, transcriptome and proteome. Also shown are the multiomic technologies which examine a combination of these omics.

To understand these systems based on our current resources and knowledge a few common themes have appeared, including:

1. Comparing cell types across datasets to better understand the uniqueness of cell types vs. others e.g. comparing monocytic population in human to those of mouse (Ramachandran et al., 2019).
2. Defining cells based ontology or lineage, shown for haematopoiesis (Weinreb et al., 2020).
3. Identifying a functional role of novel cell types.
4. Perturbation studies, by methods like CRISPR or exposing to different stimuli, to examine the stable and bounds of cellular states, i.e. how restricted are cell types

in differentiating to other states/cell types and what are the limits of this (Rubin et al., 2019).

1.6.1 The then and now of bulk RNA-Seq

Tissue or population-level transcriptomics data represents an average of signals determined by the diversity of cell populations, their numbers and their expression levels of genes/proteins. Such methodologies have helped describe differences between experimental conditions (disease and stimuli), developmental processes, subgrouping patients based on their response to disease, the functional annotation of genes and mechanisms underlying the central dogma (Wang et al., 2008, Graudenzi et al., 2018). There have been several drivers for understanding tissue-wide omics in humans, leading to the formation of several consortiums, databases and resources. Resultant studies have been crucial in explaining tissue-specific biology, including deviations in genetic architecture and gene regulation. Studies have included the Illumina Human Body Map 2.0 and GTEx (https://gtexportal.org/home/) (Lonsdale et al., 2013) from which we can characterize tissue-specific gene expression profiles (Zhang et al., 2010, Pan et al., 2013), the similarity between tissues in processes such as ageing (Yang et al., 2015), as well as functional and disease associations, as shown for Parkinson and Alzheimer's towards the brain (Sonawane et al., 2017, Greene et al., 2015). Focused efforts, such as the ABA have helped identify the biological components of the brain in mouse and humans, by generating transcriptomics data derived from numerous regions (Shen et al., 2012). Coupled with *in situ* hybridization (ISH) techniques they have provided a spatiotemporal view of the brain cytoarchitecture through spatial transcriptomics (Sunkin et al., 2013). Similar studies have generated a plethora of publicly available data for which various databases have been constructed including Gene Expression Omnibus (GEO) (Edgar et al., 2002) from the National Centre of Biotechnology and Information (NCBI) and ArrayExpress (Oezcimen et al., 2003) of the European bioinformatics institute (EBI). The data has also been summarised and visualized through resources such as BioGPS (Wu et al., 2009).

The interpretation of such data is dependent on teasing out two biological signals, cell proportions and states. Often studies focus on one depending on the research question. Accordingly, several methods/strategies have been developed to analyse this data and to address both biological signals. "Interesting" genes with respect to experimental

conditions can be identified through differential expression and GCNs. The former reveals genes whose expression changes significantly across conditions while the latter highlights groups of genes part of a common biological process through guilt-by-association as they share a common expression profile across samples. To infer the biology from these sets of genes, a supervised approach can be adopted. Numerous databases are now available like the gene ontology (GO) and molecular signature databases which describe various kinds of biology under different conditions using sets of genes or gene signatures (Liberzon et al., 2011, Ashburner et al., 2000). These gene signatures can represent cell types, cellular states, pathways, and diseases. Moreover, using transcriptomics data one can also find the directionality of a particular signature, i.e. if a process is down- or up-regulated using gene set enrichment analysis (GSEA) (Bolen et al., 2011, Verhaak et al., 2010, Shoemaker et al., 2012). Unfortunately, signatures can show enrichment for numerous terms associated with different cell types and biology due to the redundancy in known gene sets. Hence, any inferences should be made cautiously although methods are being developed to help interpret these results (Supek et al., 2011). With this redundancy in databases and their bias towards a limited number of biological processes, various resources including literature mining can aid in describing cell populations in bulk RNA-Seq data, especially cell type markers. Specific cell types may be identified using gene signatures or markers from which the relative proportion of cells can be found using methods such as CIBERSORT and ImSig (Chen et al., 2018, Nirmal et al., 2016). Such studies help describe the type of responses in disease, where the cell distribution changes, e.g. in systemic lupus erythematosus the leukocyte proportions described by transcriptomics data was shown to be associate the disease activity index as determined by clinical parameters (Abbas et al., 2009). However, a caveat of this is in segregating similar cell types. Another aspect of cell biology that can be investigated are cellular states. Some of these states are mentioned in gene set databases, e.g. activation states of immune cells in response to different stimuli (Abbas et al., 2005, Raza et al., 2014), while the degree of their activation can be explained by GSEA (Finotello and Trajanoski, 2018). Researchers have used different methods to separate the biological signal of cell activation from cell proportions in bulk-RNA-Seq data (Shen-Orr and Gaujoux, 2013). This has also been shown in an epigenome-wide study of rheumatoid arthritis, where the effects of cell number assumed to be a confounding factor, was removed from the data, revealing disease associated

methylation signatures potentially contributing to genetic risk towards the disease (Liu et al., 2013).

Methodologies developed to analyse high-dimensional bulk RNA-Seq data have laid the foundations for functional gene annotation. The advent of single cell technologies, has expanded our understanding of the complexity and heterogeneity of cells by adopting and building upon previous analytical methods (Di Carlo et al., 2012). Several aspects of these analyses come from well-defined approaches in bulk RNA-Seq including QC of reads, mapping, alignment and normalization. CPM normalization is commonly used in scRNA-Seq (Hwang et al., 2018). However, certain features are unique to scRNA-Seq which require further development of current techniques e.g. taking into account the variability of library sizes arising from differences in cell size and transcriptional activity (L. Lun et al., 2016). Though bulk RNA-Seq highlights the dominant biology across experimental conditions, there have been several cases where cell subsets and their activation states have been underappreciated. Earlier omic studies of AD pointed towards dysregulation in protein homeostasis like Aß and tau, as well as an alternate activation of microglial cells (Papassotiropoulos et al., 2006, Martiskainen et al., 2015). However, only recently scRNA-Seq studies have revealed a subpopulation of microglia, referred to as disease associated microglia (DAM), contribute to AD pathology (Keren-Shaul et al., 2017). Therefore, although bulk RNA-Seq describes key genes in the alternate activation of microglia, the approach would mask the subset of cells actually undergoing such changes. Such findings could mislead drug development which would assume a single population and may preferentially target homoeostatic microglia or those that are beneficial. Still, using both bulk- and scRNA-Seq data together has proven useful, e.g. in improving the biological signals in scRNA-Seq data through imputation using reference tissue RNA-Seq profile (Zhu et al., 2018), and estimating cell proportions from bulk RNA-Seq using cell signatures/markers identified through scRNA-Seq, as protocols for isolating different cell types are still being improved upon (Wang et al., 2019c).

1.6.2 Advances in single cell transcriptomics

Variation at the single cell level is also being studied through (ISH) or immunohistochemistry (IHC), fluorescence activated cell sorting (FACS) and new spatial transcriptomic approaches. As the initial effort in determining the transcriptomic profile single cells, Tang *et al.* were able to sequence mRNA captured from mouse oocytes and blastomeres (Tang et al., 2009). In less than a decade since then there has been an

exponential growth of single cell technologies which has spurred up different companies (10X genomics & Fluidigm), consortiums (Human Cell Atlas and Oxford Single Cell Biology Consortium), databases (Single cell portal and Single cell expression atlas) and resources (Mouse cell atlas and scQuery), all of which have contributed towards a wide range of biology at the single cell resolution (Hwang et al., 2018, Han et al., 2018, Alavi et al., 2018). Studies now include the analysis of up to 1 million cells, with technologies expanding to other omics (Svensson et al., 2018, Chappell et al., 2018). These technologies have enabled further exploration of the different aspects of cellular heterogeneity including the transitionary and transient states during development through pseudotime analysis, the relatively stable and discrete states (cell subsets and rare cell types) defined by their microenvironment, as well as the response state cells achieve based on external or internal stimuli (Farrell et al., 2018, Ledergor et al., 2018). Alternate splice forms across tissues have been well characterised, however studies from scRNA-Seq have also shown alternate isoform usage between cell types which may be a result of intercellular isoform variations as described in Drosophila, although, whether these are controlled or stochastic events is still to be known (Miura et al., 2013). Similar efforts have been made to describe the diversity in VDJ regions of B and T cell receptors, having up to 10^{18} combinations able to recognise several antigens, and directing selective responses, with applications in vaccine development (Jung and Alt, 2004, De Simone et al., 2018, Stubbington et al., 2016). Genomics studies have identified cell types specific expression quantitative trait loci (eQTLs) (van der Wijst et al., 2018) and have been able to further resolve the association of GWAS traits to cell level process like haematopoiesis (Ulirsch et al., 2019). As scRNA-Seq captures a snapshot of the tissue, differentiating cell populations are captured at various “times” of differentiation, also referred to as pseudotime (**Figure 1.9**). This has been especially useful in determining the regulators of cell fates as shown through development and cell action in disease e.g. *GATA1* and *CEBPA* in haematopoiesis (Chen et al., 2019). More recently researchers have tracked the clonal expansion of cells from which they propose a precommitment of cells as has been shown for Gata2+ cells which preferentially differentiated into megakaryocyte, basophil, mast cell and eosinophil (Weinreb et al., 2020).

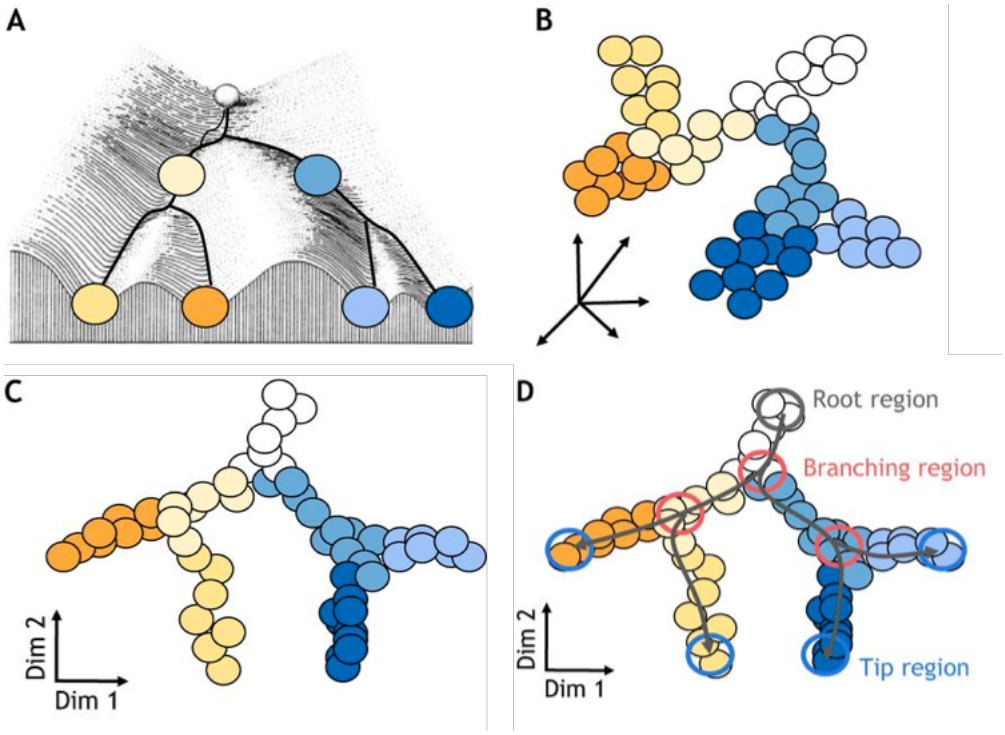


Figure 1.9. Reducing high-dimensional data for interpretation

Adapted from (Tritschler et al., 2019). We can conceptualise the differentiation of cells using Waddington's landscape where cells travel/differentiate through cellular states or valleys either stochastically or directed by stimuli to subsequently achieve more stable states. However, these paths are defined by the numerous features or dimensions that describe cells. An effort in single cell is to reduce this space while still retaining the structure of the data for our interpretation. scRNA-Seq takes a snapshot of differentiating cells which can be bioinformatically aligned in pseudotime. Such dimensionality reduction methods can aid in identifying the progenitor (root region), the branching points at which cells may diverge into multiple cell types and the stable states.

In previous sections we have examined the influence of external cues to cell identity while poorly mentioning the internal stimuli which govern cell identity and its potential to differentiate. An example of this is transcriptional noise, shown to be cell type specific and provides the transcriptional variability required for cell types to preferentially differentiate into a certain state or the other (Satija and Shalek, 2014). As mentioned previously, cellular heterogeneity is greatly influenced by the microenvironment. Such signals could be considered as extrinsic cues which orchestrate a feedback system with the gene regulatory network (Swain et al., 2002). Although gene regulatory networks

govern the type of cell, there is an intrinsic signal or noise which adds to the diversity in cells, which can be attributed to the stochastic nature of gene regulation (Swain et al., 2002). These probabilistic events have been observed at both the protein and the mRNA level, and are referred to as transcriptional bursts (Raj et al., 2006, Lubeck and Cai, 2012). The uniqueness of cell types, their stability and transiency between cellular states can be considered to be regulated by these two sources of signals or noise. Subtle differences in even clonal populations of cancer cells can result in varied drug responses (Cohen et al., 2008). In stem cells, where cell fate determination is found to be unpredictable, fluctuating between states makes cells prone to differentiate towards certain lineages (Imayoshi et al., 2013, Chalancon et al., 2012). Additionally, cell systems can minimize such effects or are made to be in sync through external cues such as cell to cell interaction (Rand et al., 2012).

Although scRNA-Seq has opened various avenues of research it also has its own technological challenges. Apart from the severe PCR duplication (Marx, 2017), the main concerns are dropout events due to the small starting material 0.01-0.25 ng mRNA and low coverage of 50,000 read on average per cell commonly used. This low starting material results in false zero expression level of lowly expressed genes. Hence though similar issues have been addressed in bulk RNA-Seq and microarrays, the frequency of dropouts or the variability of the data is more severe (Faisal and Tutz, 2017, Aittokallio, 2010). scRNA-Seq data is highly sparse where zeros account for more than ~50% of entries in the expression matrix. Methods have been developed to address this noise, such as imputation and noise models and clustering. Imputation algorithms for missing values in scRNA-Seq include global algorithms like SAVER (Huang et al., 2018) and MAGIC (van Dijk et al., 2018), or local algorithms including scImpute (Haimon et al., 2018) and DR impute (Gong et al., 2018), e.g. local algorithms evaluate a local group of similar cells to impute values across cells.

1.6.3 Network analysis of high-dimensional data

Graph-based approaches have been used in portraying the complexity of biological systems. As a result, different forms of networks have been developed, with several tools to visualize and analyse them. Notable software for network analysis in biology include Cytoscape, weighted gene co-expression network analysis (WGCNA), Biolayout, and Stanford network analysis project (SNAP), the latter adapted for large scale networks

(Leskovec and Sosi, 2016) for more than a million nodes and billion edges, a feature becoming more attractive in an age of big data. To represent the complexity of biological systems several types of networks can be constructed each emphasising different characteristics of the data. Networks can be divided based on different criteria such as, static vs. dynamic, undirected vs. directed and weighted vs. unweighted networks, a combination of which can be used to describe different biological processes like, protein-protein interactions (PPI), gene regulation, functional association and signalling. Undirected networks include PPI as shown for viruses (McCraith et al., 2000), bacteria (Wuchty and Uetz, 2014) and complex diseases (Safari-Alighiarloo et al., 2014) which have elucidated upon pathogenic mechanisms. Other undirected weighted networks for studying gene/protein associations such as coexpression, identify genes sharing the same or similar expression profiles across samples. The metrics for determining the similarity between genes (on their expression profile) can be calculated as distances such as Euclidean, Manhattan and hamming distance or similarities such as Pearson, cosine and Spearman correlation. A common metric for similarity is the Pearson correlation coefficient which ranges from -1 to 1. When comparing two genes based on their expression profile a value closer to -1 or 1 suggests that they are inversely correlated or share a positive relationship, respectively. Correlation values close to 0 indicate no similarity between genes. For omics data like that generated from RNA-Seq, pairwise comparisons can be made between genes thereby generating a gene-to-gene matrix. Thereafter, different approaches can used to construct a GCN using this matrix. The most basic algorithm for this is weighted edge pruning (WEP) where a cut-off correlation is chosen, and genes correlated beyond this threshold are connected by an edge within the network. A similar approach can be used for comparing the similarity between samples, where pairwise correlations are calculated between samples instead of genes. Although, statistical approaches such as differential expression analysis identify genes significantly different between groups, they may lose out on complex patterns within the data. Through GCNs subtle patterns are revealed including those which may not follow sample classifications required by statistical approaches. Furthermore, the method has proven to capture the structure of the data, e.g. a commonly used algorithm for network analysis in biology is WGCNA (Langfelder and Horvath, 2008) which outperforms methods like hierarchical clustering which is unaware of the data's topology thereby losing out on key information which effects its clustering. WGCNA overcomes this hurdle by applying transformations to the gene similarity matrix

to consider the topology of the data, ultimately calculating the topological overlap matrix (TOM). The TOM is then clustered using hierarchical clustering and has shown to improve clustering relative to hierarchical clustering on its own (Li et al., 2018). The caveat to GCNs however is that it is still qualitative, without a statistical approach to find significantly similar genes, thereby requiring external resources to prove their relevance, e.g. staining for proteins, correlating cell signature expression with cell numbers and including known markers genes, as have proven useful in identifying the various signatures in the brain (Hawrylycz et al., 2012).

GCNs have been essential in capturing biological signatures across experimental conditions, diseases, tissues and cells (Brazma and Vilo, 2000, Serin et al., 2016). Findings from these studies have generated various gene signature databases, which in turn help describe the biology of other studies. To interpret signalling pathways, graphs need to model directed interactions. Resources for static directed networks which model signalling pathways include Reactome (Mundo et al., 2013), KEGG pathways (Kanehisa and Goto, 2000) and Pathway Commons (Gross et al., 2010). Dynamicity can be introduced into such graphs by knowing the rate of these reactions described through differential or linear equations such as flux balance analysis (Edwards et al., 2001), or a more stochastic approach such as Petri nets can be used which is independent of kinetic parameters (Fisher and Henzinger, 2007, Livigni et al., 2018). Subsequently, these models can be applied to simulate cell biology *in silico* in Cytoscape (Arora et al., 2016), Biolayout or in MATLAB (Downes et al., 2006). An amalgamation of different types of information like gene coexpression, interaction experiments, and literature, can be incorporated into an undirected weighted association network for inference, these comprise of Search Tool for the Retrieval of Interacting Genes (STRING) (Roth et al., 2010), GeneMANIA (Warde-Farley et al., 2010) and Enrichment Map (Merico et al., 2010). These methods can be especially useful in removing redundancies and gauging a global picture of the processes involved. In contrast, although GO terms (Ashburner et al., 2000) are highly informative, their redundancy and multiplicity can prove counterintuitive, and presented through Enrichment maps the redundancies are revealed thereby highlighting components of unique biology. The setback for correlative networks such as for gene coexpression is that casual interactions can exist between nodes which can be misleading, e.g. a serial stepwise pathway would appear as an interconnect hub instead as sequentially connected nodes following the pathway. Methods are also being developed to find causative interactions which outline biological pathways (Feizi et al.,

2013) e.g. Bayesian networks inference, generate a directional network based on genes expression data, revealing the hierarchy of signalling pathways in Alzheimer's (Zhang et al., 2013).

Whilst conventional graph construction approaches like WEP have proven useful in bulk RNA-Seq, technical artefacts such as dropouts within scRNA-Seq result in poor correlation values between genes within this data. Hence, different approaches have to be used to construct graphs for this data. A popular graph construction algorithm which has proven useful in terms of clustering is the k-nearest neighbour (kNN) algorithm (Duò et al., 2018). By considering the strongest 'k' edges for each sample the approach disregards the actual strength of these edges assuming any differences to arise from noise and hence rather ranks edges for each node (sample or gene). Although this can make it difficult to differentiate between noise and actual cell subtypes when clustering, the approach has been widely adopted in the field with added variations (Levine et al., 2015, Xu and Su, 2015, Wolf et al., 2018). In addition to noise, kNN helps reveal the different levels of complexity within cells, which has been a concern in WEP. Few techniques have been developed to assess what threshold should be applied in WEP (Borate et al., 2009), however due to the hierarchical nature of biology a single threshold cannot capture all the different resolutions e.g. neurons can be differentiated from glial cells, and at a higher resolution into their various subtypes. Derived edge reductions methods such as Phenograph further the modularity of the graph by evaluating then filtering edges (Levine et al., 2015).

The biological complexity and technical artefacts in scRNA-Seq have led to the development of various methods for high-dimensional data analysis. One of the challenges has been to capture differentiation trajectories of cells for which graph-based approaches have been used, such as Single-cell trajectories reconstruction, exploration and mapping (STREAM) (Chen et al., 2019), monocle (Trapnell et al., 2014), slingshot (Street et al., 2018), p-Creode (Herring et al., 2018) and Partition-based graph abstraction (PAGA) to reduce the structure of the data to a hierarchical tree for ease of interpretation (Herring et al., 2018, Wolf et al., 2019). Such tools also provide great visualisation which has become a field on its own. Although methods such as Principal component analysis (PCA) are still being used for dimensionality reduction as well in visualising the dominant variations within the data. Newer machine learning methods such as t-distributed stochastic neighbour embedding (tSNE) (Maaten and Hinton, 2008) and Uniform manifold approximation and projection (UMAP) (Becht et al., 2019) have

also been developed to capture data topology in a specified set of dimensions, usually two. In addition, these methods optimise the visualisation to highlight differences between data-points/samples like in tSNE where discrete cell types and even differentiation trajectories appear as globular. This is not so in UMAPs and network visualisation, where the latter further retains the essence of distance between the network components, not captured in UMAP and tSNE plots. Another avenue for scRNA-Seq data analysis is in determining the gene regulatory network. As mentioned previously there are limitations to differential expression analysis, especially in determining which genes are regulated together within the set of genes differentially expressed between conditions/states. For constructing gene regulatory networks, several approaches have been developed, some of which show bias towards synthetic data such as Single cell regularized inference using time-stamped expression profiles (SINCERITIES) (Papili Gao et al., 2017), while others perform better overall like GENIE and GRNBoost (Aibar et al., 2017). Conversely, a study has also shown the inconsistency in the results of such algorithms (Chen and Mar, 2018). Network algorithms such as WGCNA are also being used to conduct such analyses and a recent study shows the promise of their application (Iacono et al., 2019). A unique network-based approach which tries to address the sparsity of the scRNA-Seq data is MetaCell. The method aims to retain the structure of the data while aggregating similar cells into MetaCells (Baran et al., 2018). This helps reduce the data sparsity by averaging reads from similar cells into so called MetaCells, thereby aiding in downstream analyses. Self-assembling Manifolds (SAM) on the other hand reduces dimensions by learning a fuzzy representation of the data through graphs, while iteratively removing genes which contribute poorly to the graph's structure. Furthermore, the number of edges for each node within the graph is also evaluated, thereby converging to the "original" structure of the data, all independent of any threshold (Tarashansky et al., 2018). In summary, graphs have infiltrated the various spheres of analysis in single cell approaches and are only moving further in their capability to handle its complexities.

1.7 Aims and objectives

In the course of my doctoral studies, I have sought to investigate different aspects of TRICs. Listed below are the various aims and questions of the work.

1.7.1 Chapter 2

Previous studies of the group have examined the transcriptional diversity of microglia during aging and in response to prion infection in mice, a well characterised model of neurodegeneration (Vincenti et al., 2016). In this chapter the core transcriptional signature of human microglia is examined through GCNs, and subsequently used to investigate the region-dependent responses of microglia in ageing and Alzheimer disease.

1. How do previously derived microglia compare, and what factors may cause them to show this consistency/inconsistency?
2. What is the core human microglia signature across brain regions?
3. How do microglial cell numbers and phenotype change in ageing and Alzheimer's depending on the region?

1.7.2 Chapter 3

The work in chapter 2 is expended upon by examining the heterogeneity of all immune cell types across tissues through different forms of RNA-Seq data including those derived from marker-defined immune cell types and those from the unbiased approach of scRNA-Seq:

1. For cell types predefined by markers, how are different immune cell types associated i.e. how similar/different are they to one another and what are the gene clusters associated with these cell types?
2. Similarly, how are immune cells associated with one another and which gene clusters represent them based on scRNA-Seq derived from different tissues?
3. How do gene clusters compare between those derived from predefined cell types and those identified in an unbiased way using scRNA-Seq?
4. What known and novel biology do these gene clusters describe, especially those concerning TRICs?

1.7.3 Chapter 4

Investigate the genes associated with TRICs in human and compare these signatures with those previously identified in mouse.

1. What are the TRIC associated gene signatures from different tissues in human?
2. What is the distribution (presence/absence) of immune cell types across tissues?

3. What biology of TRICs do these gene signatures describe and how do they signatures compare with those identified in mouse from chapter 3?

2. Chapter 2: A core transcriptional signature of human microglia: Derivation and utility in describing region-dependent alterations associated with Alzheimer's disease

2.1 Introduction

The investigation into TRICs was started by studying microglia i.e. macrophages of the nervous system known for their involvement in various neurodegenerative diseases. Investigations into microglia in different regions of the brain have shown them to exhibit different phenotypes as has been shown by our lab in a chronic neurodegenerative mouse model induced by prion inoculation (Vincenti et al., 2016). The study showed a region-dependent increase in microglial numbers post prion inoculation, an indication of microglial response towards inflammation, with those of the thalamus having the highest and the cerebellum with the lowest numbers of microglia (Vincenti et al., 2016). A similar study in mouse described the different regional profiles of microglia through ageing. Notably, microglia in the cerebellum were found to be immune-vigilant due to their higher expression of various sensory molecules, while cortical microglia highly expressed immunoregulatory molecules. The authors also found that through ageing the hippocampus began to adopt cortical microglial phenotypes (Grabert et al., 2016). Studies have also compared the transcriptomic profiles of microglia in mouse and human finding neurodegeneration disease associated genes enriched in gene modules of microglia and varying ageing transcriptomic profiles of microglia between species (Miller et al., 2010, Galatro et al., 2017). Indeed, these cells display great regional heterogeneity in health and disease, which also varies across species. To this end we sought to define a conserved set of human microglial genes in the context of the CNS using GCN analysis which could be used to investigate region-dependent changes of the cell in health and disease. Additionally, our preliminary analysis revealed great variability across previous studies defining microglia both in human and mouse. Hence, to account for possible sources of variation we considered multiple datasets for signature derivation spanning various brain regions, sequencing platforms and donors. On validating this core human microglial signatures, we could conduct qualitative (cell number) and quantitative (activation state) investigations into different brain regions through ageing and in Alzheimer’s disease. In this work I conducted all the analysis and wrote the initial draft. The co-authors were extremely helpful in discussing the types of analyses that could be conducted, understanding their results, and editing the manuscript.

Received: 25 April 2018 | Revised: 6 November 2018 | Accepted: 12 November 2018
DOI: 10.1002/glia.23572

GLIA
WILEY

RESEARCH ARTICLE

A core transcriptional signature of human microglia: Derivation and utility in describing region-dependent alterations associated with Alzheimer's disease

Anirudh Patir[1] | Barbara Shih[1] | Barry W. McColl[1,2] | Tom C. Freeman[1]

[1]The Roslin Institute, University of Edinburgh, Easter Bush, Midlothian, Scotland, United Kingdom

[2]UK Dementia Research Institute at The University of Edinburgh, Edinburgh Medical School, The Chancellor's Building, 49 Little France Crescent, Edinburgh, United Kingdom

Correspondence
Tom C. Freeman, The Roslin Institute, University of Edinburgh, Easter Bush, Midlothian, Scotland EH25 9RG, United Kingdom.
Email: tfreeman@roslin.ed.ac.uk

Funding information
Medical Research Council, Grant/Award Numbers: MR/M003833, MR/L003384/1; Biotechnology and Biological Sciences Research Council, Grant/Award Number: BB/J004227/1

Abstract

Growing recognition of the pivotal role microglia play in neurodegenerative and neuroinflammatory disorders has accentuated the need to characterize their function in health and disease. Studies in mouse have applied transcriptome-wide profiling of microglia to reveal key features of microglial ontogeny, functional profile, and phenotypic diversity. While similar, human microglia exhibit clear differences to their mouse counterparts, underlining the need to develop a better understanding of the human microglial profile. On examining published microglia gene signatures, limited consistency was observed between studies. Hence, we sought to derive a core microglia signature of the human central nervous system (CNS), through a comprehensive analysis of existing transcriptomic datasets. Nine datasets derived from cells and tissues, isolated from various regions of the CNS across numerous donors, were subjected independently to an unbiased correlation network analysis. From each dataset, a list of coexpressing genes corresponding to microglia was identified, with 249 genes highly conserved between them. This core signature included known microglial markers, and compared with other signatures provides a gene set specific to microglia in the context of the CNS. The utility of this signature was demonstrated by its use in detecting qualitative and quantitative region-specific alterations in aging and Alzheimer's disease. These analyses highlighted the reactive response of microglia in vulnerable brain regions such as the entorhinal cortex and hippocampus, additionally implicating pathways associated with disease progression. We believe this resource and the analyses described here, will support further investigations to the contribution of human microglia in CNS health and disease.

KEYWORDS

aging, Alzheimer's, microglia, neurodegenerative disease, transcriptome

1 | INTRODUCTION

Microglia are the most abundant myeloid cell type in the central nervous system (CNS), accounting for approximately 5–20% of the brain parenchyma depending on region (Lawson, Perry, Dri, & Gordon, 1990; Mittelbronn, Dietz, Schluesener, & Meyermann, 2001). These cells are phenotypically plastic and exhibit a wide spectrum of activity influenced by local and systemic factors (Cunningham, 2013; Perry & Holmes, 2014). Through development into adulthood, microglia influence the proliferation and differentiation of surrounding cells while regulating processes such as myelination, synaptic organization and synaptic signaling (Colonna & Butovsky, 2017; Hoshiko, Arnoux, Avignone, Yamamoto, & Audinat, 2012; Paolicelli et al., 2011; Prinz & Priller, 2014). As the primary immune sentinels of the CNS, microglia migrate toward lesions and sites of infection, where they attain an activated state that reflects their inflammatory environment (Leong & Ling, 1992). In these states, they can support tissue remodeling, and phagocytose cellular debris, toxic protein aggregates and microbes (Colonna & Butovsky, 2017; Li & Barres, 2017). During neuroinflammation these cells coordinate an immune response by releasing cytokines, chemoattractants and presenting antigens, thereby communicating with local immune cells and those recruited from the circulation (Hanisch & Kettenmann, 2007; Hickey & Kimura, 1988; Scholz & Woolf, 2007).

In common with mononuclear phagocyte populations throughout the body, recent studies have begun to reveal the diversity of microglial phenotypes in health, aging, and disease states, as well as their unique molecular identity in relation to other CNS resident cells and nonparenchymal macrophages (Durafourt et al., 2012; Hanisch, 2013; Li & Barres, 2017; McCarthy, 2017; Salter & Stevens, 2017). The application of transcriptomic methods has been integral to these advances by enabling an unbiased and panoramic perspective of the functional profile of microglia. In addition to an improved understanding of the variety of context-dependent microglial phenotypes, other key benefits have arisen from these studies, notably the development of new tools to label, isolate and manipulate microglia (Bennett et al., 2016; Butovsky et al., 2014; Hickman et al., 2013; Satoh et al., 2016). Although most studies have been conducted in mice, a considerable body of data is now emerging from human post-mortem and biopsy tissue (Darmanis et al., 2015; Galatro et al., 2017; Gosselin et al., 2017; Olah et al., 2018; Zhang et al., 2016). While there are many conserved features between rodent and human microglia, the importance of further refining our understanding specifically of human microglia is underscored by significant differences that have been observed between them (Butovsky et al., 2014; Galatro et al., 2017; Miller, Horvath, & Geschwind, 2010; Szulzewsky et al., 2016).

Recent transcriptomic studies have sought to characterize the human microglial transcriptomic signature from the CNS of non-neuropathologic individuals using data derived from either cells or tissue isolated from different brain regions (Darmanis et al., 2015; Galatro et al., 2017; Hawrylycz et al., 2012; Oldham et al., 2008). These analyses have been crucial in expanding our knowledge of their functional biology, however, our preliminary analysis found there to be limited inter-study agreement across the published microglia gene signatures. Such inconsistency may have arisen due to technical differences in tissue sampling, brain areas analyzed, differences in patient characteristics and biological variance, including the regional-heterogeneity of microglia in the CNS (Grabert et al., 2016; Lai, Dhami, Dibal, & Todd, 2011; Lawson et al., 1990; Vincenti et al., 2016; Yokokura et al., 2011). This highlighted a need to derive a refined human microglial signature that would enable a more precise characterization of these cells in the healthy and diseased human brain. We therefore set out to define and validate a core transcriptional signature of human microglia that is shared by all microglial populations of the human CNS, thereby refining the current understanding of the functional profile of these cells. To establish a robust gene signature we have performed an extensive network analysis of nine human cell and tissue transcriptomics datasets generated on different platforms and derived from numerous brain regions and donors. We used various approaches to validate the derived gene list and compare it to other published signatures which underscored the influence of microglial regional heterogeneity and selection of enrichment thresholds to define a core microglial signature of the CNS. The utility of our microglial signature derived herein was demonstrated by its ability to detect region-dependent changes in microglial number and activation state in aging and Alzheimer's disease.

2 | MATERIALS AND METHODS

2.1 | Comparison of published microglial signatures

To compare published microglial signatures, 15 publications, 9 in human, and 6 in mouse, were identified and the signatures compared (Table 1). Species specific comparisons were conducted as genes from each signature were converted to a common identifier, that is, HGNC (Povey et al., 2001) or MGI (Shaw, 2009) for human and mouse, respectively, using the online tool g:Profiler (Reimand et al., 2016). At the time of analysis g:Profiler used Ensembl v91 and Ensembl Genome v38 (Hubbard et al., 2002).

2.2 | Transcriptomics data acquisition and preprocessing

Tissue and cell transcriptomic datasets derived from the CNS were acquired for the derivation of a human microglial signature. These included data from the Genotype-Tissue Expression (GTEx) project (Lonsdale et al., 2013; https://gtexportal.org/), Allen Brain Atlas (ABA) (Hawrylycz et al., 2012; http://www.brain-map.org/) and from a study by Zhang et al. (2016). The GTEx data comprised of two datasets, one generated on Affymetrix microarrays (n = 207) and a second by RNA-Seq (version 6, n = 1,259). In both cases, tissue samples were isolated from 13 regions of the CNS at post-mortem, from individuals with no known neuropathology. The ABA data, generated on the Agilent microarray platform, consisted of 3,702 tissue samples taken from six individuals from up to 414 unique anatomical regions of the brain. Data from Zhang et al. (2016) consisted of RNA-Seq data (n = 41) generated from different human CNS cell types (neuronal, glial, and endothelial). For downstream analysis, data (n = 131) from immune (myeloid and lymphoid) and brain cell types (neuronal and glial) was downloaded from the Gene Expression Omnibus (GEO, GSE49910) (Mabbott, Baillie, Brown, Freeman, & Hume, 2013). Additional data from the GTEx project (v7) was used in the validation of the current signature and to compare it to published signatures. Lastly, for the analysis of microglia in aging and Alzheimer's, data (Berchtold et al., 2013) derived from post-mortem samples of Alzheimer's patients and controls from four brain regions, cortical, and hippocampal, from 85 individuals (n = 253) was downloaded. Further details of these datasets are provided in Supporting Information Table S2.

Transcriptomics data was downloaded from the appropriate sources and underwent stringent quality control. The unprocessed microarray data from GTEx (GSE45878) and the Alzheimer's dataset (GSE48350), were downloaded from GEO. Data quality was assessed using the ArrayQualityMetrics package (Kauffmann, Gentleman, & Huber, 2008) in Bioconductor, and samples failing more than one of three metrics (between arrays comparison, array intensity distribution, and variance mean difference) were removed. Subsequently, data was normalized using robust multi-array average from the oligo package (Carvalho & Irizarry, 2010) and where multiple probesets represented the same gene, the probeset with the highest average expression across donors was selected to represent the respective gene. GTEx RNA-Seq and ABA preprocessed data were downloaded directly from the GTEx and ABA websites, the latter consisting of six pre-

TABLE 1 Experimental designs of previous studies defining the microglial signature

Species	Publication	Brain region analyzed	Total samples	Sample type	Sample isolation method	Transcriptomics platform	Analysis method
Human	Oldham et al. (2008)	Regions	160	Tissue	Dissections	Microarray	Clustering
	Miller et al. (2010)	Regions	535	Tissue	Dissections	Microarray	Clustering
	Hawrylycz et al. (2012)	Regions	911	Tissue	Dissections	Microarray	Clustering
	Zhang et al. (2013)	Regions	1,647	Tissue	Dissections	Microarray	Clustering
	Darmanis et al. (2015)	Region (cortex)	332	Cell	Single cell	RNA-Seq	Fold enrichment
	Wehrspaun et al. (2015)	Region (cortex)	381	Tissue	Dissections	Microarray	Clustering
	Gosselin et al. (2017)	Region (cortex)	24	Cell and tissue	Pooled cells and dissections	RNA-Seq	Fold enrichment
	Galatro et al. (2017)	Region (cortex)	49	Cell and tissue	Pooled cells and dissections	RNA-Seq	Fold enrichment
	Olah et al. (2018)	Region (cortex)	550	Cell and tissue	Pooled cells and dissections	RNA-Seq	Fold enrichment
Mouse	Chiu et al. (2013)	Regions	42	Cell	Pooled cells	RNA-Seq	Fold enrichment
	Hickman et al. (2013)	Whole brain	17	Cell and tissue	Pooled cells and dissections	RNA-Seq	Fold enrichment
	Butovsky et al. (2014)	Whole brain	77	Cell	Pooled cells	Microarray	Fold enrichment
	Zhang et al. (2014)	Regions	17	Cell	Pooled cells	RNA-Seq	Fold enrichment
	Zeisel et al. (2015)	Regions	3,005	Cell	Single cell	RNA-Seq	Clustering
	Bennett et al. (2016)	Whole brain	11	Cell	Pooled cells	RNA-Seq	Fold enrichment

normalized datasets from the brains of separate donors. Furthermore, quality control was conducted by inspecting sample-to-sample correlation networks using Graphia Professional (Kajeka Ltd, Edinburgh, UK), revealing outlier samples or batch effects which were removed. Evident from the GTEx (v6) RNA-Seq data, early batches (LCSET-1156 to LCSET-1480) poorly correlated with other samples, forming a highly connected group separate from other samples within the network. For downstream analysis, genes were filtered out from the Affymetrix microarray data with a normalized expression level <20 and <1 RPKM/TPM for RNA-Seq.

2.3 | Derivation of core microglial gene signature by coexpression network analysis

To derive a core microglial gene signature, the nine tissues and cell transcriptomics datasets described above were analyzed using the network analysis tool Graphia Professional. For each dataset, Pearson correlations (r) were calculated between all genes to produce a gene-to-gene correlation matrix. From this matrix, a gene coexpression network (GCN) was generated, where nodes represented genes and genes whose expression profile were correlated more than a defined threshold were connected by edges. Coexpressed genes formed highly connected cliques within the overall topology of the graph, which were defined as clusters using the Markov clustering algorithm (MCL), with the default inflation value of 2.2 (Van Dongen & Abreu-Goodger, 2012). All Pearson threshold values used for individual datasets were above $r \geq 0.7$, ensuring graphs included only correlations that were highly unlikely to occur by chance (Supporting Information Figure S1). For each dataset, the correlation threshold was adjusted to achieve a single microglial cluster containing three canonical marker genes for microglia, *CX3CR1*, *AIF1*, and *CSF1R* (Elmore et al., 2014; Mittelbronn et al., 2001). The final microglial gene signature was defined as genes present in at least three of the nine dataset-derived microglial signatures. In order to annotate the function of signature genes with relevance to myeloid and immune cells, the GeneCards database (Safran et al., 2010) and literature were consulted.

2.4 | Validation of the core human microglial signature

Various lines of evidence were investigated to validate the conserved nature of the derived human microglial signature. Firstly, the average expression of signature genes was compared between myeloid and other cell types from an atlas of primary human cells (Mabbott et al., 2013), using the Mann–Whitney *U* test. Similarly, the average expression of signature genes in the GTEx RNA-Seq data (v6) and donor one of the ABA microarray datasets was also compared with the microglial cell densities in mouse, for comparable regions (Lawson et al., 1990). Where available, immunohistochemical (IHC) staining of proteins encoded by signature genes were examined in the Human Protein Atlas (HPA) (Nilsson et al., 2005) across different regions of the human CNS. Enrichment analysis for Gene Ontologies (GO), pathways and transcription factor binding sites were conducted using ToppGene (Chen, Bardes, Aronow, & Jegga, 2009).

To compare the microglial specificity of signature genes proposed by different studies, their expression profile was examined in GTEx (v7) RNA-Seq data derived from 13 CNS regions, data not used in the derivation of the current signature. A GCN was constructed ($r \geq 0.8$) from the GTEx RNA-Seq dataset and clustered. The current and published signature genes were overlaid on the network, highlighting the confidence of each gene based on how many publications they were identified in. Six clusters enriched in genes from published signatures were identified, whose expression profiles represented region-specific and nonmicroglial biology associated with the CNS. For each of these clusters, the average expression of the different signature genes was compared between the region with the highest expression versus the remaining samples, using the Mann–Whitney *U* test. Additionally, the differential expression of signature genes from each study was calculated for the cerebellum versus all other regions, using the same test.

2.5 | Analyzing microglia in aging and Alzheimer's disease

To examine the utility of the signature in the study of microglia in aging and Alzheimer's disease, samples from the study by Berchtold et al. (2008) were binned into four age groups: 20–39, 40–59, 60–79, and 80–99 years. The average expression level of microglial signature genes was calculated for each age group and comparisons were made between Alzheimer's samples (80–99 years) with age-matched controls, and between older groups (80–99 years) against younger controls using the Mann–Whitney *U* test, corrected for multiple testing. To identify genes that specifically represent microglial activation in Alzheimer's, a GCN ($r \geq 0.7$) was constructed from only those samples derived from Alzheimer's patients. The Fischer's exact test was used to identify clusters enriched (adjusted $p < 0.01$) in core signature genes. Two such clusters were identified, cluster 5 and 67, containing 333 and 18 genes, respectively. Of these genes, 165 were not part of the derived signature and were considered as potential microglial-associated genes (MAGs) in Alzheimer's. To aid the interpretation of the MAGs, their enrichment of pathways, GO annotations and associated transcription factor binding sites, were calculated using ToppGene (Chen et al., 2009). Following this, differentially expressed core genes and MAGs across brain regions were identified for the superior frontal gyrus, between old (≥60 years) versus young controls (<60 years), and also between Alzheimer's (≥60 years) versus age-matched controls, using the limma package in R (Smyth, 2005). A similar enrichment analysis was conducted for genes differentially expressed between Alzheimer's samples and age-matched controls.

3 | RESULTS

3.1 | Heterogeneity of existing microglial signatures in human and mouse

To explore the consistency of published human microglia gene signatures, previous signatures derived from human brain tissue or cells were compared (Table 1, Figure 1a). The nine human studies examined varied considerably in the number of genes they defined, ranging from 21 to 1,232 genes (Supporting Information Table S1). Of the 2,170 unique genes identified across these studies, less than half (43%, 939 genes) were present in two or more signatures, with none of the genes common to all nine publications. To explore whether these discordant results could be attributed to the genetic variation of humans or collected CNS region, the six publications reporting mouse-microglial signatures were also compared (Table 1). Altogether these included a smaller set of 689 genes (ranging from 47 to 433 genes). Here, only 26% (180 genes) of genes were reported by more than one study, with a single gene common to all (Supporting Information Table S1, Figure 1b). These observations highlight the discordance between existing microglia marker lists and the need to develop a robust and validated human microglial gene signature.

3.2 | Derivation of a conserved core human microglial signature

Observing the variability across published studies, we set out to define a microglia gene signature from human tissue and cell data using a gene coexpression-based approach (Figure 2a, Supporting Information Table S2). For this analysis, tissue datasets including the GTEx project (v6) and ABA data were chosen, which cover a larger spectrum of samples, across numerous CNS regions and donors, relative to previous studies (Hawrylycz et al., 2012; Lonsdale et al., 2013; Shen, Overly, & Jones, 2012). Additionally, transcriptomics data from different brain cell populations (neuronal, astrocytes, oligodendrocytes, microglia, and endothelial cells) was included as it provided a direct comparison of microglia with other CNS cell types (Zhang et al., 2016). After quality control, the data amounted to a total of 5,020 samples isolated from 220 donors and 427 anatomical regions of the brain. To extract a microglial cluster from individual datasets, each was analyzed independently using a GCN approach (Freeman et al., 2007; Theocharidis, Van Dongen, Enright, & Freeman, 2009). This method exploits the inherent variability between samples due to differences in sampling, donors and cellular diversity across different CNS regions. In this case, genes expressed specifically by microglia in the context of the CNS, will vary in expression according to the regional abundance of these cells and therefore correlate in their expression. For example, the poorly populated cerebellum presents a low expression of microglial genes relative to other regions. For constructing GCNs, genes are represented by nodes, and connected by edges based on the similarity between their expression profiles, as quantified by the Pearson correlation coefficient (Supporting Information Figure S1). In such networks, correlated genes, often associated with specific cell types or pathways, form highly connected cliques within the overall topology of the GCN and are defined as clusters using the MCL algorithm (Enright, Van Dongen, & Ouzounis, 2002; Shih et al., 2017). Using this approach on individual datasets, a microglial cluster containing the known marker genes *CX3CR1*, *AIF1*, and *CSF1R*, was identified for each dataset (Elmore et al., 2014; Mittelbronn et al., 2001). The final, high confidence microglia gene signature was defined as 249 genes, which were present in three or more dataset-derived clusters, so as to avoid biases toward individual datasets (Supporting Information Table S3). However, it should be noted that the 399 genes observed in at least two dataset-derived microglial clusters, also showed a strong enrichment for genes with a known immunobiological function.

3.3 | Validation and description of the core human microglial signature

To validate the microglial signature genes, various lines of evidence were examined. First, a comparison of the average expression of core signature genes across cell types revealed a significantly higher ($p < 0.001$) expression in myeloid cells relative to other immune (most of which are scarce within non-neuropathologic brain tissue; Ginhoux et al., 2010) and CNS cell types (Figure 2b). Second, the average expression of core genes across brain regions in the GTEx and ABA datasets correlated with regional microglial densities as measured in

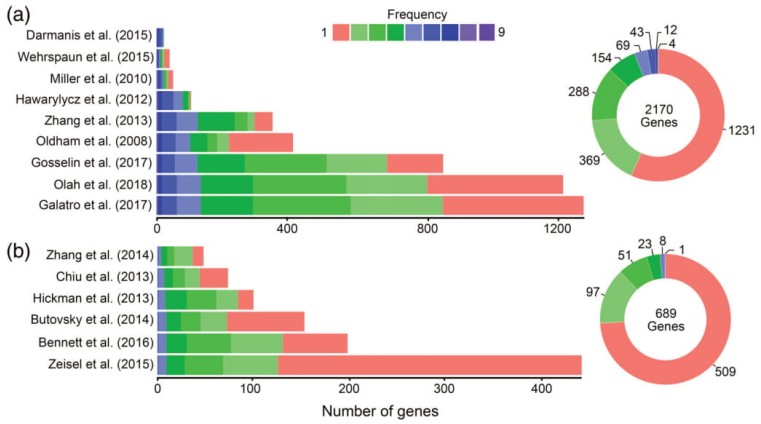


FIGURE 1 Comparison of published human and mouse microglial signatures. Comparison of signature size and gene overlaps amongst microglial signatures defined by studies in (a) human and (b) mouse (left panel). Plot (right panel) of gene overlap considering microglial genes identified in all studies for the respective species [Color figure can be viewed at wileyonlinelibrary.com]

the mouse (Lawson et al., 1990) (Figure 2c). Third, where data was available, the IHC staining of proteins encoded by signature genes was examined in the CNS. This confirmed the microglial expression of known markers for example, *AIF1*, as well as less characterized proteins in the core set, for example, *APBB1IP*, *ABI3*, *FCER1G*, and *ARHGDIB*, which specifically stained for microglia across the four

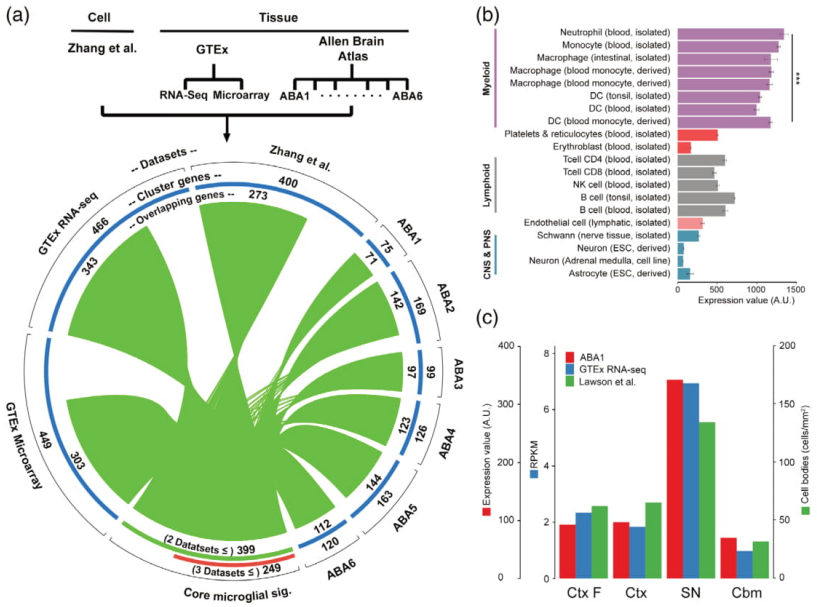


FIGURE 2 Microglial signature derivation and validation. (a) Diagram showing steps in the derivation of the core microglial signature. From each dataset (upper panel) a microglial-specific cluster was identified using coexpression network analysis (blue sectors). In comparing these gene clusters, 399 overlapped across more than one dataset (green sector). From this set of overlapping genes, green lines connect a specific gene to all datasets in which it was identified. Of the overlapping genes, those co-occurring in three or more datasets were taken to represent the core microglial signature (red sector). (b) Average expression of core signature genes in various neuronal and immune cell types selected from an expression atlas of primary cells (Mabbott et al., 2013). (c) Comparison of the average expression across tissue transcriptomics data and microglial cell numbers in mouse, for regions common to the respective studies (Lawson et al., 1990). Abbreviations: AU = arbitrary units; sig = signature; ABA = Allen brain atlas; Ctx F = frontal cortex; Ctx = cortex; SN = substantia nigra; Cbm = cerebellum; DC = dendritic cell. *** Significant at $p < 0.001$ [Color figure can be viewed at wileyonlinelibrary.com]

regions analyzed by the HPA resource (Nilsson et al., 2005) (Figure 3). Finally, GO enrichment analysis was performed and complemented by manual annotation of the core human microglial gene signature. Literature mining showed most genes in the list to have some association with microglial/macrophage biology and overall there was a significant enrichment in genes known to be associated with microglial processes (Supporting Information Tables S3 and S4). These include TLR signaling (*TLR1*, *TLR2*), complement pathway (*C3AR1*, *C1QA*, and *C2*), TYROBP signaling (*TREM2*, *TYROBP*), and cytoskeletal organization (*AIF1*, *CAPG*, and *WAS*) (Hong et al., 2016; Marinelli et al., 2015; Yeh, Hansen, & Sheng, 2017). Genes recently identified as highly enriched in human or mouse microglia, relative to other macrophages and CNS cells, were also present in the signature (e.g., *GPR34*, *P2RY12*, *P2RY13*, and *TMEM119*) (Butovsky et al., 2014).

The core signature was then compared with the published microglial signatures from human (Figure 4, Supporting Information Table S1). All genes from the derived signature were found to be supported by one or more of the previous studies (Figure 4a) with the majority of core signature genes (99%, 246 genes) identified in two or more studies. To further investigate the specificity of the current microglial signature and other human signatures across CNS regions, their expression was examined in the context of the GTEx (v7) RNA-Seq dataset, a dataset not used in the derivation of any signature including the signature proposed by this work. Since microglia are less abundant in the cerebellum, exemplified by the lower expression of *CSF1R*, *AIF1*, and *CX3CR1*, we examined the differential expression of the current and published signature genes in the cerebellum (Figure 4b). In contrast to the expected pattern, genes from several signatures were highly expressed (*FC* > 1.5, FDR < 0.05) in the cerebellum, including pyramidal neuron marker gene, *PCP2* (Lobsiger & Cleveland, 2007). To further explore the expression profile of signature genes across all regions, the coexpression patterns of signature genes was determined. On constructing a GCN from the GTEx (v7) data, microglia signature genes identified by a single study were to be found widely distributed across the network, highlighting their variable patterns of expression across different CNS regions. Conversely, genes common to a number of signatures tended to localize in the same region of the graph, indicative of their coexpression across the CNS and specific association with microglia (Figure 4c). Clustering the network, six clusters enriched in putative signature

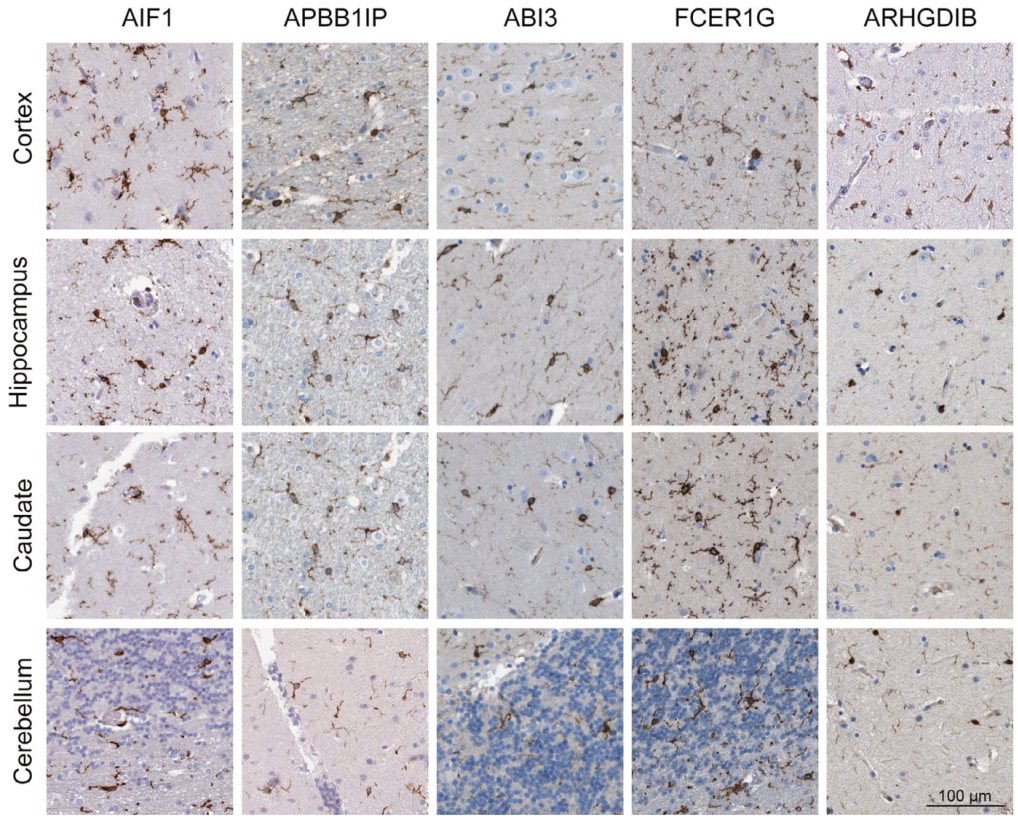


FIGURE 3 Supporting evidence for core microglial signature. IHC staining of proteins encoded by signature genes specifically staining for microglia within various regions of the CNS. Images from human protein atlas (Nilsson et al., 2005) [Color figure can be viewed at wileyonlinelibrary.com]

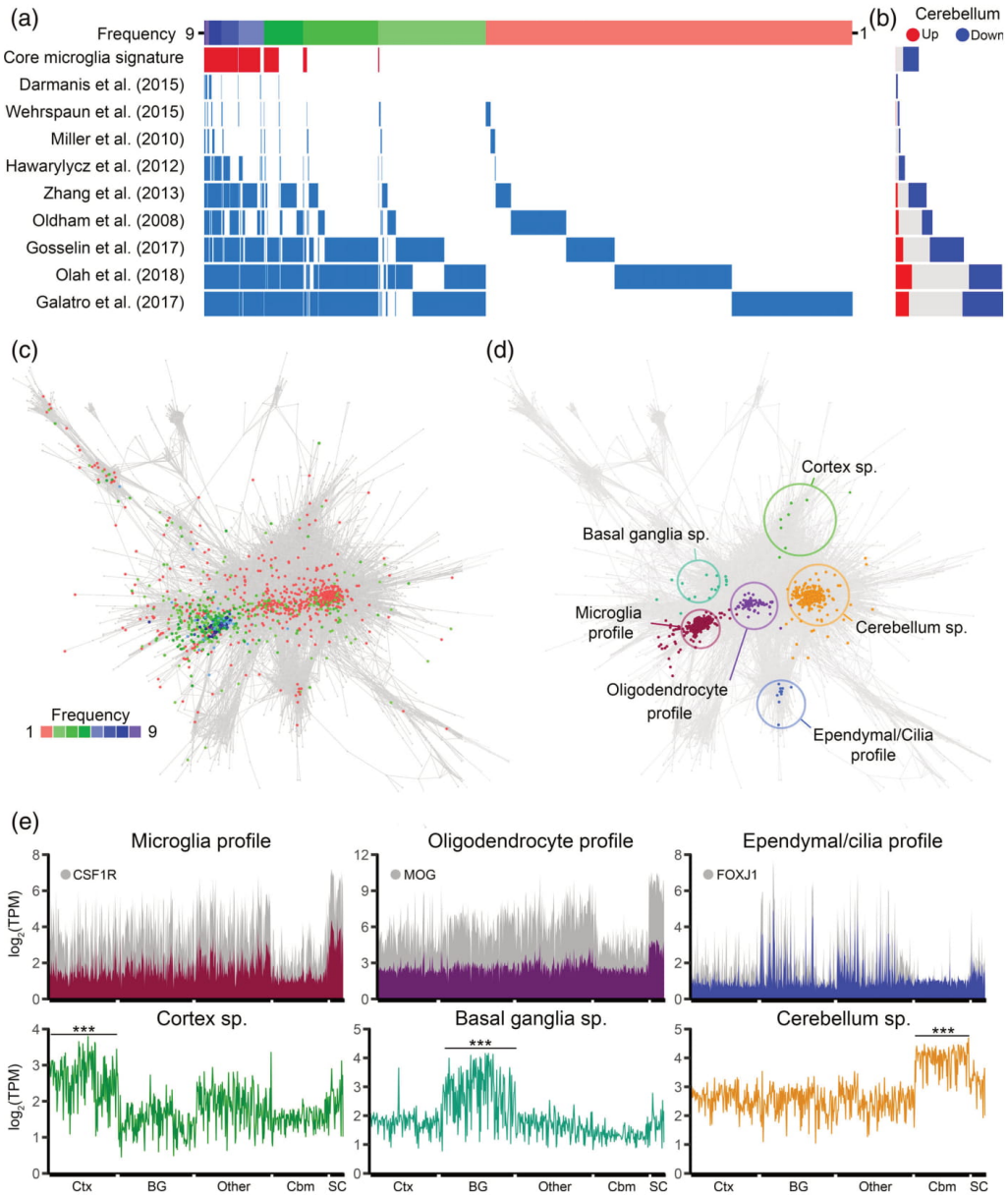


FIGURE 4 Comparison of the derived and previous human microglial signatures. (a) Comparison in the gene content of the current and published human microglial signatures. Genes have been ranked on their relative occurrence across lists. (b) The expression of genes present in individual studies were examined as to their differential expression in the cerebellum relative to other regions. The expectation is that a microglia gene is expressed at lower levels in the cerebellum compared with other regions. (c) GCN of the GTEx (v7) data, highlighting genes identified across all studies and colored by their inclusion across studies. (d) Cluster analysis of the GTEx data, showing six clusters enriched in signature genes from all studies and their (e) expression profile. Abbreviations: sp = specific; Ctx = cortex; BG = basal ganglia; Cbm = cerebellum; SC = spinal cord. ***Significant at $p < .0001$ [Color figure can be viewed at wileyonlinelibrary.com]

genes were inspected and shown to have differing regional and biological specificities associated with the CNS (Figure 4d, Supporting Information Table S5). Genes from certain clusters followed the expression profile of genes such as *MOG* and *FOXJ1*, markers of oligodendrocytes and ependymal cells within the CNS. Genes from other clusters exhibited region-specific expression (*FDR* < 0.01) for the cortex, basal ganglia and cerebellum. Analysis of the immunostaining profile of selected genes in the cerebellum-high cluster, using the HPA database suggested that whilst some were specifically expressed in microglia in other regions, they were not microglial-specific in the cerebellum (Supporting Information Figure S2). For all these clusters, the expression patterns differed from that of known microglial marker genes. The cluster containing known microglial markers was also enriched in genes identified by multiple studies. The core signature genes localized to this cluster, as did most genes from smaller signatures, with the exception of that from Wehrspaun et al. (2015). Based on this analysis, the larger, more recently published signatures, contain many genes not specifically expressed by microglia in all brain regions. Views of the distribution of genes associated with different signatures across the GTEx GCN are shown in (Supporting Information Figure S3).

3.4 | Microglia in Alzheimer's disease

We next used the 249 gene signature to assess the human microglial profile in aging and Alzheimer's disease through analysis of a transcriptomics dataset derived from cortical and hippocampal regions of Alzheimer's patients and non-neuropathologic controls (Berchtold et al., 2013; Supporting Information Table S2). As a preliminary analysis, the average expression of signature genes was used as a proxy measure of microglial number and calculated for all age groups (20-year intervals) across regions (Figure 5a). A significant increase in the expression of core genes was observed in most regions with aging. For example, in the hippocampus, a 1.6 fold change (FC) in expression (*FDR* < 0.01) was observed between the oldest and youngest control age groups. An exception was the entorhinal cortex where a lack of significance is likely attributed to the significant variation between samples across the different age groups. On comparing the average expression of core genes in Alzheimer's with age-matched controls, only the superior frontal gyrus showed a significant increase in Alzheimer's samples (FC = 1.2, *FDR* < 0.05), a region known to be significantly affected in both aging and Alzheimer's based on neuronal connectivity studies (Bakkour, Morris, Wolk, & Dickerson, 2013; Stam, 2014). Although nonsignificant, other regions also showed a consistent increasing trend in the expression of the signature genes related to age-matched controls.

Based on the hypothesis that microglia in Alzheimer's not only increase in number but are also phenotypically altered by the presence of misfolded beta-amyloid protein and other biochemical stressors (Manocha et al., 2016), we sought to identify other genes which were specifically coexpressed with the core signature genes across brain samples from Alzheimer's patients. A GCN was generated using only those samples derived from Alzheimer's patients ($r \geq 0.7$), and two clusters were found to be enriched in core microglial genes based on a Fisher's exact test (adj. $p < 0.01$). In addition to the core microglial gene signature, 165 microglia-associated genes (MAG), were present in these clusters, that is, coexpressed with the core genes, and were used for downstream analyses (Figure 5b, Supporting Information Table S6). Enrichment analyses of these MAGs conducted using ToppGene (Chen et al., 2009) revealed GO terms associated with cell activation, wound healing, angiogenesis, apoptosis, and immune defense response (Supporting Information Table S7). These analyses were complemented by an enrichment in the MAGs for pathways linked to platelet activation, NFKB signaling, TGFB-SMAD signaling and VLDL metabolism. Additionally, an enrichment of the *ETS2* binding site was observed for these genes, a transcription factor implicated in Alzheimer's and a known transactivator of the APP promoter (Wolvetang et al., 2003).

In order to identify genes specifically associated with microglia in Alzheimer's but not aging, a differential expression analysis was conducted based on the MAGs and core genes, to compare the response of microglia in aging and Alzheimer's. Thus, the expression fold change between the old (≥60 year) versus young, was compared with that of Alzheimer's versus age-matched controls (Figure 5c, Supporting Information Table S6). Reinforcing our preliminary analysis in estimating microglial numbers in Alzheimer's versus age-matched controls, the majority of differentially expressed genes (*FDR* < 0.05) were restricted to the superior frontal gyrus. Interestingly, the trends in expression for each region (represented by the regression line) matched the degree to which each region undergoes neurodegeneration in Alzheimer's. For example, the post-central gyrus, which is comparatively unaffected in Alzheimer's relative to other regions of the brain (Thompson et al., 2003), showed the least upward trend (intercept = −0.01, slope = 0.23). In contrast gene expression in the entorhinal cortex and hippocampus, regions considered vulnerable to Alzheimer's, showed an upward trend, highlighting the significance of these genes in Alzheimer's and not only aging. Although the genes differentially expressed across regions were not all the same, certain genes such as *SAMSN1* (superior frontal gyrus: FC = 1.48, *FDR* < 0.003) and *CX3CR1* (superior frontal gyrus: FC = 0.88, *FDR* < 0.707) had a consistent expression pattern across regions when comparing the expression fold change in Alzheimer's and aging.

To better characterize the microglial response in Alzheimer's, we focused on the superior frontal gyrus, having the most number of differentially expressed genes and being a region significantly affected in Alzheimer's. In identifying genes likely representing changes in activation state rather than cell number, we considered the 52 genes differentially expressed only in Alzheimer's versus age-matched controls. Genes differentially expressed in aged versus young were excluded as they are likely a result of changes in microglia abundance, not disease activation. Enrichment analysis of the genes highlighted processes related to cell activation (*PYCARD* and *PIK3CG*), wound healing (*A2M* and *SERPING1*), innate immune response (*TLR5* and *ITGAM*), and pathways associated with phagocytosis, TLR cascade, and cell activation linked with neuronal survival (Supporting Information Table S8). Moreover, several members of TYROBP signaling pathway were differentially expressed (*SAMSN1*, *SIRPB2*, *CD37*, *IL10RA*, *PIK3CG*, and *BIN2*), a pathway dysregulated in microglia during Alzheimer's (Keren-Shaul et al., 2017; Ma, Jiang, Tan, & Yu, 2015; Zhang et al., 2013). Of the differentially expressed genes, 11 were MAGs including *LYZ*,

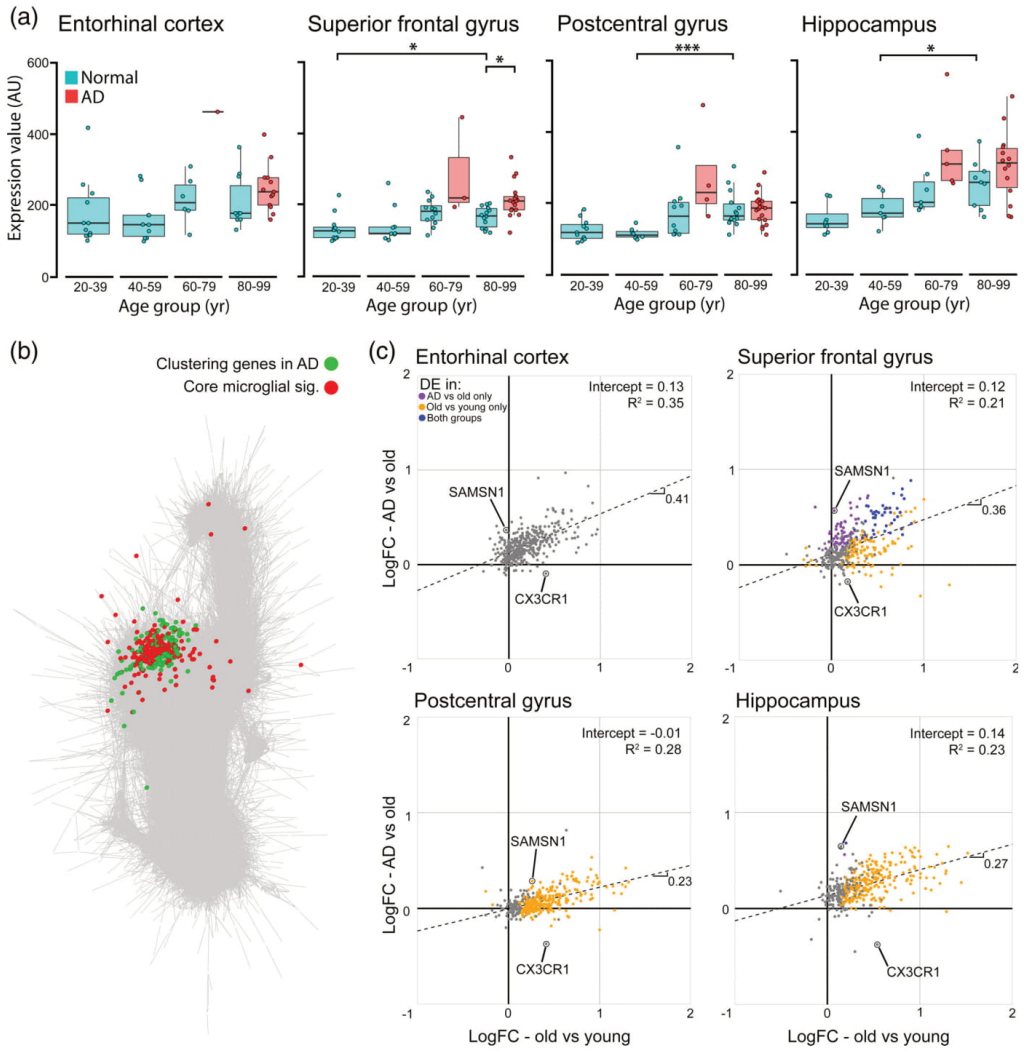


FIGURE 5 The microglial response to Alzheimer's disease. (a) Average expression of core signature genes in normal and Alzheimer's samples from different age groups. (b) Coexpression network highlighting core (red) and MAGs (green) in Alzheimer's samples. (c) For all the MAGs and core genes, a comparison is shown between the fold change of Alzheimer's versus age-matched controls (y-axis) and fold change between old versus young controls (x-axis), across regions. For each comparison differentially expressed genes are shown, in the process of aging (yellow), in Alzheimer's (purple) or differentially expressed in both processes (blue). The trend in expression of these in Alzheimer's and aging is represented by the regression line (dashed line) with their slope, intercept and R^2. Abbreviations: sig = signature; AD = Alzheimer's disease. *Significant at *FDR* < 0.05, *** significant at *FDR* < 0.001 [Color figure can be viewed at wileyonlinelibrary.com]

RPS6KA1, and *SLA*, with known associations to Alzheimer's (Ellison, Bradley-Whitman, & Lovell, 2017; Hu, Xin, Hu, Zhang, & Wang, 2017; Tuppo & Arias, 2005). Interestingly, certain classical microglial marker genes were differentially upregulated in Alzheimer's, for example, *ITGAM* and *PTPRC*, while others showed a downward trend, including *CX3CR1* and *P2RY12*; the latter consistent with a loss of homeostatic microglial identity observed in Alzheimer's mouse models (Keren-Shaul et al., 2017; Krasemann et al., 2017). Alternatively, whilst tissue gene expression can be influenced by cell activation and cell numbers, certain genes found differentially upregulated in both Alzheimer's and aging, such as *TSPO*, *MS4A6A*, and MHC class 2 genes, are known features of microglial activation (Bergen, Kaing, Jacoline, Gorgels, & Janssen, 2015; Hamelin et al., 2016; Hu et al., 2017). Overall, these observations demonstrate the value of the refined microglial signature

we have derived in deducing changes in microglial profile (numbers and functional status) in Alzheimer's and are consistent with the region-specific vulnerability and progression of Alzheimer's pathology.

4 | DISCUSSION

Recent transcriptomic studies, the majority of which have been conducted in mice, have greatly advanced our knowledge of the functional profile of microglia (Butovsky et al., 2014; Darmanis et al., 2015; Galatro et al., 2017; Hickman et al., 2013; Zeisel et al., 2015), their regional heterogeneity in the CNS (Grabert et al., 2016; McCarthy, 2017), and altered functional profile associated with neurodegeneration (Holtman et al., 2015; Keren-Shaul et al., 2017; Miller et al., 2010; Vincenti et al., 2016). However, key differences between mouse and human microglia have been suggested (Galatro et al., 2017; Olah et al., 2018), emphasizing the importance of better characterizing the functional profile of human microglia in health and disease. Our initial investigations demonstrated that published microglia gene signatures vary considerably in their size and composition relative to one another. Contributing factors to the observed discrepancy are likely the differing objectives of studies, *p*-value or fold-change enrichment thresholds used in defining signature genes, donor variability, differing analysis platforms, cell isolation methodologies (Haimon et al., 2018; Okaty, Sugino, & Nelson, 2011) and brain regions examined.

To identify a conserved human microglial signature, we employed gene correlation network (GCN) analysis of transcriptomic data ranging across CNS regions and cell types to exploit the variation across these variables. For this we harnessed the power of large publicly available brain tissue transcriptomics resources, specifically the GTEx and ABA datasets, where gene expression has been analyzed across a comprehensive range of brain regions, collected from numerous donors and analyzed on different platforms (Hawrylycz et al., 2012; Lonsdale et al., 2013). We also included data from different brain cell populations to provide additional discriminatory power to the analysis (Zhang et al., 2013). Using this approach we sought to minimize the technical and biological factors that introduce variation in smaller or more focused studies. Additionally, by using GCNs, we are able to capture the transcriptomic profile common to microglia from all regions of the CNS, as has been shown possible for other cell types and tissues using this approach (Freeman et al., 2012; Mabbott et al. 2013; Shih et al., 2017). Creating a GCN for each dataset, a microglia cluster was identified, based on the presence of canonical marker genes for these cells. The consensus from these dataset derived signatures, provided 249 genes representative of human microglia across datasets. This is the first study to deconvolute a microglia signature from the current GTEx and ABA data using this approach. The derived signature included all known markers of microglia, including *TMEM119*, *P2RY12*, and *CD68* (Bennett et al., 2016; Perego, Fumagalli, & De Simoni, 2011; Wes, Holtman, Boddeke, Möller, & Eggen, 2016) and many other genes known to be associated with microglial/macrophage biology. This includes representatives of the TLR, complement, and MHC class 2 antigen-presenting pathways. While microglia are the most abundant of the myeloid populations in the CNS, we cannot rule out that there may be some genes associated with other populations, such as perivascular macrophages. However, genes specifically expressed by these populations are unlikely to be coexpressed with the microglia cluster, as their distribution across regions is unlikely to be the same.

Validation of the derived signature, included an examination of HPA immunostaining of proteins encoded by the signature genes, demonstrating that they were specifically expressed in myeloid cell types relative to other neuronal and immune cells. On comparing the proposed core human microglial gene signature to those reported by others, it overlapped completely with genes proposed by one or more of the earlier studies, further validating the robustness of the gene list (Supporting Information Table S1). To provide an objective view of the specificity of the current and published human signatures, the coexpression and regional-specific expression of all published signatures was investigated in the context of the new (v7) brain atlas data produced by the GTEx consortium (these data had not been used in the derivation of any of the signatures). Supporting Information Figure S3 shows the results of these analyses. In general, the smaller published signatures which include many of the verified microglial markers for example, Darmanis et al. and Hawarylycz et al. were found to tightly coexpress and exhibited region-specific expression in line with microglia abundance in the CNS. In contrast, more recent larger microglia signatures defined by comparing the expression of isolated cortical microglia to intact brain, that is, Gosselin et al., Galatro et al., and Olah et al., showed a significant range in expression profiles, as illustrated by their wide distribution across the GTEx brain GCN. Whilst these genes are all likely to be expressed in cortical microglia, as this is how they were originally identified, these data suggest that they may be expressed by other cell types elsewhere in the brain. Indeed, IHC data from the HPA suggests that a number of such genes expressed highly in the cerebellum, are expressed in multiple cell types in this region, making them poor markers of microglia outside the brain regions in which they were identified. Other genes in the signatures generated from the analysis of isolated microglia include those associated with NFKB signaling, the interferon and early responses, suggesting that the act of isolation may have partially activated the cells. This isolation-related activation has recently been documented for FACS isolation of mouse microglia (Haimon et al., 2018). In summary, these analyses suggest that the majority of genes that comprise the smaller signatures are microglia-specific, but the lists represent only a fragment of the microglia functional profile. In contrast, larger signature gene lists contain some genes that are likely expressed by other cell types in brain regions different from that in which the signature was derived (cerebral cortex) or are potentially only expressed by cortical microglia. The microglia signature as defined here, attempts to address the shortcomings of previous lists, by providing a comprehensive list of microglial genes whose coexpression pattern is specifically associated with microglia across all regions of the CNS. These results underscore the importance of considering the heterogeneity of cell populations across brain regions and the selection of enrichment thresholds in deriving a conserved microglia signature.

Evidence for the central role microglia play in the pathogenesis of neurodegenerative disease continues to grow, however, the cellular

and molecular changes that occur in these cells during human brain pathologies are poorly understood. Although previous tissue transcriptomics studies have identified key changes in microglial genes, studies often do not discriminate between changes in microglial number and activation state (Miller et al., 2010; Soreq et al., 2017; Zhang et al., 2013). To validate the application of the core signature in describing region dependent alterations in health and disease, and in discriminating between the influence of cell number versus activation state, a dataset consisting of healthy aging and Alzheimer's samples, across a number of brain regions was examined (Berchtold et al., 2013). Given that the majority of microglial genes maintain their expression with age (Galatro et al., 2017; Jyothi et al., 2015; Poliani et al., 2015), we made the assumption that the average expression of the signature genes, can be used as a proxy for microglia number through aging. Supporting this, the increased average expression of signature genes with age was substantiated by studies directly measuring microglia cell numbers (Damani et al., 2011; Peters, Josephson, & Vincent, 1991; Tremblay, Zettel, Ison, Allen, & Majewska, 2012). The largest changes observed were in the hippocampus, a region particularly vulnerable to aging, and where greater microglial activation and neuronal loss have been observed with age relative to the other cortical regions (Bartsch & Wulff, 2015; Kumar et al., 2012; Raz et al., 2005). Our analyses support the fact that microglial numbers change in a region-dependent manner and that these changes correlate with age-associated regional atrophy and inflammation. When comparing Alzheimer's to age-matched controls, a similar trend toward increased expression levels of signature genes was also observed. However, a significant increase was only observed in the superior frontal gyrus, a region known to be highly susceptible to the effects of both aging and Alzheimer's, based on neuronal connectivity studies (Bakkour et al., 2013; Stam, 2014). Interestingly, the entorhinal cortex and hippocampus, whose atrophy characterize Alzheimer's pathology, showed the greatest differences between Alzheimer's and controls, although these lacked statistical significance, likely due to the relatively small number of samples and large variability between them (Khan et al., 2014; Velayudhan et al., 2013). In contrast, the post-central gyrus, a region shown to maintain its grey matter content and functional connectivity relative to other regions in late-onset Alzheimer's, showed little change in Alzheimer's derived samples versus controls (Adriaanse et al., 2014; Thompson et al., 2003). Strikingly, these findings are consistent with regional Alzheimer's progression based on tau burden, neuroinflammation, and neuronal loss, which are prominent in the entorhinal cortex and hippocampus (Cope et al., 2018; Freer et al., 2016; Kreisl et al., 2016). Overall, these data demonstrate the utility of the signature in assessing quantitative differences in microglial numbers from tissue-level expression datasets.

To gain insight into molecular pathways specifically affected in Alzheimer's, qualitative changes in the profile of microglia were examined. Coexpression analysis identified a set of 165 MAGs correlating with the core gene signature in samples isolated from Alzheimer's patients. The MAG list was enriched in genes from various pathways associated with innate immune signaling, consistent with the inflammatory environment within Alzheimer's brain tissue and the reactivity of microglia within this environment, for example *TSPO* (Kumar et al., 2012). It was particularly interesting to note that genes involved in lipid regulation and wound healing, associated with Alzheimer's disease, were over-represented in the MAGs set (Cervantes et al., 2011; Lorenzl et al., 2003; Petit-Turcotte et al., 2001; Shih et al., 2014). Members of the *APOC* gene family and *ECHDC3* are known to regulate levels of certain lipids, linked with Alzheimer's progression (Adunsky et al., 2002; Desikan et al., 2015; Lane & Farlow, 2005). Additionally, these factors are part of the wound healing cascade, which include proteins such as *TIMP1* and *PROS1*, which are key in regulating tissue integrity and plasticity, altogether pointing toward the disrupted blood–brain barrier in Alzheimer's (Bennett et al., 2016; Duits et al., 2015). These results provide some insight and support for the complexity of microglia involvement in Alzheimer's, through not only inflammatory mechanisms but also through upregulation of metabolic and tissue homeostasis/repair functions (Vincenti et al., 2016). Investigating alterations of microglial differentiation in Alzheimer's, we focused on genes differentially expressed in Alzheimer's compared with age-matched controls. Although for all regions, the majority of genes presented an upward trend of expression, most lacked significance. The exceptions were those associated with microglia of the superior frontal gyrus, which were further investigated. Genes relating to TYROBP signaling, which is implicated in Alzheimer's and regulates phagocytosis, cell proliferation, activation and survival, were significantly upregulated (Keren-Shaul et al., 2017; Landreth & Reed-Geaghan, 2009; Ma et al., 2015). Substantiating these findings *Tyrobp* knockout mice models have proven to suppress inflammation in neurodegenerative models including Alzheimer's, thereby minimizing neuronal dystrophy, implicating a failure in the resolution of inflammation in Alzheimer's (Bakker et al., 2000; Haure-Mirande et al., 2017). Interestingly mutations and expression of downstream members are also linked with Alzheimer's including *CD33*, *TREM2* and *CR3* (Hamerman, Tchao, Lowell, & Lanier, 2005; Takahashi, Rochford, & Neumann, 2005).

In summary, we have employed a coexpression analysis approach to derive a core human microglial signature under non-neuropathologic conditions that is robust to potential artifacts generated by technical and biological variation that can influence other approaches to signature derivation. Furthermore, we present the utility of this signature, demonstrating its sensitivity to detect region-specific changes in microglial alterations in aging and Alzheimer's disease, while appreciating the influence of cell number and activation in tissue transcriptomics data. We found that these responses were aligned with the known neuropathological trajectory of Alzheimer's. We propose the conserved signature described here as a specific and robust resource of gene markers that reflect the core functional profile of these cells and can aid future studies of microglial biology in the human CNS, including those using bulk and single cell transcriptomics.

ACKNOWLEDGMENTS

T.C.F. and B.W.M. are funded by an Institute Strategic Programme Grant funding from the Biotechnology and Biological Sciences Research Council [BB/J004227/1]. B.W.M. receives funding from the UK Dementia Research Institute and Medical Research Council [MR/L003384/1]. B.S. is supported by Experimental Medicine

Challenge Grant funding from the Medical Research Council [MR/M003833/1]. We would like to thank Prof. Bin Zhang from the Icahn school of medicine who generously provided data used in our analysis.

CONFLICT OF INTEREST

The authors declare that they have no conflicts of interest.

ORCID

Anirudh Patir https://orcid.org/0000-0001-6980-0100

REFERENCES

Adriaanse, S. M., Binnewijzend, M. A., Ossenkoppele, R., Tijms, B. M., van der Flier, W. M., Koene, T., … van Berckel, B. N. (2014). Widespread disruption of functional brain organization in early-onset Alzheimer's disease. *PLoS One*, *9*, e102995.

Adunsky, A., Chesnin, V., Davidson, M., Gerber, Y., Alexander, K., & Haratz, D. (2002). A cross-sectional study of lipids and ApoC levels in Alzheimer's patients with and without cardiovascular disease. *The Journals of Gerontology Series A: Biological Sciences and Medical Sciences*, *57*, M757–M761.

Bakker, A. B., Hoek, R. M., Cerwenka, A., Blom, B., Lucian, L., McNeil, T., … Lanier, L. L. (2000). DAP12-deficient mice fail to develop autoimmunity due to impaired antigen priming. *Immunity*, *13*, 345–353.

Bakkour, A., Morris, J. C., Wolk, D. A., & Dickerson, B. C. (2013). The effects of aging and Alzheimer's disease on cerebral cortical anatomy: Specificity and differential relationships with cognition. *NeuroImage*, *76*, 332–344.

Bartsch, T., & Wulff, P. (2015). The hippocampus in aging and disease: From plasticity to vulnerability. *Neuroscience*, *309*, 1–16.

Bennett, M. L., Bennett, F. C., Liddelow, S. A., Ajami, B., Zamanian, J. L., Fernhoff, N. B., … Tucker, A. (2016). New tools for studying microglia in the mouse and human CNS. *Proceedings of the National Academy of Sciences of the United States of America*, *113*, E1738–E1746.

Berchtold, N. C., Coleman, P. D., Cribbs, D. H., Rogers, J., Gillen, D. L., & Cotman, C. W. (2013). Synaptic genes are extensively downregulated across multiple brain regions in normal human aging and Alzheimer's disease. *Neurobiology of Aging*, *34*, 1653–1661.

Berchtold, N. C., Cribbs, D. H., Coleman, P. D., Rogers, J., Head, E., Kim, R., … Trojanowski, J. Q. (2008). Gene expression changes in the course of normal brain aging are sexually dimorphic. *Proceedings of the National Academy of Sciences of the United States of America*, *105*, 15605–15610.

Bergen, A. A., Kaing, S., Jacoline, B., Gorgels, T. G., & Janssen, S. F. (2015). Gene expression and functional annotation of human choroid plexus epithelium failure in Alzheimer's disease. *BMC Genomics*, *16*, 956.

Butovsky, O., Jedrychowski, M. P., Moore, C. S., Cialic, R., Lanser, A. J., Gabriely, G., … Doykan, C. E. (2014). Identification of a unique TGF-[beta]-dependent molecular and functional signature in microglia. *Nature Neuroscience*, *17*, 131–143.

Carvalho, B. S., & Irizarry, R. A. (2010). A framework for oligonucleotide microarray preprocessing. *Bioinformatics*, *26*, 2363–2367.

Cervantes, S., Samaranch, L., Vidal-Taboada, J. M., Lamet, I., Bullido, M. J., Frank-García, A., … Lorenzo, E. (2011). Genetic variation in APOE cluster region and Alzheimer's disease risk. *Neurobiology of Aging*, *32*, 2107.e7–2107.e17.

Chen, J., Bardes, E. E., Aronow, B. J., & Jegga, A. G. (2009). ToppGene suite for gene list enrichment analysis and candidate gene prioritization. *Nucleic Acids Research*, *37*, W305–W311.

Chiu, I. M., Morimoto, E. T., Goodarzi, H., Liao, J. T., O'Keeffe, S., Phatnani, H. P., … Maniatis, T. (2013). A neurodegeneration-specific gene-expression signature of acutely isolated microglia from an amyotrophic lateral sclerosis mouse model. *Cell Reports*, *4*, 385–401.

Colonna, M., & Butovsky, O. (2017). Microglia function in the central nervous system during health and neurodegeneration. *Annual Review of Immunology*, *35*, 441–468.

Cope, T. E., Rittman, T., Borchert, R. J., Jones, P. S., Vatansever, D., Allinson, K., … O'Brien, J. T. (2018). Tau burden and the functional connectome in Alzheimer's disease and progressive supranuclear palsy. *Brain*, *141*, 550–567.

Cunningham, C. (2013). Microglia and neurodegeneration: The role of systemic inflammation. *Glia*, *61*, 71–90.

Damani, M. R., Zhao, L., Fontainhas, A. M., Amaral, J., Fariss, R. N., & Wong, W. T. (2011). Age-related alterations in the dynamic behavior of microglia. *Aging Cell*, *10*, 263–276.

Darmanis, S., Sloan, S. A., Zhang, Y., Enge, M., Caneda, C., Shuer, L. M., … Quake, S. R. (2015). A survey of human brain transcriptome diversity at the single cell level. *Proceedings of the National Academy of Sciences of the United States of America*, *112*, 7285–7290.

Desikan, R. S., Schork, A. J., Wang, Y., Thompson, W. K., Dehghan, A., Ridker, P. M., … Chen, C.-H. (2015). Polygenic overlap between C-reactive protein, plasma lipids and Alzheimer's disease. *Circulation*, *131*, 2061–2069.

Duits, F. H., Hernandez-Guillamon, M., Montaner, J., Goos, J. D., Montañola, A., Wattjes, M. P., … van der Flier, W. M. (2015). Matrix metalloproteinases in Alzheimer's disease and concurrent cerebral microbleeds. *Journal of Alzheimer's Disease*, *48*, 711–720.

Durafourt, B. A., Moore, C. S., Zammit, D. A., Johnson, T. A., Zaguia, F., Guiot, M. C., … Antel, J. P. (2012). Comparison of polarization properties of human adult microglia and blood-derived macrophages. *Glia*, *60*, 717–727.

Ellison, E. M., Bradley-Whitman, M. A., & Lovell, M. A. (2017). Single-Base resolution mapping of 5-Hydroxymethylcytosine modifications in hippocampus of Alzheimer's disease subjects. *Journal of Molecular Neuroscience*, *63*, 185–197.

Elmore, M. R., Najafi, A. R., Koike, M. A., Dagher, N. N., Spangenberg, E. E., Rice, R. A., … West, B. L. (2014). Colony-stimulating factor 1 receptor signaling is necessary for microglia viability, unmasking a microglia progenitor cell in the adult brain. *Neuron*, *82*, 380–397.

Enright, A. J., Van Dongen, S., & Ouzounis, C. A. (2002). An efficient algorithm for large-scale detection of protein families. *Nucleic Acids Research*, *30*, 1575–1584.

Freeman, T. C., Goldovsky, L., Brosch, M., Van Dongen, S., Mazière, P., Grocock, R. J., … Enright, A. J. (2007). Construction, visualisation, and clustering of transcription networks from microarray expression data. *PLoS Computational Biology*, *3*, e206–e2042.

Freeman, T. C., Ivens, A., Baillie, J. K., Beraldi, D., Barnett, M. W., Dorward, D., … Hume, D. A. (2012). A gene expression atlas of the domestic pig. *BMC Biology*, *10*, 90.

Freer, R., Sormanni, P., Vecchi, G., Ciryam, P., Dobson, C. M., & Vendruscolo, M. (2016). A protein homeostasis signature in healthy brains recapitulates tissue vulnerability to Alzheimer's disease. *Science Advances*, *2*, e1600947.

Galatro, T. F., Holtman, I. R., Lerario, A. M., Vainchtein, I. D., Brouwer, N., Sola, P. R., … Eggen, B. J. L. (2017). Transcriptomic analysis of purified human cortical microglia reveals age-associated changes. *Nature Neuroscience*, *20*, 1162.

Ginhoux, F., Greter, M., Leboeuf, M., Nandi, S., See, P., Gokhan, S., … Stanley, E. R. (2010). Fate mapping analysis reveals that adult microglia derive from primitive macrophages. *Science*, *330*, 841–845.

Gosselin, D., Skola, D., Coufal, N. G., Holtman, I. R., Schlachetzki, J. C., Sajti, E., … Pasillas, M. P. (2017). An environment-dependent transcriptional network specifies human microglia identity. *Science*, *356*, eaal3222.

Grabert, K., Michoel, T., Karavolos, M. H., Clohisey, S., Baillie, J. K., Stevens, M. P., … McColl, B. W. (2016). Microglial brain region-dependent diversity and selective regional sensitivities to aging. *Nature Neuroscience*, *19*, 504–516.

Haimon, Z., Volaski, A., Orthgiess, J., Boura-Halfon, S., Varol, D., Shemer, A., … Jung, S. (2018). Re-evaluating microglia expression profiles using RiboTag and cell isolation strategies. *Nature Immunology*, *19*, 636–644.

Hamelin, L., Lagarde, J., Dorothée, G., Leroy, C., Labit, M., Comley, R. A., … Bertoux, M. (2016). Early and protective microglial activation in

Alzheimer's disease: A prospective study using 18 F-DPA-714 PET imaging. *Brain*, *139*, 1252–1264.

Hamerman, J. A., Tchao, N. K., Lowell, C. A., & Lanier, L. L. (2005). Enhanced toll-like receptor responses in the absence of signaling adaptor DAP12. *Nature Immunology*, *6*, 579–586.

Hanisch, U.-K. (2013). Functional diversity of microglia – How heterogeneous are they to begin with? *Frontiers in Cellular Neuroscience*, *7*, 65.

Hanisch, U.-K., & Kettenmann, H. (2007). Microglia: Active sensor and versatile effector cells in the normal and pathologic brain. *Nature Neuroscience*, *10*, 1387–1394.

Haure-Mirande, J.-V., Audrain, M., Fanutza, T., Kim, S. H., Klein, W. L., Glabe, C., … Wang, M. (2017). Deficiency of TYROBP, an adapter protein for TREM2 and CR3 receptors, is neuroprotective in a mouse model of early Alzheimer's pathology. *Acta Neuropathologica*, *134*, 769–788.

Hawrylycz, M. J., Lein, E. S., Guillozet-Bongaarts, A. L., Shen, E. H., Ng, L., Miller, J. A., … Riley, Z. L. (2012). An anatomically comprehensive atlas of the adult human brain transcriptome. *Nature*, *489*, 391–399.

Hickey, W. F., & Kimura, H. (1988). Perivascular microglial cells of the CNS are bone marrow-derived and present antigen in vivo. *Science*, *239*, 290–292.

Hickman, S. E., Kingery, N. D., Ohsumi, T. K., Borowsky, M. L., Wang, L.-C., Means, T. K., & El Khoury, J. (2013). The microglial sensome revealed by direct RNA sequencing. *Nature Neuroscience*, *16*, 1896–1905.

Holtman, I. R., Raj, D. D., Miller, J. A., Schaafsma, W., Yin, Z., Brouwer, N., … Eggen, B. J. (2015). Induction of a common microglia gene expression signature by aging and neurodegenerative conditions: A co-expression meta-analysis. *Acta Neuropathologica Communications*, *3*, 31.

Hong, S., Beja-Glasser, V. F., Nfonoyim, B. M., Frouin, A., Li, S., Ramakrishnan, S., … Barres, B. A. (2016). Complement and microglia mediate early synapse loss in Alzheimer mouse models. *Science*, *352*, 712–716.

Hoshiko, M., Arnoux, I., Avignone, E., Yamamoto, N., & Audinat, E. (2012). Deficiency of the microglial receptor CX3CR1 impairs postnatal functional development of thalamocortical synapses in the barrel cortex. *Journal of Neuroscience*, *32*, 15106–15111.

Hu, Y.-S., Xin, J., Hu, Y., Zhang, L., & Wang, J. (2017). Analyzing the genes related to Alzheimer's disease via a network and pathway-based approach. *Alzheimer's Research & Therapy*, *9*, 29.

Hubbard, T., Barker, D., Birney, E., Cameron, G., Chen, Y., Clark, L., … Down, T. (2002). The Ensembl genome database project. *Nucleic Acids Research*, *30*, 38–41.

Jyothi, H., Vidyadhara, D., Mahadevan, A., Philip, M., Parmar, S. K., Manohari, S. G., … Alladi, P. A. (2015). Aging causes morphological alterations in astrocytes and microglia in human substantia nigra pars compacta. *Neurobiology of Aging*, *36*, 3321–3333.

Kauffmann, A., Gentleman, R., & Huber, W. (2008). arrayQualityMetrics—A bioconductor package for quality assessment of microarray data. *Bioinformatics*, *25*, 415–416.

Keren-Shaul, H., Spinrad, A., Weiner, A., Matcovitch-Natan, O., Dvir-Szternfeld, R., Ulland, T. K., … Amit, I. (2017). A unique microglia type associated with restricting development of Alzheimer's disease. *Cell*, *169*, 1276–1290.

Khan, U. A., Liu, L., Provenzano, F. A., Berman, D. E., Profaci, C. P., Sloan, R., … Small, S. A. (2014). Molecular drivers and cortical spread of lateral entorhinal cortex dysfunction in preclinical Alzheimer's disease. *Nature Neuroscience*, *17*, 304–311.

Krasemann, S., Madore, C., Cialic, R., Baufeld, C., Calcagno, N., El Fatimy, R., … Butovsky, O. (2017). The TREM2-APOE pathway drives the transcriptional phenotype of dysfunctional microglia in neurodegenerative diseases. *Immunity*, *47*, 566–581.

Kreisl, W. C., Lyoo, C. H., Liow, J.-S., Wei, M., Snow, J., Page, E., … Pike, V. W. (2016). 11 C-PBR28 binding to translocator protein increases with progression of Alzheimer's disease. *Neurobiology of Aging*, *44*, 53–61.

Kumar, A., Muzik, O., Shandal, V., Chugani, D., Chakraborty, P., & Chugani, H. T. (2012). Evaluation of age-related changes in translocator protein (TSPO) in human brain using 11C-[R]-PK11195 PET. *Journal of Neuroinflammation*, *9*, 232.

Lai, A. Y., Dhami, K. S., Dibal, C. D., & Todd, K. G. (2011). Neonatal rat microglia derived from different brain regions have distinct activation responses. *Neuron Glia Biology*, *7*, 5–16.

Landreth, G. E., & Reed-Geaghan, E. G. (2009). Toll-like receptors in Alzheimer's disease. In T. Kielian (Ed.), *Toll-like receptors: Roles in infection and neuropathology* (pp. 137–153). Berlin, Heidelberg: Springer.

Lane, R. M., & Farlow, M. R. (2005). Lipid homeostasis and apolipoprotein E in the development and progression of Alzheimer's disease. *Journal of Lipid Research*, *46*, 949–968.

Lawson, L., Perry, V., Dri, P., & Gordon, S. (1990). Heterogeneity in the distribution and morphology of microglia in the normal adult mouse brain. *Neuroscience*, *39*, 151–170.

Leong, S. K., & Ling, E. A. (1992). Amoeboid and ramified microglia: Their interrelationship and response to brain injury. *Glia*, *6*, 39–47.

Li, Q., & Barres, B. A. (2017). Microglia and macrophages in brain homeostasis and disease. *Nature Reviews Immunology*, *18*, 225–242.

Lobsiger, C. S., & Cleveland, D. W. (2007). Glial cells as intrinsic components of non-cell-autonomous neurodegenerative disease. *Nature Neuroscience*, *10*, 1355–1360.

Lonsdale, J., Thomas, J., Salvatore, M., Phillips, R., Lo, E., Shad, S., … Young, N. (2013). The genotype-tissue expression (GTEx) project. *Nature Genetics*, *45*, 580–585.

Lorenzl, S., Albers, D. S., Relkin, N., Ngyuen, T., Hilgenberg, S. L., Chirichigno, J., … Beal, M. F. (2003). Increased plasma levels of matrix metalloproteinase-9 in patients with Alzheimer's disease. *Neurochemistry International*, *43*, 191–196.

Ma, J., Jiang, T., Tan, L., & Yu, J.-T. (2015). TYROBP in Alzheimer's disease. *Molecular Neurobiology*, *51*, 820–826.

Mabbott, N. A., Baillie, J. K., Brown, H., Freeman, T. C., & Hume, D. A. (2013). An expression atlas of human primary cells: Inference of gene function from coexpression networks. *BMC Genomics*, *14*, 632.

Manocha, G. D., Floden, A. M., Rausch, K., Kulas, J. A., McGregor, B. A., Rojanathammanee, L., … Nichols, M. R. (2016). APP regulates microglial phenotype in a mouse model of Alzheimer's disease. *Journal of Neuroscience*, *36*, 8471–8486.

Marinelli, C., Di Liddo, R., Facci, L., Bertalot, T., Conconi, M. T., Zusso, M., … Giusti, P. (2015). Ligand engagement of toll-like receptors regulates their expression in cortical microglia and astrocytes. *Journal of Neuroinflammation*, *12*, 244.

McCarthy, M. M. (2017). Location, location, location: Microglia are where they live. *Neuron*, *95*, 233–235.

Miller, J. A., Horvath, S., & Geschwind, D. H. (2010). Divergence of human and mouse brain transcriptome highlights Alzheimer disease pathways. *Proceedings of the National Academy of Sciences of the United States of America*, *107*, 12698–12703.

Mittelbronn, M., Dietz, K., Schluesener, H. J., & Meyermann, R. (2001). Local distribution of microglia in the normal adult human central nervous system differs by up to one order of magnitude. *Acta Neuropathologica*, *101*, 249–255.

Nilsson, P., Paavilainen, L., Larsson, K., Ödling, J., Sundberg, M., Andersson, A. C., … Ottosson, J. (2005). Towards a human proteome atlas: High-throughput generation of mono-specific antibodies for tissue profiling. *Proteomics*, *5*, 4327–4337.

Okaty, B. W., Sugino, K., & Nelson, S. B. (2011). A quantitative comparison of cell-type-specific microarray gene expression profiling methods in the mouse brain. *PLoS One*, *6*, e16493.

Olah, M., Patrick, E., Villani, A.-C., Xu, J., White, C. C., Ryan, K. J., … Cimpean, M. (2018). A transcriptomic atlas of aged human microglia. *Nature Communications*, *9*, 539.

Oldham, M. C., Konopka, G., Iwamoto, K., Langfelder, P., Kato, T., Horvath, S., & Geschwind, D. H. (2008). Functional organization of the transcriptome in human brain. *Nature Neuroscience*, *11*, 1271–1282.

Paolicelli, R. C., Bolasco, G., Pagani, F., Maggi, L., Scianni, M., Panzanelli, P., … Dumas, L. (2011). Synaptic pruning by microglia is necessary for normal brain development. *Science*, *333*, 1456–1458.

Perego, C., Fumagalli, S., & De Simoni, M.-G. (2011). Temporal pattern of expression and colocalization of microglia/macrophage phenotype markers following brain ischemic injury in mice. *Journal of Neuroinflammation*, *8*, 174.

Perry, V. H., & Holmes, C. (2014). Microglial priming in neurodegenerative disease. *Nature Reviews Neurology*, *10*, 217–224.

Peters, A., Josephson, K., & Vincent, S. L. (1991). Effects of aging on the neuroglial cells and pericytes within area 17 of the rhesus monkey cerebral cortex. *The Anatomical Record*, *229*, 384–398.

Petit-Turcotte, C., Stohl, S. M., Beffert, U., Cohn, J. S., Aumont, N., Tremblay, M., … Shachter, N. S. (2001). Apolipoprotein CI expression in the brain in Alzheimer's disease. *Neurobiology of Disease*, *8*, 953–963.

Poliani, P. L., Wang, Y., Fontana, E., Robinette, M. L., Yamanishi, Y., Gilfillan, S., & Colonna, M. (2015). TREM2 sustains microglial expansion during aging and response to demyelination. *The Journal of Clinical Investigation*, *125*, 2161–2170.

Povey, S., Lovering, R., Bruford, E., Wright, M., Lush, M., & Wain, H. (2001). The HUGO gene nomenclature committee (HGNC). *Human Genetics*, *109*, 678–680.

Prinz, M., & Priller, J. (2014). Microglia and brain macrophages in the molecular age: From origin to neuropsychiatric disease. *Nature Reviews Neuroscience*, *15*, 300–312.

Raz, N., Lindenberger, U., Rodrigue, K. M., Kennedy, K. M., Head, D., Williamson, A., … Acker, J. D. (2005). Regional brain changes in aging healthy adults: General trends, individual differences and modifiers. *Cerebral Cortex*, *15*, 1676–1689.

Reimand, J., Arak, T., Adler, P., Kolberg, L., Reisberg, S., Peterson, H., & Vilo, J. (2016). Profiler-a web server for functional interpretation of gene lists (2016 update). *Nucleic Acids Research*, *44*, W83–W89.

Safran, M., Dalah, I., Alexander, J., Rosen, N., Iny Stein, T., Shmoish, M., … Lancet, D. (2010). GeneCards version 3: The human gene integrator. *Database*, *2010*, baq020.

Salter, M. W., & Stevens, B. (2017). Microglia emerge as central players in brain disease. *Nature Medicine*, *23*, 1018–1027.

Satoh, J., Kino, Y., Asahina, N., Takitani, M., Miyoshi, J., Ishida, T., & Saito, Y. (2016). TMEM119 marks a subset of microglia in the human brain. *Neuropathology*, *36*, 39–49.

Scholz, J., & Woolf, C. J. (2007). The neuropathic pain triad: Neurons, immune cells and glia. *Nature Neuroscience*, *10*, 1361–1368.

Shaw, D. R. (2009). Searching the mouse genome informatics (MGI) resources for information on mouse biology from genotype to phenotype. *Current Protocols in Bioinformatics*, *56*, 1.7.1–1.7.16.

Shen, E. H., Overly, C. C., & Jones, A. R. (2012). The Allen human brain atlas: Comprehensive gene expression mapping of the human brain. *Trends in Neurosciences*, *35*, 711–714.

Shih, B. B., Nirmal, A. J., Headon, D. J., Akbar, A. N., Mabbott, N. A., & Freeman, T. C. (2017). Derivation of marker gene signatures from human skin and their use in the interpretation of the transcriptional changes associated with dermatological disorders. *The Journal of Pathology*, *241*, 600–613.

Shih, Y.-H., Tsai, K.-J., Lee, C.-W., Shiesh, S.-C., Chen, W.-T., Pai, M.-C., & Kuo, Y.-M. (2014). Apolipoprotein C-III is an amyloid-β-binding protein and an early marker for Alzheimer's disease. *Journal of Alzheimer's Disease*, *41*, 855–865.

Smyth, G. K. (2005). Limma: Linear models for microarray data. In R. Gentleman, V. J. Carey, W. Huber, R. A. Irizarry, & S. Dudoit (Eds.), *Bioinformatics and computational biology solutions using R and bioconductor* (pp. 397–420). New York, NY: Springer New York.

Soreq, L., Rose, J., Soreq, E., Hardy, J., Trabzuni, D., Cookson, M. R., … Ule, J. (2017). Major shifts in glial regional identity are a transcriptional hallmark of human brain aging. *Cell Reports*, *18*, 557–570.

Stam, C. J. (2014). Modern network science of neurological disorders. *Nature Reviews Neuroscience*, *15*, 683–695.

Szulzewsky, F., Arora, S., de Witte, L., Ulas, T., Markovic, D., Schultze, J. L., … Kettenmann, H. (2016). Human glioblastoma-associated microglia/monocytes express a distinct RNA profile compared to human control and murine samples. *Glia*, *64*, 1416–1436.

Takahashi, K., Rochford, C. D., & Neumann, H. (2005). Clearance of apoptotic neurons without inflammation by microglial triggering receptor expressed on myeloid cells-2. *Journal of Experimental Medicine*, *201*, 647–657.

Theocharidis, A., Van Dongen, S., Enright, A. J., & Freeman, T. C. (2009). Network visualization and analysis of gene expression data using BioLayout Express3D. *Nature Protocols*, *4*, 1535–1550.

Thompson, P. M., Hayashi, K. M., De Zubicaray, G., Janke, A. L., Rose, S. E., Semple, J., … Doddrell, D. M. (2003). Dynamics of gray matter loss in Alzheimer's disease. *Journal of Neuroscience*, *23*, 994–1005.

Tremblay, M. È., Zettel, M. L., Ison, J. R., Allen, P. D., & Majewska, A. K. (2012). Effects of aging and sensory loss on glial cells in mouse visual and auditory cortices. *Glia*, *60*, 541–558.

Tuppo, E. E., & Arias, H. R. (2005). The role of inflammation in Alzheimer's disease. *The International Journal of Biochemistry & Cell Biology*, *37*, 289–305.

Van Dongen, S., & Abreu-Goodger, C. (2012). Using MCL to extract clusters from networks. In J. van Helden, A. Toussaint, & D. Thieffry (Eds.), *Bacterial molecular networks: Methods and protocols* (pp. 281–295). New York, NY: Springer New York.

Velayudhan, L., Proitsi, P., Westman, E., Muehlboeck, J., Mecocci, P., Vellas, B., … Spenger, C. (2013). Entorhinal cortex thickness predicts cognitive decline in Alzheimer's disease. *Journal of Alzheimer's Disease*, *33*, 755–766.

Vincenti, J. E., Murphy, L., Grabert, K., McColl, B. W., Cancellotti, E., Freeman, T. C., & Manson, J. C. (2016). Defining the microglia response during the time course of chronic neurodegeneration. *Journal of Virology*, *90*, 3003–3017.

Wehrspaun, C. C., Haerty, W., & Ponting, C. P. (2015). Microglia recapitulate a hematopoietic master regulator network in the aging human frontal cortex. *Neurobiology of Aging*, *36*, 2443.e9–2443.e20.

Wes, P. D., Holtman, I. R., Boddeke, E. W., Möller, T., & Eggen, B. J. (2016). Next generation transcriptomics and genomics elucidate biological complexity of microglia in health and disease. *Glia*, *64*, 197–213.

Wolvetang, E., Bradfield, O., Hatzistavrou, T., Crack, P., Busciglio, J., Kola, I., & Hertzog, P. (2003). Overexpression of the chromosome 21 transcription factor Ets2 induces neuronal apoptosis. *Neurobiology of Disease*, *14*, 349–356.

Yeh, F. L., Hansen, D. V., & Sheng, M. (2017). TREM2, microglia, and neurodegenerative diseases. *Trends in Molecular Medicine*, *23*, 512–533.

Yokokura, M., Mori, N., Yagi, S., Yoshikawa, E., Kikuchi, M., Yoshihara, Y., … Suda, S. (2011). In vivo changes in microglial activation and amyloid deposits in brain regions with hypometabolism in Alzheimer's disease. *European Journal of Nuclear Medicine and Molecular Imaging*, *38*, 343–351.

Zeisel, A., Muñoz-Manchado, A. B., Codeluppi, S., Lönnerberg, P., La Manno, G., Juréus, A., … Betsholtz, C. (2015). Cell types in the mouse cortex and hippocampus revealed by single-cell RNA-seq. *Science*, *347*, 1138–1142.

Zhang, B., Gaiteri, C., Bodea, L.-G., Wang, Z., McElwee, J., Podtelezhnikov, A. A., … Dobrin, R. (2013). Integrated systems approach identifies genetic nodes and networks in late-onset Alzheimer's disease. *Cell*, *153*, 707–720.

Zhang, Y., Sloan, S. A., Clarke, L. E., Caneda, C., Plaza, C. A., Blumenthal, P. D., … Barres, B. A. (2016). Purification and characterization of progenitor and mature human astrocytes reveals transcriptional and functional differences with mouse. *Neuron*, *89*, 37–53.

SUPPORTING INFORMATION

Additional supporting information may be found online in the Supporting Information section at the end of the article.

How to cite this article: Patir A, Shih B, McColl BW, Freeman TC. A core transcriptional signature of human microglia: Derivation and utility in describing region-dependent alterations associated with Alzheimer's disease. *Glia*. 2019;67:1240–1253. https://doi.org/10.1002/glia.23572

2.2 Discussion

In this chapter we studied the human microglia in ageing and Alzheimer's disease which was motivated by the discrepancies observed between other studies defining this cell signature, the regional heterogeneity of microglia and the susceptibility of human microglia towards diseases like Alzheimer's. At the time, bulk tissue transcriptomics was the dominant source of gene expression data from post-mortem human brain samples. Today, four years later this resource continues to expand, however with increasing single cell and nuclear RNA-Seq data e.g. ABA recently generated the spatially-resolved transcript amplicon readout mapping (STARmap) which includes single cell spatial transcriptomics of the mouse brain (Wang et al., 2018). Using the selected bulk transcriptomics datasets, we identified core signature genes of the human microglia while trying to take account of technical (sequencing platforms) and biological factors (donors and brain regions). Another factor for future analyses would be the methodologies used for signature derivation as these methods have tuneable parameters which influence the sensitivity and size of the signature. A meta-analysis of current single cell and bulk RNA-Seq datasets could definitely help describe a core microglial signature as well as regional homeostatic signatures and activation states. The advantage of deriving a core signature was that it would enable the investigation into regional responses of microglia as shown for ageing and in Alzheimer's disease. Since the study was published it has also been used in determining the microglial state in Huntington's disease (Al-Dalahmah et al., 2020). Whilst, GCN analysis was ideal for deriving a core human microglia signature as the cell type represents the dominant macrophage or immune cell type within the brain, it could be challenging to derive such TRIC signatures in other tissues where TRICs represented a small proportion of the cell within the tissue and shared their biology/pathways with other resident cells. Hence, single cell or pooled cell derived RNA-Seq would be preferred in more complex tissues. A recent single nuclear RNA-Seq (snRNA-Seq) study of Alzheimer's disease identified microglial states of which associated genes matched those identified here in, including the downregulation of homeostatic genes (*CX3CR1* and *P2RY12*) and upregulation of immune related inflammatory genes (*HLA-DRA* and *SPP1*) (Mathys et al., 2019). However, as a note of caution a recent snRNA-Seq study has found poor capture efficiency of certain key genes characteristic of mouse disease associated microglia (Thrupp et al., 2020) which highlights the importance of high quality (coverage and sensitivity) sequencing data such as bulk transcriptomics. Hopefully in the future the

quality of data generated from single cell technologies would compare with that of bulk RNA-Seq.

From a biological standpoint we have examined region-dependent microglial responses in ageing and Alzheimer's. These results match those from single cell studies however similar region-dependent analyses are still required to describe potential microglia subpopulations. In terms of signature derivation, the analyses showed the potential of using bulk transcriptomics data. Altogether the study shows the relevance of deriving homeostatic signatures to understand cell type biology under physiological conditions and using them to understand these cells in pathology, while highlighting the possible technical challenges in deriving cell signatures.

3. Chapter 3: Diversity of tissue-resident immune cells in mouse and their gene signatures

3.1 Introduction

In chapter 2, the transcriptomic profile of human microglia was investigated in health, ageing and Alzheimer's disease. Like other studies our analyses revealed the region-dependent heterogeneity of microglia. Indeed, the transcriptional signature of these cells has shown to be different from tissue-naïve macrophages and they have specialised functionality associated with CNS homeostasis, such as synaptic organization (Galatro et al., 2017). By understanding the biology of homeostatic microglia, their role in pathological conditions such as ageing, and Alzheimer's disease could be investigated. Hence, an understanding of the homeostatic signature of TRICs can help describe their role in disease. With this in mind, the analyses in this chapter expands from microglia to immune cells across different tissues under homeostatic conditions.

In addition to their protective role, TRICs are specialised towards maintaining homoeostasis of particular tissues. In the introduction, the role of TRICs both in health and disease has been described. Briefly, immune cells exposed to certain environmental cues from the surrounding tissue can differentiate into TRICs, which involves remodelling of their transcriptomic and epigenetic landscape (Yoshida et al., 2019, Lavin et al., 2014a). As a result, these cells adapt to their environment, gaining tissue-specific phenotypes to maintain tissue homeostasis, metabolism, facilitate organogenesis, and during inflammation TRICs also display alternate activation states relative to their tissue-naïve counterparts (Ransohoff, 2016). Furthermore, TRICs can have different ontogenies and be sustained through slow proliferation of local progenitors instead of repopulation through circulating immune cells. TRICs are not only relevant in physiological conditions but play a key role in disease pathology, e.g. a reduction in the phagocytic capacity of alveolar macrophages has been observed in COPD patients, making them ineffective in clearing lung infections (Jubrail et al., 2017). Appreciation of such functionalities and their relevance to disease has pushed forth therapeutic strategies which try and regulate TRIC phenotypes, e.g. an increase in microglial activation is associated with AD, hence various therapeutic strategies have been proposed to inhibit microglial activation in patients (Biber et al., 2019).

The importance of TRICs in physiological and pathological conditions has led to an interest in understanding their biology. Researchers have examined various aspects of

these cell subtypes, including their ontogeny (Mass, 2018, Vasanthakumar et al., 2015, Mackay et al., 2013), renewal capacity (Hashimoto et al., 2013), homeostatic role in tissues (Nguyen et al., 2011, Sathi et al., 2017), how they differ from other immune cell subtypes (Lavin et al., 2014a, Sojka et al., 2014a) and their contribution to disease (Shaw et al., 2018, Bergsbaken and Bevan, 2015). Cross-tissue comparisons allow us to appreciate the degree of immune cell heterogeneity explained by tissues, i.e. tissue dependent and independent regulatory networks of the immune landscape. However, to date there are very few reports of cross-tissue comparisons of TRICs, and the majority of these reports are reviews (Fan and Rudensky, 2016, Weller and Spencer, 2017, Masopust and Soerens, 2019). Alternatively, studying TRIC biology in the context of a single tissue, cross-tissue comparisons would enable us to gauge the degree of immune cell heterogeneity explained by tissues i.e. tissue dependent and independent regulatory network of the immune landscape. In this chapter cross-tissue comparisons are conducted in order to describe immune cell type and TRIC diversity.

Previous studies have investigated immune cells using predefined sets of markers TRICs or using an unbiased approach such as scRNA-Seq. Subsequently, cells/samples have been profiled using omics technologies to understand the regulatory networks defining these cell types. Considering a few examples where researchers have adopted markers to isolate cell types; Lavin et al. described the regulatory landscape (enhancers and transcription factors) of isolated macrophages from different tissues (lung, brain, gut, spleen and liver) (Lavin et al., 2014a). This is further expanded upon by the ImmGen consortium which analysed 86 cell populations across lineages, tissues and maturation states (Yoshida et al., 2019). Although cell type markers are a great resource for cell identification, the field has begun to appreciate the heterogeneity of cells, and hence recent studies have adopted single cell omics to identify novel TRICs. These analyses investigate different immune cell types across tissues (three to five) including T cells (Szabo et al., 2019), Tregs (Miragaia et al., 2019), NK cells (Sagebiel et al., 2019) and macrophages (Mass et al., 2016). However, with an ever increasing number of scRNA-Seq datasets and atlases (Schaum et al., 2018) there is great scope of investigating TRICS across a larger set of tissues.

The primary aims of this chapter were to:

1) Explore how cell types are associated based on their transcriptomics profile, and how well does this correlate with our understanding of marker defined cell types. Thereby,

appreciating how one defines cell types and the extent to which they can be separated. For this, the similarity between marker-defined cell types from the ImmGen resources were studied based on their transcriptomic profile and gene modules uniquely associated with some but not all cell types were identified.

2) In order to conduct an unbiased investigation of TRICs Tabula Muris scRNA-Seq dataset was considered, a compendium of 100,605 cells from 20 mouse tissues across seven adult mice (3 months), sequenced using FACS- and microfluidic-based single cell technologies (Schaum et al., 2018). Using such large datasets would help identify and characterise the hierarchy of cell ontology, ranging from cell-lineages to -subtypes, including TRICs.

3) Lastly, to examine the commonality between the biased (ImmGen) and unbiased (Tabula Muris) approach to observe the advantages and disadvantages of each.

3.2 Material and methods

3.2.1 ImmGen RNA-Seq data pre-processing and analysis

In order to provide a set of reference gene expression profiles for a wide range of immune cell populations, two RNA-Seq datasets from the ImmGen resource were identified, including the “System-wide RNA-Seq profiles” (GSE109125) and “OpenSource Mononuclear Phagocytes Project” (GSE122108) datasets. Cell populations represented within these datasets were of lymphoid, myeloid and mesenchymal origin, isolated from a range of embryonic and adult mouse tissues. Both datasets were generated from pooled cells isolated by FACS, each cell population being defined by a set of markers with some samples derived from stimulated cells. The Mononuclear Phagocyte dataset consisted of 412 samples and the System-wide dataset of 157 samples. With the current study focusing on the biology of unstimulated adult cell populations, samples derived from cells that were treated/stimulated and those isolated from embryos or neonates were excluded from this analysis. This resulted in 384 samples in total being taken forward with 139 samples being from the System-wide dataset and 245 samples from the Mononuclear Phagocyte dataset. Represented in this combined dataset were 128 combinations of cell-markers used for FACS isolation of cells from 25 tissues, each assumed to describe a distinct cell type. The dataset comprised of various cell types with one or more subsets (number shown in parenthesis) dependent on the localisation and maturation state, these included B cells (18), DCs (20), endothelial cells (2), epithelial

cells (1), fibroblasts (1), neutrophils (2), ILCs (5), macrophages (27), mast cells (1), monocytes (5), NK cells (6), NK T cells (1), pericytes (1), stem cells (6) and T cell (22) (**Table S3.1**). There were 1 to 23 replicates for each cell type with majority having 2 replicates. Datasets were individually normalized for TPM-like values using a scaling factor of 10^4. Datasets were merged based on 12,144 genes having an expression value greater than 0.5 TPM in at least one sample.

To analyse the similarity between samples, a sample-to-sample correlation network was generated from the combined dataset using the open-source network analysis software, Graphia (https://graphia.app/). A pairwise Pearson correlation matrix was calculated between samples based on all the 12,144 genes within the dataset. A graph was then constructed using a Pearson similarity coefficient threshold of $r \geq 0.87$, generating a network of 379 samples/nodes and 3,739 edges. At this threshold, similar groups of cells formed cliques characterised by dense interconnectivity between similar samples. The graph was clustered using Markov cluster algorithm (MCL) with an inflation of 2.0, resulting in 30 cell clusters. To analyse genes underpinning the cellular diversity a GCN was generated using a Pearson threshold of $r \geq 0.8$. The resultant graph comprising of 6,307 genes and 193k edges was clustered using MCL with an inflation of 2.0 (**Table S3.2**). Of the 473 gene clusters, the 175 possessing five or more genes were annotated based on known cell markers and manual inspection of the cell types displaying the highest average expression of clustered genes relative to other cell types.

3.2.2 Tabula Muris single cell RNA-Seq data pre-processing and quality control

To investigate TRIC types in mouse, scRNA-Seq data from the Tabula Muris dataset (Quake et al., 2017) was selected for analysis. This dataset comprised of cells dissociated from 20 individual mouse tissues prepared through the 10X Chromium and Smartseq2 platforms. For this study, only immune cells sequenced using the SmartSeq2 platform were selected as a wider range of tissues were analysed using this platform and the approach also provided higher quality data in terms of the number of genes identified per cell (Wang et al., 2019b). In addition, cells from the bone marrow were not considered for this study, as the focus was on mature TRIC types, not their differentiating states from progenitors. In total 10,860 cells were selected, derived from 12 tissues of both male and female mice (n = 7). This data was then pre-processed using the Seurat pipeline (Satija et al., 2015) as described in the Tabula Muris study. This included filtering: 1) any gene expressed in less than five cells or having fewer than five reads in total; 2) any cell that

did not have 500 to 25k detectable genes and 50k to 2M reads. The quality-controlled data, which comprised of 10,792 cells and 16,091 genes was log-normalized and scaled to regress out the effects of library size, ribosomal content and Rn45s. Based on the most variable genes (standard deviation ≥ 0.5), principal components were calculated and tested for using the Jackstraw approach (Chung and Storey, 2015). The first 42 significant principal components were selected (empirical *P* value < 0.05). These were then used for visualisation and subsequent quality control and clustering using Graphia.

Pairwise correlations were calculated between each cell based on the 42 significant principal components to generate a cell-to-cell network. To reveal distinct cell populations, nodes and edges were filtered based on a correlation threshold of $r \geq 0.7$, a minimum node degree of 30 and the k-NN algorithm was applied with k = 25. The resultant network consisted of 6,536 cells and was clustered using the MCL algorithm at an inflation of 1.7, giving a total of 53 clusters. For downstream statistical analyses, a sufficient number of cells were ensured for each grouping. 1) At the cluster level, small clusters having less than 10 cells were removed (cluster 52 & 53). 2) Cells belonging to the same cluster and tissue, having less than 5 cells were removed. 3) The nine clusters identified as microglial (based on the expression of *Cx3cr1*, *P2ry12* and *Tmem119*) were collapsed into a single cluster as they captured donor differences rather than differences due to brain regions. This filtering left a total of 43 clusters comprising a total of 6,355 cells.

3.2.3 Cluster based annotation of single cells

Single cells from the Tabula Muris dataset were annotated by comparing their expression profile with FACS isolated cell types from the ImmGen data. To determine the similarity between cells from the Tabula Muris dataset with those of the ImmGen dataset, only the most variable genes (standard deviation > 0.5) identified for each dataset were considered amounting to 6,399 genes. A Pearson correlation was then calculated for each cell of the Tabula Muris dataset with samples of the ImmGen dataset. Considering similarity coefficients above the threshold of r ≥ 0.3, cells were provisionally annotated based on the most similar cell type they associated with from the ImmGen. For clusters where more than 50% of cells were of the same cell type annotation and found enriched using the Fisher exact test, all cells within that cluster were annotated as that cell type.

3.2.4 Approaches to construct gene coexpression networks from single cell RNA-Seq data and their comparison

GCNs were constructed from the scRNA-Seq data of the Tabula Muris. Three methods were used to generate GCNs, these included the “Average”, “Subcell”, and “Filter-average” method. The first approach averaged gene expression across cells of each cluster resulting in a cluster vs. gene expression matrix. The second approach further subdivided clusters into 5 subclusters or Subcells, each having a minimum of 5 cells. The subclustering was done by examining cells from a given cluster and iteratively clustering them using the Louvain clustering algorithm with each iteration clustering the graph with a lower k (starting from k = number of cells -1) using a kNN (lower values of k resulted in more clusters) until the following thresholds were crossed, that is 1) maximum of five subclusters to a cluster and 2) each subcluster having a minimum of five cells. Through each iteration the largest subcluster was subdivided first and if not a lower value of k was chosen for clustering. After the Subcells were constructed a similar averaging of gene expression was done across cells of the subcluster which gave a Subcell vs. gene expression matrix. The third method constructed a GCN in a similar way to the first, however before averaging, lower expression values were filtered. Here, each cell-cluster was examined one at a time and those genes were considered which showed 1) expression in more than 5% of cells, 2) had maximum expression beyond 0.5 log TPM and 3) were expressed in a minimum of 3 cells. Genes which did not meet these criteria were given a zero-expression value for the cluster being filtered. Subsequently, gene expression was averaged for cells across each subcluster or Subcell.

The three methods used to construct GCNs were compared using two approaches. First, the top 4,000 correlated genes from each method were examined for their correlation values across the three methods and if they changed from one method to the other. For this a single representative metric was required to estimate how correlated a gene was within a dataset generated from a method. For this, the kNNcorrelation of a gene was considered, which was the average correlation the gene had with its ten nearest neighbours. The metric was used to examine the distribution of correlations for the top 4000 correlated genes in one method and how this distribution changed in other methods i.e. if highly correlated genes in one method were similarly correlated in the others. This approach was taken further by ranking the genes based on their kNNcorrelation and comparing them across datasets. Comparing rankings would account for the differences in the ranges of the Pearson correlations between methods. Second, the top 50 (by size,

number of genes > 10) gene clusters unique to each approach were investigated and examined if they were a result of technical artefacts. For this, gene clusters unique to a certain method were identified and studied. To estimate the uniqueness of a gene cluster for each method, each cluster was compared to all the top 4000 correlated genes from the other two methods based on their gene overlap using the Jaccard index (from 0 to 1, with fully overlapping gene lists having a value of 1) and the two values were averaged to estimate the final gene cluster uniqueness. Hence, gene clusters unique to a method had the least overlap i.e. low average Jaccard index with the top 4,000 correlated genes from the other two methods. Of the top 50 clusters from each method the two most unique clusters (lowest Jaccard index) from a method were examined for their GO enrichment, estimated using clusterProfiler (Yu et al., 2012).

3.3 Results

3.3.1 Mouse immune cell gene signatures from RNA-Seq analysis of FACS cell populations

The ImmGen project provides an extensive resource of transcriptomics and epigenetic data derived from pooled FACS cells spanning a variety of immune cell types from different tissues of the mouse. For this study, two RNA-Seq datasets were examined the “System-wide” and “Mononuclear phagocytes” collection. These datasets comprised of cells isolated by FACS from different mouse tissues using various markers. As the focus of the work was on homeostatic TRICs from adult mice, samples derived from embryonic stages and stimulated conditions were removed. However, a small number of non-immune cells, endothelial cells and fibroblasts were left in to provide contrast. The datasets were merged totalling to 384 samples derived from FACS cells based on 128 combinations of markers from 25 different tissues which were analysed together. To investigate cell groupings, a sample-to-sample network (**Figure 3.1**) was generated based on the Pearson similarity coefficient. The *r* threshold was set to ≥ 0.87, the highest value at which majority of samples were connected within the graph. The resultant network comprised of 379 cell samples which were connected by 3,739 edges. Upon clustering the graph with MCL (inflation = 2), samples grouped into 30 ImmGen cell cluster (ImmCC), with five singletons, i.e. unconnected cell samples which were not similar to any other sample at this threshold (**Table S3.1**).

All replicate samples were connected, and similar cell types tended to co-cluster or reside in the same vicinity of the graph, e.g. several myeloid cells (macrophages and DCs) were part of the same component, while cells of the lymphoid lineage formed a separate component. In certain cases, however, cells formed separate components, suggesting they had a distinct expression profile. These included, fibroblasts (ImmCC 25), endothelial cells (ImmCC 24), certain types of macrophages (alveolar, ImmCC 13 & 22; microglia, ImmCC 4; peritoneal, ImmCC 2 & 17; and splenic, ImmCC 30 & 29) and plasma cells (ImmCC 20). Certain macrophages and DC types were part of the same component but clustered independently into tissue-resident cell types, such as Kupffer cells (ImmCC 14) and aortic macrophages (ImmCC 12), as well as immature cell types like pre-DCs (ImmCC 10). The analysis confirmed the quality of the data, as in the majority of cases, graph visualisation and clustering reaffirmed the known similarities

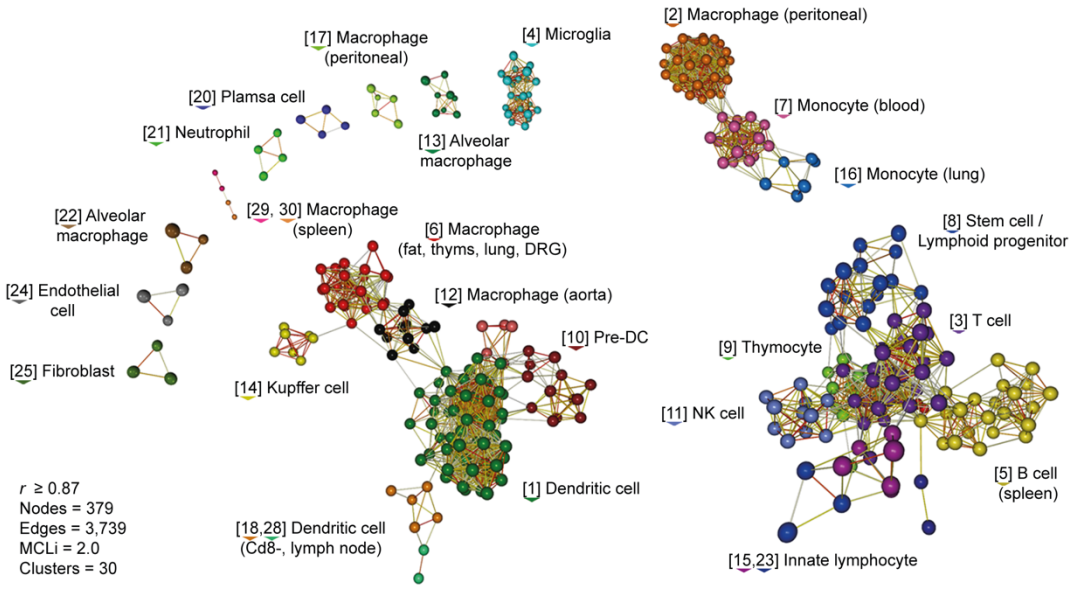


Figure 3.1. Sample-to-sample correlation network of ImmGen data

A sample-to-sample correlation network was constructed from the 384 ImmGen samples selected for this study based on the similarity between their gene expression profile. Samples correlated at $r \geq 0.87$ are connected by a weighted non-directional edge. The resulting network was clustered into 30 groups using MCL (inflation = 2). Cell samples clustered together were given the same colour and labelled with the cluster number (in parenthesis) and cell types within it (as described by the authors). Distinct components can be observed for lymphoid cells (bottom right), certain macrophages and DCs (bottom centre), while several macrophage and other immune subtypes exist as separate graph components.

between immune cell types. To study subtle associations/differences between cell types, the GCN was studied.

With few biological replicates included in the ImmGen data it is difficult to conduct statistical analyses across cell types as is common practice in differential expression analyses. Therefore, GCN analysis was adopted to reveal genes that shared a common expression profile across cell populations, a methodology known to capture coexpressing genes associated with a given biology (Xue et al., 2014, Nirmal et al., 2018, Patir et al., 2019, Patir et al., 2020). Similar to the network generated for samples, a gene-to-gene network was constructed based on the Pearson correlation coefficient, using a threshold of $r \geq 0.8$ and subsequently clustered using MCL (inflation = 2). The resultant graph comprised of more than 193k edges that connected 6,307 genes (nodes), which clustered into 608 gene clusters (**Figure 3.2**). The 175 largest clusters (n $\geq$ 5 genes) were manually annotated based on the average expression profile of genes within the cluster. ImmGen gene clusters (ImmGC) were annotated for a certain cell type if the average expression of its genes were highly expressed in that cell population (**Table S3.2**). Additionally, gene clusters were annotated based on the cell type(s) they were associated with and to what degree/specificity. Gene clusters were annotated as "General" towards a certain cell type if they showed a high average expression of clustered genes relative to other samples and "Specific" if other cells showed negligible expression these genes. Where multiple cell types showed a comparable average expression, the gene cluster was annotated for cell types from the highest to lowest average expression. For example, cluster 1 was annotated as "General [Lymphoid progenitors = Stem cells], because the clustered genes were expressed in several cell populations, with the highest expression in lymphoid progenitors and stem cells which had similar average expression levels. In this case, the cluster was associated with the cell cycle, as its genes had a high expression in proliferative cells and included many well-known cell cycle gene families such as cyclins (*Ccnf, h* & *e1*), Cdk (*Cdk*1, 2, 4 & 7) and Cdc (*Cdc6*, 7 & 45). Annotating the gene clusters revealed various aspects of TRIC biology including cell lineages (myeloid - ImmGC 52 & lymphoid ImmGC 110), cell types (T cell - ImmGC 24, Treg ImmGC 205, NK cells – ImmGC 14 & ILC – ImmGC 60), developmental stages (preDC ImmGC 25/169, lymphoid progenitors ImmGC 9/84/92 & mature lymphocytes ImmGC 42), and TRIC types. Gene clusters representative of cell types included known transcriptional regulators, e.g. *Eomes* for NK cells (Zhang et al., 2018b) and *Rora* for ILCs (Walker et al., 2019). Of the clusters representing TRIC

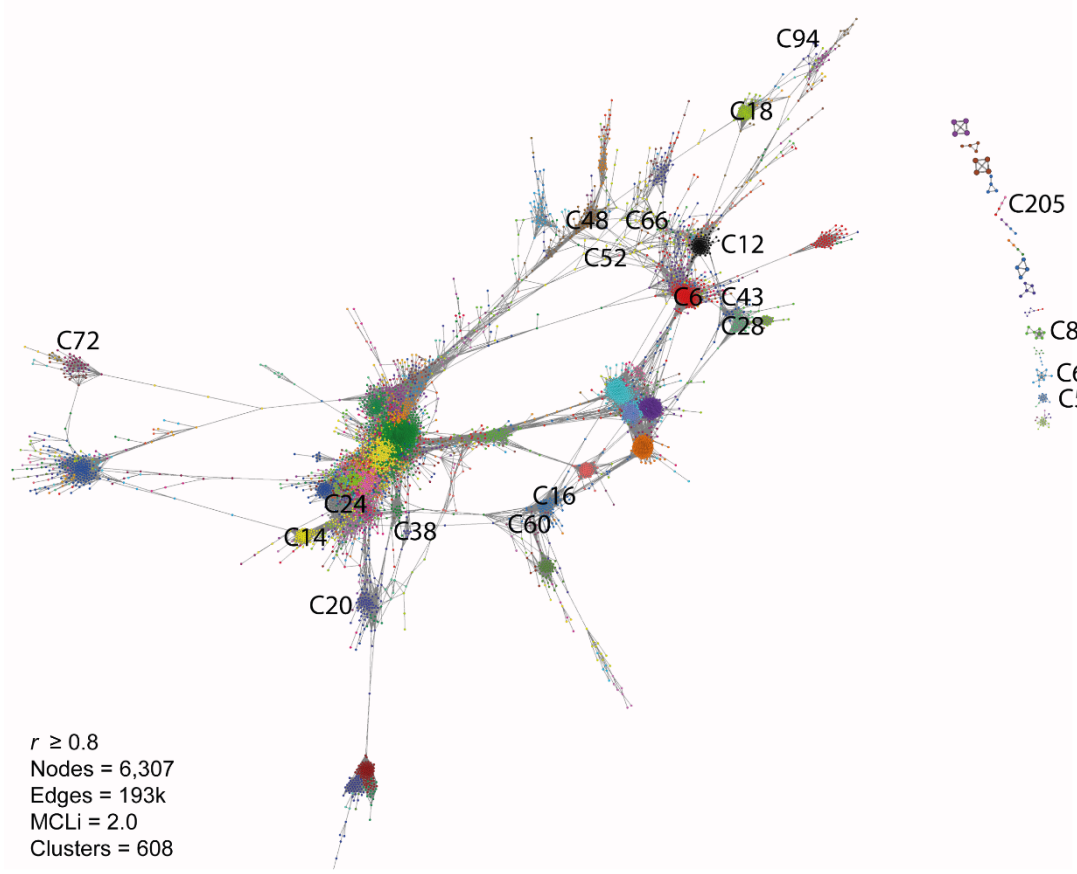


Figure 3.2. ImmGen gene correlation network.

GCN constructed based on the similarity of expression profiles across samples from the ImmGen data ($r \geq 0.8$). The GCN was clustered into 608 clusters using MCL. The graph is broadly segmented into myeloid-specific clusters (top right) and lymphoid-specific clusters (bottom left), with the biology of other cell types and subtypes being captured by gene clustering.

signatures, most numerous were those for macrophages which contained well established transcriptional regulators, e.g. *Gata6*, *Pparg*, *Sall1* and *Spic* for peritoneal, alveolar, microglial and splenic macrophages, respectively (Mass, 2018, T'Jonck et al., 2018, Davies et al., 2013). Other TRIC associated genes included *Mybl1* a marker for germinal centre B cells (Ding et al., 2015) and *Tnfrsf9,* expressed in non-lymphoid tissue Tregs (Miragaia et al., 2019). Interestingly, *Ctla4* which was present in the gene cluster associated with colonic Tregs has been shown to increase Treg cell numbers in the colon and not in other tissues during Foxp3 induction (Barnes et al., 2013). In addition to the supplementary tables for gene coexpression clusters the cell annotations from the top

50 gene clusters (based on size) have been summarised, which is an indicator of cellular heterogeneity across the different cell types and as expected dominant in macrophages (**Table 3.1**) Overall, the GCN analysis revealed known TRIC associated genes, as well as others which are potentially related.

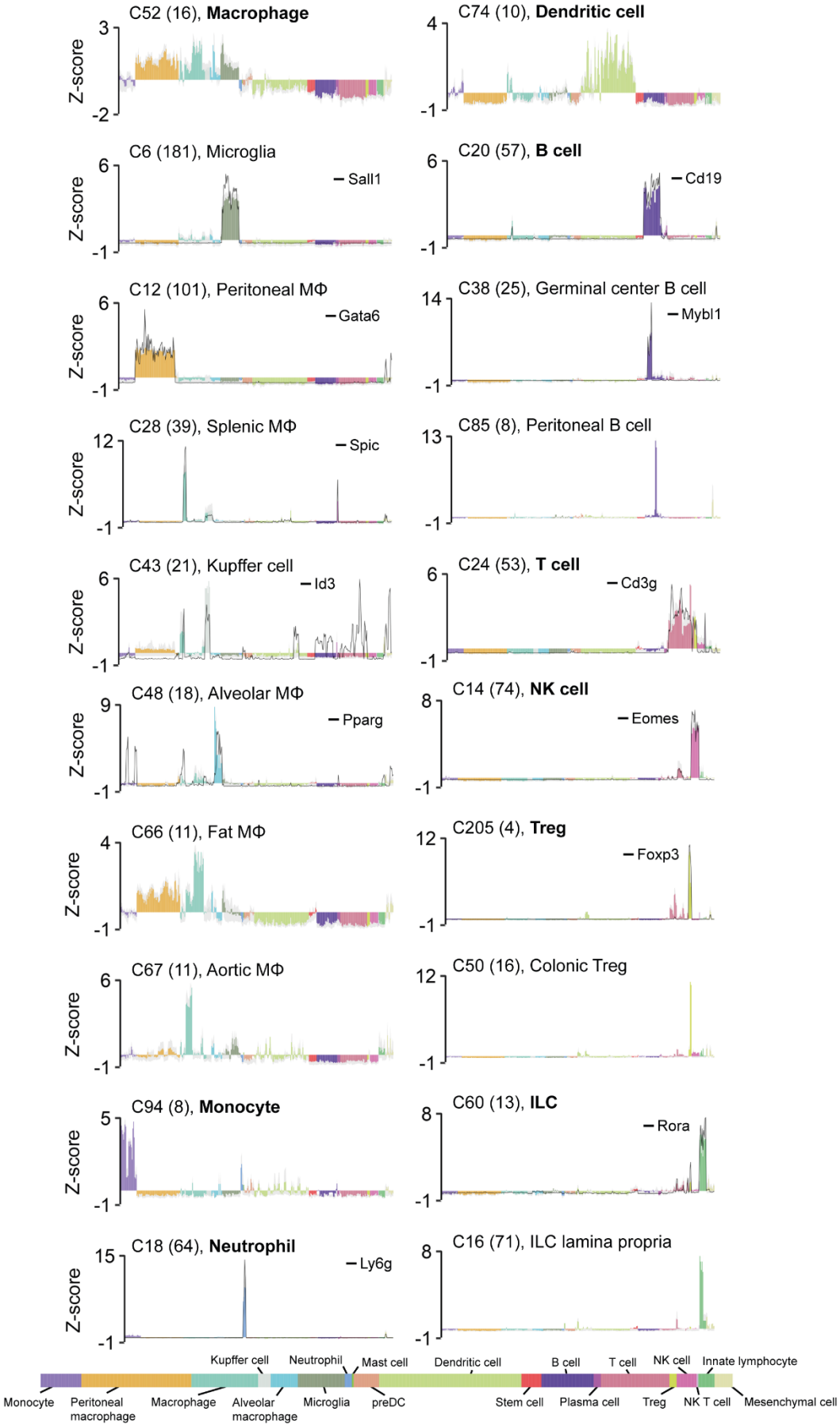

C52 (16), Macrophage
C74 (10), Dendritic cell
C6 (181), Microglia
Sall1
C20 (57), B cell
Cd19
C12 (101), Peritoneal MΦ
Gata6
C38 (25), Germinal center B cell
Mybl1
C28 (39), Splenic MΦ
Spic
C85 (8), Peritoneal B cell
C43 (21), Kupffer cell
Id3
C24 (53), T cell
Cd3g
C48 (18), Alveolar MΦ
Pparg
C14 (74), NK cell
Eomes
C66 (11), Fat MΦ
C205 (4), Treg
Foxp3
C67 (11), Aortic MΦ
C50 (16), Colonic Treg
C94 (8), Monocyte
C60 (13), ILC
Rora
C18 (64), Neutrophil
Ly6g
C16 (71), ILC lamina propria
Z-score
Kupffer cell
Neutrophil
Mast cell
Dendritic cell
B cell
T cell
NK cell
Innate lymphocyte
Monocyte
Peritoneal macrophage
Macrophage
Alveolar macrophage
Microglia
preDC
Stem cell
Plasma cell
Treg
NK T cell
Mesenchymal cell

Figure 3.3 The average expression profile of gene coexpression clusters identified in ImmGen.

Each panel within the layout represents the average expression profile (z-score) of a gene cluster identified from the GCN shown in **Figure 3.2**. The panels are titled with the cluster number, number of genes in the cluster (shown in parenthesis) and the cell type(s) they are highly expressed in. The x axis constitutes of samples which are arranged and coloured based on the bottom legend. The y axis shows the average expression profile (z-score) of all the genes within that cluster. In addition, for certain cell type-associated gene clusters the expression profile (z-score) of marker genes is also shown (black line). The panels are layered out such that the gene cluster associated with a cell type (in bold) is displayed first, after which related TRIC associated gene clusters are shown.

Table 3.1 Summary of immune cell type associated coexpression clusters (top 50)

Lineage	Cell ontology	Cell ontology hierarchy	Cell subtype	Tissue	Gene Cluster
Myeloid	Myeloid	Lineage			Cluster 32
	Myeloid (macrophage and DCs)	Cell type	multiple		Cluster 36
	Macrophage	TRIC		Brain	Cluster 6
				Peritoneal	Cluster 12
				Lung	Cluster 23
				Spleen	Cluster 28
				multiple	Cluster 34
				Lung	Cluster 37
				Brain	Cluster 39
				Liver	Cluster 43
				Liver	Cluster 44
				Spleen	Cluster 45
				Lung	Cluster 48
	Dendritic cell	Cell type	multiple		Cluster 35
		Cell subtype	multiple		Cluster 8
		Cell subtype and TRIC	multiple		Cluster 25
	Neutrophil	Cell type			Cluster 18
	Mast cell	Cell type			Cluster 19
Lymphoid	Lymphoid (excluding B cells)	Cell type	multiple		Cluster 30
					Cluster 24
	B cell	Cell type			Cluster 20
			Plasmablast		Cluster 27

		TRIC		Germinal centre Spleen	Cluster 38 Cluster 49
	Plasma cell	Cell type			Cluster 10
	Thymocyte	Cell subtype	Double +ve Double +ve		Cluster 22 Cluster 46
	T cell	Cell type			Cluster 42
		Cell subtype	Naïve		Cluster 31
	Treg	TRIC		Colon	Cluster 50
	NK cell	Cell type			Cluster 14
	ILC	Cell subtype	ILC3		Cluster 16
			ILC2		Cluster 26

3.3.2 Clustering and classification of single cells

In order to identify TRICs unbiased by predefined cell type, scRNA-Seq from the Tabula Muris dataset (Quake et al., 2017) was examined. The dataset is derived from cells isolated from 20 different mouse tissues. To investigate the immune component, immune cells (as annotated by the authors) spanning 12 tissues were selected. Furthermore, given the focus on the biology of TRICs, not haematopoiesis, all cells from the bone marrow were excluded. The selected cells were pre-processed using the Tabula Muris pipeline and quality controlled for through a graph-based approach which identified outliers based on their lack of similarity to neighbouring cells (**Table S3.3**). To calculate the similarity between cells, the most variable genes across these cells were identified and through PCA the 42 most significant (*P value* < 0.05) PCs were calculated for each cell. Subsequently, a Pearson correlation matrix was calculated to define the similarity between cells based on their PCA profile. The matrix was used to construct a sample-to-sample graph at a threshold of $r \geq 0.7$, applying the kNN algorithm with the k = 25. Within this network, outliers were poorly connected to neighbouring cells and those connected to less than 30 cells were excluded from the network. The remaining graph of 6,536 cells was then clustered using MCL (inflation = 1.7). Although microglial cells formed five clusters, this division represented donor variation rather than regional variation. Hence, the microglial clusters were aggregated into a single cluster. For reliable statistical analysis, small groups based on clustering and tissue of origin were filtered for those having less than five cells. The remaining cells totalled to 6,355 cells spanning 12 tissues (**Table S3.4, Figure 3.4**) and formed 43 Tabula Muris cell clusters (TMCC) (**Figure 3.5**).

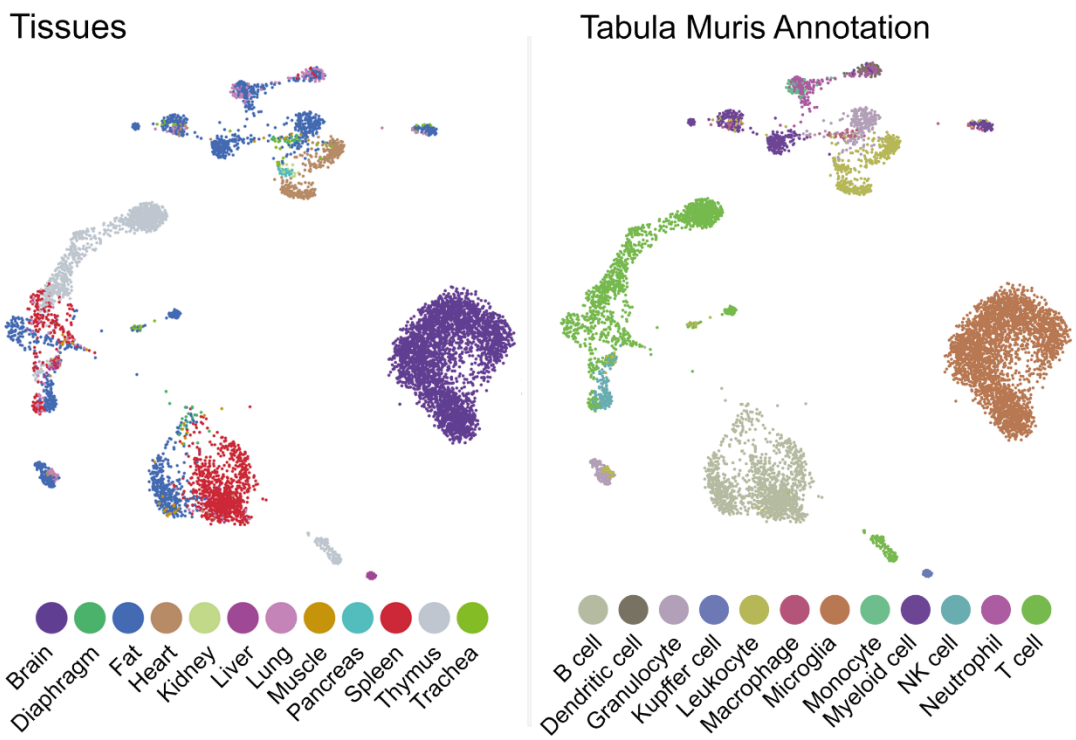


Figure 3.4 Tissue of origin and original annotation of Tabula Muris immune cells.

Visualisation of the final quality controlled 6,355 immune cells from Tabula Muris, where similar cells are proximal to one another. This cell layout is constructed only for visualisation using a kNN graph with k = 20, followed by the Fruchterman-Reingold force directed algorithm to layout the network in two dimensions (edges not displayed). In the left panel cells are coloured according to the 12 tissues from which they were isolated. The right panel displays the 12 original cell annotations defined by the authors. Broadly, the larger cell populations which segregate out include the different myeloid populations (top component), B cells (bottom, centre), T cells (left) and microglia (right). The shortcomings of the cell annotation can be observed in the interspersed myeloid cell types, especially the monocyte and neutrophil populations (top).

On inspecting the cell annotaions provided by the authors there were two main issues (for downstream analysis in this chapter), 1) many cells were broadly classified into leukocytes and myeloid cells; 2) myeloid cell types like monocytes and neutrophils were difficult to differentiate as they did not form separate clusters and instead aggregated together. To improve upon the cell annotation, cells were reannotated using the ImmGen resource as a reference, since the resource includes transcriptomics data of a wide variety of predefined immune cell types. Each cell from the processed Tabula Muris data

was compared with samples from the ImmGen dataset using the most variable genes across each dataset. Cells were temporarily annotated for the cell types they were most similar to from the ImmGen. Where this similarity was poor $r < 0.3$ cells were annotated as “low confidence”. Finally, TMCCs enriched in cells of a certain cell type’s annotation were annotated as such. This resulted in TMCCs being grouped into 20 broad cell types (**Table S3.4**, **Figure 3.5**). Relative to the original annotation there was a greater granularity and distinction of myeloid and lymphoid populations. In addition to the correct cell type annotation certain TMCCs were correctly annotated for their tissue of origin. For example, TMCC 24 and 26 comprised of cells isolated from the heart and were annotated as aortic macrophage from the ImmGen data. Such cases also included splenic B cells, microglia and Kupffer cells, hence, providing evidence that the approach worked well.

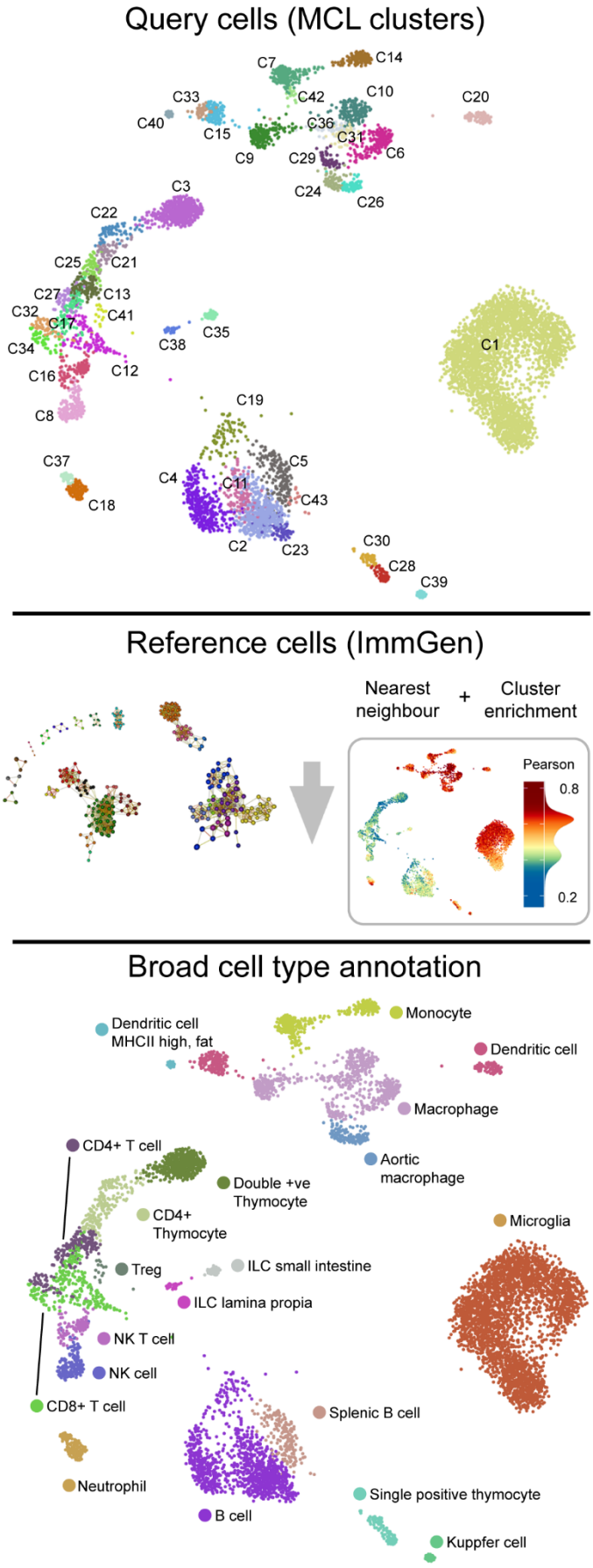

Query cells (MCL clusters)
C7
C14
C33
C42
C10
C20
C40
C15
C36
C31
C9
C29
C6
C24
C26
C3
C22
C25
C21
C27
C13
C32
C17
C41
C34
C38
C35
C1
C16
C12
C8
C19
C37
C5
C4
C11
C43
C18
C2
C23
C30
C28
C39
Reference cells (ImmGen)
Nearest neighbour
+
Cluster enrichment
Pearson
0.8
0.2
Broad cell type annotation
Dendritic cell MHCII high, fat
Monocyte
Dendritic cell
Macrophage
Aortic macrophage
CD4+ T cell
Double +ve Thymocyte
CD4+ Thymocyte
Microglia
Treg
ILC small intestine
ILC lamina propia
NK T cell
NK cell
CD8+ T cell
Splenic B cell
Neutrophil
Single positive thymocyte
B cell
Kuppfer cell

Figure 3.5 ImmGen-based annotation of Tabula Muris scRNA-Seq data.

The figure shows the different steps involved in the broad annotation of the Tabula Muris immune cell populations based on the known cell classifications from ImmGen. This was done in three steps: 1) Tabula Muris cells were clustered into 43 groups (top panel); 2) Each cell was then annotated for the most similar cell type from ImmGen, based on their transcriptomic profile. This is shown in the middle panel where cells from the Tabula Muris are coloured based on their Pearson correlation coefficient, with the most similar sample from ImmGen; 3) Lastly, cell clusters enriched in cells of a given annotation were then annotated as such resulting in the final cluster level annotation (bottom panel).

3.3.3 Constructing gene coexpression networks from scRNA-Seq data

After broadly classifying each cell from the Tabula Muris immune subset, genes that gene associated with them were identified. Due to the strong influence of technical and biological factors associated with scRNA-Seq, including dropout effects and the apparent transcriptional heterogeneity within cell types, it is a challenge to capture the complexity of gene expression patterns which underlie the biology within this data. Although differential expression analysis helps identify genes highly expressed in a cell type, how specific the genes are, and the genes shared across different cell types may go unseen. To overcome this, I looked at aggregating reads from cells within a cluster. This would reduce the variation within the data attributed by dropouts or variation within cell types, and would focus on the variation between clusters, ideally representing the differences between cell types or cell states, e.g. activation or differentiation. The approach would help improve modelling gene expression patterns like those captured by gene coexpression analysis. To test this theory, three approaches were explored (**Figure 3.6**). For the first approach (Average method), the log normalized counts for each gene across cells within a cluster were averaged, thereby generating a cluster vs. gene expression matrix. In the second method (Subcell method) each cluster was divided into ‘subclusters’ or ‘Subcells’ using the Louvain clustering algorithm (Blondel et al., 2008) and in a similar way gene expression values were averaged across cells of the subclusters thus generating a Subcell vs. gene expression matrix. The approach was motivated by the need for replicates for each cluster. For the third method (Filter-average

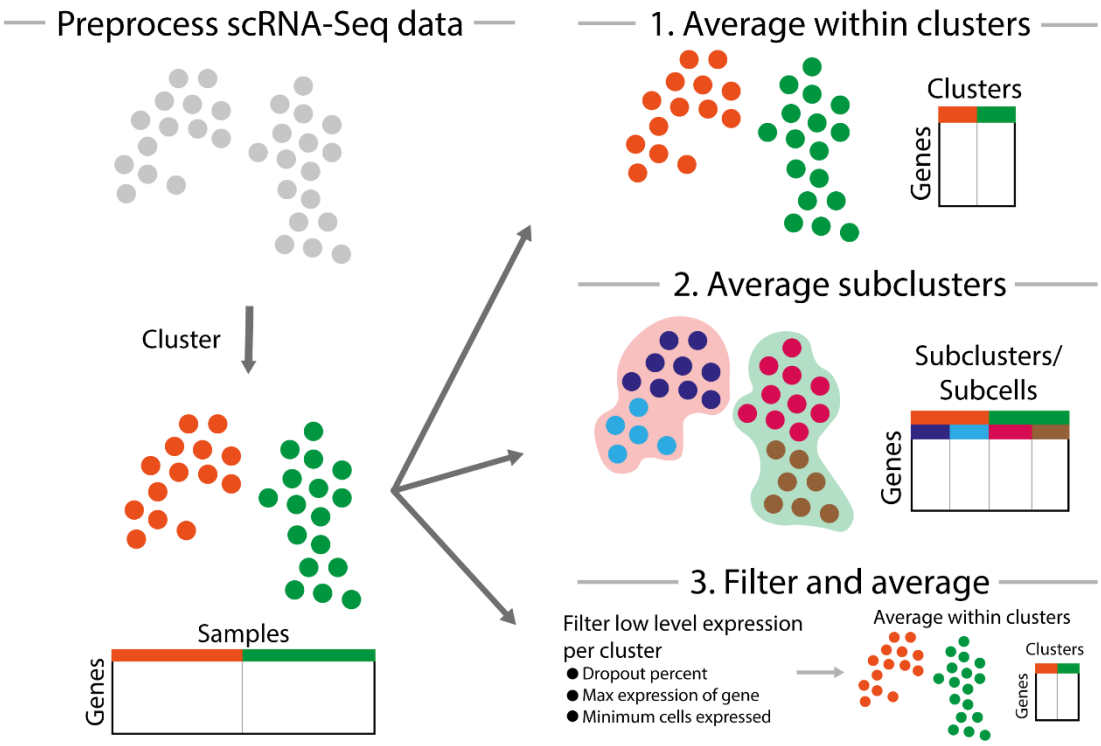


Figure 3.6 Depiction of the three approaches used for sample reduction of scRNA-Seq data

The three approaches used to stabilise signals within scRNA-Seq data to construct a GCN. All approaches require the pre-processed expression matrix with samples clustered. The approaches include 1) averaging normalised reads of each gene across samples of a given cluster. 2) Subclustering the data by further clustering samples from each cluster and averaging within these subclusters. 3) Applying different low-end expression filters for each cluster and subsequently averaging gene expression across each cluster.

method), it was hypothesized that some of the averaged signals from the above approaches might be unduly influenced by noise or spikes in expression contributed by a few cells. These signals would thereby serve to falsely represent the pattern of coexpression of the cluster. Hence, similar to the first approach, the third method used the average signal of genes within clusters; however, before averaging each gene from the original cell expression matrix, lowly expressed genes with spikes in expression were replaced with zero expression values. The thresholds were applied on clusters individually for each gene. These thresholds included: 1) genes expressed a minimum of 5% of cells within the cluster; 2) at least one cell with an expression above 0.5 log TPM and; 3) a minimum of 3 cells expressing a gene. Additionally, to exclude spikes observed in RNA-Seq data, the expression values of cells within a cluster were capped

at 95% confidence interval for each gene. The approach would potentially highlight strong biological signals while removing low level signals, the majority of which are likely due to technical factors. Using the expression matrices generated from each method, GCNs were constructed at different Pearson correlation thresholds optimised to include ~4,000 genes within each network (**Table S3.5**). First, the overlap between these top 4,000 correlated genes from each network was compared (**Figure 3.7A & B**). The overlap in gene content between networks was 2,307 genes (37%), with approximately a third of the total number of genes (found in each network) unique to each method. Next, I wanted to know if highly correlated genes (the top ~4,000 correlated genes) from one dataset were differently correlated in another dataset, e.g. from one method certain genes could be highly correlated to one another, however on using a different method they may be poorly correlated. This comparison does not comment on the performance of methods but reveals how similar methods are. For this comparison, a metric to measure whether a gene was highly/poorly correlated within a dataset was required. This was measured for individual genes from each dataset by taking the average correlation of the gene with its ten nearest neighbours, referred to as "kNNcorrelation". Resultantly, each gene had a single kNNcorrelation value, a measure of its correlation within the dataset. Looking at the top correlated genes from the Average method (~4,000 genes from the Average method network), the kNNcorrelation of genes showed a similar distribution to the Filter-average method, while the Subcell method had a long tail. This indicated that highly correlated genes in method 1 and 3 were poorly correlated in the Subcell method, likely due to the replicates (Subcells) for each cell cluster unique to the approach. As a secondary measure, genes were ranked within a dataset based on the kNNcorrelation, such that highly correlated genes had higher rank values and the least correlated gene had a rank of 1 (**Figure 3.7C**). On examining the ranks of the top 4,000 genes from each method, the Average and Subcell method showed similar distributions. This suggested that both methods identified similar genes to be highly correlated. Strikingly, the distribution of Filter-average gene ranks differed from both the Average and Subcell method, where top genes from the Filter-average approach had lower ranks in the Average and Subcell method, signified by their tailed distribution. The inverse was also observed as highly correlated genes in the other two methods presented lower ranks in the Filter-average approach. Overall the analysis revealed the similarity and differences between the three approaches, in that: 1) genes from the Subcell method had lower correlation values overall likely due to the number of replicates (Subcells)

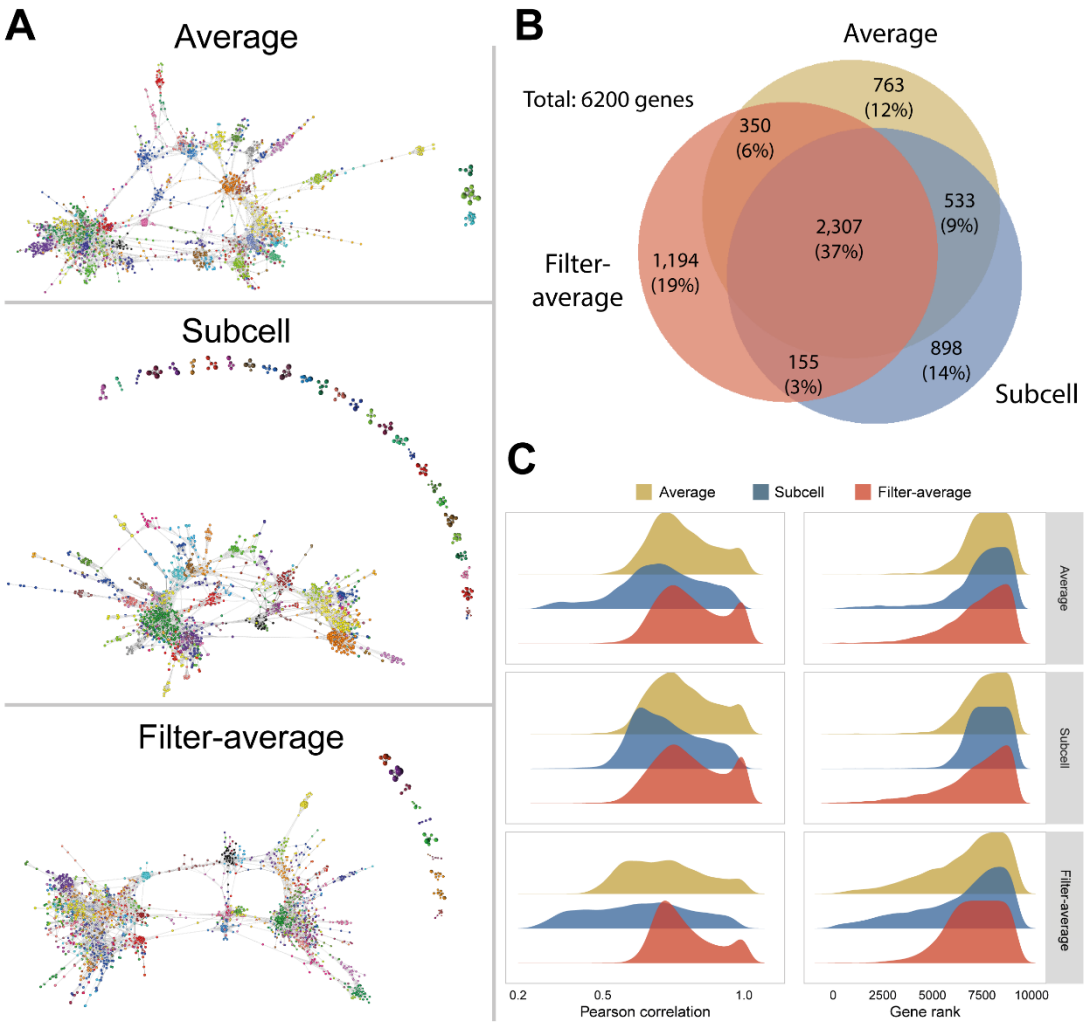


Figure 3.7. Comparison of gene coexpression networks from signal stabilising approaches.

A) Comparison of gene coexpression networks generated from three approaches used to transform scRNA-Seq data prior to GCN construction. Each network comprises of close to 4,000 genes connected by the highest correlations between them. **B**) Overlap between the network genes produced by the three methods. **C**) Metrics to compare the distribution of gene-to-gene correlations based on the different gene summary algorithms. Each plot represents a comparison of the gene networks produced from each method, e.g. the top left panel shows the proportion of network genes from the "Average" method based on their mean correlation with their ten closest members calculated for each the three datasets (coloured). The right panel ranks genes based on the above metric and makes a similar comparison for the distribution of genes from a given gene set across the three methods.

representing each cluster in contrast to the other methods which had a single sample representing a cell cluster; 2) the Filter-average method identified certain genes as highly correlated which were poorly correlated genes in the other two methods and vice versa.

As one of the most pronounced differences between methods were the genes unique to each, they examined for their association with any known biology rather than technical artefacts. By analysing gene clusters unique to a given approach one could determine which methods were influenced by technical artefacts as they would not be enriched in any known biology. First, the top 50 (based on size) clusters were considered from each method as these represented large gene clusters which were likely to be enriched for a known biological function (GO terms). Next, those clusters were examined which were conserved across methods and which of those were unique. To quantify the uniqueness of gene clusters to a given method its overlap as measured by the Jaccard index was calculated for each cluster with the top 4,000 correlated gene from the other two methods (**Table S3.6**). The Jaccard index is calculated by taking the number of genes in the cluster that overlap with the top 4,000 network genes generated from the other method, divided by the total number of genes in that cluster. I measured the combined overlap for each cluster with other methods by averaging the Jaccard indices. This measure was then used to rank individual clusters thereby highlighting poorly overlapping clusters. Clusters having a lower value were more unique to the method, and those whose genes are highly conserved had higher values. For comparison, two of the most unique clusters (lowest average Jaccard Index) from each method were analysed. From the Average approach, clusters 31 and 40 (**Figure 3.8**) were underrepresented in the other methods. It was assumed that the unique signals captured by the “better” method would represent a biological signal rather than a technical artefact, such as a spike in expression. On analysing the two clusters from the average approach GO enrichment analysis showed no significant enrichment for GO terms (**Table 3.2**). This was also observed for cluster 48 and 49 from the Subcell method (**Figure 3.9**). On investigating the expression pattern of genes from these gene clusters, for the Average method, they displayed high expression in only one of the samples (representing a cell cluster) and similarly a high expression in a single Subcell from the Subcell approach. Hence, these genes coexpressed due to the relatively high expression in a single Subcell. It was argued that if this signal described real biology it would be highly expressed by the majority of cells representing that cell cluster (in the Average approach) or Subcell. In both cases, it was found that only a few cells (mostly one cell) showed a high expression in the cell grouping

(cell cluster or Subcell), and therefore this signal likely represented a random spike in expression due to technical factors. In contrast, the Filter-average method gene cluster 15 and 44 (**Figure 3.10**) were enriched in processes associated with RNA processing including “Mitochondrial translation” (*adj. P* value 9.7×10^{-4}) and “rRNA processing” (*adj. P* value 4.6×10^{-4}). Consistent with GO terms for gene-clusters examined in the Filter-average method, the clusters were proximal to the cell cycle gene-cluster 1. Unlike the other methods, genes from these clusters were expressed in several samples (cell-clusters). These results indicated that the Average and Subcell method were prone to capturing false signals introduced by technical noise, while these signals are removed in the Filter-average approach.

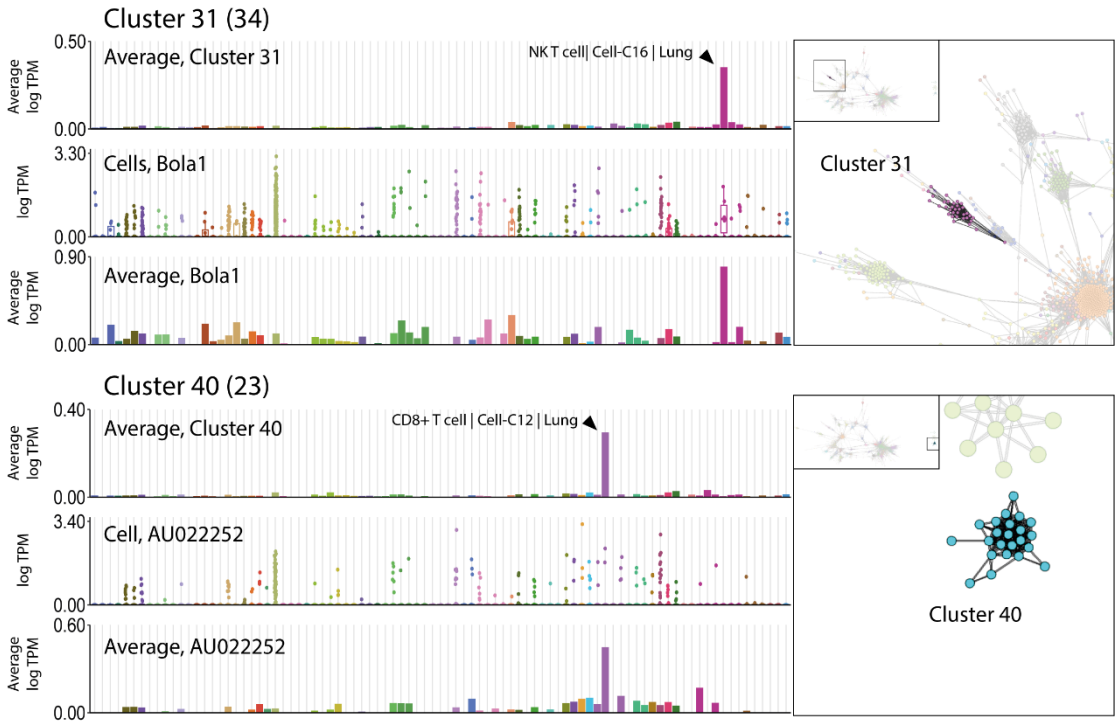


Figure 3.8. Gene clusters from Average method.

Shown are the expression profiles for two gene clusters, cluster 31 (upper panel) and 40 (lower panel) from the Average method. These clusters had the least overlap of genes with networks from the other two methods. Each panel has three expression profiles, the first, being the average expression (y axis) profile of genes from that gene cluster as observed in each of the different cell clusters (x axis). The second shows the expression profile of the genes with the highest average expression in that gene cluster for each cell (dots) across the different cell

clusters. The third shows the average expression profile of the same genes by the average method. The networks on the right shows the location of these clusters within the network.

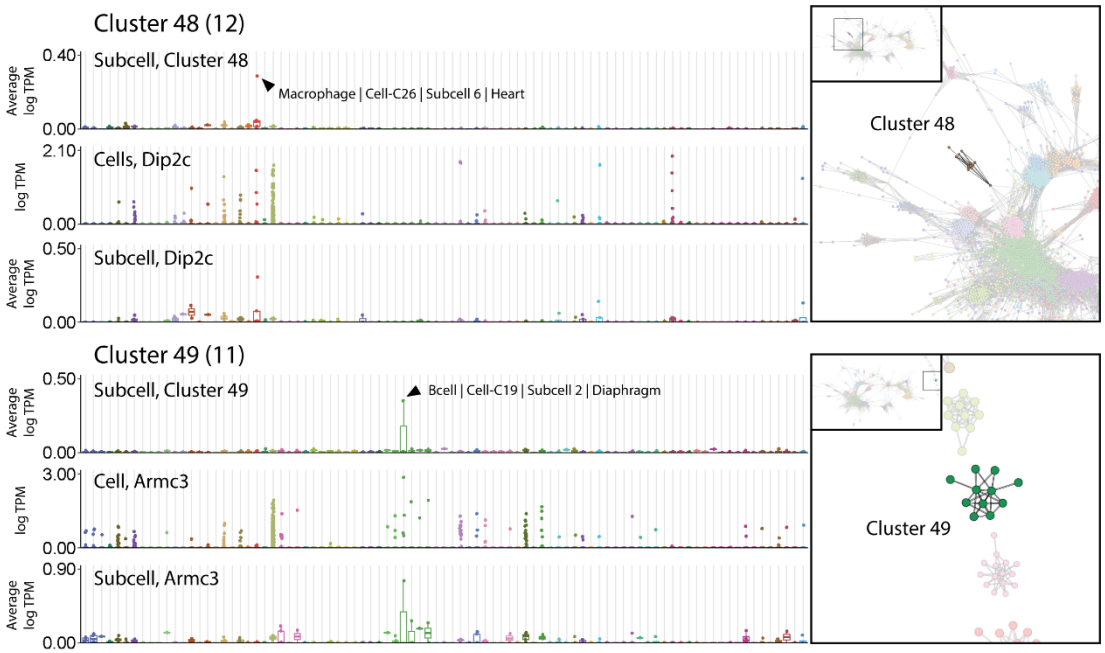


Figure 3.9 Gene clusters from Subcell method.

Shown are the expression profiles for two gene clusters, cluster 48 (upper panel) and 49 (lower panel) from the Subcell method. These clusters were found to have the least overlap with networks from the other two methods. Each panel has three expression profiles, the first, being the average expression (y axis) profile of genes from that gene-cluster as observed in each of the Subcells (dots) of the different cell-clusters (x axis). The second shows the expression profile of the genes with the highest counts in that gene cluster for each cell (dots) within the different cell clusters. The third, shows the average expression profile of the same genes by the Subcell method. Here, the gene expression of Subcells are shown as boxplots, as multiple Subcells are part of a cell-cluster, unlike the Average and Filter-average approach where there is a single sample/bin for a cell-cluster, hence displayed with histograms. The networks on the right shows the location of these clusters within the network.

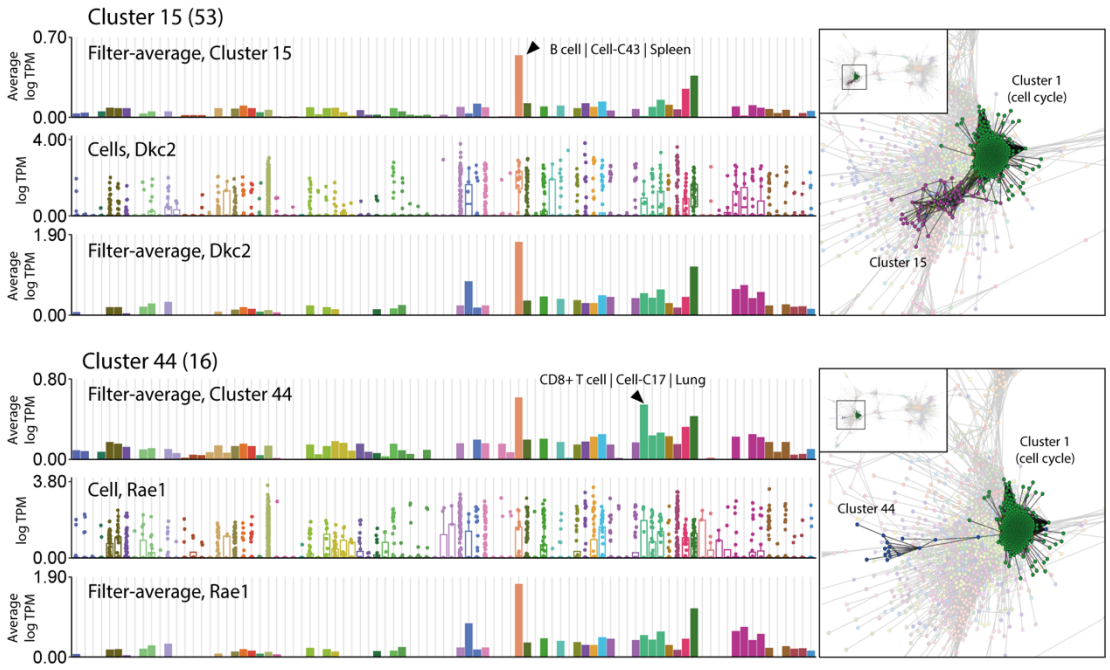


Figure 3.10. Gene clusters from Filter-average method.

Shown are the expression profiles for two gene clusters, 15 and 44 from the Filter-average method. These clusters had the least overlap with networks from the other two methods. Each panel has three expression profiles, the first, being the average expression (y axis) profile of genes from that gene-cluster as observed in each of the different cell-clusters (x axis). The second shows the expression profile of the genes with the highest counts in that gene cluster for each cell (dots) within the different cell clusters. The third, shows the average expression profile of the same genes by the average method. The networks on the right shows the location of these clusters within the network.

Table 3.2. GO enrichment of gene clusters compared between methods

Method	Cluster	ID	Description	Adjusted P value	Count
Average	Cluster 40	GO:0060215	Primitive hemopoiesis	8.98E-02	1
		GO:0018342	Protein prenylation	8.98E-02	1
		GO:0097354	Prenylation	8.98E-02	1
		GO:0007183	SMAD protein complex assembly	8.98E-02	1
		GO:0007351	Tripartite regional subdivision	8.98E-02	1
	Cluster 31	GO:0071712	ER-associated misfolded protein catabolic process	4.81E-02	2
		GO:0071218	Cellular response to misfolded protein	6.32E-02	2
		GO:0051788	Response to misfolded protein	6.32E-02	2
		GO:0006515	Protein quality control for misfolded or incompletely synthesized proteins	6.32E-02	2
		GO:0043901	Negative regulation of multi-organism process	1.10E-01	3
Subcell	Cluster 49	GO:0060502	Epithelial cell proliferation involved in lung morphogenesis	6.22E-02	1
		GO:0090494	Dopamine uptake	6.22E-02	1
		GO:0061140	Lung secretory cell differentiation	6.22E-02	1
		GO:0071679	Commissural neuron axon guidance	6.22E-02	1
		GO:0090493	Catecholamine uptake	6.22E-02	1
	Cluster 48	GO:0098739	Import across plasma membrane	5.55E-02	2
		GO:0098656	Anion transmembrane transport	5.55E-02	2
		GO:0015893	Drug transport	5.55E-02	2
		GO:0015809	Arginine transport	5.55E-02	1
		GO:0007548	Sex differentiation	5.55E-02	2
Filter-average	Cluster 44	GO:0032543	Mitochondrial translation	9.69E-04	3
		GO:0140053	Mitochondrial gene expression	1.26E-03	3
		GO:0051028	mRNA transport	4.78E-02	2
		GO:0034660	ncRNA metabolic process	4.78E-02	3
		GO:0050657	Nucleic acid transport	4.78E-02	2
		GO:0042254	Ribosome biogenesis	7.12E-05	8

	Cluster 15	GO:0022613	Ribonucleoprotein complex biogenesis	7.12E-05	9
		GO:0034660	ncRNA metabolic process	7.98E-05	9
		GO:0006364	rRNA processing	4.65E-04	6
		GO:0034470	ncRNA processing	7.25E-04	7

3.3.4 Gene coexpression network from tissue-resident immune cells

To identify gene coexpression clusters associated with TRICs from Tabula Muris, the GCN constructed from Filter-average approach in the previous section was used. Relative to the other two methods for constructing GCNs, the Filter-average approach had few false positive gene clusters i.e. Tabula. Muris Gene Clusters (TMGC) represented known biology rather than technical artefacts e.g. spikes in expression. The GCN constructed from the previous section was used, which composed of 4,006 genes. Individual TMGCs, where possible, were given a functional annotation based on literature mining and enrichment analysis using the ImmGC annotation and GO terms (**Table S3.7, Figure 3.11**). 195 TMGCs were found to be enriched in ImmGCs. This included TMGCs representative of pathway biology, cell lineages, cell types and TRICs. The pathway signatures identified included cell cycle (TMGC 1), ribosomal (TMGC 5), and class 2 antigen presentation (TMGC 35). Cell cycle genes from TMGC 1 (including *Cdk1, Cdk2*, and *Mcm* gene family) were highly expressed in single and double positive thymocytes (TMCC 3, TMCC 28 and TMCC 30) which are known to be highly proliferative during thymopoiesis. Ribosomal genes were highly expressed across cell types, excluding neutrophils which have relatively little transcriptional activity to other immune cells during homeostasis (Monaco et al., 2019). As expected, class 2 antigen-presentation genes (*H2-Aa, H2-Ab1* and *H2-Eb1*) were highly expressed in B cells, macrophages and DC populations. Lineage associated signatures were identified for myeloid (TMGC 2, 281 genes and TMGC 11, 66 genes) and lymphoid (TMGC 45, 15 genes) lineages. Lymphoid associated genes included *Ets1* and *Ptprcap*, known regulators of lymphoid differentiation and activation. While, the myeloid signature comprised of known regulators (*Sfpi1*, *Cebpa* and *Cebpb*), certain genes within these TMGCs were highly expressed in specific cell types, e.g. the complement family (*C1qa*, *C1qb* and *C1qc*) were highly expressed only in macrophages. Such cases where TMGC include gene with multiple expression patterns highlights the influence of clustering and network construction parameters on the final TMGC.

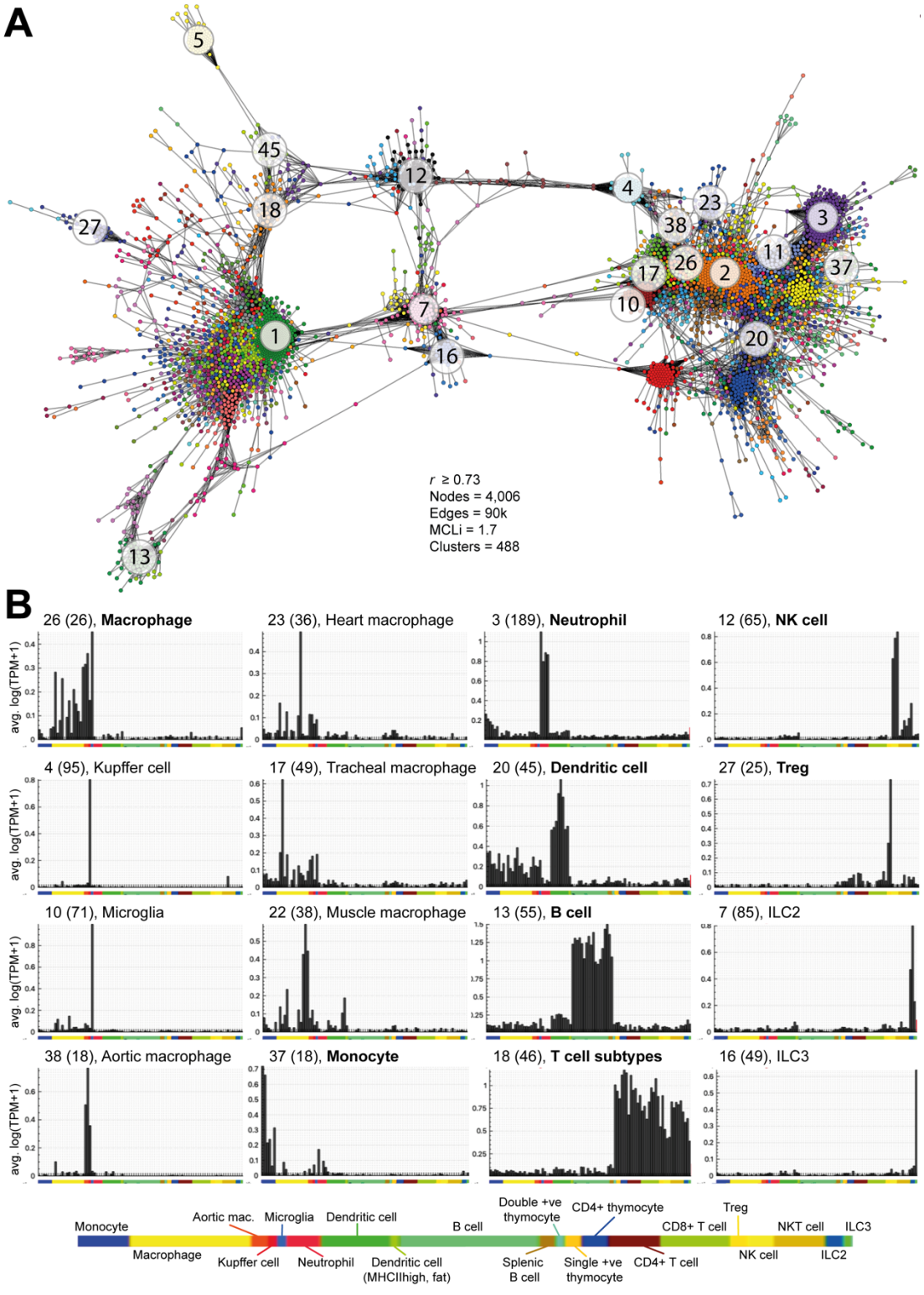


Figure 3.11. Tabula Muris gene coexpression patterns.

A) The gene coexpression network constructed from the Filter-average method (generated and studied in the previous section). Briefly, the GCN was constructed from the expression matrix calculated from the Filter-average approach. Here, a correlation threshold was chosen such that ~4,000 genes were connected within the graph. For each of the clusters shown in A), there are also B) the average expression profiles across Tabula Muris cell-clusters (x axis). Each panel represents gene signatures for either immune cell lineages, cell types or TRICs. The samples of each plot are ordered and coloured based on the bottom legend. Headings in bold represent the major cell types after which follow the expression profiles of related subtypes. mac: macrophage

Cell type-associated gene signatures were identified for macrophages (TMGC 26), monocytes (TMGC 37), neutrophils (TMGC 3), DCs (TMGC 20), B cells (TMGC 13), T cells (TMGC 20), Tregs (TMGC 27), NK cells (TMGC 12) and ILCs (TMGC 7 & 16). Except for NKT cells, all aforementioned gene signatures had a comparable ImmGC (enrichment *adj. P value* < 0.05). Genes associated with NKT cell were split across TMGCs, much like gating strategies adopted for flow-cytometry which require a series of markers to specify cell types, e.g. NKT cells showed a high expression of signature genes from both T cells (TMGC 18, 46 genes) and NK cells (TMGC 12, 65 genes). For similar reasons, TMGCs were identified for some cell subtypes including some TRICs, but not all the cell subsets.

TMGCs were associated with ten macrophage subsets. Well-studied macrophage subtypes associated with relatively large TMGCs (indicative of their unique biology) were Kupffer cells (TMGC 4, 95 genes) and microglia (TMGC 10, 71 genes). Kupffer cells showed a high expression of known markers, e.g. *Clec4f*, *Alb* and *Nr1h3*, while microglia clusters included *Cx3cr1*, *P2ry12*, *Tyrobp* and key regulator *Sall1*. TMGC, 23 (36 genes) and 38 (18 genes) corresponded to the aortic macrophage TMCC 31 and a combination of TMCC 24 and 26, respectively. TMGC 23 contained aorta-associated genes including *Fktn* (highly expressed in the heart and mutations in which cause Fukuyama muscular dystrophy) (Ujihara et al., 2019), *Hrh1* (G protein mediating muscle contraction and capillary permeability) (Bhuiyan et al., 2011), and *Pppp2r5b* (involved in myocyte function)(Lubbers and Mohler, 2016). TMGC 38 comprised of important myocyte homeostatic genes such as *Nppa* (required for myocyte growth and differentiation) (Kny et al., 2019) and *Actc1* (cardiomyocyte survival) (Forte et al., 2018). TMGC 17 and 28 showed a high average expression in tracheal macrophages of TMCC 6 and 27. TMGC 17 comprised of various lung and endothelium associated homeostatic genes like *Aqp1*

(lung fluid transport) (Gao et al., 2013), *Bpifa1* (upper airway immunity) (Tsou et al., 2018), *Eng*, *Egfr* (Brechbuhl et al., 2014), and *Cav1* (Krasteva et al., 2006). Certain genes within this TMGC were also linked with lung diseases, such as *Kank2* (Zhang et al., 2018a), *Prmt1* (Avasarala et al., 2015) and *Cul4* (Wang et al., 2015) were associated with non-small cell lung cancer, and *Tns1* with COPD (Soler Artigas et al., 2011). TMGC 28 was enriched in the ImmGC 37 (adj. *P* value < $2.8x10^{-3}$) and found to have high expression in alveolar macrophages. The TMGC comprised of several chemokines (*Ccl3*, *Ccl4* and *Ccl12*) supported by enrichment of "Leukocyte migration" GO term (*adj. P value* < $4.4x10^{-11}$). Macrophage TMCC 36 comprised of cells derived from various muscle-containing tissues, i.e. diaphragm, trachea and skeletal muscle. The group of cells showed high expression of genes from TMGC 22, which was enriched in muscle-associated (*adj. P value* < $1.2x10^{-2}$) GO terms. Cell types spanning different tissues were also identified, like ILCs of TMCC 35 and 38. Corresponding to these TMCCs were the TMGC 16 and 7, respectively. In the Tabula Muris these cells originated from adipose and trachea, suggesting these TMCCs represented ILC subtypes, rather than tissue-resident populations. TMGC 16 included *Il17a* known to be expressed in ILC3 *and Il5* associated with ILC2 was present in TMGC 7. Whilst, not all TMCCs or TRICs identified from the sample-to-sample network had a corresponding gene coexpression module, some were identified and are presented in **Table 3.3**.

Table 3.3. Cell-clusters and associated gene coexpression clusters.

Cell ontology	Cell ontology hierarchy	Cell Cluster	Tissue	Gene Cluster
Myeloid	Cell lineage			2 11
Monocyte	Cell subtype	7 14		37 14
	TRICS	42	Fat	33
Macrophage	Cell type			26
	TRICS	36 1 1 6 9 29 29 31 39 39 multiple multiple	multiple Brain Brain Trachea Fat Kidney Trachea Heart Liver Liver Aortic Fat	22 10 47 17 55 64 28 23 4 60 38 21

Neutrophil	Cell type			3
Dendritic cell	Cell type			20 51 56
	Cell subtype	15 20 15, 33		50 8 40
	TRICS	15 33 2	Lung Trachea Fat	49 29 6
Dendritic cell, Monocyte	Cell type			48 58
Lymphoid	Cell lineage			45
B cell	Cell type			13
	TRICS	2 2 23 43 multiple	Liver Lung Lung Spleen Spleen	46 54 31 19 30
T cell populations	Cell type			18 39
CD4+ T cell	TRICS	27	Lung	59
CD8+ T cell	Cell subtype	17,34		62
	TRICS	17	Lung	67
Treg	Cell subtype	41		27
	TRICS	41	Fat	57
NK cell	Cell type			12
NK and NKT cell	Cell type			32
NKT cell	TRICS	16 16	Liver Fat	36 43
ILC	Cell subtype	38 35 38		7 16 41

3.4 Discussion

The protective role of the immune system against infection is well understood, with many studies characterising their responses to activation in disease, injury and to artificial stimuli. In the past decade, efforts have also been made to understand tissue-specific roles of the immune system. TRICs maintain tissue homeostasis by acquiring various functions specific to the needs of the host tissue, e.g. phagocytosing surrounding debris/necrotic cells, or facilitating organogenesis during early development (Caputa et al., 2019, Munro and Hughes, 2017). These adaptations occur at the transcriptomic and epigenetic level and influence the cell's response to inflammation which differs from their tissue naïve cells of the same cell type. The importance of these TRICs has been shown

in various diseases, e.g. microglia in neurodegenerative diseases as well as tumour-associated ILCs and memory T cells (Dadi and Li, 2017, Djenidi et al., 2015, Colonna and Butovsky, 2017). Consequently, several studies are being conducted to further our understanding of their role in these tissues in health and disease. Recent efforts have investigated the cellular heterogeneity of a specific cell types across tissues (Mass, 2018, Sagebiel et al., 2019, Miragaia et al., 2019, Szabo et al., 2019), while others have compared immune cell populations (Lavin et al., 2014a, Yoshida et al., 2019). Such studies have generally only examined a few tissues, and while some focus on a single immune cell type, others investigate specific cell types using predefined markers. In this study, I utilise the growing public resources available from bulk and scRNA-Seq data derived from various tissues to compare immune cell populations. Through this analysis I also explore the contribution of tissues to immune cell heterogeneity. In the past, classification of immune cell subtypes has largely focused on their response to stimuli, and whilst some have been named based on their tissue of origin, e.g. Kupffer cells, microglia, little attention has been given to their tissue-specific function under physiological conditions. This also raises the question of how we define cell types, for which in this chapter differences between cell types were examined, as identified by markers and those through unbiased clustering of single cells based on their transcriptomic profiles. Furthermore, through these analyses I developed techniques to identify cells based on their transcriptomes and construct GCNs to find coexpressing gene clusters associated to the cell classifications.

To understand how well cell markers differentiate cell types two ImmGen datasets were studied, the “Mononuclear phagocyte” dataset (the ImmGen and Benoist, 2016) which has only very recently been examined for the metabolic diversity of mononuclear phagocytes (Gainullina et al., 2020), and the “System wide” dataset used to define cis-regulatory elements of the mouse immune system (Yoshida et al., 2019). On analysing the sample-to-sample network the observed grouping of samples was influenced by the cell type, tissue of origin and developmental state, as might be expected. Distinct groups were observed for certain macrophages such as alveolar, microglial, splenic and peritoneal macrophages. Furthermore, peritoneal macrophages were found to be proximal to monocytes. In contrast, Kupffer cells and aortic macrophages were closely related to one another. These associations could hint at the ontogeny of cell types, e.g. both liver and heart macrophage have a significant contribution from yolk-sac-derived macrophages of E8.5 (Epelman et al., 2014b) while peritoneal macrophages are

maintained through circulating monocytes in adults (Cassado Ados et al., 2015). Unlike myeloid cell types, all those from the lymphoid lineage were connected as part of a single component and clustered together based on cell types. 156 gene modules were found associated with different cell biology which ranged from cell lineages to cell subtypes such as TRICs. Such an analysis to identify TRICs signatures has not been conducted using these datasets and provide gene signatures for several known TRICs as defined by markers. These gene signatures highlighted TRIC associated homeostatic genes including transcription factors e.g. *Sall1* (microglia), *Gata6* (peritoneal macrophage), *Spic* (splenic macrophage) and *Pparg* (alveolar macrophage) (Mass, 2018). Interestingly, only a few gene clusters were specific to cell types as defined by markers, e.g. a subset of splenic macrophages. Hence, instead of defining cell types from cell-cell clustering based on their entire transcriptome, a better way to define cell types/clusters may be to use gene coexpression patterns, especially for those cell types with few genes differing in expression from other cell types. The gene cluster corresponding to Kupffer cells (cluster 43), did not contain the newly identified transcriptional regulator *Id3* (Mass et al., 2016), as the gene displayed expression in some lymphoid cells as well. Indicative of a multifunctional role of transcriptional factors that are maybe required in combination to prompt a unique cell identity.

The above analysis showed that cell types defined by markers did not always separate well when considering the whole transcriptome. This could suggest: 1) that certain markers may not be specific and instead encompass several cell types e.g. F4/80 was used to isolate microglia (Lawson et al., 1990) in the CNS under the assumption of a lack of other CNS macrophages (Greter et al., 2015). Therefore current markers to isolate microglia are more specific like Cx3cr1 and Tmem119 (Bennett et al., 2016). The use of scRNA-Seq to identify cell types in an unbiased way should help mitigate such misconceptions. 2) Using the entire transcriptome may not be sufficient/appropriate in differentiating cell types, especially for those cell types distinguished by a few marker genes, especially in the case of closely connected cell types observed in the ImmGen sample-to-sample network, including DC subtypes and lymphoid cells, which in GCNs show specific gene coexpression clusters. Furthermore, classifying cell types may require other omics and functional studies.

As an unbiased approach, immune cells from the Tabula Muris dataset were also studied. The dataset comprises of scRNA-Seq data derived from 20 different tissues. Here, immune cells were selected for analysis based on the author's original cell

annotation. However, examination of these cells using a cell-to-cell similarity network, suggested the original annotation could be improved upon, e.g. different myeloid cells like monocytes and neutrophils could not be differentiated within the network, and the annotation included broad groupings such as leukocytes and myeloid cells. Here, I developed a novel method which utilized gene expression profiles of known cell types to help annotate unknown cells based on their transcriptomic profile. Unlike other methodologies, the approach annotated entire cell clusters to reduce the influence of misannotated cells, i.e. only when the majority of cells within a cluster were of the same annotation was the cluster annotated as such. The approach added significant detail to the original annotation, and in certain cases revealed TRICs, such as aortic macrophages and splenic B cells. Further granularity was added to this by clustering the data. However, much like our conclusion from ImmGen, transcriptome wide comparisons of cell types could be limited to broad cell type annotations, hence further investigations were made for cell subtypes like TRICs by examining gene coexpression modules.

The field of scRNA-Seq has brought forth several methodologies to account for the noise and heterogeneity within the data, involving clustering (Lin et al., 2017), imputation (Eraslan et al., 2019), dimensionality reduction (Sun et al., 2019) and trajectory analysis (Wolf et al., 2019). However, few studies have analysed coexpressing genes (Chen and Mar, 2018, Pratapa et al., 2020) and rather rely on differential expression analysis to define cell types. Here, three approaches to construct GCNs from scRNA-Seq data were developed and evaluated to help identify gene coexpression modules supporting the cell classifications from Tabula Muris. The methods extended from the idea of MetaCells (Baran et al., 2018), which aggregates reads across near identical groups of cells to improve the signal within scRNA-Seq data. Here, similar approach was adopted to reduce the noise and variation within the data to help construct GCNs with a higher distribution of correlation values relative to that generated from the entire dataset of cells. By aggregating gene expression values over clusters one can capture the variation across cell types rather than cells. Since, each cell type is defined by one bin/sample the correlation space is balanced giving equal weightage to each cell type, especially in cases where there is a large discrepancy in the number of cells between cell types. Although the approach is yet to be compared with other methodologies, it addresses two issues, the need for analysing inter-cell type variation rather than intra-cell type variation and having a balanced correlation space. To construct GCNs from Tabula Muris by aggregating data across cell types/clusters two factors were additionally considered,

filtering expression outlier (including spikes and dropouts) and introducing “replicates” or Subcells by subclustering cells within a cluster. These two ideas addressed key issues in defining correlations between samples from RNA-Seq data, the noise (spikes and dropouts) and variation (intra-cell types variation). Accordingly, three approaches were employed 1) simple aggregation/averaging read across cells in clusters, 2) averaging of reads within subclusters and 3) filtering the expression before averaging within clusters. Preliminary analysis for the distribution of correlations found similarities between the average and Subcell approach while also revealing lower correlation values for the Subcell method. This was understandable as the method involved a greater number of bins/Subcells to represent a cluster relative to the other two approaches. A simple averaging within clusters and subclusters revealed a great number of gene modules associated with spikes in expression of few cells. This effect was more pronounced in clusters with a small number of cells as each cell’s contribution was higher relative to larger clusters having hundreds of cells. Some of these gene clusters were further shown to be absent of any known biology based on GO annotations. In contrast, filtering and then averaging within clusters resulted in fewer gene modules associated with technical artefacts. The factors discussed herein could influence other GCN construction approaches which may be biased by dominant (cell numbers) cell types within the dataset and to technical artefacts. However, this would require a formal comparison. The Filter-average approach was taken forward to examine the coexpressing genes and biology of TRICs from the Tabula Muris.

To support the cell classifications, gene coexpression clusters were annotated based on their expression profile and validated using the results from the ImmGen analysis and GO annotations. 44 gene clusters showed high expression for either certain cell linages, cell types, cell clusters or TRICs. Several of these signatures were found to overlap with the ImmGen gene clusters for the same cell type. In line with our current knowledge of TRICs the most diverse population was of macrophages having 11 gene clusters, including associated with the brain, trachea, fat, kidney, liver and aorta. Second, were DCs with 7 gene clusters (lung, trachea and fat) and third B cells with 6 gene clusters (liver, lung and spleen). Through literature mining several homeostatic genes were identified supporting gene clusters associated with eight of the tissue-resident macrophage populations. Furthermore, signatures were found for cell types dependent and independent of tissue, e.g. DC clusters from the lung and trachea shared a common biological gene signature while also having signatures unique to their tissue. While few

such cases were observed cell types largely segregated on tissue, signifying this to be the dominant factor for cell type variation. Although a further comparison is required with cell types identified in literature this study covers a wide range of cell types across tissues and supports them with relevant coexpressing gene clusters.

In summary, this chapter has used both biased and unbiased approaches to examine the tissue dependence of immune cell diversity based on their transcriptomic profile. The unbiased Tabula Muris analysis revealed 44 gene clusters associated with the hierarchy of immune cell types from lineages to cell subtypes, with the most numerous being those associated with TRICs, indicative of their unique biology. Furthermore, these observations aligned with well characterised marker-defined cell types from the ImmGen. Indeed, similar analyses on larger datasets will help us appreciate the full spectrum of TRICs to ultimately construct an immune cell type map.

4. Chapter 4: Human tissue-resident immune cell signatures from bulk tissue RNA-Seq data

4.1 Introduction

In previous chapters TRIC signatures were derived using data generated from various transcriptomic platforms and across species. The utility of these signatures was also demonstrated. For instance, it was in shown in chapter 2 that by deriving the homeostatic signature of human microglia, one could then examine the cell's profile and proportion across different brain regions and conditions in health and disease. Accordingly, in chapter 3 the analysis was extended to derive homeostatic TRIC signatures across different mouse tissues at the cellular level. This resolution helped determine the biology of these immune cell types across tissues and the contribution of tissue-dependent variation towards cellular heterogeneity. Together these analyses in mouse and human pointed towards the potential to deconvolute TRIC signatures from bulk RNA-Seq data for human tissues.

Similar to studies in mouse, researchers have investigated the heterogeneity of human derived TRICs. These studies span several cell types including tissue-resident memory T cells (Kumar et al., 2017), regulatory T cells (Niedzielska et al., 2018), ILCs (Ricardo-Gonzalez et al., 2018), NK cells (Crinier et al., 2018), macrophages (De Schepper et al., 2018), B cells (Zhao et al., 2020) and DCs (Alcántara-Hernández et al., 2017), derived from the lung, liver, spleen, gut, skin, adipose, blood, bone marrow, brain and thymus. Different methodologies have been adopted to study the transcriptomic, proteomic and epigenetic profiles of these cells, e.g. bulk-RNA-Seq, scRNA-Seq, CyTOF and ATAC-Seq. The experimental design of these studies can be generalised into one of two approaches; the investigation of multiple or all cell types within a tissue, and the examination of a single cell type across multiple tissues. The first provides information on the biology of individual tissue cell types and accordingly helps understand the tissue as a whole. For example, analysis of each cell type from the liver (hepatocytes, liver endothelial cells, cholangiocytes, stellate cells, Kupffer cells, monocytes, T cells, NK-like cells and B cells) using scRNA-Seq provided a description of their function, thereby informing us of the different biology within the tissue (MacParland et al., 2018). However, in this chapter the focus was on studying the variability across TRIC types across tissues. The aforementioned studies which analysed a single cell type across tissues tended to describe the cell's core signature and its tissue specific functions, as have been described for tissue-resident Tregs in non-lymphoid tissues (Miragaia et al., 2019). As a

reference for tissue-resident populations, many studies follow a common procedure of comparing TRICs with the corresponding cell types in blood or lymphatic tissues (for lymphocytes). Such studies have highlighted tissue specific properties such as genes associated with cell migration which are required to establish tissue residency (Miragaia et al., 2019). In contrast to such exploratory analyses, certain studies have focussed on the relevance of TRICs in pathological conditions (Savas et al., 2018) or their response to stimuli (Woodward Davis et al., 2019). Finally, TRICs have also been compared across species to reveal evolutionary changes in cellular heterogeneity and their functions. A notable example being the comprehensive examination of microglia across eight species (human, macaque, marmoset, sheep, mouse, hamster, chicken, and zebrafish) to examine the changes within microglial populations through evolution (Geirsdottir et al., 2019). In summation, these studies inform us of TRICs with respect to their core signature, tissue specific functions, tissue-dependent subsets, their abundance, and their conservation across species. This study addresses similar research questions by utilising GTEx, a human tissue expression atlas.

Similar to our analysis of human microglia in chapter 2, a thorough examination of human TRIC expression signatures was conducted in this chapter by utilizing the GTEx resource (Lonsdale et al., 2013). The resource comprises of bulk RNA-Seq data derived from 33 tissues across 52 regions and is the most comprehensive human tissue transcriptomic resource currently available. Using this dataset, the following questions were investigated: 1) the distribution immune cell types across tissues; 2) genes unique to cell types and those that are shared amongst them across tissues; and 3) how they compare to mouse TRIC signatures. To do so, a set of reference immune signatures (RIS) which represent core genes for cell types (macrophage, B cell, T cell, NK cell, neutrophils and plasma cell) were first defined. These gene sets guided the derivation of TRIC signatures within individual tissues, which was performed using GCNs. Subsequently, these signatures were compared with one another and with TRIC signatures previously derived from the mouse RNA-Seq data in chapter 3.

4.2 Material and methods

4.2.1 Data preprocessing

Human tissue transcriptomics data was downloaded from the GTEx portal (https://www.gtexportal.org/) (Lonsdale et al., 2013). The dataset (version 7) is a

comprehensive collection of RNA-Seq data from various human tissues. It includes samples from 33 tissues which are further sub-divided into 52 regions. For these analyses, tissues with a minimum of 20 samples were considered, thus excluding samples from the bladder (n = 11), cervix (n = 11) and fallopian tube (n = 7). Furthermore, samples derived from Epstein-Barr virus transformed lymphocytes were excluded as they were not considered relevant to this work. The filtered dataset was derived from 28 tissues (n = 11,529, across 48 regions), where the number of samples per tissue varied from 45 to 564. Due to the large number of brain regions, samples from similar regions were merged into a single dataset, e.g. all cortex-related regions were treated as one. This was done for the cerebellum, basal ganglia and cortex, producing a single expression matrix for each. Two rounds of quality control were conducted for each of the tissues. First, a sample-to-sample network was generated at a Pearson correlation coefficient threshold of $r \geq 0.9$ for each tissue region, with lowly expressed genes (maximum expression <1 TPM) filtered from the input gene expression matrix. In each case the majority of samples were connected as a single network component. Outlier samples that were singlets i.e. unconnected to other samples were filtered out. Second, samples from each tissue region were filtered on the basis of their expression of ImSig gene signatures (Macrophages, Monocytes, Neutrophils, B cells, Plasma cells, NK cells, T cells, and Interferon). Samples were removed if they had an average expression for any of the signatures greater than 3 standard deviations away from the linear or the geometric mean. Resultantly an expression matrix was produced for each of the 28 tissues which included 10,461 samples.

4.2.2 Deriving the reference immune signature

To generate RISs a GCN was generated from a dataset comprising of a subset of samples from each tissue of the GTEx. To balance the contribution of each tissue to this combined dataset an equal distribution of 100 samples per tissue type (where available) were selected. In tissues where fewer than 100 samples were available, all the samples from these tissues were considered. Additionally, where a tissue was represented by more than one region, equal number of samples from each tissue region were selected such that the samples from the tissue totalled hundred. For example, 100 samples were selected from the colon where 50 samples were taken from each of the regions, the “Sigmoid” and “Traverse” region. To decide which samples should be selected for the combined dataset, samples from each tissue region were first ranked based on their average expression of the individual ImSig signatures. As a result, each sample had nine

rankings corresponding to the individual ImSig signatures and samples showing the highest variation in their ranking across signatures were preferred. For instance, samples with a high average expression of T cell signature genes and a low expression of those for B cells would be preferentially selected over samples with a high expression of both cell types. Theoretically, such an approach would include samples across which the proportions of immune cell types varied the most. Therefore, this would be ideal for gene coexpression analysis which is driven by the variation of biological signals/gene expression across samples. The resultant combined dataset amounted to 2,719 samples.

To identify genes associated with potential cell subtypes, coexpressed genes went through two rounds of clustering (**Table S4.1 & 4.2**). First, genes associated with the immune system were identified i.e. clusters enriched in ImSig genes. For this a GCN was generated from the combined dataset at a Pearson correlation threshold of $r \geq 0.83$ in order to accommodate close to 15,000 genes. The resultant GCN was clustered using MCL (Dongen, 2000) with an inflation of 1.7 Each cluster was tested for an enrichment in ImSig gene signatures using the Fischer's exact test. The majority of enriched (*adj. P value* < 0.05) clusters formed a subnetwork which were separate from other components of the graph. These clusters comprised of 1,090 genes and were considered as the immune component of the network. Using this subset of genes from the combined dataset a GCN was reconstructed using the kNN algorithm (k = 5) and a correlation threshold of $r \geq 0.82$. The graph was clustered into 64 clusters using an inflation of 1.7. These clusters were annotated for a given cell type depending on the ImSig cell signatures in which they were enriched (*adj. P value* < 0.05) , and marker genes e.g. for B cells (CD19 and BLK), endothelial cells (ICAM2), macrophages (complement family, CSF1R and AIF1), NK cells (KLRK and GZM family), neutrophils (CSF3R and FCGR3B), plasma cells (IGH family) and T cells (CD3 chains). Gene clusters associated with certain cell types of the myeloid lineage, including macrophages, monocytes and DCs could not be differentiated and were hence classified as ""Mac-Mono-Den". The genes with the "Mac-Mono-Den" annotation were divided into two groups as some of these clusters were segregated i.e. not directly connected with one another in the graph. A similar case was observed for "T cells". Clusters which included genes from multiple ImSig signatures were annotated as "Immune". Cluster 18 which comprised of mitochondrial and ribosomal genes was not considered for downstream analyses.

4.2.3 Derivation of human tissue-resident immune signatures

To derive TRIC signatures a GCN was constructed for each of the 21 tissues individually. Here, the quality-controlled data without down-sampling was considered. A GCN was constructed for each tissue at a Pearson correlation threshold at which ~15,000 genes were present in the graph (**Table S4.3**). For the artery, liver, pancreas, pituitary, small intestine, spleen and stomach, genes representing multiple RIS clustered together, hence, GCNs were reconstructed at a higher Pearson correlation threshold to include ~10,000 genes within the graph. Clustering was carried out for each tissue using MCL with an inflation of 1.7. The annotation of gene clusters for a given cell type was carried out in three steps. Broadly, the strategy first identified which immune cell types were present in the tissue i.e. a confident set of core clusters representative of a specific immune cell type, followed by the aggregation of neighbouring clusters which likely consisted of immune genes. The three steps involved, 1) constructing the GCN and clustering it, as mentioned above. 2) Identifying gene clusters that confidently represented immune cell types i.e. clusters having a significant enrichment in a particular RIS cell type (adj. *P* value $<10^{-10}$) and included cell type associated markers. In majority of tissues NK cell and T cell markers clustered together these were annotated as “T cell – NK cell”. 3) Those clusters neighbouring the core were annotated for that cell type if they were enriched in RIS or if they were not connected to any other cluster. In cases where the clusters were neighbouring core clusters associated with multiple immune cell types, they were left unannotated.

4.3 Results

4.3.1 Reference immune signatures for human tissues

TRIC signatures were derived using GCNs from human bulk tissue transcriptomics data drawn from the GTEx project (v7). The dataset was subsetted by including tissues which had more than 20 samples, while excluding samples derived from EBV-transformed lymphocytes, thus leaving 11,529 samples from 28 tissues (43 tissue regions). Quality control was performed on each of the tissue regions by removing outlier samples which poorly correlated with other samples from that region. Additionally, we sought to remove samples with elevated levels of inflammation, as their presence in the dataset would drive inflammation associated immune signatures and potentially obscure homeostatic TRIC signatures. To identify such signatures, we used ImSig, a collection of cell/pathway

signature gene lists representative of different immune cells derived from samples under pathological conditions. The average expression of each signature was calculated for individual samples and those beyond 3 standard deviations of the geometric and linear mean for a given tissue were removed. The quality-controlled data comprised of expression matrices for each of the tissues amounting to 10,461 samples (**Table 4.1**).

Table 4.1 Sample distribution after quality control and in the combined dataset.

Tissue	Subregion	Number of RNA-Seq Samples	QC Samples	Combined dataset
Adipose	Subcutaneous	442	400	50
	Visceral (Omentum)	355	310	50
Adrenal Gland	Adrenal Gland	190	170	100
Artery	Aorta	299	267	33
	Coronary	173	153	33
	Tibial	441	403	33
Brain	Amygdala	100	94	12
	Caudate (basal ganglia)	160	148	6
	Nucleus accumbens (basal ganglia)	147	134	4
	Putamen (basal ganglia)	124	118	2
	Cerebellar Hemisphere	136	123	6
	Cerebellum	173	167	6
	Anterior cingulate cortex (BA24)	121	111	4
	Cortex	158	146	3
	Frontal Cortex (BA9)	129	122	5
	Hippocampus	123	117	12
	Hypothalamus	121	112	12
	Spinal cord (cervical c-1)	91	86	12
	Substantia nigra	88	81	12
Breast	Mammary Tissue	290	266	100
Colon	Sigmoid	233	212	50
	Transverse	274	248	50
Esophagus	Gastroesophageal Junction	244	223	33
	Mucosa	407	377	33
	Muscularis	370	342	33
Fibroblast	Transformed fibroblasts	343	306	100
Heart	Atrial Appendage	297	273	50
	Left Ventricle	303	277	50
Kidney	Cortex	45	39	39
Liver	Liver	175	161	100
Lung	Lung	427	376	100
Minor Salivary Gland	Minor Salivary Gland	97	86	86
Muscle	Skeletal	564	501	100
Nerve	Tibial	414	373	100
Ovary	Ovary	133	116	100
Pancreas	Pancreas	248	230	100
Pituitary	Pituitary	183	165	100
Prostate	Prostate	152	137	100

Skin	Not Sun Exposed (Suprapubic)	387	356	50
	Sun Exposed (Lower leg)	473	426	50
Small Intestine	Terminal Ileum	137	121	100
Spleen	Spleen	162	139	100
Stomach	Stomach	262	235	100
Testis	Testis	259	234	100
Thyroid	Thyroid	446	406	100
Uterus	Uterus	111	100	100
Vagina	Vagina	115	106	100
Whole Blood	Whole Blood	407	368	100
Bladder	Bladder	11	NA	NA
Cells	Cells - EBV-transformed lymphocytes	130	NA	NA
Cervix	Ectocervix	6	NA	NA
	Endocervix	5	NA	NA
Fallopian Tube	Fallopian Tube	7	NA	NA

Total number of samples downloaded from the GTEx (v7) portal (33 tissues, 52 regions). Those tissues which have not been considered for the study were not QC'd and accordingly are marked as "NA". Downstream GCN analysis for deriving TRIC signature was conducted for individual tissues highlighted in green. The data was also downsampled to produce the combined dataset used to identify the RIS downstream.

To guide the derivation of immune cell signatures for each tissue, we required a set of markers genes or signatures for these cell types. Although, the ImSig signatures characterised these cell types, they were derived from samples under pathological conditions. Therefore, a new immune cell signatures were derived from a combined dataset of the different GTEx tissues using GCN analysis, ergo these would represent homeostatic RIS. To have an equal contribution from each tissue to maintain a balanced correlation space the data was down sampled to 100 samples per tissue (where possible) and equally sampling regions for that tissue where more than one was defined. Each tissue was downsampled independently, first, for a given ImSig signature all samples were ranked, and this was done for each of the seven ImSig cell signatures thus generating seven ranks for each sample. Subsequently, samples with the most variation between their rankings were given priority. Such an approach would select samples across which the expression of immune signatures varied the most i.e. the proportion (cell abundance) of different cell types varied the most across these samples. This procedure of downsampling would be ideal for GCN analysis, as it would balance the contribution of tissues to the Pearson correlation between genes.

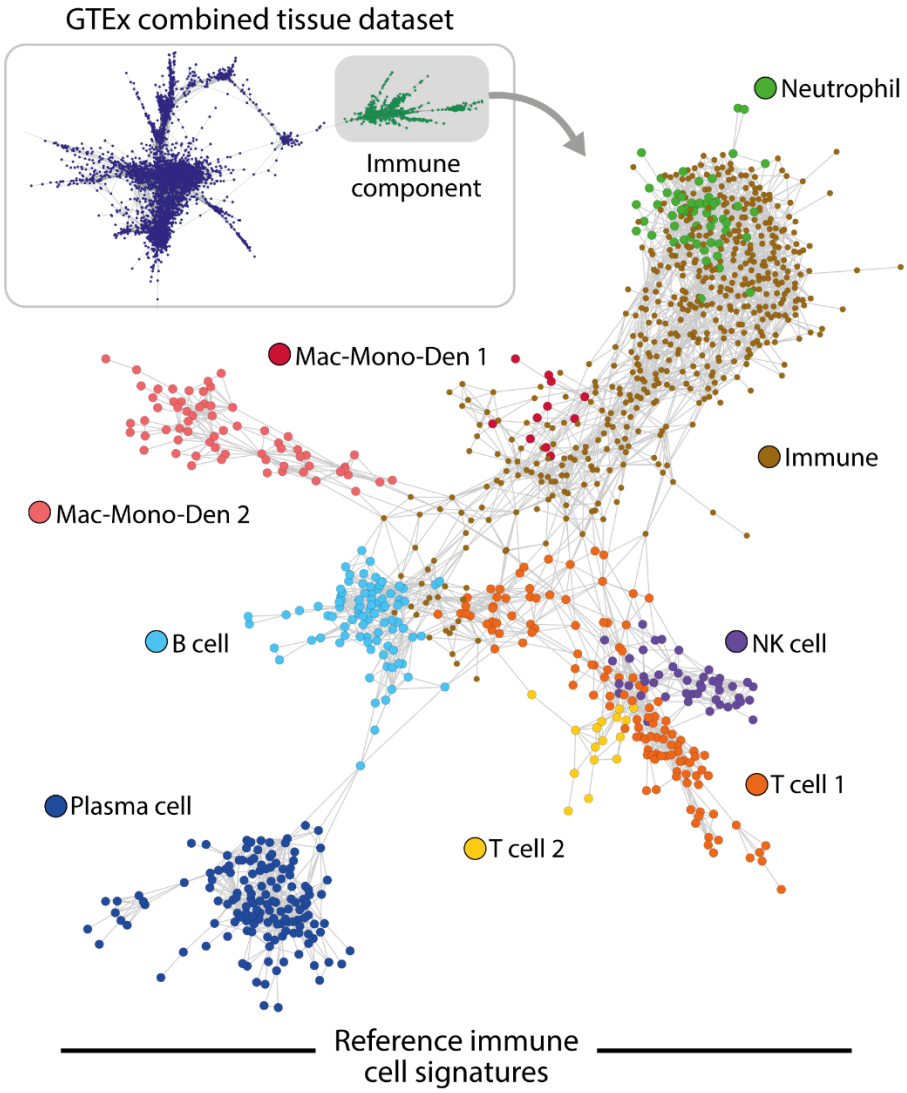


Figure 4.1. Derivation of a human reference immune signature.

Procedure for identifying homeostatic immune cell signatures from a subset of GTEx samples from each tissue. A combined dataset was produced by down-sampling to 100 samples from each of 28 human tissues. First (top left), a GCN of the top ~15,000 correlated genes was constructed (~8.6M edges) using $r \geq 0.83$ and clustered using MCL (inflation = 1.7). A distinct region of the network (highlighted in green) was enriched in ImSig genes. These 1,090 genes were reanalysed (centre panel) by constructing a GCN (4,156 edges) using a kNN (k = 5), $r \geq 0.82$ and clustered using MCL (inflation = 2.2). Highlighted, are the subsequent annotations for gene clusters of the network for immune cell types/subtypes.

From the combined dataset, RIS were derived in two stages (**Figure 4.1**). First, a GCN was constructed from the combined tissue dataset using an *r* threshold value of ≥ 0.83. The 15,573 gene network connected by 8,648,165 edges was clustered using MCL with

an inflation of 1.7. Interestingly, the majority of clusters enriched (*adj. P value* < 0.05) in ImSig signatures localised to a distinct area of the graph (**Table S4.1**). This “immune segment" which comprised of 1,090 genes was re-examined for potential cell subtype signatures using a GCN. Using this gene-set a kNN graph was constructed at a correlation threshold of $r \geq 0.82$ and k = 5. The graph consisted of all the genes from the immune segment which were connected through 4,156 edges and clustered using MCL with an inflation of 2.2. Each of the resulting 64 clusters were annotated as being associated with a given cell type based on their enrichment in ImSig signatures and immune marker genes (**Table S4.2**). Genes were annotated for one of the 12 classes including nine immune cell types, MHC class 2, mitochondrial (cluster 18) and “immune” clusters. These “immune” clusters were annotated as such because they consisted of genes from multiple ImSig signatures, hence making it difficult to provide a single cell type annotation for these genes. As signatures of the different myeloid cell types (monocyte, macrophage and DC) were difficult to differentiate in the GCN, they were classified under a single annotation of macrophage-monocytes-dendritic cells (Mac-Mono-Den). For the majority of cell types, their associated gene clusters were concentrated within a region of the network. However, clusters for certain cell types were segregated, i.e. disconnected within the graph, suggesting that they could represent different cell subtypes. This was observed for Mac-Mono-Den and T cells where each had two groups of gene clusters spatially separated within the network. This distinction was considered for annotating signatures, e.g. T cells were identified by two signatures “T cell 1” and “T cell 2”. Such granularity in defining cell signatures could aid in distinguishing cell subtypes in downstream analyses by increasing the statistical power of enrichment analyses where an immune cell subtype maybe represented by only a few genes. The final RIS included 1,073 genes annotated for 11 signatures, as the mitochondrial gene cluster was excluded (**Table S4.2**).

4.3.2 Deriving tissues-resident immune cell signatures from human tissue transcriptomics

To derive TRIC signatures a GCN was constructed for each of the 21 tissues and genes were annotated based on the following criteria. For every tissue a GCN was constructed using a specific Pearson correlation threshold (**Table S4.3**) to produce networks with approximately 15,000 genes, which were subsequently clustered using MCL with an inflation of 1.7. For certain tissues immune signatures were difficult to segregate from

the GCN (artery, liver, pancreas, pituitary, small intestine, spleen and stomach) i.e. known markers of different cell types clustered together within the GCN such as T cell (*CD3E* and *CD3D*) and Mac-Mono-Den (*AIF1* and *CSF1R*). To help segregate such clusters into clusters associated with distinct cell types, GCNs were reconstructed at higher thresholds to include approximately 10,000 genes. The parameters for each network are described in **Table S4.3**. Gene clusters were annotated for immune cell types in three steps. As shown for the liver (**Figure 4.2**), first GCNs were constructed, clustered and clusters enriched for RIS were identified (*adj. P values* < 0.05). Second, as different tissues have distinct combinations of immune cell types present within them it was first necessary to determine which immune cell types were present in the tissue by identifying core gene clusters for each immune cell type. A core cluster associated to a certain cell type was identified based on its enrichment (*adj. P value* < 10^{-10}) of that RIS and known marker genes for that cell type. Third, clusters neighbouring core gene clusters were annotated for the associated cell type provided they were enriched in RIS or were only connected to the clusters considered from step 2. However, in cases where a cluster was neighbouring core clusters of multiple cell types they were left unannotated. Thus, TRIC signatures were derived from each of the 21 tissues by constructing a GCN for each (**Figure 4.3 & Table S4.3**).

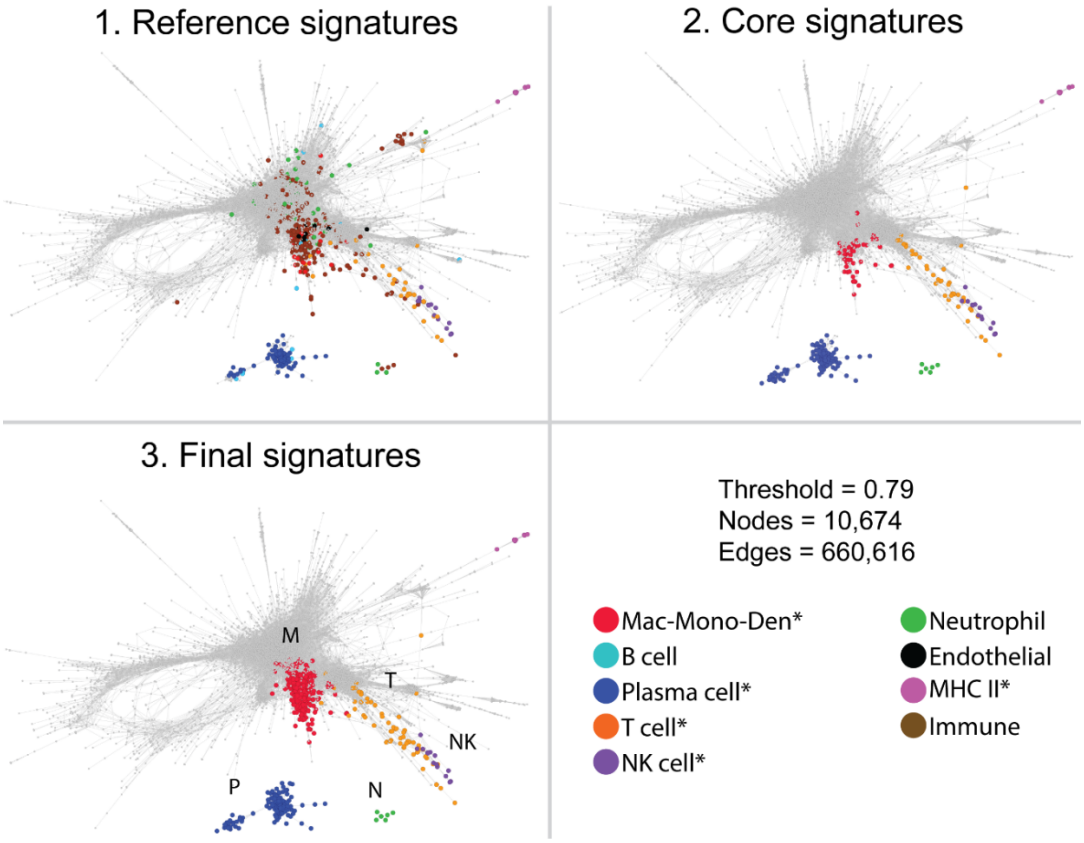
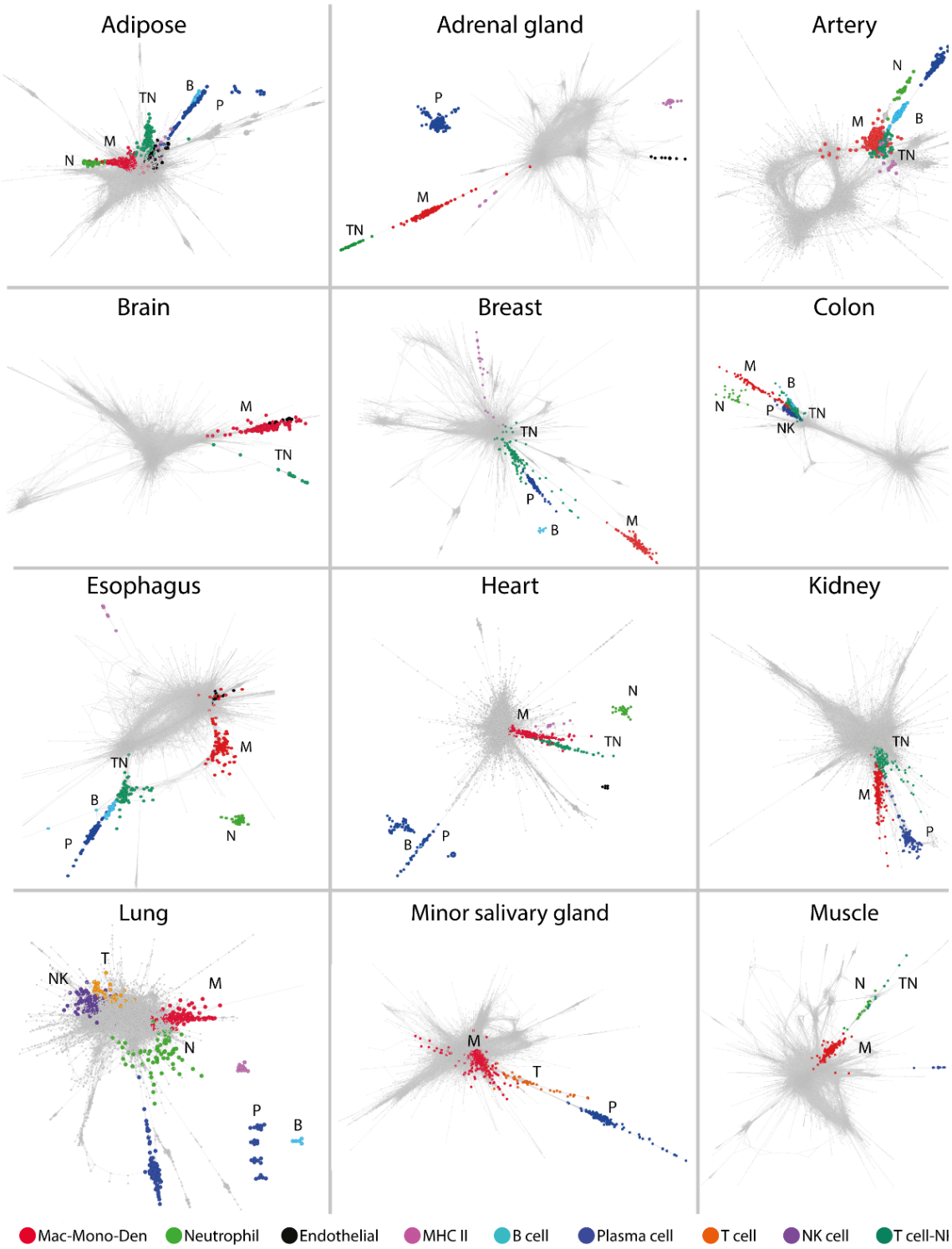


Figure 4.2. Derivation strategy of tissue-resident immune cell signatures from individual human tissues shown for liver.

The three steps used to identify immune cell signatures for a given tissue. 1) A GCN was constructed for each tissue with correlation threshold set to keep ~15,000 genes within the network. In certain tissues, where signatures of different immune cell types could not be easily discerned, GCNs with ~10,000 genes were constructed. All graphs were clustered (inflation = 1.7) and clusters examined for their enrichment of RIS. Gene clusters annotated for a certain biology have all genes coloured based on the bottom right panel. **2**) To identify which immune cells were present in the tissue a set of confident core clusters were selected for each cell type based on their enrichment and cell type marker genes. **3**) The final annotation included clusters neighbouring the core cluster, which were also enriched in RIS or were not connected to any other cluster.


Adipose
TN
B
P
M
N
Adrenal gland
P
M
TN
Artery
N
M
B
TN
Brain
M
TN
Breast
TN
P
B
M
Colon
M
B
N
P
TN
NK
Esophagus
TN
M
B
P
N
Heart
N
M
TN
B
P
Kidney
TN
M
P
Lung
T
NK
M
N
P
B
Minor salivary gland
M
T
P
Muscle
N
TN
M
Mac-Mono-Den
Neutrophil
Endothelial
MHC II
B cell
Plasma cell
T cell
NK cell
T cell-NI

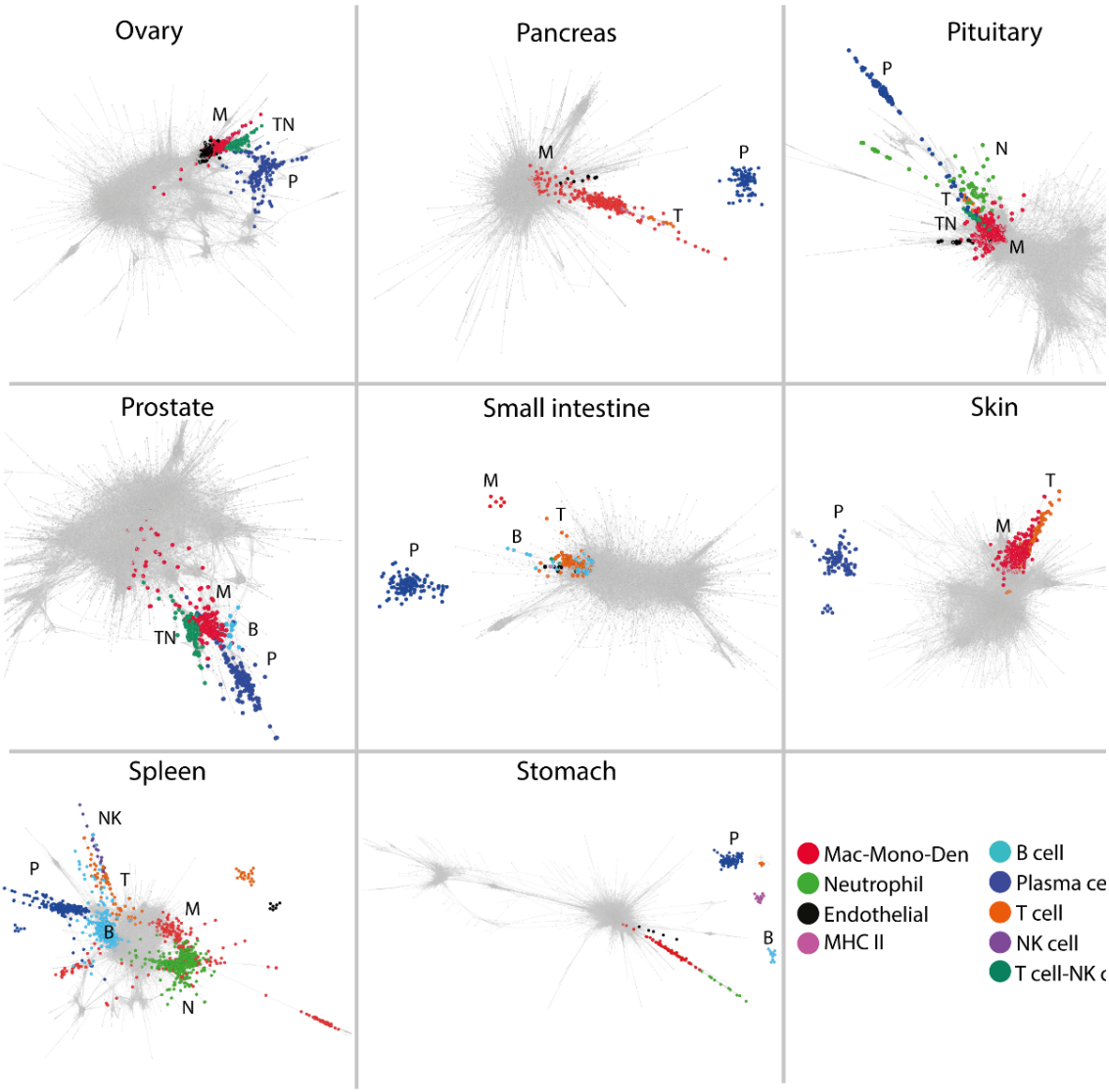


Figure 4.3. Immune cell signatures from human tissues.

TRIC signatures are highlighted across GCNs constructed for each of the 21 tissues from GTEx. Each panel consists of a GCN constructed from gene expression data of the respective tissue. Networks were constructed using a correlation threshold such that the resultant network included 15,000 genes, while for certain tissues (artery, liver, pancreas, pituitary, small intestine, spleen and stomach) as clusters associated with different cell types were not easily differentiated a higher correlation threshold was used to accommodate 10,000 genes in the network. Within the network of genes (grey) the derived immune cell signature genes are highlighted (coloured) and annotated. B: B cell, NK: natural killer cell, N: neutrophil, TN: T cell – NK cell, M: Mac-Mono-Den, P: plasma cell and T: T cell.

After the analysis, 91 signatures were identified across 21 tissues describing 7 TRIC annotations (Mac-Mono-Den, neutrophil, plasma cell, B cell, T cell, NK cell and T cell-NK cell). In majority of tissues, marker genes for T cells and NK cells co-clustered and were instead annotated as “T cell-NK cell”. Signature sizes varied from 4 genes including the B cell signature from the heart (*CD19, BLK, FCRL2* and *MS4A1*) to Mac-Mono-Den signature from the minor salivary gland containing 437 genes. Not all immune cell type signatures were identified across tissues. The Mac-Mono-Den and T cell-NK cell associated signatures were identified in all 21 tissues, followed by plasma cells (20 tissues), neutrophils (14 tissues) and B cells (11 tissues) (**Figure 4.4**). Certain tissues including colon, lung and spleen, contained all the six types of immune cell signatures (6 signatures). The least number of TRIC signatures were identified for the brain, this included two signatures, the Mac-Mono-Den and T cells-NK cells.

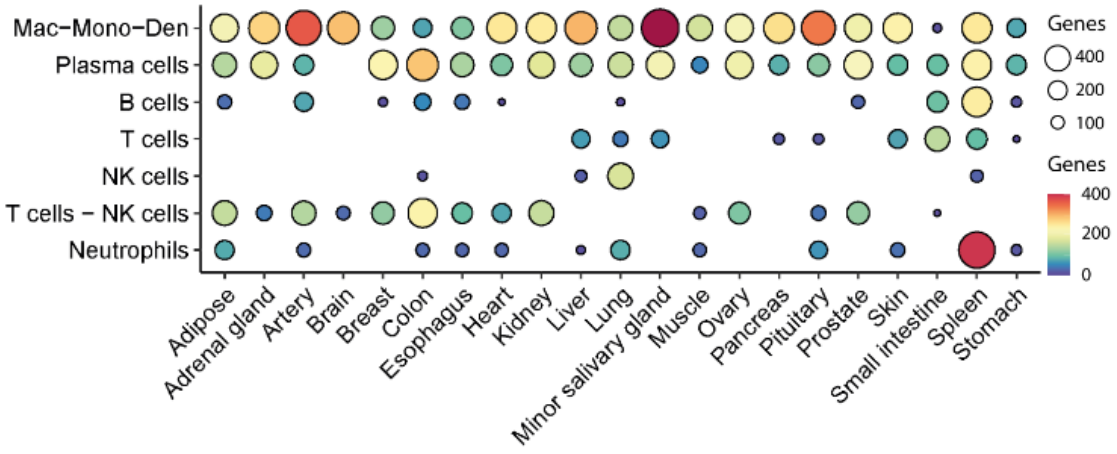


Figure 4.4. Immune cell signature sizes across tissues

The number of genes (size and colour) annotated for a given immune cell type (y axis) across the 21 tissues of the GTEx (x axis) (Lonsdale et al., 2013).

The overlapping biology between T and NK cells was also observed when comparing the number of genes overlapping between immune cell signatures across tissues (**Figure 4.5**). Signatures associated with T cell, NK cell and T cell–NK cell showed preferential overlap relative to other cell types as they had higher values (close to 1) of Jaccard index. Based on the same index Mac-Mono-Den largely shared signature genes with neutrophils and T cell-NK cells while B cell signatures overlapped most with those of T cell-NK cells. The number of genes unique to a certain cell classification varied from

30 genes identified for NK cells to 1,012 genes for Mac-Mono-Den. Across tissues plasma cell signature genes overlapped the least with any of the other signatures (93 overlapping genes), while Mac-Mono-Den signatures genes (393 genes) overlapped the most with other cell classifications.

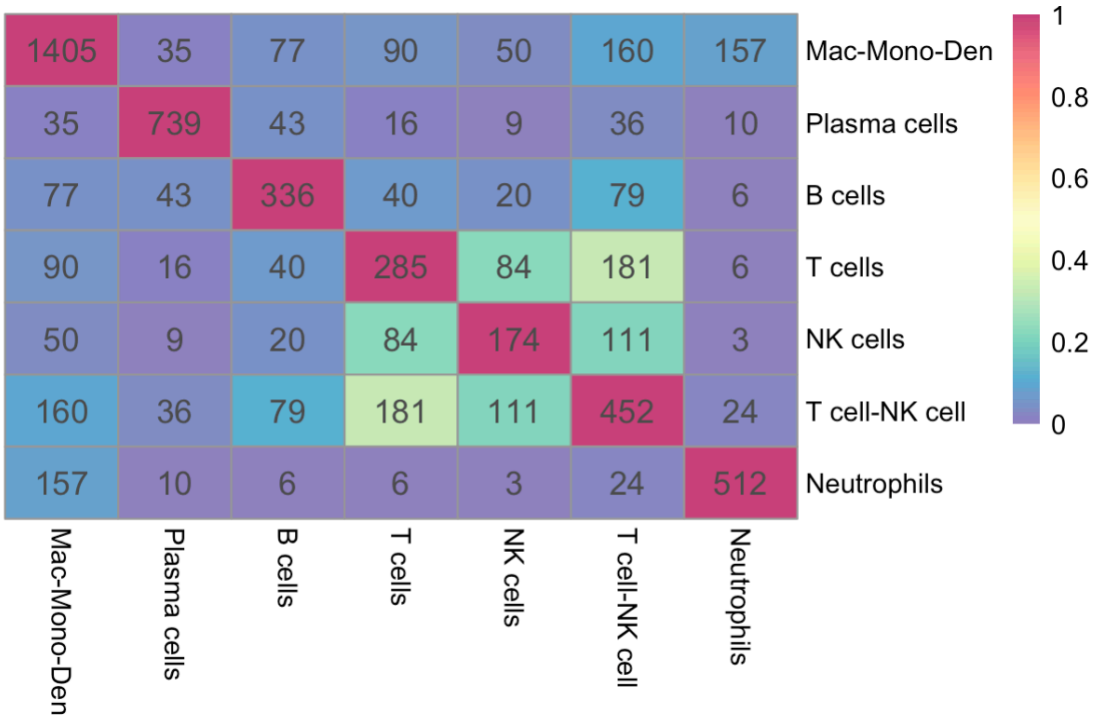


Figure 4.5. Gene overlap across tissue resident immune cell signatures.

The number of genes annotated for each of the seven immune cell signatures across 21 tissue transcriptomics datasets from the GTEx. The matrix shows the number of common genes between the different immune cell type signatures. Additionally, the colour of the cells represents a metric of comparing inter-signature overlap, that is the Jaccard index where a value of 1 (red) signifies perfect overlap between gene lists and that of 0 (purple) represents no overlap.

Next, we examined the genes associated with the different immune cell types and how strong this association was based on the number of times they associated with an immune cell signatures across the various tissues. This analysis revealed high confident core genes for cell types while also highlighting genes uniquely associated with an immune cell type from a certain tissue (**Table 4.2**). For example, four genes (*C1QA*, *C1QB*, *C1QC*, and *MS4A6A*) were found associated with Mac-Mono-Den in all 21 tissues, again highlighting the ubiquitous distribution of these cell types relative to others.

Other cell types had fewer signatures derived across the same tissues, indicative of their presence in limited number of tissues as was observed in **Figure 4.5B**. High confident genes or core genes of an immune cell types were considered as those which were associated with the cell types in more than half the number of tissues the cell type signature was derived in e.g. Mac-Mono-Den signatures were identified in all 21 tissues and accordingly the 124 genes associated with Mac-Mono-Den in 11 tissues or more were considered as high confidence. Interestingly, for a given cell type only a few core genes were identified across all the tissues. In contrast, several genes were associated with a cell type in a single tissue e.g. 142 genes were associated with NK cell signatures only in single tissues while 828 genes were associated with Mac-Mono-Den. Further examination of high confidence genes revealed known markers and gene families associated with the respective immune cell type (**Table 4.3**). High confidence genes for Mac-Mono-Den included those from gene families such as Fc fragment of IgG receptors (*FCER1G, FCGR1A, FCGR1B, FCGR2A* and *FCGR3A*), scavenger receptors (*CLEC7A, CD163, CD68, MSR1* and *STAB1*) and the complement system (*C1QA, C1QB, C1QC, C3AR1* and *C5AR1*) (Hirayama et al., 2017). The majority of genes associated with plasma cells were of the Immunoglobulin family while those for B cells included known markers (*CD19*, *CD79A* and *BLK*) and members of the Fc Receptor-Like (*FCRL1, FCRL2*, and *FCRLA*) (Capone et al., 2016). T cells were associated with CD3 molecules (*CD3D, CD3E* and *CD3G*) and associated proteins (*CD8A, ITK, ICOS,* and *LCK*) (Smith-Garvin et al., 2009). NK cell high confident genes included *EOMES*, which is a known transcription factor for the cell type (Simonetta et al., 2016). Additionally, genes included killer cell lectin like receptors (*KLRB1, KLRC2, KLRC3, KLRC4, KLRD1, KLRF1, KLRG1* and *KLRK1*) and granzymes (*GZMA, GZMH* and *GZMM*) associated with NK cells development and cytotoxic activity, respectively (Abel et al., 2018). A combination of high confident genes for T cells and NK cells was observed for the T cell-NK cell classification, suggesting that the classification did capture signatures of both cell types where it was difficult to separate them. Genes associated with neutrophils included chemotactic receptors like formyl peptide receptors and C-X-C motif chemokine receptor (Leoni et al., 2015) and L-selectin required for neutrophil migration towards sites of inflammation (Sheshachalam et al., 2014). Thus, the GCN approach was able to capture core immune cell type signatures from the 21 different tissues. In contrast, identifying TRIC associated genes was more challenging as it was difficult to segregate true TRIC associated genes from false positive and false negatives based on the GCN

analysis alone. Although, on examining certain known TRIC genes such as microglial genes, *CX3CR1* and *P2RY12* were observed in Mac-Mono-Den signatures of the brain.

Table 4.2 Frequency of genes associated with cell types across tissues

Number of tissues	Mac-Mono-Den	Plasma cells	B cells	T cells	NK cells	T cell-NK cell	Neutrophils
21	4	0	0	0	0	0	0
20	5	8	0	0	0	0	0
19	13	12	0	0	0	0	0
18	5	8	0	0	0	0	0
17	13	10	0	0	0	0	0
16	17	15	0	0	0	0	0
15	13	6	0	0	0	0	0
14	12	12	0	0	0	0	0
13	16	5	0	0	0	4	0
12	13	10	0	0	0	3	0
11	13	14	2	0	0	9	2
10	18	13	2	0	0	9	1
9	25	9	2	0	0	11	1
8	25	19	4	1	0	13	3
7	21	12	1	3	0	10	3
6	24	21	8	2	0	20	11
5	33	28	6	6	0	15	7
4	48	20	4	17	0	29	6
3	97	39	17	19	6	29	19
2	162	60	33	56	26	53	46
1	828	418	257	181	142	247	413

The table shows the number of genes which are associated with a certain immune signature (column 2 to 7) and over how many tissues (column 1: number of tissues). The colours represent the number of genes from green to red representing low to high counts.

Table 4.3 High confidence immune cell associated genes

Cell type	Genes
Mac-Mono-Den	ADAP2, AIF1, ALOX5AP, AOAH, APOBR, ARHGAP30, ARHGAP4, ARRB2, BIN2, BTK, C1orf162, C1QA, C1QB, C1QC, C3AR1, C5AR1, CCR1, CD163, CD180, CD300A, CD300C, CD33, CD37, CD4, CD53, CD68, CD84, CD86, CLEC7A, CORO1A, CSF1R, CTSS, CYBB, CYTH4, DOCK2, DOK2, DOK3, EBI3, ENSG00000268802.1, EVI2A, EVI2B, F13A1, FCER1G, FCGR1A, FCGR1B, FCGR2A, FCGR3A, FERMT3, FGD2, FOLR2, FPR3, GPR34, GPSM3, HAVCR2, HCK, HCLS1, HCST, HK3, IFI30, IGSF6, IL10RA, IRF5,

	ITGAM, ITGB2, LAIR1, LAPTM5, LAT2, LCP1, LILRA2, LILRA6, LILRB4, LINC01272, LRRC25, LST1, LY86, MNDA, MPEG1, MS4A4A, MS4A6A, MS4A7, MSR1, MYO1F, NCF2, NCF4, NCKAP1L, NFAM1, OSCAR, PARVG, PCED1B-AS1, PIK3R5, PLCB2, PLEK, PTAFR, RAC2, RASGRP4, RNASE6, RP11-1334A24.6, RP11-750H9.5, SASH3, SCIMP, SELPLG, SIGLEC1, SIGLEC7, SIGLEC9, SIRPB2, SLA, SLC11A1, SLC7A7, SLCO2B1, SNX20, SPI1, STAB1, TBXAS1, TFEC, THEMIS2, TLR2, TLR8, TMIGD3, TNFAIP8L2, TYROBP, VAV1, VSIG4, WAS, WDFY4
Plasma cells	ENSG00000211671.2, ENSG00000211936.2, ENSG00000211940.2, ENSG00000211953.2, ENSG00000231486.3, ENSG00000235896.2, ENSG00000253705.1, FCRL5, IGHA1, IGHA2, IGHG1, IGHG2, IGHG3, IGHGP, IGHJ1, IGHJ2, IGHJ3, IGHJ4, IGHJ5, IGHJ6, IGHM, IGHV1-14, IGHV1-18, IGHV1-2, IGHV1-24, IGHV1-3, IGHV1-46, IGHV1-67, IGHV1-69, IGHV2-26, IGHV2-5, IGHV2-70, IGHV3-11, IGHV3-13, IGHV3-15, IGHV3-21, IGHV3-23, IGHV3-33, IGHV3-43, IGHV3-48, IGHV3-49, IGHV3-53, IGHV3-6, IGHV3-60, IGHV3-66, IGHV3-7, IGHV3-71, IGHV3-72, IGHV3-74, IGHV3OR16-8, IGHV3OR16-9, IGHV4-28, IGHV4-31, IGHV4-34, IGHV4-39, IGHV4-55, IGHV4-59, IGHV4-61, IGHV5-51, IGHV6-1, IGKJ3, IGKJ4, IGKJ5, IGKV1-12, IGKV1-16, IGKV1-17, IGKV1-27, IGKV1-33, IGKV1-39, IGKV1-5, IGKV1-6, IGKV1-9, IGKV1D-33, IGKV1D-39, IGKV2-28, IGKV2-30, IGKV2D-28, IGKV3-11, IGKV3-15, IGKV3-20, IGKV3-7, IGKV3D-11, IGKV3D-15, IGKV3D-20, IGKV3D-7, IGKV4-1, IGKV7-3, IGLC2, IGLC3, IGLC6, IGLC7, IGLJ1, IGLJ2, IGLJ3, IGLL5, IGLV1-40, IGLV1-44, IGLV1-47, IGLV1-51, IGLV2-11, IGLV2-14, IGLV2-23, IGLV3-1, IGLV3-10, IGLV3-16, IGLV3-19, IGLV3-21, IGLV3-25, IGLV4-69, IGLV6-57, JCHAIN, MZB1, POU2AF1
B cells	BLK, CD19, CD79A, CLEC17A, CNR2, ENSG00000260655.1, FAM129C, FCRL1, FCRL2, FCRLA, MS4A1, P2RX5, PAX5, RP11-148O21.2, RP11-148O21.4, TBC1D27, TLR10, TNFRSF13B, TNFRSF13C
T cells	ACAP1, CARD11, CARMIL2, CD2, CD3D, CD3E, CD3G, CD40LG, CD6, CD8A, CD96, ENSG00000229164.5, GPR171, ICOS, ITGAL, ITK, LCK, MAP4K1, NLRC3, PCED1B-AS1, RASAL3, SH2D1A, SIRPG, SIT1, THEMIS, TIGIT, TMC8, TRAT1, TRBC2
NK cells	CD160, CD247, ENSG00000161570.4, EOMES, FASLG, FCRL6, GZMA, GZMH, GZMM, IL2RB, KLRB1, KLRC2, KLRC3, KLRC4, KLRD1, KLRF1, KLRG1, KLRK1, LINC00861, NKG7, PRF1, PTGDR, PYHIN1, S1PR5, SAMD3, SH2D1A, SH2D2A, SLA2, STAT4, TBX21, TRDC, ZAP70
T cell-NK cell	ACAP1, CARD11, CARMIL2, CD2, CD247, CD3D, CD3E, CD3G, CD48, CD5, CD6, CD69, CD8A, CD8B, CD96, CTSW, ENSG00000161570.4, ENSG00000264198.1, EOMES, GFI1, GPR171, GPR174, GZMA, GZMH, GZMK, GZMM, IL2RB, IL2RG, IL7R, ITGAL, ITK, KLRB1, KLRK1, LCK, LINC00861, LY9, MAP4K1, NKG7, P2RY10, PYHIN1, RASAL3, RUNX3, SCML4, SH2D1A, SIRPG, SIT1, SLA2, SLAMF6, SYTL1, TBC1D10C, TBX21, THEMIS, TIGIT, TMC8, TRAF3IP3, TRAT1, TRBC2, UBASH3A, ZAP70

Neutrophils	AC007278.3, ADGRG3, AQP9, BCL2A1, CLEC4D, CSF3R, CTB-61M7.2, CXCR1, CXCR2, FCAR, FCGR3B, FPR1, FPR2, IL18RAP, LILRA5, MCEMP1, MEFV, NFE2, S100A12, SELL, VNN2

Each of the rows represent immune cell annotations for which gene signatures were derived from 21 human tissues. For each annotation column 2 shows all the high confident genes which were repeatedly associated with the corresponding cell type in more than 50% of the tissues the cell type was identified in.

4.3.3 Comparing tissue-resident immune cells from mouse and human

To try and help identify potential TRIC signatures and compare them with mouse, TRIC signatures from human (GTEx) were compared with the corresponding TRIC signatures from mouse (ImmGen and Tabula Muris) (**Table S4.4**). For each GTEx gene cluster (GTExGC) we obtained overlapping gene clusters from mouse (ImmGen and Tabula Muris) (**Table 4.4**). Three main observations were made from these analyses. First, 39 GTExGCs matched gene clusters from mouse with respect to their cell type and tissue. These clusters were associated with twelve TRICs classifications in the GTEx including six gene clusters associated with Mac-Mono-Den, four for T cell-NK cell and two for B cells. Second, unlike the TRIC annotation from the mouse datasets, those from the human were much broader and majority of the GTEx gene clusters annotated as Mac-Mono-Den overlapped with tissue-resident macrophage signatures from mouse, and those annotated as T cell-NK cells or T cells overlapped with mouse tissue-resident regulatory T cell signatures. Third, on closer investigation of the genes, we identified both known marker genes for B cells (*SPIB*) and Tregs (*CCR4*, *CTLA4*, *GATA3* and *TIGIT*). In addition, seven of the TRICs had some of their signature genes supported by literature in representing those TRICs e.g. *CX3CR1*, *P2RY12* and *TREM2* for microglia. Other examples of tissue-specific genes included, *BLVRA*, *BLVRB*, *HMOX1*, *SCL48A1* and *VCAM1* for splenic macrophages, which help in regulating the cell's inflammatory phenotype (Wegiel and Otterbein, 2012), function in maintaining iron homeostasis (Sukhbaatar and Weichhart, 2018) and migration (Dutta et al., 2015). The literature also supported the role of *CR2* and *DTX1* for splenic B cells (Descatoire et al., 2014); *C5AR1* (Wiese et al., 2017) and *FOLR2* (Schyns et al., 2019) for oesophageal macrophages; *MCOLN1* (associated with adipose tissue) (Wiklund et al., 2016) and *CLEC10A* (Hill et al., 2014) for adipose macrophages; *CCR4* (aids in Treg migration and establishment in the tissue)(Guo et al., 2008) for colonic Tregs. A GTEx gene cluster annotated for Kupffer

cells (cluster 28) was enriched in ImmGen gene cluster 44 defined by a high expression in Kupffer cells.

Table 4.4 Overlap between human and mouse tissue-resident immune cells.

GTEx			Datasets		P value	Gene ID
Cluster	Tissue	Annotation	Dataset	Dataset: (Cluster Annotation)		
14	Adipose	Monocyte/Macrophage	ImmGen	154_General [Macrophage: Mesenteric sheet == Macrophage: Alveolar CD45+ CD11blo CD64+ CD11c+ SiglecF+ >> Macrophage Fat]	0.106	MCOLN1
14	Adipose	Monocyte/Macrophage	ImmGen	34_Specific [Macrophage: Aortic == Macrophage: Fat >> Macrophage: Alveolar Cd11b+ Cd64+ Ly6c- CD206+ MHCII+]	0.086	CLEC10A/FCGRT
17	Adipose	T cells_NK cells	TM	41_Innate lymphocyte (type LP): C38, Fat	0.050	SYTL1
17	Adipose	T cells_NK cells	TM	43_NK T cell: C16, Fat	0.043	TMC8
17	Adipose	T cells_NK cells	TM	57_Treg: C41, Fat	0.033	TIGIT
169	Brain	Monocyte/Macrophage	TM	10_Microglia	0.007	P2RY12/P2RY13
7	Brain	Monocyte/Macrophage	TM	10_Microglia	0.030	GPR34/LHFPL2/SLC2A5/TMEM119/TREM2
7	Brain	Monocyte/Macrophage	ImmGen	119_Specific [Macrophage: Microglia >> Macrophage: Peritoneal]	0.126	PLXDC2
169	Brain	Monocyte/Macrophage	TM	47_Microglia	0.024	CX3CR1
7	Brain	Monocyte/Macrophage	TM	47_Microglia	0.032	LPCAT2/PLXDC2
169	Brain	Monocyte/Macrophage	ImmGen	6_Specific [Macrophage: Microglia]	0.001	CX3CR1/P2RY12/P2RY13
7	Brain	Monocyte/Macrophage	ImmGen	6_Specific [Macrophage: Microglia]	0.005	GPR34/HPGDS/LHFPL2/LPCAT2/SELPLG/SLC2A5/SLCO2B1/SUSD3/TMEM119/TREM2
8	Colon	T cells_NK cells	ImmGen	113_Specific [Treg: Colon == Innate lymphocyte: Colon] >> General [Innate lymphocyte == Treg == Epithelial cell]	0.060	CCR4
8	Colon	T cells_NK cells	ImmGen	26_Specific [Innate lymphocyte: Small intestine >> Treg: Colon] >> General [Innate lymphocyte]	0.433	GATA3
8	Colon	T cells_NK cells	ImmGen	50_Specific [Treg: Colon] >> General [Treg == NK cell == Types of Dendritic cells]	0.001	CTLA4/TIGIT/TNFRSF4
31	Esophagus	B cells	TM	31_B cell: C23, Lung	0.063	PAX5
16	Esophagus	Monocyte/Macrophage	ImmGen	142_Specific [Macrophage: Aorta >> Macrophage: Lung CD11b+ CD64+ Ly6C- CD206+ MHCII-]	0.110	FOLR2
16	Esophagus	Monocyte/Macrophage	ImmGen	161_Specific [Macrophage: Lung CD27+ SiglecF+ >> Macrophage: Spleen == Macrophage: Kupffer cell] >> General [Monocyte == Macrophage == Fibroblast == Endothelial cell]	0.092	MYO7A

16	Esophagus	Monocyte/Macrophage	TM	17_Macrophage: C6 , Trachea	0.618	TNFSF13
16	Esophagus	Monocyte/Macrophage	ImmGen	23_Specific [Macrophage: Alveolar Cd11c+ SiglecF+]	0.626	RNASE2
16	Esophagus	Monocyte/Macrophage	TM	28_Macrophage: C29, Trachea	0.385	TLR2
16	Esophagus	Monocyte/Macrophage	ImmGen	34_Specific [Macrophage: Aortic == Macrophage: Fat >> Macrophage: Alveolar Cd11b+ Cd64+ Ly6c- CD206+ MHCII+]	0.013	C5AR1/CLEC10A/STAB1
398	Kidney	Monocyte/Macrophage	TM	64_Macrophage: C29, Kidney	0.000	LILRA6/LILRB3
5	Kidney	Monocyte/Macrophage	TM	64_Macrophage: C29, Kidney	0.007	LILRA2/LILRB4/LILRB5
6	Liver	Monocyte/Macrophage	ImmGen	25_Specific [Dendritic cell: Liver preDC >> Dendritic cell: Plasmacytoid] >> General [Dendritic cell: Bone Marrow, Dendritic cell: Spleen]	0.373	CD300C/TMEM229B
28	Liver	Monocyte/Macrophage	ImmGen	44_Specific [Macrophage: liver >> Endothelial cell: CD45- CD31+ PDPN+] >> General [Endothelial cell == Kupffer cell]	0.000	CLEC4G/EHD3/GPR182/OIT3/STAB2
1	Small intestine	T cells	ImmGen	50_Specific [Treg: Colon] >> General [Treg == NK cell == Types of Dendritic cells]	0.054	TIGIT
256	Spleen	B cells	TM	19_B cell: C43: Spleen	0.030	POLR1E
8	Spleen	B cells	TM	30_B cell Spleen	0.001	ARHGAP24/CR2/DTX1
8	Spleen	B cells	ImmGen	38_Specific [B cell: Germinal centre]	0.218	MBD4
110	Spleen	B cells	ImmGen	97_Specific [B cell: Fr E. >> B cell: Splenic memory == B cell: Splenic T1] >> General [B cell == Dendritic cell: PreDC == Dendritic cell: Plasmacytoid]	0.007	SPIB
400	Spleen	Monocyte/Macrophage	ImmGen	108_Specific [Macrophage: Mesenteric == Macrophage: Spleen]	0.011	BLVRB
105	Spleen	Monocyte/Macrophage	ImmGen	129_Specific [Macrophage: Spleen >> Macrophage: Thymus 64+ == Macrophage: Lung CD27+ SiglecF+ >> Macrophage: Fat]	0.018	PLD3
105	Spleen	Monocyte/Macrophage	ImmGen	161_Specific [Macrophage: Lung CD27+ SiglecF+ >> Macrophage: Spleen == Macrophage: Kupffer cell] >> General [Monocyte == Macrophage == Fibroblast == Endothelial cell]	0.015	BLVRA
400	Spleen	Monocyte/Macrophage	ImmGen	28_Specific [Macrophage: Spleen]	0.002	HMOX1/SLC48A1
125	Spleen	Monocyte/Macrophage	ImmGen	28_Specific [Macrophage: Spleen]	0.029	CREG1
139	Spleen	Monocyte/Macrophage	ImmGen	28_Specific [Macrophage: Spleen]	0.084	VCAM1
118	Spleen	T cells	TM	27_Tregs: C41, Spleen >> Tregs: C41, Fat	0.015	CTLA4
59	Spleen	T cells	TM	27_Tregs: C41, Spleen >> Tregs: C41, Fat	0.015	MATK

Shown are the 39 GTExGCs which overlap with gene clusters from mouse (Tabula Muris and ImmGen dataset) and are also associated with the same cell types and tissues from those derived from mouse. The first three columns refer to the GTExGC, their cluster number, the tissue they

were derived from and the cell type they represent. The remaining columns refer to gene clusters from mouse and which were also enriched for certain GTExGCs. Included are the mouse gene clusters, their annotation, *P value* and the overlapping gene. Significant *P values* < 0.05 are in green.

4.4 Discussion

To understand the functions of TRICs in maintaining tissue homeostasis in health and death, in chapter 3 TRIC signatures were derived from two data sources derived from mouse tissue, the ImmGen (bulk RNA-Seq of known FACS isolated cell types) and Tabula Muris (scRNA-seq data of cells from different tissues). In this chapter, we sought to extend these efforts to human tissues by utilising the extensive GTEx resource. GCN analysis was used to derive TRIC signature from each of the 21 tissues considered, and subsequently they were compared with those derived from mouse (in chapter 3), for validation and to estimate the conservation of TRIC signatures across species.

To derive TRIC signatures each tissue was examined for coexpressing genes using GCNs, however for this a reference immune signature was required to determine which gene clusters were representative of immune cell types. A recent publication showed the importance of deriving immune signatures directly from tissues so as to identify immune cell types (Nirmal et al., 2018). This set of signatures, called ImSig, was in contrast to blood-derived immune signatures for deconvoluting immune cell types from tissues, e.g. CIBERSORT (Chen et al., 2018). Although, ImSig was a great resource for immune cell derivation from human tissue, the approach was directed towards identifying immune cell signatures under pathological conditions especially focussing on different types of cancers. Therefore, in order to define TRIC signatures from homeostatic tissues, a set of RIS were derived from human tissue samples having no known pathology. For this analysis the GTEx resource was used, a comprehensive human tissue atlas of gene expression data. This extensive dataset was first downsampled to attain an equal number of samples from different regions, thus balancing the correlation space for GCN analysis. Such an approach would give equal weightage to tissues for the correlations calculated between genes. After downsampling the dataset, RISs were derived using GCN analysis where gene clusters were annotated for a given cell type based on their enrichment for ImSig signatures. The RIS signatures defined broad immune cell classifications, including neutrophils, B cells, plasma cells, T cells, NK cell and in the case of macrophages, monocyte and DCs where no cell type specific clusters could be

identified, associated gene clusters were classified as ‘Mac-Mono-Den’. Altogether, the RIS represented the core signatures of these broadly defined cell classifications that are conserved across the 28 tissues of GTEx.

Using the RIS, TRIC signatures were defined for each of the 21 selected tissues of the GTEx. The RIS were used to annotate gene clusters generated from the GCNs for each tissue. For constructing GCNs an appropriate threshold was applied to retain 15,000 genes to make GCNs and gene clusters comparable across tissue. Alternatively, disparities in GCN sizes could lead to numerous small clusters in GCNs with few genes due to the sparsity of the network while bigger but fewer clusters would be present in larger GCNs. To annotate tissue derived gene clusters for immune cell types, first, core gene clusters were identified based on the enrichment of RIS and cell types makers. Assuming that not all immune cell types are present in every tissue, the core clusters revealed the presence or absence of cell types within each tissue. This step was followed by incorporating neighbouring gene clusters enriched in RIS genes, with that of core clusters. In majority of tissues distinct signatures were obtained for each cell type. However, similar to how *ImSig* genes associated with macrophages, monocytes and DCs co-clustered while deriving RIS, on deriving TRIC signatures from individual tissues, certain tissues had markers genes associated with T cells and NK cells co-cluster. Hence, in certain cases it was difficult to identify gene signatures uniquely defining a single cell type. This is likely due to the amalgamation of biological signals in bulk RNA-Seq data, also touched upon in chapter 2. Signals in tissue transcriptomics data are dependent on the sample’s cell composition, and their expression levels of genes. An ideal GCN analysis to identify cell signatures would have cell types vary in different proportions across samples with a high expression of marker genes, likely the case for the colon, lung and spleen where all cell types (those considered in the RIS) are identified with distinct signatures. In contrast, there are two possibilities where marker genes for different cell types may co-cluster when deriving cell signature from bulk RNA-Seq data using GCN analysis. First, if different cell type associated genes would coexpress with one another (as a function of cell type proportion and expression level of their signature genes) across samples, they would appear to represent a single biology when using GCN analysis. Second, if the signal of a certain cell type is low relative to other cell types within the tissue sample, associated genes of the cell type may co-cluster with genes associated with similar cell types. These reasons possibly contribute to the combined signatures observed for “T cell-NK cell” and “Mac-Mono-Den”. In certain cases where

genes have similar coexpression patterns but are associated with different cell types, one can adjust the correlation threshold to construct GCNs to help segregate these genes into separate clusters. This was observed in certain tissues where genes for Mac-Mono-Den co-clustered with those of neutrophils and T cells. In such cases, a higher correlation threshold was used to construct the GCN for a tissue thus segregating immune cell types signatures. Therefore, even though the number of genes was kept consistent for each tissue GCN, the degree to which TRIC associated coexpression patterns differed from one another varied across tissues. This variation is likely a function of the cellular composition of the tissue. Thus, constructing GCNs based on known biology would be appropriate as genes from each tissue would have different distributions of correlation values dependent on the cellular composition of the tissue. Altogether, the aforementioned procedure with the use of RIS could be used to identify immune cell signatures across other bulk RNA-Seq datasets derived from healthy human tissues.

The derived signatures were then studied for their distribution across tissues. From the brain T cell-NK cells and microglial signatures were derived. This observation was supported by the fact that majority of immune cells in the brain are microglia (80% of immune cell types), second to which are various types of T cells (Korin et al., 2017). Although other immune cell types have been found in the brain through FACS, due to their lower cell abundance, and accordingly fewer reads relative to other brain cell types, their signatures would not be captured using GCN analysis. In other tissues all 6 immune cell types were derived, and for these tissues mass cytometry studies have shown their respective cell numbers to represent approximately more than 10% of cells within the tissues, this includes the colon (van Unen et al., 2016), lung (Hanidziar et al., 2020) and spleen (Goltsev et al., 2018). Hence, the analysis is able to capture the presence/absence of immune cells across tissues, although differentiating amongst certain signatures was challenging, e.g. for myeloid cell types as well as T cells and NK cells. The TRIC signatures were compared with those in mouse (ImmGen and Tabula Muris TRIC signatures derived in chapter 3) using enrichment analysis. The mouse signatures provided a higher resolution of cell classifications many of which were supported through predefined markers from the ImmGen. 39 gene clusters overlapped across similar TRICs signatures from mouse to human which described 12 TRIC populations in human. Interestingly, for broad cell classifications such as Mac-Mono-Den, associated gene clusters mainly overlapped with macrophage signatures from the

mouse, rather than monocytes or DCs. A similar observation was seen in the T cell-NK cell classification in human which pointed to Treg gene signatures in mouse. This could suggest that these cell types were abundant, highly expressed marker genes, or expressed a unique set of genes in the context of the tissue. There has been an increased appreciation for tissue-resident Tregs in maintaining tissue homeostasis and as a therapeutic target (Lu et al., 2015, Sharma and Rudra, 2018, Zhou et al., 2015). Through literature mining supporting genes were found for seven of the TRIC populations, including macrophages (brain, liver, spleen, oesophagus and adipose), B cells (splenic) and Tregs (colonic). Thus, the analysis revealed TRIC associated genes conserved across species, however few TRIC signatures from human were enriched for those from mouse. One reason could be the lack of conservation of TRICs across species. A more likely scenario is that TRICs would likely share genes with other tissue-resident cells to maintain tissue homeostasis and GCN analysis would capture associated gene modules, however, be unable to differentiate between the cell types utilising the pathways. Hence, GCN analysis enables the derivation of different immune cell signatures across tissues, revealing core genes associated with these cell types. In contrast, identifying and supporting genes associated with tissue homeostasis was more challenging, because such genes would need to be uniquely expressed by the cell type or be abundantly expressed (a combination of higher cell abundance or expression of the gene) to be captured by GCN analysis. Furthermore, whilst certain signature genes were supported by literature, an indication of true positive TRIC associated genes, further experimental validation would be required to separate them from the false positive genes. In such cases where we study functional gene modules shared across cell types their derivation from scRNA-Seq may be more informative as it would describe which cells utilise those modules and how highly they are expressed within them.

In summary, human TRIC signatures have been characterised across 21 tissues of the GTEx resource and have compared them with those from the mouse. In the process 1) gene signatures associated with 6 immune cell types were derived under physiological conditions, referred to as RIS. 2) Furthermore, with the methodologies adopted here in these RIS were further used to derive TRIC signatures from bulk RNA-Seq data derived from individual tissues of the GTEx. Using the RIS, 91 TRIC signatures were derived from the 21 tissues through GCN analysis. 3) These signatures comprised of known markers which were conserved across tissues for the respective cell type. 4) Immune cell signatures overlapped across similar cell types e.g. lymphoid immune cell signatures

shared several genes. 5) Not all immune cell types had signatures in all tissues, and their presence/absence across tissues. 6) 39 signatures mapped between common TRICs across mouse (ImmGen and Tabula Muris) and human. 7) Of these, seven signatures included genes supported by literature in representing the respective cell type. In future studies one would hope to conduct similar analyses for deriving TRIC signature with the power and resolution of scRNA-Seq data.

5. Chapter 5: Conclusions

In this thesis TRIC gene signatures from mouse and human were investigated using transcriptomic data from different sources and platform. Through these signatures the biology of TRICs and their diversity was investigated. In the process, the advantages and disadvantages of using different transcriptomic technologies for cell signature derivation were discussed, while also developing novel approaches to analyse single cell transcriptomics data. Sequentially, through the chapters various aspects of TRIC biology are studied at different resolutions (tissue or cell) and species.

This work began with the investigation of microglia, following on from the work of the Freeman and McColl labs (Grabert et al., 2016, Vincenti et al., 2016) and their relevance in to mouse models of mimicking the pathology of Alzheimer's disease. This work showed the complexity of neurodegenerative models where microglia displayed regional heterogeneity in their numbers and activation state. The following questions on microglia were examined in chapter 2:

1. Can we define an expression signature for microglia in the context of human brain tissue?
2. Does this signature show any quantitative or qualitative differences based on the of brain?
3. How can the signature be used in studying microglia behaviour (cell abundance and activation state) in the context of aging and neurodegenerative disease?

Initially, existing microglial expression signatures were examined, however, little consistency was observed between these signatures both in mouse and human. It was argued that this could be due to the different transcriptomic platforms used, donor differences, sample types (tissues and cells), brain regions sampled across studies or the methods used to derive them. The reality is that it is probably a combination of all factors. To see if we could do better, nine transcriptomic datasets were used spanning different platforms (microarray and RNA-Seq), 427 brain regions, 220 donors and sample types (pooled cells and brain tissue). GCN analysis of each dataset produced microglial signatures for each from which a core human microglia signature was generated with the 298 genes which overlapped across more than two of the dataset derived signatures. Signature genes were validated through various public resources, examining staining of their encoded proteins across brain regions (from the HPA), higher average expression in myeloid cell types relative to other immune and CNS cell types

(primary cell atlas) (Mabbott et al., 2013), and the correlation between their average expression with microglial density across regions (Lawson et al., 1990). The signature was then used to determine microglial response in different brain regions during ageing and in Alzheimer's disease. In relation to ageing, the proportion of microglia as determined by the average expression of signature genes correlated with the distribution of microglial inflammation and neuronal loss across the brain regions examined, while for Alzheimer's it correlated with neuroinflammation, tau burden, and neuronal loss. Furthermore, the analysis of the microglial transcriptomic profile in Alzheimer disease revealed the downregulation of homeostatic genes and upregulation of neuroinflammatory genes as well as genes from the TYROBP-TREM2 pathway implicated in the disease. A similar phenotype was observed in a recent snRNA-Seq study that examined brain cells from 48 donors balanced for donors with no neuropathology and those with Alzheimer's disease (Mathys et al., 2019). However, this study also revealed that human microglia are not entirely similar to mouse disease associated microglia and instead present an ageing like profile. An example of this is *TREM2* which is upregulated in mouse neurodegenerative models while remaining unchanged in human Alzheimer's disease (Keren-Shaul et al., 2017). We too observed a similar profile for *TREM2*, however further analyses are required to confirm these findings as several studies have also shown how current single cell technologies may still lack the sensitivity of detecting certain genes (Thrupp et al., 2020). In the Alzheimer's study, Mathys et al. also voiced concerns for the small number of cells examined in their study. Hence, even though an ideal experiment for chapter 2 would be to analyse snRNA-Seq data derived from cells of different brain regions under physiological and neurodegenerative conditions, from a technical standpoint the current technologies still require further improvements in their sensitivity to capturing and quantifying reads, as well as their throughput as having more cells aids downstream analyses. This will become possible in the future with the continuous development of single cell omics, e.g. SmartSeq3 (Hagemann-Jensen et al., 2020). In certain cases, bulk RNA-Seq data may be a better choice over (more cost effective and sensitivity towards reads) scRNA-Seq when deriving a core signature as has been conducted for microglia which have a distinct transcriptional program relative to other CNS cells across brain regions. The core human microglial signature derived by this work has already been utilised in characterising the microglial response in Huntington's disease (Al-Dalahmah et al., 2020). Where scRNA-Seq takes the upper hand is in tracking the various subpopulations for that cell type,

which in bulk RNA-Seq would amalgamate into a single signature, i.e. appear as single cell type if they are not sufficiently different based on their transcriptome. The field of microglial biology continues to expand, studying the many phenotypes of microglia. For example, the use of CyTOF to describe homeostatic microglia across five brain regions (Böttcher et al., 2019). Alternatively, to study the activation states of microglia researchers have been generating and exploring neurodegenerative mouse models (Masuda et al., 2020) while others have compared these states across species (Geirsdottir et al., 2019). As we expand on the number of scRNA-Seq datasets derived from a combination of different models, stimuli, species and diseases we would essentially possess a host of cell perturbations from which to deconstruct their gene regulatory network ultimately enabling in silico perturbations to the regulatory network as has been proposed (Lotfollahi et al., 2018). Whilst, studying the brain transcriptome, one other gene module caught my interest and this led to recent publication of a signature for motile cilia, as seen in the brain associated with ependymal cells (Patir et al., 2020).

The microglial signature from chapter 2 highlighted the potential use of gene signatures and the importance of studying homeostatic TRIC signatures. This work was expanded upon in chapter 3 where several immune cell types including various TRIC types were examined based on their transcriptomic profile. In this chapter the following questions were addressed:

1. How similar are immune cell types with respect to their lineage and tissue of residence?
2. To what extent does tissue determine the diversity of cell types?
3. What are the signatures associated with different TRICs?
4. How well do the signatures between marker defined cell types compare with cell types defined through scRNA-Seq?

Three datasets aided in addressing these questions. Two of the datasets were from the ImmGen resource where samples were derived from pooled cells of FACS isolated immune cells. Here, cell types were predefined based on markers. In contrast the third dataset, the Tabula Muris comprised of scRNA-Seq data derived from different mouse tissues, from which immune cell types, as annotated by the authors, were considered. Generated by what might be considered a ‘biased’ approach, the ImmGen datasets revealed lymphoid cell types to be largely similar to one another while for the myeloid system cell types were more heterogeneous including the different types of tissue-

resident macrophages (alveolar, microglial, splenic and peritoneal). Hence, at least in this dataset, the sample-to-sample network suggested that cell types within a lineage are very similar (interconnected within a single graph component) and that tissue-resident populations, mostly with respect to macrophages are more distinct. One would assume that based on markers, replicates for the same cell type would form separate clusters. In contrast, a continuity was observed from one cell type to the other. This lack of distinction between certain cell types was also observed in the Tabula Muris, where different cell types were part of a continuum of states, e.g. aortic macrophages clustered together however they were still part of the macrophage component with no obvious boundaries distinguishing them from other tissue-resident macrophages, hence requiring approaches such as graph-based clustering to distinguish them. 156 coexpressing gene modules were found associated with a cell type or groups of cell types, ranging from cell lineages to TRICs. Several signatures associated with TRICs contained known markers and transcription factors. For the unbiased analysis the SmartSeq2 derived scRNA-Seq data from Tabula Muris was used as it is known to have a better-read coverage of genes with fewer dropouts than the 10x Chromium system, thereby improving downstream analyses. Furthermore, there is also the issue of the non-uniformity in isolating cells and the susceptibility of certain cells which likely explain the absence of certain obvious cell types, e.g. splenic macrophages in Tabula Muris. Improving such protocols would make single cell omics data more closely resemble the tissue composition and improve read capture sensitivity. Analysis of scRNA-Seq data from Tabula Muris found the similarity/differences between cell types matching those observed from the marker-based ImmGen analysis, i.e. lymphoid cell types showed great similarity in comparison to the more heterogenous myeloid cell types, especially macrophages. These observations suggest an agreement between cell types identified by unbiased approaches such as scRNA-Seq and biased approaches like those defined by markers, highlighting how far the field has come in defining cell types. However, a more refined evaluation of marker-based cell types would involve the investigation of the marker-defined cell types using scRNA-Seq as has been recently examined for DCs and monocytes (Villani et al., 2017).

Forty-four gene clusters were identified associated with different cell classifications, several of which mapped to ImmGen gene clusters for the same cell type. For both ImmGen and Tabula Muris, the majority of the cell types were defined by their tissue of origin, as opposed to a cell subtype (a cell type present in several tissues), e.g. the

different ILC subsets. Furthermore, both datasets showed macrophages to be the most heterogeneous, for which associated signature genes found were supported by literature for similar TRICs. This would suggest that these gene signatures closely represent the TRICs identified here in, however, a further evaluation based on known cell types would help support the analysis, especially for the other TRICs which have not been as thoroughly examined as tissue-resident macrophages. The annotations of cell types and their corresponding gene clusters are a crucial part of omic analyses, as has been conducted for the Tabula Muris dataset using the immune cell signatures derived from ImmGen. The use of the ImmGen signatures in annotating Tabula Muris gene clusters highlights the importance of the context in which the reference signatures are derived to annotate either the gene or cell clusters of the query dataset. For example, the ImmGen signatures were a perfect reference for annotating the Tabula Muris gene clusters as both datasets shared similar homeostatic immune cell types and were derived in the context of other immune cell types. In contrast, if either of the datasets also included non-immune cell types, the resultant immune cell signatures (derived in the context of immune and non-immune cell types) would represent biological differences between immune and non-immune cells rather than TRIC biology. Current signature databases define cell type signatures across various tissues and conditions; however, they do not account for the variation in signatures due to the context they are derived from, i.e. the types of cells/tissues they are derived from. Hence, a flexibility is needed in such databases to define the context of the signature which is especially important in studies where cell types are enriched for.

Researchers are continuing to generate cell type signatures to better understand the biology of cells and for annotating unknown cells by aggregating markers from the various stages of development and across human tissues (Zhang et al., 2018c). Unfortunately, even with these numerous variables, the current nomenclature or definition of cell types has still been restricted to a few markers as is evident from the ImmGen resource. Even single cell studies which describe cell biology at the highest resolution using tens of thousands of variables still name cell types based on 2-3 markers (Villani et al., 2017). Due to a lack of a comprehensive nomenclature, characterising novel cell types is done through manual scanning of literature of markers/signatures, enrichment of signatures from databases and comparison with omic profiles of previously identified cell types. This was especially an issue in chapter 3 as one of the future analyses for the Tabula Muris dataset would be to compare cells based on known cell

types. A structured nomenclature would certainly facilitate cell annotations through comparisons across studies. Such a naming system would have to be flexible, able to accommodate several variables, potentially include even those variables which are to come through future technologies. Such a cell ontology would help annotate novel cell types and place them within the context of existing cell types simply based on the naming scheme. A potential answer to this could be the use of QR codes which can potentially accommodate more than 30,000 modules of information.

Through chapter 3 two methods were also developed. The first being, annotation of single cells based on their transcriptomic profile through comparison with the transcriptomic profiles of known cell types. Unlike other available approaches this process was done at the cell-cluster level so as to annotate groups of cells which reduced the chances of incorrect annotations of cells due to the influence of technical artefacts. 2) The second method looked at constructing GCNs from scRNA-Seq data, where reads across cells from the same cell cluster were averaged thereby constructing a balanced GCN where each cell type/cluster would have an equal weight for the correlation calculated between genes. For constructing GCNs, two factors were evaluated which could potentially influence the GCN, lowly expressed genes and the number of cells. These factors were considered by examining three approaches (Average, Subcell and Filter-Average). Whilst, all methods captured similar biology (gene clusters), the Filter-average approach comprised of fewer false positives as gene clusters unique to the approach were enriched for GO terms unlike the other methods. The approach involved averaging of reads across cells of a given cell type and filtering genes based on potential technical artefacts, e.g. dropouts and spikes in expression. Hence, removal of certain signals in gene expression helped reduce gene clusters due to technical factors and helped reduce the variation within the data, enabling the GCN to focus on inter-cell cluster variation. Although, several methods have been developed to capture regulatory networks in the single cell omics space such factors have not yet been considered. Whilst, methodologies are being developed using various benchmarks (Pratapa et al., 2020) the field is still lacking statistical approaches for such an analysis. Certain methods have tried to address this, although unpublished the Seurat package in R enables hierarchical clustering on which differential expression analysis can be done for grouping at the different levels of the dendrogram, comparing two branches/groupings at a time. Such an analysis would greatly help in differentiating differences due to the different levels of cell hierarchy, from cell lineage to cell subsets. Another way of imagining such

a hierarchy of cell types is as a mixture of cellular states, e.g. studies are also finding immune like phenotypes in structural tissue cells (Krausgruber et al., 2020). This area, however, requires further investigation to determine conserved state associated gene modules across cell types. Apart from the need to develop methods to determine gene regulatory networks, another factor is the experimental design. Currently we rely on disease models to study the breath of immune cell diversity, however, this also limits us to the states that are attained in these diseases. Therefore, to better describe all potential states a cell can attain, and the pathways used to generate them, perturbation studies such as scRNA-Seq in combination with CRISPR would prove immensely useful (Duan et al., 2019). This would allow researchers to perturb the gene regulatory network of a certain cell type which in combination with probabilistic models could help deconstruct causative networks.

In chapter 3, a comprehensive set of immune cell types were examined at the resolution of single cells, however, only in mouse. Hence, in chapter 4 this effort was extended to human. For this the GTEx dataset was considered as it is the largest and most diverse resource of human tissue transcriptomics data. In chapter 4, the following questions were investigated:

1. Can bulk transcriptomics data be used to capture immune cell signatures across tissues using GCN analysis? If so, to what resolution (lineage, cell types, subtypes or TRICs)?
2. What is the distribution (presence or absence) of immune cell types across tissues?
3. How do TRIC signatures compare across tissues, i.e. which cell types are more similar to one another based on their signature genes?
4. What are the genes that define human TRICs, i.e. the core set of genes which are conserved across the class of immune cell types and those genes which are specific to TRIC from a certain tissue?
5. How do the TRIC signatures from human (GTEx) compare with that of mouse (ImmGen and Tabula Muris)?

To derive TRIC signature from each of the 21 tissues considered, first a set of RISs were derived from a combination of downsampled data from all tissues. The RISs described six broad classifications of immune cell types of both lymphoid (B cells, plasma cells, T cells and NK cells) and myeloid (Mac-Mono-Den and neutrophils) lineages. These RIS

genes were then used to capture TRIC signatures from each tissue based on a pipeline developed herein using GCNs. The method accounted for the fact that not all immune cell types are present in every tissue. Approaches such as the use of RIS are necessary as the context in which we derive markers or signatures determines how we can use them, e.g. if we used signatures derived from a single tissue such as blood to annotate gene clusters in other tissues, some of the genes from blood immune cell signatures could represent a different biology in other tissues. The RIS helped identify the presence and absence of immune cell types across tissues thereby capturing 91 TRIC signatures, with Mac-Mono-Den clusters being present in all tissues. The analysis also described the distribution of immune cell types across tissues, the most evident example being for the brain which comprises dominantly of microglia and secondly T cells as also captured by the analysis. For the majority of cell types their gene signatures did not overlap, i.e. marker genes for different cell types did not co-cluster, however, for T cells and NK cells this was sometimes the case likely due to similarity between cell types. For certain tissues gene signatures for particular cell types like those of the myeloid lineage were not separable within the GCN. Therefore, higher thresholds were required to distinguish cell type gene signatures. This is likely due to the overlapping gene usage or biology across cell types and their comparable compositions within the tissue. The analysis also underscored the unique biology of each tissue which could be seen in the distinct topology of each GCN. The topology of GCNs from each tissue is likely a function of cell heterogeneity, cell proportions and the levels of genes they express, however, it is challenging to account for these factors and determine the “correct” threshold for GCN construction. Methods are being developed to determine a relevant threshold (Gobbi and Jurman, 2015). On examining the gene content of immune cell classifications across tissues, genes which were repeatedly associated with a certain immune cell classification included known marker genes for that cell type. Hence, the TRIC signatures derived for each of the 21 tissues did represent the broad classifications of immune cell types. For those genes which were unique to TRIC from a specific tissue, their validation was more challenging. The signatures were compared with the TRIC gene signatures obtained from the analysis of the ImmGen and Tabula Muris datasets in chapter 3. 39 gene clusters mapped across human (GTEx) and mouse (ImmGen and Tabula Muris) for the same TRIC. These clusters were associated with 12 TRICs of which literature was found supporting genes for seven from their associated gene clusters. Although, these low confidence genes (associated with few TRICs) likely comprised of a number

of false positives, the aforementioned comparison between mouse and human helped validate certain TRIC signatures. The number of gene clusters that mapped from mouse to human were still few relative to the total number of gene clusters identified, the likely reason being the resolution of bulk transcriptomics rather than interspecies differences. Unlike the microglia signature which have genes specific to it which describe their homeostatic functions in the brain, other tissues may comprise of several cell types which share such genes that are responsible for tissue homeostatic functions. Assuming that the TRICs studied here express such genes, they may not coexpress with the immune cell signatures as the expression of these genes is also a function of other cell types (their portions and the levels at which they express these genes). Hence, although it is easier to identify the core signatures for immune cell types within bulk transcriptomics while finding TRIC associated genes is still more challenging.

Building on initiatives such as GTEx and ImmGen, are the Human Cell Atlas (Regev et al., 2017) and human cell landscape (Han et al., 2020), the latter consisting of more than half a million cells generated from 60 tissues. This would be a great resource for the final analysis between mouse and human TRICs in this thesis which was otherwise limited by the resolution of bulk tissue transcriptomic data. To address some of the issues in deriving cell signatures from bulk transcriptomics data, recent studies have moved towards scRNA-Seq which enables the examination of cell proportions and cell types independently. Now, bulk transcriptomics is instead being used to estimate cell proportions based on signatures or markers identified from single cell omics, as single cell isolation protocols may not capture different cell types with the same efficacy (Jew et al., 2020, Wang et al., 2019c, Dong et al., 2020, Kang et al., 2019). However, even with the development of such methodologies it will be interesting to see whether the error rate of estimating cell types and their proportions in bulk RNA-Seq data vary across tissues. Nevertheless, the approach is important in treatments, e.g. in determining the tumour cell profiles for prognosis (Racle et al., 2017). In addition to tissue-derived bulk transcriptomics there is also transcriptomics data derived from pooled cells such as the ImmGen data examined in chapter 3. Whilst, the signal from this data was less affected by technical artefacts specific to scRNA-Seq such as dropouts, the approach still relied on FACS which is reliant on a limited number of markers for defined cell populations. Hence, single cell based approaches would still likely be preferred especially with future improvements in isolation protocols and technologies such as the Smart-Seq3 which improves upon the read coverage, sensitivity of read detection and library cost per cell

relative to previous single cell long-read technologies (Hagemann-Jensen et al., 2020). Recent studies have also used bulk RNA-Seq data as a reference to help impute values in scRNA-Seq data thereby improving downstream analyses including the detection of distinct cell types (Hou et al., 2020). As of now bulk RNA-Seq still holds a significant place in the field due to the quality of signal and its relatively low cost.

With the diversity and development of transcriptomic technologies as described herein, the field of cell biology continues to develop single cell omic technologies which our furthers our understanding of their heterogeneity. As a result of this complexity we are delving deeper into the fundamental and conceptual questions of what a cell type is. My outlook of this discussion, also shared by Michael Elowits and his colleagues (2017), is to consider cells as human beings who through life make various decisions thereby determining their final personality and position in the world. Furthermore, these decisions are dependent on those made previously while also relying on the environment or situations they live through. For example, in a study of the spleen and kidney, macrophage-derived from embryonic progenitors could be discerned from adult monocyte-derived macrophages (Epelman et al., 2014a). Accordingly several methods are now being developed for tracking cell phenotypes (Wagner and Klein, 2020) thereby determining what is involved in the decisions cells make that ultimately defines their identity. In addition to the gene regulatory networks of cell types described in most studies included in this thesis, recent methodologies for capturing cell-to-cell interaction networks hold great potential in understanding the complex dynamics of cells within tissues. Apart from the cocktail of ligands which make up environmental queues that are required for cell development, to adopt tissue resident phenotypes and activations states, cell-to-cell interactions are also important signals to study. For example, activation of T cells is dependent on the context of TCR engagement with MHC (Kim and Williams, 2010). Now, using bioinformatic techniques on non-enriched samples one can determine the potential interactors from scRNA-Seq with tools such as CellPhoneDB (Efremova et al., 2020). Only recently have methods been developed to capture these instances of cell-to-cell interaction (Giladi et al., 2020). The approach labels distinct cell types which are then FACS sorted into singlets as well as interacting cells which are then analysed. Examining the omics of interacting cells would help identify which cells interact more often and what responses are induced. In addition to the network of pathways that define cell types, such experiments would provide the network of cell-to-cell interactions which determine the immune response. Hence, cell interaction networks may reveal

subtle or transient events between cell types relevant in disease pathology which would open various therapeutic avenues.

In summary, this thesis identifies various novel immune cell signatures. In the case of human microglia, the signature was thoroughly validated and used to determine region dependent alterations in ageing vs. Alzheimer's. Such immune cell signatures were also derived across tissues thereby dissecting lineage and tissue dependence. Furthermore, TRIC signatures were derived using various transcriptomic platforms finally comparing across species and RNA-Seq technologies. These analyses emphasised the limited resolution of bulk transcriptomics data and the complexities of single cell transcriptomics for which techniques for cell annotation and GCN analysis were developed. The thesis discusses the complexities of cellular heterogeneity and how we might better understand them in terms of their definition/nomenclature and the importance of databases to encourage cross study comparisons in this era of big data.

6. Bibliography

2017. What Is Your Conceptual Definition of "Cell Type" in the Context of a Mature Organism? *Cell Syst,* 4**,** 255-259.

ABBAS, A. R., BALDWIN, D., MA, Y., OUYANG, W., GURNEY, A., MARTIN, F., FONG, S., VAN LOOKEREN CAMPAGNE, M., GODOWSKI, P., WILLIAMS, P. M., CHAN, A. C. & CLARK, H. F. 2005. Immune response in silico (IRIS): immune-specific genes identified from a compendium of microarray expression data. *Genes Immun,* 6**,** 319-31.

ABBAS, A. R., WOLSLEGEL, K., SESHASAYEE, D., MODRUSAN, Z. & CLARK, H. F. 2009. Deconvolution of Blood Microarray Data Identifies Cellular Activation Patterns in Systemic Lupus Erythematosus. *PLOS ONE,* 4**,** e6098.

ABEL, A. M., YANG, C., THAKAR, M. S. & MALARKANNAN, S. 2018. Natural Killer Cells: Development, Maturation, and Clinical Utilization. *Frontiers in Immunology,* 9.

ADOLFSSON, J., MANSSON, R., BUZA-VIDAS, N., HULTQUIST, A., LIUBA, K., JENSEN, C. T., BRYDER, D., YANG, L., BORGE, O. J., THOREN, L. A., ANDERSON, K., SITNICKA, E., SASAKI, Y., SIGVARDSSON, M. & JACOBSEN, S. E. 2005. Identification of Flt3+ lympho-myeloid stem cells lacking erythro-megakaryocytic potential a revised road map for adult blood lineage commitment. *Cell,* 121**,** 295-306.

ADROVER, J. M., DEL FRESNO, C., CRAINICIUC, G., CUARTERO, M. I., CASANOVA-ACEBES, M., WEISS, L. A., HUERGA-ENCABO, H., SILVESTRE-ROIG, C., ROSSAINT, J., COSSÍO, I., LECHUGA-VIECO, A. V., GARCÍA-PRIETO, J., GÓMEZ-PARRIZAS, M., QUINTANA, J. A., BALLESTEROS, I., MARTIN-SALAMANCA, S., AROCA-CREVILLEN, A., CHONG, S. Z., EVRARD, M., BALABANIAN, K., LÓPEZ, J., BIDZHEKOV, K., BACHELERIE, F., ABAD-SANTOS, F., MUÑOZ-CALLEJA, C., ZARBOCK, A., SOEHNLEIN, O., WEBER, C., NG, L. G., LOPEZ-RODRIGUEZ, C., SANCHO, D., MORO, M. A., IBÁÑEZ, B. & HIDALGO, A. 2019. A Neutrophil Timer Coordinates Immune Defense and Vascular Protection. *Immunity,* 50**,** 390-402.e10.

AIBAR, S., GONZÁLEZ-BLAS, C. B., MOERMAN, T., HUYNH-THU, V. A., IMRICHOVA, H., HULSELMANS, G., RAMBOW, F., MARINE, J.-C., GEURTS, P., AERTS, J., VAN DEN OORD, J., ATAK, Z. K., WOUTERS, J. & AERTS, S. 2017. SCENIC: single-cell regulatory network inference and clustering. *Nature Methods,* 14**,** 1083-1086.

AITTOKALLIO, T. 2010. Dealing with missing values in large-scale studies: microarray data imputation and beyond. *Brief Bioinform,* 11**,** 253-64.

AJAMI, B., BENNETT, J. L., KRIEGER, C., MCNAGNY, K. M. & ROSSI, F. M. V. 2011. Infiltrating monocytes trigger EAE progression, but do not contribute to the resident microglia pool. *Nature Neuroscience,* 14**,** 1142.

AJAMI, B., BENNETT, J. L., KRIEGER, C., TETZLAFF, W. & ROSSI, F. M. V. 2007. Local self-renewal can sustain CNS microglia maintenance and function throughout adult life. *Nature Neuroscience,* 10**,** 1538-1543.

AL-DALAHMAH, O., SOSUNOV, A. A., SHAIK, A., OFORI, K., LIU, Y., VONSATTEL, J. P., ADORJAN, I., MENON, V. & GOLDMAN, J. E. 2020. Single-nucleus RNA-seq identifies Huntington disease astrocyte states. *Acta Neuropathologica Communications,* 8**,** 19.

ALAVI, A., RUFFALO, M., PARVANGADA, A., HUANG, Z. & BAR-JOSEPH, Z. 2018. A web server for comparative analysis of single-cell RNA-seq data. *Nature Communications,* 9**,** 4768.

ALCÁNTARA-HERNÁNDEZ, M., LEYLEK, R., WAGAR, L. E., ENGLEMAN, E. G., KELER, T., MARINKOVICH, M. P., DAVIS, M. M., NOLAN, G. P. & IDOYAGA, J. 2017. High-Dimensional Phenotypic Mapping of Human Dendritic Cells Reveals Interindividual Variation and Tissue Specialization. *Immunity,* 47**,** 1037-1050.e6.

ALDER, M. N., OPOKA, A. M., LAHNI, P., HILDEMAN, D. A. & WONG, H. R. 2017. Olfactomedin-4 Is a Candidate Marker for a Pathogenic Neutrophil Subset in Septic Shock. *Crit Care Med,* 45**,** e426-e432.

ANCUTA, P., RAO, R., MOSES, A., MEHLE, A., SHAW, S. K., LUSCINSKAS, F. W. & GABUZDA, D. 2003. Fractalkine preferentially mediates arrest and migration of CD16+ monocytes. *J Exp Med,* 197**,** 1701-7.

ANTONIANI, C., ROMANO, O. & MICCIO, A. 2017. Concise Review: Epigenetic Regulation of Hematopoiesis: Biological Insights and Therapeutic Applications. *STEM CELLS Translational Medicine,* 6**,** 2106-2114.

ARANGO DUQUE, G. & DESCOTEAUX, A. 2014. Macrophage cytokines: involvement in immunity and infectious diseases. *Front Immunol,* 5**,** 491.

ARCHAMBAUD, C., SALCEDO, S. P., LELOUARD, H., DEVILARD, E., DE BOVIS, B., VAN ROOIJEN, N., GORVEL, J. P. & MALISSEN, B. 2010. Contrasting roles of macrophages and dendritic cells in controlling initial pulmonary Brucella infection. *Eur J Immunol,* 40**,** 3458-71.

ARINOBU, Y., IWASAKI, H., GURISH, M. F., MIZUNO, S., SHIGEMATSU, H., OZAWA, H., TENEN, D. G., AUSTEN, K. F. & AKASHI, K. 2005. Developmental checkpoints of the basophil/mast cell lineages in adult murine hematopoiesis. *Proc Natl Acad Sci U S A,* 102**,** 18105-10.

ARORA, D., SINGH, V. & SINGH, A. Modeling and simulation analysis of Salmonella typhimurium inside human epithelial cells: Host-pathogen relationship analysis by system biology. 2016 International Conference on Bioinformatics and Systems Biology (BSB), 4-6 March 2016 2016. 1-4.

ARRENBERG, P., MARICIC, I. & KUMAR, V. 2011. Sulfatide-mediated activation of type II natural killer T cells prevents hepatic ischemic reperfusion injury in mice. *Gastroenterology,* 140**,** 646-55.

ASHBURNER, M., BALL, C. A., BLAKE, J. A., BOTSTEIN, D., BUTLER, H., CHERRY, J. M., DAVIS, A. P., DOLINSKI, K., DWIGHT, S. S., EPPIG, J. T., HARRIS, M. A., HILL, D. P., ISSEL-TARVER, L., KASARSKIS, A., LEWIS, S., MATESE, J. C., RICHARDSON, J. E., RINGWALD, M., RUBIN, G. M. & SHERLOCK, G. 2000. Gene Ontology: tool for the unification of biology. *Nature Genetics,* 25**,** 25.

AVASARALA, S., VAN SCOYK, M., KARUPPUSAMY RATHINAM, M. K., ZERAYESUS, S., ZHAO, X., ZHANG, W., PERGANDE, M. R., BORGIA, J. A., DEGREGORI, J., PORT, J. D., WINN, R. A. & BIKKAVILLI, R. K. 2015. PRMT1 Is a Novel Regulator of Epithelial-Mesenchymal-Transition in Non-small Cell Lung Cancer. *The Journal of biological chemistry,* 290**,** 13479-13489.

AW YEANG, H. X., PIERSMA, S. J., LIN, Y., YANG, L., MALKOVA, O. N., MINER, C., KRUPNICK, A. S., CHAPMAN, W. C. & YOKOYAMA, W. M. 2017. Cutting Edge: Human CD49e- NK Cells Are Tissue Resident in the Liver. *J Immunol,* 198**,** 1417-1422.

BAIN, C. C., BRAVO-BLAS, A., SCOTT, C. L., GOMEZ PERDIGUERO, E., GEISSMANN, F., HENRI, S., MALISSEN, B., OSBORNE, L. C., ARTIS, D. & MOWAT, A. M. 2014. Constant replenishment from circulating monocytes maintains the macrophage pool in the intestine of adult mice. *Nature Immunology,* 15**,** 929.

BAIN, C. C., SCOTT, C. L., URONEN-HANSSON, H., GUDJONSSON, S., JANSSON, O., GRIP, O., GUILLIAMS, M., MALISSEN, B., AGACE, W. W. & MOWAT, A. M. 2013. Resident and pro-inflammatory macrophages in the colon represent alternative context-dependent fates of the same Ly6Chi monocyte precursors. *Mucosal Immunol,* 6**,** 498-510.

BAL, S. M., GOLEBSKI, K. & SPITS, H. 2020. Plasticity of innate lymphoid cell subsets. *Nature Reviews Immunology*.

BAR-SHAVIT, Z. 2007. The osteoclast: a multinucleated, hematopoietic-origin, bone-resorbing osteoimmune cell. *J Cell Biochem,* 102**,** 1130-9.

BARAN, Y., SEBE-PEDROS, A., LUBLING, Y., GILADI, A., CHOMSKY, E., MEIR, Z., HOICHMAN, M., LIFSHITZ, A. & TANAY, A. 2018. MetaCell: analysis of single cell RNA-seq data using k-NN graph partitions. *bioRxiv***,** 437665.

BARNES, M. J., GRISERI, T., JOHNSON, A. M., YOUNG, W., POWRIE, F. & IZCUE, A. 2013. CTLA-4 promotes Foxp3 induction and regulatory T cell accumulation in the intestinal lamina propria. *Mucosal Immunol,* 6**,** 324-34.

BECHER, B., SCHLITZER, A., CHEN, J., MAIR, F., SUMATOH, H. R., TENG, K. W. W., LOW, D., RUEDL, C., RICCARDI-CASTAGNOLI, P., POIDINGER, M., GRETER, M., GINHOUX, F. & NEWELL, E. W. 2014. High-dimensional analysis of the murine myeloid cell system. *Nature Immunology,* 15**,** 1181-1189.

BECHT, E., MCINNES, L., HEALY, J., DUTERTRE, C.-A., KWOK, I. W. H., NG, L. G., GINHOUX, F. & NEWELL, E. W. 2019. Dimensionality reduction for visualizing single-cell data using UMAP. *Nature Biotechnology,* 37**,** 38-44.

BEHAM, A. W., PUELLMANN, K., LAIRD, R., FUCHS, T., STREICH, R., BREYSACH, C., RADDATZ, D., ONIGA, S., PECCERELLA, T., FINDEISEN, P., KZHYSHKOWSKA, J., GRATCHEV, A., SCHWEYER, S., SAUNDERS, B., WESSELS, J. T., MÖBIUS, W., KEANE, J., BECKER, H., GANSER, A., NEUMAIER, M. & KAMINSKI, W. E. 2011. A TNF-regulated recombinatorial macrophage immune receptor implicated in granuloma formation in tuberculosis. *PLoS Pathog,* 7**,** e1002375.

BENDELAC, A., LANTZ, O., QUIMBY, M., YEWDELL, J., BENNINK & BRUTKIEWICZ, R. 1995. CD1 recognition by mouse NK1+ T lymphocytes. *Science,* 268**,** 863-865.

BENNETT, M. L., BENNETT, F. C., LIDDELOW, S. A., AJAMI, B., ZAMANIAN, J. L., FERNHOFF, N. B., MULINYAWE, S. B., BOHLEN, C. J., ADIL, A. & TUCKER, A. 2016. New tools for studying microglia in the mouse and human CNS. *Proceedings of the National Academy of Sciences,* 113**,** E1738-E1746.

BERGSBAKEN, T. & BEVAN, M. J. 2015. Proinflammatory microenvironments within the intestine regulate the differentiation of tissue-resident CD8+ T cells responding to infection. *Nature Immunology,* 16**,** 406-414.

BERTRAND, J. Y., JALIL, A., KLAINE, M., JUNG, S., CUMANO, A. & GODIN, I. 2005. Three pathways to mature macrophages in the early mouse yolk sac. *Blood,* 106**,** 3004-3011.

BEURA, L. K., HAMILTON, S. E., BI, K., SCHENKEL, J. M., ODUMADE, O. A., CASEY, K. A., THOMPSON, E. A., FRASER, K. A., ROSATO, P. C., FILALI-MOUHIM, A., SEKALY, R. P., JENKINS, M. K., VEZYS, V., HAINING, W. N., JAMESON, S. C. & MASOPUST, D. 2016. Normalizing the environment recapitulates adult human immune traits in laboratory mice. *Nature,* 532**,** 512-6.

BHUIYAN, M. E., WAKI, H., GOURAUD, S. S., TAKAGISHI, M., KOHSAKA, A. & MAEDA, M. 2011. Histamine receptor H1 in the nucleus tractus solitarii regulates arterial pressure and heart rate in rats. *American journal of physiology-heart and circulatory physiology,* 301**,** H523-H529.

BIBER, K., BHATTACHARYA, A., CAMPBELL, B. M., PIRO, J. R., ROHE, M., STAAL, R. G. W., TALANIAN, R. V. & MÖLLER, T. 2019. Microglial Drug Targets in AD: Opportunities and Challenges in Drug Discovery and Development. *Frontiers in Pharmacology,* 10.

BLASCHITZ, C. & RAFFATELLU, M. 2010. Th17 cytokines and the gut mucosal barrier. *J Clin Immunol,* 30**,** 196-203.

BLONDEL, V. D., GUILLAUME, J.-L., LAMBIOTTE, R. & LEFEBVRE, E. 2008. Fast unfolding of communities in large networks. *Journal of statistical mechanics: theory and experiment,* 2008**,** P10008.

BOLEN, C. R., UDUMAN, M. & KLEINSTEIN, S. H. 2011. Cell subset prediction for blood genomic studies. *BMC Bioinformatics,* 12**,** 258.

BORATE, B. R., CHESLER, E. J., LANGSTON, M. A., SAXTON, A. M. & VOY, B. H. 2009. Comparison of threshold selection methods for microarray gene co-expression matrices. *BMC Research Notes,* 2**,** 240.

BÖTTCHER, C., SCHLICKEISER, S., SNEEBOER, M. A. M., KUNKEL, D., KNOP, A., PAZA, E., FIDZINSKI, P., KRAUS, L., SNIJDERS, G. J. L., KAHN, R. S., SCHULZ, A. R., MEI, H. E., HOL, E. M., SIEGMUND, B., GLAUBEN, R., SPRUTH, E. J., DE WITTE, L. D., PRILLER, J. & PSY, N. B. B. 2019. Human microglia regional heterogeneity and phenotypes determined by multiplexed single-cell mass cytometry. *Nature Neuroscience,* 22**,** 78-90.

BOYETTE, L. B., MACEDO, C., HADI, K., ELINOFF, B. D., WALTERS, J. T., RAMASWAMI, B., CHALASANI, G., TABOAS, J. M., LAKKIS, F. G. & METES, D. M. 2017. Phenotype, function, and differentiation potential of human monocyte subsets. *PLOS ONE,* 12**,** e0176460.

BRANDES, M., WILLIMANN, K. & MOSER, B. 2005. Professional Antigen-Presentation Function by Human γδ T Cells. *Science,* 309**,** 264-268.

BRAZMA, A. & VILO, J. 2000. Gene expression data analysis. *FEBS Letters,* 480**,** 17-24.

BRECHBUHL, H. M., LI, B., SMITH, R. W. & REYNOLDS, S. D. 2014. Epidermal growth factor receptor activity is necessary for mouse basal cell proliferation. *American journal of physiology. Lung cellular and molecular physiology,* 307**,** L800-L810.

BUETTNER, M. & LOCHNER, M. 2016. Development and Function of Secondary and Tertiary Lymphoid Organs in the Small Intestine and the Colon. *Frontiers in Immunology,* 7.

CACCAMO, N., LA MENDOLA, C., ORLANDO, V., MERAVIGLIA, S., TODARO, M., STASSI, G., SIRECI, G., FOURNIÉ, J. J. & DIELI, F. 2011. Differentiation, phenotype, and function of interleukin-17–producing human Vγ9Vδ2 T cells. *Blood,* 118**,** 129-138.

CAMINATI, M., PHAM, D. L., BAGNASCO, D. & CANONICA, G. W. 2018. Type 2 immunity in asthma. *World Allergy Organization Journal,* 11**,** 13.

CAPONE, M., BRYANT, J. M., SUTKOWSKI, N. & HAQUE, A. 2016. Fc Receptor-Like Proteins in Pathophysiology of B-cell Disorder. *J Clin Cell Immunol,* 7.

CAPUTA, G., CASTOLDI, A. & PEARCE, E. J. 2019. Metabolic adaptations of tissue-resident immune cells. *Nature Immunology,* 20**,** 793-801.

CASANOVA-ACEBES, M., NICOLÁS-ÁVILA, J. A., LI, J. L., GARCÍA-SILVA, S., BALACHANDER, A., RUBIO-PONCE, A., WEISS, L. A., ADROVER, J. M., BURROWS, K., N, A. G., BALLESTEROS, I., DEVI, S., QUINTANA, J. A., CRAINICIUC, G., LEIVA, M., GUNZER, M., WEBER, C., NAGASAWA, T., SOEHNLEIN, O., MERAD, M., MORTHA, A., NG, L. G., PEINADO, H. & HIDALGO, A. 2018. Neutrophils instruct homeostatic and pathological states in naive tissues. *J Exp Med,* 215**,** 2778-2795.

CASSADO ADOS, A., D'IMPÉRIO LIMA, M. R. & BORTOLUCI, K. R. 2015. Revisiting mouse peritoneal macrophages: heterogeneity, development, and function. *Front Immunol,* 6**,** 225.

CASSETTA, L. & KITAMURA, T. 2018. Targeting Tumor-Associated Macrophages as a Potential Strategy to Enhance the Response to Immune Checkpoint Inhibitors. *Front Cell Dev Biol,* 6**,** 38.

CHALANCON, G., RAVARANI, C. N. J., BALAJI, S., MARTINEZ-ARIAS, A., ARAVIND, L., JOTHI, R. & BABU, M. M. 2012. Interplay between gene expression noise and regulatory network architecture. *Trends in Genetics,* 28**,** 221-232.

CHALLEN, G. A., BOLES, N. C., CHAMBERS, S. M. & GOODELL, M. A. 2010. Distinct hematopoietic stem cell subtypes are differentially regulated by TGF-beta1. *Cell Stem Cell,* 6**,** 265-78.

CHAPLIN, D. D. 2010. Overview of the immune response. *J Allergy Clin Immunol,* 125**,** S3-23.

CHAPPELL, L., RUSSELL, A. J. C. & VOET, T. 2018. Single-Cell (Multi)omics Technologies. *Annual Review of Genomics and Human Genetics,* 19**,** 15-41.

CHÁVEZ-GALÁN, L., OLLEROS, M. L., VESIN, D. & GARCIA, I. 2015. Much More than M1 and M2 Macrophages, There are also CD169+ and TCR+ Macrophages. *Frontiers in Immunology,* 6.

CHEN, B., KHODADOUST, M. S., LIU, C. L., NEWMAN, A. M. & ALIZADEH, A. A. 2018. Profiling Tumor Infiltrating Immune Cells with CIBERSORT. *In:* VON STECHOW, L. (ed.) *Cancer Systems Biology: Methods and Protocols.* New York, NY: Springer New York.

CHEN, H., ALBERGANTE, L., HSU, J. Y., LAREAU, C. A., LO BOSCO, G., GUAN, J., ZHOU, S., GORBAN, A. N., BAUER, D. E., ARYEE, M. J., LANGENAU, D. M., ZINOVYEV, A., BUENROSTRO, J. D., YUAN, G.-C. & PINELLO, L. 2019. Single-cell trajectories reconstruction, exploration and mapping of omics data with STREAM. *Nature Communications,* 10**,** 1903.

CHEN, S. & MAR, J. C. 2018. Evaluating methods of inferring gene regulatory networks highlights their lack of performance for single cell gene expression data. *BMC Bioinformatics,* 19**,** 232.

CHERRIER, M., SAWA, S. & EBERL, G. 2012. Notch, Id2, and RORγt sequentially orchestrate the fetal development of lymphoid tissue inducer cells. *J Exp Med,* 209**,** 729-40.

CHUNG, J. B., SILVERMAN, M. & MONROE, J. G. 2003. Transitional B cells: step by step towards immune competence. *Trends in Immunology,* 24**,** 342-348.

CHUNG, N. C. & STOREY, J. D. 2015. Statistical significance of variables driving systematic variation in high-dimensional data. *Bioinformatics,* 31**,** 545-54.

COHEN, A. A., GEVA-ZATORSKY, N., EDEN, E., FRENKEL-MORGENSTERN, M., ISSAEVA, I., SIGAL, A., MILO, R., COHEN-SAIDON, C., LIRON, Y., KAM, Z., COHEN, L., DANON, T., PERZOV, N. & ALON, U. 2008. Dynamic proteomics of individual cancer cells in response to a drug. *Science,* 322**,** 1511-6.

COLONNA, M. & BUTOVSKY, O. 2017. Microglia function in the central nervous system during health and neurodegeneration. *Annual Review of Immunology,* 35**,** 441-468.

COONS, A. H., CREECH, H. J. & JONES, R. N. 1941. Immunological properties of an antibody containing a fluorescent group. *Proceedings of the society for experimental biology and medicine,* 47**,** 200-202.

CRINIER, A., MILPIED, P., ESCALIÈRE, B., PIPEROGLOU, C., GALLUSO, J., BALSAMO, A., SPINELLI, L., CERVERA-MARZAL, I., EBBO, M., GIRARD-MADOUX, M., JAEGER, S., BOLLON, E., HAMED, S., HARDWIGSEN, J., UGOLINI, S., VÉLY, F., NARNI-MANCINELLI, E. & VIVIER, E. 2018. High-Dimensional Single-Cell Analysis Identifies Organ-Specific Signatures and Conserved NK Cell Subsets in Humans and Mice. *Immunity,* 49**,** 971-986.e5.

CROTTY, S. 2011. Follicular Helper CD4 T Cells (TFH). *Annual Review of Immunology,* 29**,** 621-663.

CUPEDO, T., CRELLIN, N. K., PAPAZIAN, N., ROMBOUTS, E. J., WEIJER, K., GROGAN, J. L., FIBBE, W. E., CORNELISSEN, J. J. & SPITS, H. 2009. Human fetal lymphoid tissue-inducer cells are interleukin 17-producing precursors to RORC+ CD127+ natural killer-like cells. *Nat Immunol,* 10**,** 66-74.

DADI, S. & LI, M. O. 2017. Tissue-resident lymphocytes: sentinel of the transformed tissue. *Journal for ImmunoTherapy of Cancer,* 5**,** 41.

DAVIES, L. C., JENKINS, S. J., ALLEN, J. E. & TAYLOR, P. R. 2013. Tissue-resident macrophages. *Nature Immunology,* 14**,** 986.

DE SCHEPPER, S., VERHEIJDEN, S., AGUILERA-LIZARRAGA, J., VIOLA, M. F., BOESMANS, W., STAKENBORG, N., VOYTYUK, I., SCHMIDT, I., BOECKX, B., DIERCKX DE CASTERLÉ, I., BAEKELANDT, V., GONZALEZ DOMINGUEZ, E., MACK, M., DEPOORTERE, I., DE STROOPER, B., SPRANGERS, B., HIMMELREICH, U., SOENEN, S., GUILLIAMS, M., VANDEN BERGHE, P., JONES, E., LAMBRECHTS, D. & BOECKXSTAENS, G. 2018. Self-Maintaining Gut Macrophages Are Essential for Intestinal Homeostasis. *Cell,* 175**,** 400-415.e13.

DE SIMONE, M., ROSSETTI, G. & PAGANI, M. 2018. Single Cell T Cell Receptor Sequencing: Techniques and Future Challenges. *Frontiers in Immunology,* 9.

DEHO, L., LEONI, C., BRODIE, T. M., MONTAGNER, S., DE SIMONE, M., POLLETTI, S., BAROZZI, I., NATOLI, G. & MONTICELLI, S. 2014. Two functionally distinct subsets of mast cells discriminated By IL-2-independent CD25 activities. *J Immunol,* 193**,** 2196-206.

DENG, H., HU, N., WANG, C., CHEN, M. & ZHAO, M.-H. 2018. Interaction between CD177 and platelet endothelial cell adhesion molecule-1 downregulates membrane-bound proteinase-3 (PR3) expression on neutrophils and attenuates neutrophil activation induced by PR3-ANCA. *Arthritis Research & Therapy,* 20**,** 213.

DENNING, T. L., NORRIS, B. A., MEDINA-CONTRERAS, O., MANICASSAMY, S., GEEM, D., MADAN, R., KARP, C. L. & PULENDRAN, B. 2011. Functional Specializations of Intestinal Dendritic Cell and Macrophage Subsets That Control Th17 and Regulatory T Cell Responses Are Dependent on the T Cell/APC Ratio, Source of Mouse Strain, and Regional Localization. *The Journal of Immunology,* 187**,** 733.

DESCATOIRE, M., WELLER, S., IRTAN, S., SARNACKI, S., FEUILLARD, J., STORCK, S., GUIOCHON-MANTEL, A., BOULIGAND, J., MORALI, A., COHEN, J., JACQUEMIN, E., IASCONE, M., BOLE-FEYSOT, C., CAGNARD, N., WEILL, J. C. & REYNAUD, C. A. 2014. Identification of a human splenic marginal zone B cell precursor with NOTCH2-dependent differentiation properties. *J Exp Med,* 211**,** 987-1000.

DESCH, A. N., RANDOLPH, G. J., MURPHY, K., GAUTIER, E. L., KEDL, R. M., LAHOUD, M. H., CAMINSCHI, I., SHORTMAN, K., HENSON, P. M. & JAKUBZICK, C. V. 2011. CD103+ pulmonary dendritic cells preferentially acquire and present apoptotic cell-associated antigen. *J Exp Med,* 208**,** 1789-97.

DI CARLO, D., TSE, H. T. & GOSSETT, D. R. 2012. Introduction: why analyze single cells? *Methods Mol Biol,* 853**,** 1-10.

DIEU, M. C., VANBERVLIET, B., VICARI, A., BRIDON, J. M., OLDHAM, E., AÏT-YAHIA, S., BRIÈRE, F., ZLOTNIK, A., LEBECQUE, S. & CAUX, C. 1998. Selective recruitment of immature and mature dendritic cells by distinct chemokines expressed in different anatomic sites. *J Exp Med,* 188**,** 373-86.

DING, J., DIRKS, W. G., EHRENTRAUT, S., GEFFERS, R., MACLEOD, R. A., NAGEL, S., POMMERENKE, C., ROMANI, J., SCHERR, M., VAAS, L. A., ZABORSKI, M., DREXLER, H. G. & QUENTMEIER, H. 2015. BCL6--regulated by AhR/ARNT and wild-type MEF2B--drives expression of germinal center markers MYBL1 and LMO2. *Haematologica,* 100**,** 801-9.

DIXON, L. J., BARNES, M., TANG, H., PRITCHARD, M. T. & NAGY, L. E. 2013. Kupffer cells in the liver. *Compr Physiol,* 3**,** 785-97.

DJENIDI, F., ADAM, J., GOUBAR, A., DURGEAU, A., MEURICE, G., DE MONTPREVILLE, V., VALIDIRE, P., BESSE, B. & MAMI-CHOUAIB, F. 2015. CD8+CD103+ tumor-infiltrating lymphocytes are tumor-specific tissue-resident memory T cells and a prognostic factor for survival in lung cancer patients. *J Immunol,* 194**,** 3475-86.

DOEING, D. C., BOROWICZ, J. L. & CROCKETT, E. T. 2003. Gender dimorphism in differential peripheral blood leukocyte counts in mice using cardiac, tail, foot, and saphenous vein puncture methods. *BMC Clin Pathol,* 3**,** 3.

DONG, M., THENNAVAN, A., URRUTIA, E., LI, Y., PEROU, C. M., ZOU, F. & JIANG, Y. 2020. SCDC: bulk gene expression deconvolution by multiple single-cell RNA sequencing references. *Briefings in Bioinformatics*.

DONGEN, S. 2000. A cluster algorithm for graphs.

DOUGHERTY, R. H., SIDHU, S. S., RAMAN, K., SOLON, M., SOLBERG, O. D., CAUGHEY, G. H., WOODRUFF, P. G. & FAHY, J. V. 2010. Accumulation of intraepithelial mast cells with a unique protease phenotype in T(H)2-high asthma. *J Allergy Clin Immunol,* 125**,** 1046-1053.e8.

DOWDEN, H. & MUNRO, J. 2019. Trends in clinical success rates and therapeutic focus. *Nat Rev Drug Discov,* 18**,** 495-496.

DOWNES, C. S., GILBERT, D., FUS, H., KURTH, M. J., ORTON, R., ROBINSON, S., VYSHEMIRSKY, V., DUBITZKY, W. & GU, X. 2006. Computational methodologies for modelling, analysis and simulation of signalling networks. *Briefings in Bioinformatics,* 7**,** 339-353.

DUAN, B., ZHOU, C., ZHU, C., YU, Y., LI, G., ZHANG, S., ZHANG, C., YE, X., MA, H., QU, S., ZHANG, Z., WANG, P., SUN, S. & LIU, Q. 2019. Model-based understanding of single-cell CRISPR screening. *Nature Communications,* 10**,** 2233.

DUHEN, T., DUHEN, R., MONTLER, R., MOSES, J., MOUDGIL, T., DE MIRANDA, N. F., GOODALL, C. P., BLAIR, T. C., FOX, B. A., MCDERMOTT, J. E., CHANG, S. C., GRUNKEMEIER, G., LEIDNER, R., BELL, R. B. & WEINBERG, A. D. 2018. Co-expression of CD39 and CD103 identifies tumor-reactive CD8 T cells in human solid tumors. *Nat Commun,* 9**,** 2724.

DUNKELBERGER, J. R. & SONG, W.-C. 2010. Complement and its role in innate and adaptive immune responses. *Cell Research,* 20**,** 34-50.

DUÒ, A., ROBINSON, M. D. & SONESON, C. 2018. A systematic performance evaluation of clustering methods for single-cell RNA-seq data. *F1000Res,* 7**,** 1141.

DUTTA, P., HOYER, F. F., GRIGORYEVA, L. S., SAGER, H. B., LEUSCHNER, F., COURTIES, G., BORODOVSKY, A., NOVOBRANTSEVA, T., RUDA, V. M., FITZGERALD, K., IWAMOTO, Y., WOJTKIEWICZ, G., SUN, Y., DA SILVA, N., LIBBY, P., ANDERSON, D. G., SWIRSKI, F. K., WEISSLEDER, R. & NAHRENDORF, M. 2015. Macrophages retain hematopoietic stem cells in the spleen via VCAM-1. *J Exp Med,* 212**,** 497-512.

DWYER, D. F., BARRETT, N. A., AUSTEN, K. F., DWYER, D. F., BARRETT, N. A., AUSTEN, K. F., KIM, E. Y., BRENNER, M. B., SHAW, L., YU, B., GOLDRATH, A., MOSTAFAVI, S., REGEV, A., RHOADES, A., MOODLEY, D., YOSHIDA, H., MATHIS, D., BENOIST, C., NABEKURA, T., LAM, V., LANIER, L. L., BROWN, B., MERAD, M., CREMASCO, V., TURLEY, S., MONACH, P., DUSTIN, M. L., LI, Y., SHINTON, S. A., HARDY, R. R., SHAY, T., QI, Y., SYLVIA, K., KANG, J., FAIRFAX, K., RANDOLPH, G. J., ROBINETTE, M. L., FUCHS, A., COLONNA, M. & THE IMMUNOLOGICAL GENOME PROJECT, C. 2016. Expression

profiling of constitutive mast cells reveals a unique identity within the immune system. *Nature Immunology,* 17**,** 878-887.

EDGAR, R., DOMRACHEV, M. & LASH, A. E. 2002. Gene Expression Omnibus: NCBI gene expression and hybridization array data repository. *Nucleic Acids Research,* 30**,** 207-210.

EDWARDS, J. S., IBARRA, R. U. & PALSSON, B. O. 2001. In silico predictions of Escherichia coli metabolic capabilities are consistent with experimental data. *Nature Biotechnology,* 19**,** 125.

EFREMOVA, M., VENTO-TORMO, M., TEICHMANN, S. A. & VENTO-TORMO, R. 2020. CellPhoneDB: inferring cell–cell communication from combined expression of multi-subunit ligand–receptor complexes. *Nature Protocols,* 15**,** 1484-1506.

ENSAN, S., LI, A., BESLA, R., DEGOUSEE, N., COSME, J., ROUFAIEL, M., SHIKATANI, E. A., EL-MAKLIZI, M., WILLIAMS, J. W., ROBINS, L., LI, C., LEWIS, B., YUN, T. J., LEE, J. S., WIEGHOFER, P., KHATTAR, R., FARROKHI, K., BYRNE, J., OUZOUNIAN, M., ZAVITZ, C. C., LEVY, G. A., BAUER, C. M., LIBBY, P., HUSAIN, M., SWIRSKI, F. K., CHEONG, C., PRINZ, M., HILGENDORF, I., RANDOLPH, G. J., EPELMAN, S., GRAMOLINI, A. O., CYBULSKY, M. I., RUBIN, B. B. & ROBBINS, C. S. 2016. Self-renewing resident arterial macrophages arise from embryonic CX3CR1(+) precursors and circulating monocytes immediately after birth. *Nat Immunol,* 17**,** 159-68.

EPELMAN, S., LAVINE, KORY J., BEAUDIN, ANNA E., SOJKA, DOROTHY K., CARRERO, JAVIER A., CALDERON, B., BRIJA, T., GAUTIER, EMMANUEL L., IVANOV, S., SATPATHY, ANSUMAN T., SCHILLING, JOEL D., SCHWENDENER, R., SERGIN, I., RAZANI, B., FORSBERG, E. C., YOKOYAMA, WAYNE M., UNANUE, EMIL R., COLONNA, M., RANDOLPH, GWENDALYN J. & MANN, DOUGLAS L. 2014a. Embryonic and Adult-Derived Resident Cardiac Macrophages Are Maintained through Distinct Mechanisms at Steady State and during Inflammation. *Immunity,* 40**,** 91-104.

EPELMAN, S., LAVINE, KORY J. & RANDOLPH, GWENDALYN J. 2014b. Origin and Functions of Tissue Macrophages. *Immunity,* 41**,** 21-35.

ERASLAN, G., SIMON, L. M., MIRCEA, M., MUELLER, N. S. & THEIS, F. J. 2019. Single-cell RNA-seq denoising using a deep count autoencoder. *Nature Communications,* 10**,** 390.

ERICKSON, M. A., DOHI, K. & BANKS, W. A. 2012. Neuroinflammation: a common pathway in CNS diseases as mediated at the blood-brain barrier. *Neuroimmunomodulation,* 19**,** 121-30.

FAISAL, S. & TUTZ, G. 2017. Missing value imputation for gene expression data by tailored nearest neighbors. *Statistical applications in genetics and molecular biology,* 16**,** 95-106.

FAN, X. & RUDENSKY, ALEXANDER Y. 2016. Hallmarks of Tissue-Resident Lymphocytes. *Cell,* 164**,** 1198-1211.

FARIA, A. M. C., REIS, B. S. & MUCIDA, D. 2017. Tissue adaptation: Implications for gut immunity and tolerance. *The Journal of Experimental Medicine,* 214**,** 1211-1226.

FARRELL, J. A., WANG, Y., RIESENFELD, S. J., SHEKHAR, K., REGEV, A. & SCHIER, A. F. 2018. Single-cell reconstruction of developmental trajectories during zebrafish embryogenesis. *Science,* 360**,** eaar3131.

FEIZI, S., MARBACH, D., MÉDARD, M. & KELLIS, M. 2013. Network deconvolution as a general method to distinguish direct dependencies in networks. *Nature Biotechnology,* 31**,** 726.

FERRERO, G., MAHONY, C. B., DUPUIS, E., YVERNOGEAU, L., DI RUGGIERO, E., MISEROCCHI, M., CARON, M., ROBIN, C., TRAVER, D., BERTRAND, J. Y. & WITTAMER, V. 2018. Embryonic Microglia Derive from Primitive Macrophages and Are Replaced by cmyb-Dependent Definitive Microglia in Zebrafish. *Cell Rep,* 24**,** 130-141.

FIGUERES-OÑATE, M., GUTIÉRREZ, Y. & LÓPEZ-MASCARAQUE, L. 2014. Unraveling Cajal's view of the olfactory system. *Frontiers in Neuroanatomy,* 8.

FINOTELLO, F. & TRAJANOSKI, Z. 2018. Quantifying tumor-infiltrating immune cells from transcriptomics data. *Cancer Immunol Immunother,* 67**,** 1031-1040.

FISHER, J. & HENZINGER, T. A. 2007. Executable cell biology. *Nature Biotechnology,* 25**,** 1239.

FORBES, L. R. & HACZKU, A. 2010. SP-D and regulation of the pulmonary innate immune system in allergic airway changes. *Clin Exp Allergy,* 40**,** 547-62.

FORTE, E., FURTADO, M. B. & ROSENTHAL, N. 2018. The interstitium in cardiac repair: role of the immune–stromal cell interplay. *Nature Reviews Cardiology,* 15**,** 601-616.

FUCHS, T., PUELLMANN, K., EMMERT, A., FLEIG, J., ONIGA, S., LAIRD, R., HEIDA, N. M., SCHÄFER, K., NEUMAIER, M., BEHAM, A. W. & KAMINSKI, W. E. 2015. The macrophage-TCRαβ is a cholesterol-responsive combinatorial immune receptor and implicated in atherosclerosis. *Biochemical and Biophysical Research Communications,* 456**,** 59-65.

FUCHS, T., PUELLMANN, K., HAHN, M., DOLLT, C., PECHLIVANIDOU, I., OVSIY, I., KZHYSHKOWSKA, J., GRATCHEV, A., FLEIG, J., EMMERT, A., NEUMAIER, M., BEHAM, A. W. & KAMINSKI, W. E. 2013. A second combinatorial immune receptor in monocytes/macrophages is based on the TCRγδ. *Immunobiology,* 218**,** 960-968.

GAINULLINA, A., HUANG, L.-H., TODOROV, H., KIM, K., YNG, L. S., KENT, A., JIA, B., SEDDU, K., KRCHMA, K., WU, J., CROZAT, K., TOMASELLO, E., NARANG, V., DRESS, R., SEE, P., SCOTT, C., GIBBINGS, S., BAJPAI, G., DESAI, J. V., MAIER, B., THIS, S., WANG, P., AGUILAR, S. V., POUPEL, L., DUSSAUD, S., ZHOU, T.-A., ANGELI, V., BLANDER, J. M., CHOI, K., DALOD, M., DZHAGALOV, I., GAUTIER, E. L., JAKUBZICK, C., LAVINE, K., LIONAKIS, M.

S., PAIDASSI, H., SIEWEKE, M. H., GINHOUX, F., GUILLIAMS, M., BENOIST, C., MERAD, M., RANDOLPH, G. J., SERGUSHICHEV, A. & ARTYOMOV, M. N. 2020. Open Source ImmGen: network perspective on metabolic diversity among mononuclear phagocytes. *bioRxiv*, 2020.07.15.204388.

GALATRO, T. F., HOLTMAN, I. R., LERARIO, A. M., VAINCHTEIN, I. D., BROUWER, N., SOLA, P. R., VERAS, M. M., PEREIRA, T. F., LEITE, R. E. P., MÖLLER, T., WES, P. D., SOGAYAR, M. C., LAMAN, J. D., DEN DUNNEN, W., PASQUALUCCI, C. A., OBA-SHINJO, S. M., BODDEKE, E. W. G. M., MARIE, S. K. N. & EGGEN, B. J. L. 2017. Transcriptomic analysis of purified human cortical microglia reveals age-associated changes. *Nature Neuroscience,* 20**,** 1162.

GAO, X., WANG, G., ZHANG, W., PENG, Q., XUE, M. & JINHONG, H. 2013. Expression of pulmonary aquaporin 1 is dramatically upregulated in mice with pulmonary fibrosis induced by bleomycin. *Archives of medical science : AMS,* 9**,** 916-921.

GAUSE, W. C., WYNN, T. A. & ALLEN, J. E. 2013. Type 2 immunity and wound healing: evolutionary refinement of adaptive immunity by helminths. *Nature Reviews Immunology,* 13**,** 607-614.

GAUTIER, E. L., SHAY, T., MILLER, J., GRETER, M., JAKUBZICK, C., IVANOV, S., HELFT, J., CHOW, A., ELPEK, K. G., GORDONOV, S., MAZLOOM, A. R., MA'AYAN, A., CHUA, W.-J., HANSEN, T. H., TURLEY, S. J., MERAD, M., RANDOLPH, G. J., THE IMMUNOLOGICAL GENOME, C., GAUTIER, E. L., JAKUBZICK, C., RANDOLPH, G. J., BEST, A. J., KNELL, J., GOLDRATH, A., MILLER, J., BROWN, B., MERAD, M., JOJIC, V., KOLLER, D., COHEN, N., BRENNAN, P., BRENNER, M., SHAY, T., REGEV, A., FLETCHER, A., ELPEK, K., BELLEMARE-PELLETIER, A., MALHOTRA, D., TURLEY, S., JIANU, R., LAIDLAW, D., COLLINS, J., NARAYAN, K., SYLVIA, K., KANG, J., GAZIT, R., GARRISON, B. S., ROSSI, D. J., KIM, F., RAO, T. N., WAGERS, A., SHINTON, S. A., HARDY, R. R., MONACH, P., BEZMAN, N. A., SUN, J. C., KIM, C. C., LANIER, L. L., HENG, T., KRESLAVSKY, T., PAINTER, M., ERICSON, J., DAVIS, S., MATHIS, D. & BENOIST, C. 2012. Gene-expression profiles and transcriptional regulatory pathways that underlie the identity and diversity of mouse tissue macrophages. *Nature Immunology,* 13**,** 1118.

GEBHARDT, T., WAKIM, L. M., EIDSMO, L., READING, P. C., HEATH, W. R. & CARBONE, F. R. 2009. Memory T cells in nonlymphoid tissue that provide enhanced local immunity during infection with herpes simplex virus. *Nat Immunol,* 10**,** 524-30.

GEIRSDOTTIR, L., DAVID, E., KEREN-SHAUL, H., WEINER, A., BOHLEN, S. C., NEUBER, J., BALIC, A., GILADI, A., SHEBAN, F., DUTERTRE, C.-A., PFEIFLE, C., PERI, F., RAFFO-ROMERO, A., VIZIOLI, J., MATIASEK, K., SCHEIWE, C., MECKEL, S., MÄTZ-RENSING, K., VAN DER MEER, F., THORMODSSON, F. R., STADELMANN, C., ZILKHA, N., KIMCHI, T., GINHOUX, F., ULITSKY, I., ERNY, D., AMIT, I. & PRINZ, M. 2019. Cross-Species Single-Cell Analysis Reveals Divergence of the Primate Microglia Program. *Cell,* 179**,** 1609-1622.e16.

GERMAIN, R. N. 2002. T-cell development and the CD4–CD8 lineage decision. *Nature Reviews Immunology,* 2**,** 309-322.

GIBBINGS, S. L., GOYAL, R., DESCH, A. N., LEACH, S. M., PRABAGAR, M., ATIF, S. M., BRATTON, D. L., JANSSEN, W. & JAKUBZICK, C. V. 2015. Transcriptome analysis highlights the conserved difference between embryonic and postnatal-derived alveolar macrophages. *Blood,* 126**,** 1357-66.

GILADI, A., COHEN, M., MEDAGLIA, C., BARAN, Y., LI, B., ZADA, M., BOST, P., BLECHER-GONEN, R., SALAME, T.-M., MAYER, J. U., DAVID, E., RONCHESE, F., TANAY, A. & AMIT, I. 2020. Dissecting cellular crosstalk by sequencing physically interacting cells. *Nature Biotechnology,* 38**,** 629-637.

GILLIET, M., CAO, W. & LIU, Y.-J. 2008. Plasmacytoid dendritic cells: sensing nucleic acids in viral infection and autoimmune diseases. *Nature Reviews Immunology,* 8**,** 594-606.

GINHOUX, F., GRETER, M., LEBOEUF, M., NANDI, S., SEE, P., GOKHAN, S., MEHLER, M. F., CONWAY, S. J., NG, L. G. & STANLEY, E. R. 2010. Fate mapping analysis reveals that adult microglia derive from primitive macrophages. *Science,* 330**,** 841-845.

GINHOUX, F. & GUILLIAMS, M. 2016. Tissue-Resident Macrophage Ontogeny and Homeostasis. *Immunity,* 44**,** 439-449.

GOBBI, A. & JURMAN, G. 2015. A Null Model for Pearson Coexpression Networks. *PLOS ONE,* 10**,** e0128115.

GOLTSEV, Y., SAMUSIK, N., KENNEDY-DARLING, J., BHATE, S., HALE, M., VAZQUEZ, G., BLACK, S. & NOLAN, G. P. 2018. Deep Profiling of Mouse Splenic Architecture with CODEX Multiplexed Imaging. *Cell,* 174**,** 968-981.e15.

GOLUBOVSKAYA, V. & WU, L. 2016. Different Subsets of T Cells, Memory, Effector Functions, and CAR-T Immunotherapy. *Cancers (Basel),* 8.

GONG, W., KWAK, I.-Y., POTA, P., KOYANO-NAKAGAWA, N. & GARRY, D. J. 2018. DrImpute: imputing dropout events in single cell RNA sequencing data. *BMC Bioinformatics,* 19**,** 220.

GOSSELIN, D., LINK, V. M., ROMANOSKI, C. E., FONSECA, G. J., EICHENFIELD, D. Z., SPANN, N. J., STENDER, J. D., CHUN, H. B., GARNER, H., GEISSMANN, F. & GLASS, C. K. 2014. Environment drives selection and function of enhancers controlling tissue-specific macrophage identities. *Cell,* 159**,** 1327-40.

GRABERT, K., MICHOEL, T., KARAVOLOS, M. H., CLOHISEY, S., BAILLIE, J. K., STEVENS, M. P., FREEMAN, T. C., SUMMERS, K. M. & MCCOLL, B. W. 2016. Microglial brain region-dependent diversity and selective regional sensitivities to aging. *Nat Neurosci,* 19**,** 504-516.

GRAUDENZI, A., MASPERO, D., DI FILIPPO, M., GNUGNOLI, M., ISELLA, C., MAURI, G., MEDICO, E., ANTONIOTTI, M. & DAMIANI, C. 2018. Integration of transcriptomic data and metabolic networks in cancer samples reveals highly significant prognostic power. *J Biomed Inform,* 87**,** 37-49.

GREENE, C. S., KRISHNAN, A., WONG, A. K., RICCIOTTI, E., ZELAYA, R. A., HIMMELSTEIN, D. S., ZHANG, R., HARTMANN, B. M., ZASLAVSKY, E., SEALFON, S. C., CHASMAN, D. I., FITZGERALD, G. A., DOLINSKI, K., GROSSER, T. & TROYANSKAYA, O. G. 2015. Understanding multicellular function and disease with human tissue-specific networks. *Nature Genetics,* 47**,** 569.

GREN, S. T., RASMUSSEN, T. B., JANCIAUSKIENE, S., HÅKANSSON, K., GERWIEN, J. G. & GRIP, O. 2015. A Single-Cell Gene-Expression Profile Reveals Inter-Cellular Heterogeneity within Human Monocyte Subsets. *PLoS One,* 10**,** e0144351.

GRETER, M., LELIOS, I. & CROXFORD, A. L. 2015. Microglia Versus Myeloid Cell Nomenclature during Brain Inflammation. *Front Immunol,* 6**,** 249.

GRISERI, T., ARNOLD, I. C., PEARSON, C., KRAUSGRUBER, T., SCHIERING, C., FRANCHINI, F., SCHULTHESS, J., MCKENZIE, B. S., CROCKER, P. R. & POWRIE, F. 2015. Granulocyte Macrophage Colony-Stimulating Factor-Activated Eosinophils Promote Interleukin-23 Driven Chronic Colitis. *Immunity,* 43**,** 187-99.

GROSS, B. E., SANDER, C., DEMIR, E., BADER, G. D., RODCHENKOV, I., ANWAR, N., SCHULTZ, N., BABUR, Ö. & CERAMI, E. G. 2010. Pathway Commons, a web resource for biological pathway data. *Nucleic Acids Research,* 39**,** D685-D690.

GUILLIAMS, M., CROZAT, K., HENRI, S., TAMOUTOUNOUR, S., GRENOT, P., DEVILARD, E., DE BOVIS, B., ALEXOPOULOU, L., DALOD, M. & MALISSEN, B. 2010. Skin-draining lymph nodes contain dermis-derived CD103(-) dendritic cells that constitutively produce retinoic acid and induce Foxp3(+) regulatory T cells. *Blood,* 115**,** 1958-68.

GUO, F., LI, L., LI, J., WU, X., HU, B., ZHU, P., WEN, L. & TANG, F. 2017. Single-cell multi-omics sequencing of mouse early embryos and embryonic stem cells. *Cell Research,* 27**,** 967-988.

GUO, Q., WANG, Y., XU, D., NOSSENT, J., PAVLOS, N. J. & XU, J. 2018. Rheumatoid arthritis: pathological mechanisms and modern pharmacologic therapies. *Bone Res,* 6**,** 15.

GUO, Z., JANG, M. H., OTANI, K., BAI, Z., UMEMOTO, E., MATSUMOTO, M., NISHIYAMA, M., YAMASAKI, M., UEHA, S., MATSUSHIMA, K., HIRATA, T. & MIYASAKA, M. 2008. CD4+CD25+ regulatory T cells in the small intestinal lamina propria show an effector/memory phenotype. *Int Immunol,* 20**,** 307-15.

GURISH, MICHAEL F. & AUSTEN, K. F. 2012. Developmental Origin and Functional Specialization of Mast Cell Subsets. *Immunity,* 37**,** 25-33.

HAAS, K. M., POE, J. C., STEEBER, D. A. & TEDDER, T. F. 2005. B-1a and B-1b Cells Exhibit Distinct Developmental Requirements and Have Unique Functional Roles in Innate and Adaptive Immunity to S. pneumoniae. *Immunity,* 23**,** 7-18.

HAGEMANN-JENSEN, M., ZIEGENHAIN, C., CHEN, P., RAMSKÖLD, D., HENDRIKS, G.-J., LARSSON, A. J. M., FARIDANI, O. R. & SANDBERG, R. 2020. Single-cell RNA counting at allele and isoform resolution using Smart-seq3. *Nature Biotechnology,* 38**,** 708-714.

HAIMON, Z., VOLASKI, A., ORTHGIESS, J., BOURA-HALFON, S., VAROL, D., SHEMER, A., YONA, S., ZUCKERMAN, B., DAVID, E., CHAPPELL-MAOR, L., BECHMANN, I., GERICKE, M., ULITSKY, I. & JUNG, S. 2018. Re-evaluating microglia expression profiles using RiboTag and cell isolation strategies. *Nature Immunology,* 19**,** 636-644.

HALEY, P. J. 2003. Species differences in the structure and function of the immune system. *Toxicology,* 188**,** 49-71.

HAMPTON, H. R., BAILEY, J., TOMURA, M., BRINK, R. & CHTANOVA, T. 2015. Microbe-dependent lymphatic migration of neutrophils modulates lymphocyte proliferation in lymph nodes. *Nat Commun,* 6**,** 7139.

HAN, X., WANG, R., ZHOU, Y., FEI, L., SUN, H., LAI, S., SAADATPOUR, A., ZHOU, Z., CHEN, H., YE, F., HUANG, D., XU, Y., HUANG, W., JIANG, M., JIANG, X., MAO, J., CHEN, Y., LU, C., XIE, J., FANG, Q., WANG, Y., YUE, R., LI, T., HUANG, H., ORKIN, S. H., YUAN, G. C., CHEN, M. & GUO, G. 2018. Mapping the Mouse Cell Atlas by Microwell-Seq. *Cell,* 172**,** 1091-1107.e17.

HAN, X., ZHOU, Z., FEI, L., SUN, H., WANG, R., CHEN, Y., CHEN, H., WANG, J., TANG, H., GE, W., ZHOU, Y., YE, F., JIANG, M., WU, J., XIAO, Y., JIA, X., ZHANG, T., MA, X., ZHANG, Q., BAI, X., LAI, S., YU, C., ZHU, L., LIN, R., GAO, Y., WANG, M., WU, Y., ZHANG, J., ZHAN, R., ZHU, S., HU, H., WANG, C., CHEN, M., HUANG, H., LIANG, T., CHEN, J., WANG, W., ZHANG, D. & GUO, G. 2020. Construction of a human cell landscape at single-cell level. *Nature,* 581**,** 303-309.

HANIDZIAR, D., NAKAHORI, Y., CAHILL, L. A., GALLO, D., KEEGAN, J. W., NGUYEN, J. P., OTTERBEIN, L. E., LEDERER, J. A. & ROBSON, S. C. 2020. Characterization of pulmonary immune responses to hyperoxia by high-dimensional mass cytometry analyses. *Scientific Reports,* 10**,** 4677.

HASHIMOTO, D., CHOW, A., NOIZAT, C., TEO, P., BEASLEY, MARY B., LEBOEUF, M., BECKER, CHRISTIAN D., SEE, P., PRICE, J., LUCAS, D., GRETER, M., MORTHA, A., BOYER, SCOTT W., FORSBERG, E. C., TANAKA, M., VAN ROOIJEN, N., GARCÍA-SASTRE, A., STANLEY, E. R., GINHOUX, F., FRENETTE, PAUL S. & MERAD, M. 2013. Tissue-Resident Macrophages Self-Maintain Locally throughout Adult Life with Minimal Contribution from Circulating Monocytes. *Immunity,* 38**,** 792-804.

HAWRYLYCZ, M. J., LEIN, E. S., GUILLOZET-BONGAARTS, A. L., SHEN, E. H., NG, L., MILLER, J. A., VAN DE LAGEMAAT, L. N., SMITH, K. A., EBBERT, A. & RILEY, Z. L. 2012. An anatomically comprehensive atlas of the adult human brain transcriptome. *Nature,* 489**,** 391-399.

HENRI, S., POULIN, L. F., TAMOUTOUNOUR, S., ARDOUIN, L., GUILLIAMS, M., DE BOVIS, B., DEVILARD, E., VIRET, C., AZUKIZAWA, H., KISSENPFENNIG, A.

& MALISSEN, B. 2010. CD207+ CD103+ dermal dendritic cells cross-present keratinocyte-derived antigens irrespective of the presence of Langerhans cells. *J Exp Med,* 207**,** 189-206.

HENRI, S., VREMEC, D., KAMATH, A., WAITHMAN, J., WILLIAMS, S., BENOIST, C., BURNHAM, K., SAELAND, S., HANDMAN, E. & SHORTMAN, K. 2001. The Dendritic Cell Populations of Mouse Lymph Nodes. *The Journal of Immunology,* 167**,** 741.

HERNDLER-BRANDSTETTER, D., ISHIGAME, H., SHINNAKASU, R., PLAJER, V., STECHER, C., ZHAO, J., LIETZENMAYER, M., KROEHLING, L., TAKUMI, A., KOMETANI, K., INOUE, T., KLUGER, Y., KAECH, S. M., KUROSAKI, T., OKADA, T. & FLAVELL, R. A. 2018. KLRG1(+) Effector CD8(+) T Cells Lose KLRG1, Differentiate into All Memory T Cell Lineages, and Convey Enhanced Protective Immunity. *Immunity,* 48**,** 716-729.e8.

HERRING, C. A., BANERJEE, A., MCKINLEY, E. T., SIMMONS, A. J., PING, J., ROLAND, J. T., FRANKLIN, J. L., LIU, Q., GERDES, M. J., COFFEY, R. J. & LAU, K. S. 2018. Unsupervised Trajectory Analysis of Single-Cell RNA-Seq and Imaging Data Reveals Alternative Tuft Cell Origins in the Gut. *Cell Systems,* 6**,** 37-51.e9.

HILL, A. A., REID BOLUS, W. & HASTY, A. H. 2014. A decade of progress in adipose tissue macrophage biology. *Immunol Rev,* 262**,** 134-52.

HIRAYAMA, D., IIDA, T. & NAKASE, H. 2017. The Phagocytic Function of Macrophage-Enforcing Innate Immunity and Tissue Homeostasis. *Int J Mol Sci,* 19.

HOBAN, M. D., ORKIN, S. H. & BAUER, D. E. 2016. Genetic treatment of a molecular disorder: gene therapy approaches to sickle cell disease. *Blood,* 127**,** 839-48.

HODGE, R. D., BAKKEN, T. E., MILLER, J. A., SMITH, K. A., BARKAN, E. R., GRAYBUCK, L. T., CLOSE, J. L., LONG, B., JOHANSEN, N., PENN, O., YAO, Z., EGGERMONT, J., HÖLLT, T., LEVI, B. P., SHEHATA, S. I., AEVERMANN, B., BELLER, A., BERTAGNOLLI, D., BROUNER, K., CASPER, T., COBBS, C., DALLEY, R., DEE, N., DING, S.-L., ELLENBOGEN, R. G., FONG, O., GARREN, E., GOLDY, J., GWINN, R. P., HIRSCHSTEIN, D., KEENE, C. D., KESHK, M., KO, A. L., LATHIA, K., MAHFOUZ, A., MALTZER, Z., MCGRAW, M., NGUYEN, T. N., NYHUS, J., OJEMANN, J. G., OLDRE, A., PARRY, S., REYNOLDS, S., RIMORIN, C., SHAPOVALOVA, N. V., SOMASUNDARAM, S., SZAFER, A., THOMSEN, E. R., TIEU, M., QUON, G., SCHEUERMANN, R. H., YUSTE, R., SUNKIN, S. M., LELIEVELDT, B., FENG, D., NG, L., BERNARD, A., HAWRYLYCZ, M., PHILLIPS, J. W., TASIC, B., ZENG, H., JONES, A. R., KOCH, C. & LEIN, E. S. 2019. Conserved cell types with divergent features in human versus mouse cortex. *Nature,* 573**,** 61-68.

HOEFFEL, G., CHEN, J., LAVIN, Y., LOW, D., ALMEIDA, FRANCISCA F., SEE, P., BEAUDIN, ANNA E., LUM, J., LOW, I., FORSBERG, E. C., POIDINGER, M., ZOLEZZI, F., LARBI, A., NG, LAI G., CHAN, JERRY K. Y., GRETER, M., BECHER, B., SAMOKHVALOV, IGOR M., MERAD, M. & GINHOUX, F. 2015. C-Myb+ Erythro-Myeloid Progenitor-Derived Fetal Monocytes Give Rise to Adult Tissue-Resident Macrophages. *Immunity,* 42**,** 665-678.

HOEFFEL, G., WANG, Y., GRETER, M., SEE, P., TEO, P., MALLERET, B., LEBOEUF, M., LOW, D., OLLER, G., ALMEIDA, F., CHOY, S. H., GRISOTTO, M., RENIA, L., CONWAY, S. J., STANLEY, E. R., CHAN, J. K., NG, L. G., SAMOKHVALOV, I. M., MERAD, M. & GINHOUX, F. 2012. Adult Langerhans cells derive predominantly from embryonic fetal liver monocytes with a minor contribution of yolk sac-derived macrophages. *J Exp Med,* 209**,** 1167-81.

HOGAN, S. P., ROSENBERG, H. F., MOQBEL, R., PHIPPS, S., FOSTER, P. S., LACY, P., KAY, A. B. & ROTHENBERG, M. E. 2008. Eosinophils: biological properties and role in health and disease. *Clin Exp Allergy,* 38**,** 709-50.

HONDA, K. & TANIGUCHI, T. 2006. IRFs: master regulators of signalling by Toll-like receptors and cytosolic pattern-recognition receptors. *Nat Rev Immunol,* 6**,** 644-58.

HOU, W., JI, Z., JI, H. & HICKS, S. C. 2020. A Systematic Evaluation of Single-cell RNA-sequencing Imputation Methods. *bioRxiv***,** 2020.01.29.925974.

HOU, Y., GUO, H., CAO, C., LI, X., HU, B., ZHU, P., WU, X., WEN, L., TANG, F., HUANG, Y. & PENG, J. 2016. Single-cell triple omics sequencing reveals genetic, epigenetic, and transcriptomic heterogeneity in hepatocellular carcinomas. *Cell Research,* 26**,** 304-319.

HSIAO, C. J., TUNG, P., BLISCHAK, J. D., BURNETT, J. E., BARR, K. A., DEY, K. K., STEPHENS, M. & GILAD, Y. 2020. Characterizing and inferring quantitative cell cycle phase in single-cell RNA-seq data analysis. *Genome Res,* 30**,** 611-621.

HU, Y., AN, Q., SHEU, K., TREJO, B., FAN, S. & GUO, Y. 2018. Single Cell Multi-Omics Technology: Methodology and Application. *Frontiers in Cell and Developmental Biology,* 6.

HUANG, M., WANG, J., TORRE, E., DUECK, H., SHAFFER, S., BONASIO, R., MURRAY, J. I., RAJ, A., LI, M. & ZHANG, N. R. 2018. SAVER: gene expression recovery for single-cell RNA sequencing. *Nat Methods,* 15**,** 539-542.

HWANG, B., LEE, J. H. & BANG, D. 2018. Single-cell RNA sequencing technologies and bioinformatics pipelines. *Experimental & Molecular Medicine,* 50**,** 96.

IACONO, G., MASSONI-BADOSA, R. & HEYN, H. 2019. Single-cell transcriptomics unveils gene regulatory network plasticity. *Genome Biology,* 20**,** 110.

IMAYOSHI, I., ISOMURA, A., HARIMA, Y., KAWAGUCHI, K., KORI, H., MIYACHI, H., FUJIWARA, T., ISHIDATE, F. & KAGEYAMA, R. 2013. Oscillatory Control of Factors Determining Multipotency and Fate in Mouse Neural Progenitors. *Science,* 342**,** 1203.

JEW, B., ALVAREZ, M., RAHMANI, E., MIAO, Z., KO, A., GARSKE, K. M., SUL, J. H., PIETILÄINEN, K. H., PAJUKANTA, P. & HALPERIN, E. 2020. Accurate estimation of cell composition in bulk expression through robust integration of single-cell information. *Nature Communications,* 11**,** 1971.

JIANG, X., PARK, C. O., GEDDES SWEENEY, J., YOO, M. J., GAIDE, O. & KUPPER, T. S. 2017. Dermal γδ T Cells Do Not Freely Re-Circulate Out of Skin and Produce IL-17 to Promote Neutrophil Infiltration during Primary Contact Hypersensitivity. *PLoS One,* 12**,** e0169397.

JONES, E. G. 1999. Colgi, Cajal and the Neuron Doctrine. *J Hist Neurosci,* 8**,** 170-8.

JONGBLOED, S. L., KASSIANOS, A. J., MCDONALD, K. J., CLARK, G. J., JU, X., ANGEL, C. E., CHEN, C.-J. J., DUNBAR, P. R., WADLEY, R. B. & JEET, V. 2010. Human CD141+ (BDCA-3)+ dendritic cells (DCs) represent a unique myeloid DC subset that cross-presents necrotic cell antigens. *Journal of Experimental Medicine,* 207**,** 1247-1260.

JUBRAIL, J., KURIAN, N. & NIEDERGANG, F. 2017. Macrophage phagocytosis cracking the defect code in COPD. *Biomedical journal,* 40**,** 305-312.

JUNG, D. & ALT, F. W. 2004. Unraveling V(D)J Recombination: Insights into Gene Regulation. *Cell,* 116**,** 299-311.

KANEHISA, M. & GOTO, S. 2000. KEGG: Kyoto Encyclopedia of Genes and Genomes. *Nucleic Acids Research,* 28**,** 27-30.

KANG, K., MENG, Q., SHATS, I., UMBACH, D. M., LI, M., LI, Y., LI, X. & LI, L. 2019. CDSeq: A novel complete deconvolution method for dissecting heterogeneous samples using gene expression data. *PLOS Computational Biology,* 15**,** e1007510.

KAWASAKI, T. & KAWAI, T. 2014. Toll-Like Receptor Signaling Pathways. *Frontiers in Immunology,* 5.

KEMP, R. A., BÄCKSTRÖM, B. T. & RONCHESE, F. 2005. The phenotype of type 1 and type 2 CD8+ T cells activated in vitro is affected by culture conditions and correlates with effector activity. *Immunology,* 115**,** 315-24.

KEREN-SHAUL, H., SPINRAD, A., WEINER, A., MATCOVITCH-NATAN, O., DVIR-SZTERNFELD, R., ULLAND, T. K., DAVID, E., BARUCH, K., LARA-ASTAISO, D., TOTH, B., ITZKOVITZ, S., COLONNA, M., SCHWARTZ, M. & AMIT, I. 2017. A Unique Microglia Type Associated with Restricting Development of Alzheimer's Disease. *Cell,* 169**,** 1276-1290.e17.

KIM, C. & WILLIAMS, M. A. 2010. Nature and nurture: T-cell receptor-dependent and T-cell receptor-independent differentiation cues in the selection of the memory T-cell pool. *Immunology,* 131**,** 310-7.

KNY, M., CSÁLYI, K. D., KLAESKE, K., BUSCH, K., MEYER, A. M., MERKS, A. M., DARM, K., DWORATZEK, E., FLIEGNER, D., BACZKO, I., REGITZ-ZAGROSEK, V., BUTTER, C., LUFT, F. C., PANÁKOVÁ, D. & FIELITZ, J. 2019. Ninjurin1 regulates striated muscle growth and differentiation. *PLOS ONE,* 14**,** e0216987.

KOH, J., KIM, S., KIM, M. Y., GO, H., JEON, Y. K. & CHUNG, D. H. 2017. Prognostic implications of intratumoral CD103+ tumor-infiltrating lymphocytes in pulmonary squamous cell carcinoma. *Oncotarget,* 8**,** 13762-13769.

KOHYAMA, M., ISE, W., EDELSON, B. T., WILKER, P. R., HILDNER, K., MEJIA, C., FRAZIER, W. A., MURPHY, T. L. & MURPHY, K. M. 2009. Role for Spi-C in the development of red pulp macrophages and splenic iron homeostasis. *Nature,* 457**,** 318-21.

KORIN, B., BEN-SHAANAN, T. L., SCHILLER, M., DUBOVIK, T., AZULAY-DEBBY, H., BOSHNAK, N. T., KOREN, T. & ROLLS, A. 2017. High-dimensional, single-cell characterization of the brain's immune compartment. *Nature Neuroscience,* 20**,** 1300-1309.

KRASTEVA, G., PFEIL, U., DRAB, M., KUMMER, W. & KÖNIG, P. 2006. Caveolin-1 and -2 in airway epithelium: expression and in situ association as detected by FRET-CLSM. *Respiratory Research,* 7**,** 108.

KRAUSGRUBER, T., FORTELNY, N., FIFE-GERNEDL, V., SENEKOWITSCH, M., SCHUSTER, L. C., LERCHER, A., NEMC, A., SCHMIDL, C., RENDEIRO, A. F., BERGTHALER, A. & BOCK, C. 2020. Structural cells are key regulators of organ-specific immune responses. *Nature,* 583**,** 296-302.

KRIEG, C., NOWICKA, M., GUGLIETTA, S., SCHINDLER, S., HARTMANN, F. J., WEBER, L. M., DUMMER, R., ROBINSON, M. D., LEVESQUE, M. P. & BECHER, B. 2018. High-dimensional single-cell analysis predicts response to anti-PD-1 immunotherapy. *Nature Medicine,* 24**,** 144-153.

KUMAR, B. V., MA, W., MIRON, M., GRANOT, T., GUYER, R. S., CARPENTER, D. J., SENDA, T., SUN, X., HO, S.-H., LERNER, H., FRIEDMAN, A. L., SHEN, Y. & FARBER, D. L. 2017. Human Tissue-Resident Memory T Cells Are Defined by Core Transcriptional and Functional Signatures in Lymphoid and Mucosal Sites. *Cell Reports,* 20**,** 2921-2934.

L. LUN, A. T., BACH, K. & MARIONI, J. C. 2016. Pooling across cells to normalize single-cell RNA sequencing data with many zero counts. *Genome Biology,* 17**,** 75.

LANDRITH, T. A., SURESHCHANDRA, S., RIVERA, A., JANG, J. C., RAIS, M., NAIR, M. G., MESSAOUDI, I. & WILSON, E. H. 2017. CD103(+) CD8 T Cells in the Toxoplasma-Infected Brain Exhibit a Tissue-Resident Memory Transcriptional Profile. *Front Immunol,* 8**,** 335.

LANGFELDER, P. & HORVATH, S. 2008. WGCNA: an R package for weighted correlation network analysis. *BMC Bioinformatics,* 9**,** 559.

LAVIN, Y., WINTER, D., BLECHER-GONEN, R., DAVID, E., KEREN-SHAUL, H., MERAD, M., JUNG, S. & AMIT, I. 2014a. Tissue-resident macrophage enhancer landscapes are shaped by the local microenvironment. *Cell,* 159**,** 1312-26.

LAVIN, Y., WINTER, D., BLECHER-GONEN, R., DAVID, E., KEREN-SHAUL, H., MERAD, M., JUNG, S. & AMIT, I. 2014b. Tissue-Resident Macrophage Enhancer Landscapes Are Shaped by the Local Microenvironment. *Cell,* 159**,** 1312-1326.

LAWSON, L., PERRY, V., DRI, P. & GORDON, S. 1990. Heterogeneity in the distribution and morphology of microglia in the normal adult mouse brain. *Neuroscience,* 39**,** 151-170.

LEBIEN, T. W. & TEDDER, T. F. 2008. B lymphocytes: how they develop and function. *Blood,* 112**,** 1570-1580.

LEDERGOR, G., WEINER, A., ZADA, M., WANG, S. Y., COHEN, Y. C., GATT, M. E., SNIR, N., MAGEN, H., KOREN-MICHOWITZ, M., HERZOG-TZARFATI, K., KEREN-SHAUL, H., BORNSTEIN, C., ROTKOPF, R., YOFE, I., DAVID, E., YELLAPANTULA, V., KAY, S., SALAI, M., BEN YEHUDA, D., NAGLER, A., SHVIDEL, L., ORR-URTREGER, A., HALPERN, K. B., ITZKOVITZ, S., LANDGREN, O., SAN-MIGUEL, J., PAIVA, B., KEATS, J. J., PAPAEMMANUIL, E., AVIVI, I., BARBASH, G. I., TANAY, A. & AMIT, I. 2018. Single cell dissection of plasma cell heterogeneity in symptomatic and asymptomatic myeloma. *Nat Med,* 24**,** 1867-1876.

LEIPOLD, M. D. 2015. Another step on the path to mass cytometry standardization. *Cytometry Part A,* 87**,** 380-382.

LEONE, C., LE PAVEC, G., MEME, W., PORCHERAY, F., SAMAH, B., DORMONT, D. & GRAS, G. 2006. Characterization of human monocyte-derived microglia-like cells. *Glia,* 54**,** 183-92.

LEONI, G., GRIPENTROG, J., LORD, C., RIESSELMAN, M., SUMAGIN, R., PARKOS, C. A., NUSRAT, A. & JESAITIS, A. J. 2015. Human neutrophil formyl peptide receptor phosphorylation and the mucosal inflammatory response. *J Leukoc Biol,* 97**,** 87-101.

LESKOVEC, J. & SOSI, R. 2016. SNAP: A General-Purpose Network Analysis and Graph-Mining Library. *ACM Trans. Intell. Syst. Technol.,* 8**,** 1-20.

LEVINE, JACOB H., SIMONDS, ERIN F., BENDALL, SEAN C., DAVIS, KARA L., AMIR, E.-AD D., TADMOR, MICHELLE D., LITVIN, O., FIENBERG, HARRIS G., JAGER, A., ZUNDER, ELI R., FINCK, R., GEDMAN, AMANDA L., RADTKE, I., DOWNING, JAMES R., PE'ER, D. & NOLAN, GARRY P. 2015. Data-Driven Phenotypic Dissection of AML Reveals Progenitor-like Cells that Correlate with Prognosis. *Cell,* 162**,** 184-197.

LI, J., CAO, F., YIN, H.-L., HUANG, Z.-J., LIN, Z.-T., MAO, N., SUN, B. & WANG, G. 2020. Ferroptosis: past, present and future. *Cell Death & Disease,* 11**,** 88.

LI, J., ZHOU, D., QIU, W., SHI, Y., YANG, J.-J., CHEN, S., WANG, Q. & PAN, H. 2018. Application of Weighted Gene Co-expression Network Analysis for Data from Paired Design. *Scientific Reports,* 8**,** 622.

LI, X., LIU, R., SU, X., PAN, Y., HAN, X., SHAO, C. & SHI, Y. 2019. Harnessing tumor-associated macrophages as aids for cancer immunotherapy. *Molecular Cancer,* 18**,** 177.

LI, Y., QI, X., LIU, B. & HUANG, H. 2015. The STAT5-GATA2 pathway is critical in basophil and mast cell differentiation and maintenance. *J Immunol,* 194**,** 4328-38.

LIBERZON, A., SUBRAMANIAN, A., PINCHBACK, R., THORVALDSDOTTIR, H., TAMAYO, P. & MESIROV, J. P. 2011. Molecular signatures database (MSigDB) 3.0. *Bioinformatics,* 27**,** 1739-40.

LIN, P., TROUP, M. & HO, J. W. K. 2017. CIDR: Ultrafast and accurate clustering through imputation for single-cell RNA-seq data. *Genome Biology,* 18**,** 59.

LIN, Y., XU, J. & LAN, H. 2019. Tumor-associated macrophages in tumor metastasis: biological roles and clinical therapeutic applications. *Journal of Hematology & Oncology,* 12**,** 76.

LIU, S., CAI, X., WU, J., CONG, Q., CHEN, X., LI, T., DU, F., REN, J., WU, Y.-T., GRISHIN, N. V. & CHEN, Z. J. 2015. Phosphorylation of innate immune adaptor proteins MAVS, STING, and TRIF induces IRF3 activation. *Science,* 347**,** aaa2630.

LIU, W., YAN, M., LIU, Y., MCLEISH, K. R., COLEMAN, W. G., JR. & RODGERS, G. P. 2012. Olfactomedin 4 inhibits cathepsin C-mediated protease activities, thereby modulating neutrophil killing of Staphylococcus aureus and Escherichia coli in mice. *J Immunol,* 189**,** 2460-7.

LIU, Y., ARYEE, M. J., PADYUKOV, L., FALLIN, M. D., HESSELBERG, E., RUNARSSON, A., REINIUS, L., ACEVEDO, N., TAUB, M., RONNINGER, M., SHCHETYNSKY, K., SCHEYNIUS, A., KERE, J., ALFREDSSON, L., KLARESKOG, L., EKSTRÖM, T. J. & FEINBERG, A. P. 2013. Epigenome-wide association data implicate DNA methylation as an intermediary of genetic risk in rheumatoid arthritis. *Nature Biotechnology,* 31**,** 142.

LIVIGNI, A., O'HARA, L., POLAK, M. E., ANGUS, T., WRIGHT, D. W., SMITH, L. B. & FREEMAN, T. C. 2018. A graphical and computational modeling platform for biological pathways. *Nature Protocols,* 13**,** 705-722.

LONSDALE, J., THOMAS, J., SALVATORE, M., PHILLIPS, R., LO, E., SHAD, S., HASZ, R., WALTERS, G., GARCIA, F. & YOUNG, N. 2013. The genotype-tissue expression (GTEx) project. *Nat Genet,* 45**,** 580-585.

LOTFOLLAHI, M., WOLF, F. A. & THEIS, F. J. 2018. Generative modeling and latent space arithmetics predict single-cell perturbation response across cell types, studies and species. *bioRxiv***,** 478503.

LU, J., MENG, H., ZHANG, A., YANG, J. & ZHANG, X. 2015. Phenotype and function of tissue-resident unconventional Foxp3-expressing CD4+ regulatory T cells. *Cellular Immunology,* 297**,** 53-59.

LUBBERS, E. R. & MOHLER, P. J. 2016. Roles and regulation of protein phosphatase 2A (PP2A) in the heart. *J Mol Cell Cardiol,* 101**,** 127-133.

LUBECK, E. & CAI, L. 2012. Single-cell systems biology by super-resolution imaging and combinatorial labeling. *Nature Methods,* 9**,** 743.

LUND, H., PIEBER, M., PARSA, R., HAN, J., GROMMISCH, D., EWING, E., KULAR, L., NEEDHAMSEN, M., ESPINOSA, A., NILSSON, E., OVERBY, A. K., BUTOVSKY, O., JAGODIC, M., ZHANG, X. M. & HARRIS, R. A. 2018. Competitive repopulation of an empty microglial niche yields functionally distinct subsets of microglia-like cells. *Nat Commun,* 9**,** 4845.

LURIER, E. B., DALTON, D., DAMPIER, W., RAMAN, P., NASSIRI, S., FERRARO, N. M., RAJAGOPALAN, R., SARMADY, M. & SPILLER, K. L. 2017. Transcriptome analysis of IL-10-stimulated (M2c) macrophages by next-generation sequencing. *Immunobiology,* 222**,** 847-856.

LYNCH, L., MICHELET, X., ZHANG, S., BRENNAN, P. J., MOSEMAN, A., LESTER, C., BESRA, G., VOMHOF-DEKREY, E. E., TIGHE, M., KOAY, H. F., GODFREY, D. I., LEADBETTER, E. A., SANT'ANGELO, D. B., VON ANDRIAN, U. & BRENNER, M. B. 2015. Regulatory iNKT cells lack expression of the transcription factor PLZF and control the homeostasis of T(reg) cells and macrophages in adipose tissue. *Nat Immunol,* 16**,** 85-95.

MAATEN, L. V. D. & HINTON, G. 2008. Visualizing data using t-SNE. *Journal of machine learning research,* 9**,** 2579-2605.

MABBOTT, N. A., BAILLIE, J. K., BROWN, H., FREEMAN, T. C. & HUME, D. A. 2013. An expression atlas of human primary cells: inference of gene function from coexpression networks. *BMC Genomics,* 14**,** 632.

MACKAY, L. K., MINNICH, M., KRAGTEN, N. A., LIAO, Y., NOTA, B., SEILLET, C., ZAID, A., MAN, K., PRESTON, S., FREESTONE, D., BRAUN, A., WYNNE-JONES, E., BEHR, F. M., STARK, R., PELLICCI, D. G., GODFREY, D. I., BELZ, G. T., PELLEGRINI, M., GEBHARDT, T., BUSSLINGER, M., SHI, W., CARBONE, F. R., VAN LIER, R. A., KALLIES, A. & VAN GISBERGEN, K. P. 2016. Hobit and Blimp1 instruct a universal transcriptional program of tissue residency in lymphocytes. *Science,* 352**,** 459-63.

MACKAY, L. K., RAHIMPOUR, A., MA, J. Z., COLLINS, N., STOCK, A. T., HAFON, M. L., VEGA-RAMOS, J., LAUZURICA, P., MUELLER, S. N., STEFANOVIC, T., TSCHARKE, D. C., HEATH, W. R., INOUYE, M., CARBONE, F. R. & GEBHARDT, T. 2013. The developmental pathway for CD103(+)CD8+ tissue-resident memory T cells of skin. *Nat Immunol,* 14**,** 1294-301.

MACPARLAND, S. A., LIU, J. C., MA, X.-Z., INNES, B. T., BARTCZAK, A. M., GAGE, B. K., MANUEL, J., KHUU, N., ECHEVERRI, J., LINARES, I., GUPTA, R., CHENG, M. L., LIU, L. Y., CAMAT, D., CHUNG, S. W., SELIGA, R. K., SHAO, Z., LEE, E., OGAWA, S., OGAWA, M., WILSON, M. D., FISH, J. E., SELZNER, M., GHANEKAR, A., GRANT, D., GREIG, P., SAPISOCHIN, G., SELZNER, N., WINEGARDEN, N., ADEYI, O., KELLER, G., BADER, G. D. & MCGILVRAY, I. D. 2018. Single cell RNA sequencing of human liver reveals distinct intrahepatic macrophage populations. *Nature Communications,* 9**,** 4383.

MAKINO, Y., KANNO, R., ITO, T., HIGASHINO, K. & TANIGUCHI, M. 1995. Predominant expression of invariant V alpha 14+ TCR alpha chain in NK1.1+ T cell populations. *Int Immunol,* 7**,** 1157-61.

MARTINEZ, F. O., SICA, A., MANTOVANI, A. & LOCATI, M. 2008. Macrophage activation and polarization. *Front Biosci,* 13**,** 453-61.

MARTINEZ-POMARES, L. & GORDON, S. 2012. CD169+ macrophages at the crossroads of antigen presentation. *Trends Immunol,* 33**,** 66-70.

MARTISKAINEN, H., VISWANATHAN, J., NYKÄNEN, N.-P., KURKI, M., HELISALMI, S., NATUNEN, T., SARAJÄRVI, T., KURKINEN, K. M. A., PURSIHEIMO, J.-P., RAURAMAA, T., ALAFUZOFF, I., JÄÄSKELÄINEN, J. E., LEINONEN, V., SOININEN, H., HAAPASALO, A., HUTTUNEN, H. J. & HILTUNEN, M. 2015. Transcriptomics and mechanistic elucidation of Alzheimer's disease risk genes in the brain and in vitro models. *Neurobiology of Aging,* 36**,** 1221.e15-1221.e28.

MARX, V. 2017. How to deduplicate PCR. *Nature Methods,* 14**,** 473.

MASOPUST, D. & SOERENS, A. G. 2019. Tissue-Resident T Cells and Other Resident Leukocytes. *Annual Review of Immunology,* 37**,** 521-546.

MASS, E. 2018. Delineating the origins, developmental programs and homeostatic functions of tissue-resident macrophages. *International Immunology,* 30**,** 493-501.

MASS, E., BALLESTEROS, I., FARLIK, M., HALBRITTER, F., GÜNTHER, P., CROZET, L., JACOME-GALARZA, C. E., HÄNDLER, K., KLUGHAMMER, J., KOBAYASHI, Y., GOMEZ-PERDIGUERO, E., SCHULTZE, J. L., BEYER, M., BOCK, C. & GEISSMANN, F. 2016. Specification of tissue-resident macrophages during organogenesis. *Science,* 353**,** aaf4238.

MASSENA, S., CHRISTOFFERSSON, G., VÅGESJÖ, E., SEIGNEZ, C., GUSTAFSSON, K., BINET, F., HERRERA HIDALGO, C., GIRAUD, A., LOMEI, J., WESTRÖM, S., SHIBUYA, M., CLAESSON-WELSH, L., GERWINS, P., WELSH, M., KREUGER, J. & PHILLIPSON, M. 2015. Identification and characterization of VEGF-A-responsive neutrophils expressing CD49d, VEGFR1, and CXCR4 in mice and humans. *Blood,* 126**,** 2016-26.

MASUDA, T., SANKOWSKI, R., STASZEWSKI, O. & PRINZ, M. 2020. Microglia Heterogeneity in the Single-Cell Era. *Cell Reports,* 30**,** 1271-1281.

MATHYS, H., DAVILA-VELDERRAIN, J., PENG, Z., GAO, F., MOHAMMADI, S., YOUNG, J. Z., MENON, M., HE, L., ABDURROB, F., JIANG, X., MARTORELL, A. J., RANSOHOFF, R. M., HAFLER, B. P., BENNETT, D. A., KELLIS, M. & TSAI, L.-H. 2019. Single-cell transcriptomic analysis of Alzheimer's disease. *Nature,* 570**,** 332-337.

MCCRAITH, S., HOLTZMAN, T., MOSS, B. & FIELDS, S. 2000. Genome-wide analysis of vaccinia virus protein–protein interactions. *Proceedings of the National Academy of Sciences,* 97**,** 4879-4884.

MCGRATH, KATHLEEN E., FRAME, JENNA M., FEGAN, KATHERINE H., BOWEN, JAMES R., CONWAY, SIMON J., CATHERMAN, SEANA C., KINGSLEY, PAUL D., KONISKI, ANNE D. & PALIS, J. 2015. Distinct Sources of Hematopoietic Progenitors Emerge before HSCs and Provide Functional Blood Cells in the Mammalian Embryo. *Cell Reports,* 11**,** 1892-1904.

MCKENNA, H. J., STOCKING, K. L., MILLER, R. E., BRASEL, K., DE SMEDT, T., MARASKOVSKY, E., MALISZEWSKI, C. R., LYNCH, D. H., SMITH, J., PULENDRAN, B., ROUX, E. R., TEEPE, M., LYMAN, S. D. & PESCHON, J. J. 2000. Mice lacking flt3 ligand have deficient hematopoiesis affecting hematopoietic progenitor cells, dendritic cells, and natural killer cells. *Blood,* 95**,** 3489-97.

MEHTA, D., JACKSON, R., PAUL, G., SHI, J. & SABBAGH, M. 2017. Why do trials for Alzheimer's disease drugs keep failing? A discontinued drug perspective for 2010-2015. *Expert Opin Investig Drugs,* 26**,** 735-739.

MENEZES, S., MELANDRI, D., ANSELMI, G., PERCHET, T., LOSCHKO, J., DUBROT, J., PATEL, R., GAUTIER, E. L., HUGUES, S., LONGHI, M. P., HENRY, J. Y., QUEZADA, S. A., LAUVAU, G., LENNON-DUMÉNIL, A. M., GUTIÉRREZ-MARTÍNEZ, E., BESSIS, A., GOMEZ-PERDIGUERO, E., JACOME-GALARZA, C. E., GARNER, H., GEISSMANN, F., GOLUB, R., NUSSENZWEIG, M. C. & GUERMONPREZ, P. 2016. The Heterogeneity of Ly6C(hi) Monocytes Controls Their Differentiation into iNOS(+) Macrophages or Monocyte-Derived Dendritic Cells. *Immunity,* 45**,** 1205-1218.

MERAD, M., MANZ, M. G., KARSUNKY, H., WAGERS, A., PETERS, W., CHARO, I., WEISSMAN, I. L., CYSTER, J. G. & ENGLEMAN, E. G. 2002. Langerhans cells renew in the skin throughout life under steady-state conditions. *Nature Immunology,* 3**,** 1135.

MERICO, D., ISSERLIN, R., STUEKER, O., EMILI, A. & BADER, G. D. 2010. Enrichment Map: A Network-Based Method for Gene-Set Enrichment Visualization and Interpretation. *PLOS ONE,* 5**,** e13984.

MILDNER, A., SCHÖNHEIT, J., GILADI, A., DAVID, E., LARA-ASTIASO, D., LORENZO-VIVAS, E., PAUL, F., CHAPPELL-MAOR, L., PRILLER, J., LEUTZ, A., AMIT, I. & JUNG, S. 2017. Genomic Characterization of Murine Monocytes Reveals C/EBPβ Transcription Factor Dependence of Ly6C− Cells. *Immunity,* 46**,** 849-862.e7.

MILLER, J. A., HORVATH, S. & GESCHWIND, D. H. 2010. Divergence of human and mouse brain transcriptome highlights Alzheimer disease pathways. *Proceedings of the National Academy of Sciences,* 107**,** 12698-12703.

MILLS, C. D., KINCAID, K., ALT, J. M., HEILMAN, M. J. & HILL, A. M. 2000. M-1/M-2 macrophages and the Th1/Th2 paradigm. *J Immunol,* 164**,** 6166-73.

MIRAGAIA, R. J., GOMES, T., CHOMKA, A., JARDINE, L., RIEDEL, A., HEGAZY, A. N., WHIBLEY, N., TUCCI, A., CHEN, X., LINDEMAN, I., EMERTON, G., KRAUSGRUBER, T., SHIELDS, J., HANIFFA, M., POWRIE, F. & TEICHMANN,

S. A. 2019. Single-Cell Transcriptomics of Regulatory T Cells Reveals Trajectories of Tissue Adaptation. *Immunity,* 50**,** 493-504.e7.

MISIAK, A., WILK, M. M., RAVERDEAU, M. & MILLS, K. H. 2017. IL-17-Producing Innate and Pathogen-Specific Tissue Resident Memory γδ T Cells Expand in the Lungs of Bordetella pertussis-Infected Mice. *J Immunol,* 198**,** 363-374.

MIURA, SATORU K., MARTINS, A., ZHANG, KELVIN X., GRAVELEY, BRENTON R. & ZIPURSKY, S. L. 2013. Probabilistic Splicing of Dscam1 Establishes Identity at the Level of Single Neurons. *Cell,* 155**,** 1166-1177.

MONACO, G., LEE, B., XU, W., MUSTAFAH, S., HWANG, Y. Y., CARRE, C., BURDIN, N., VISAN, L., CECCARELLI, M. & POIDINGER, M. 2019. RNA-Seq signatures normalized by mRNA abundance allow absolute deconvolution of human immune cell types. *Cell reports,* 26**,** 1627-1640. e7.

MOSSER, D. M. & EDWARDS, J. P. 2008. Exploring the full spectrum of macrophage activation. *Nature Reviews Immunology,* 8**,** 958-969.

MUNDO, A. F., JASSAL, B., MAY, B., CROFT, D., BIRNEY, E., WU, G., SONG, H., WEISER, J., ROTHFELS, K., MATTHEWS, L., GILLESPIE, M., MILACIC, M., WILLIAMS, M., KAMDAR, M. R., CAUDY, M., GARAPATI, P., HAW, R., PALATNIK, S., JUPE, S., SHAMOVSKY, V., HERMJAKOB, H., STEIN, L. & D'EUSTACHIO, P. 2013. The Reactome pathway knowledgebase. *Nucleic Acids Research,* 42**,** D472-D477.

MUNRO, D. A. D. & HUGHES, J. 2017. The Origins and Functions of Tissue-Resident Macrophages in Kidney Development. *Front Physiol,* 8**,** 837.

MURPHY, T. L., GRAJALES-REYES, G. E., WU, X., TUSSIWAND, R., BRISEÑO, C. G., IWATA, A., KRETZER, N. M., DURAI, V. & MURPHY, K. M. 2016. Transcriptional control of dendritic cell development. *Annual review of immunology,* 34**,** 93-119.

NAIK, S. H., PERIÉ, L., SWART, E., GERLACH, C., VAN ROOIJ, N., DE BOER, R. J. & SCHUMACHER, T. N. 2013. Diverse and heritable lineage imprinting of early haematopoietic progenitors. *Nature,* 496**,** 229-32.

NARAN, K., NUNDALALL, T., CHETTY, S. & BARTH, S. 2018. Principles of Immunotherapy: Implications for Treatment Strategies in Cancer and Infectious Diseases. *Front Microbiol,* 9**,** 3158.

NGUYEN, K. D., QIU, Y., CUI, X., GOH, Y. P., MWANGI, J., DAVID, T., MUKUNDAN, L., BROMBACHER, F., LOCKSLEY, R. M. & CHAWLA, A. 2011. Alternatively activated macrophages produce catecholamines to sustain adaptive thermogenesis. *Nature,* 480**,** 104-8.

NIEDZIELSKA, M., ISRAELSSON, E., ANGERMANN, B., SIDDERS, B. S., CLAUSEN, M., CATLEY, M., MALHOTRA, R. & DUMONT, C. 2018. Differential gene expression in human tissue resident regulatory T cells from lung, colon, and blood. *Oncotarget,* 9**,** 36166-36184.

NIRMAL, A. J., REGAN, T., SHIH, B. B., HUME, D. A., SIMS, A. H. & FREEMAN, T. C. 2018. Immune Cell Gene Signatures for Profiling the Microenvironment of Solid Tumors. *Cancer Immunol Res,* 6**,** 1388-1400.

NIRMAL, A. J., REGAN, T., SHIH, B. B.-J., HUME, D. A., SIMS, A. H. & FREEMAN, T. C. 2016. ImSig: A resource for the identification and quantification of immune signatures in blood and tissue transcriptomics data. *bioRxiv***,** 077487.

NOTTA, F., ZANDI, S., TAKAYAMA, N., DOBSON, S., GAN, O. I., WILSON, G., KAUFMANN, K. B., MCLEOD, J., LAURENTI, E., DUNANT, C. F., MCPHERSON, J. D., STEIN, L. D., DROR, Y. & DICK, J. E. 2016. Distinct routes of lineage development reshape the human blood hierarchy across ontogeny. *Science,* 351**,** aab2116.

ODEGAARD, J. I. & CHAWLA, A. 2015. Type 2 responses at the interface between immunity and fat metabolism. *Curr Opin Immunol,* 36**,** 67-72.

OEZCIMEN, A., BRAZMA, A., HOLLOWAY, E., LARA, G. G., PARKINSON, H., VILO, J., KAPUSHESKY, M., SHOJATALAB, M., ABEYGUNAWARDENA, N., KEMMEREN, P., ROCCA-SERRA, P., SANSONE, S.-A. & SARKANS, U. 2003. ArrayExpress—a public repository for microarray gene expression data at the EBI. *Nucleic Acids Research,* 31**,** 68-71.

OGURO, H., DING, L. & MORRISON, S. J. 2013. SLAM family markers resolve functionally distinct subpopulations of hematopoietic stem cells and multipotent progenitors. *Cell Stem Cell,* 13**,** 102-16.

OLSSON, A., VENKATASUBRAMANIAN, M., CHAUDHRI, V. K., ARONOW, B. J., SALOMONIS, N., SINGH, H. & GRIMES, H. L. 2016. Single-cell analysis of mixed-lineage states leading to a binary cell fate choice. *Nature,* 537**,** 698-702.

PALIS, J., ROBERTSON, S., KENNEDY, M., WALL, C. & KELLER, G. 1999. Development of erythroid and myeloid progenitors in the yolk sac and embryo proper of the mouse. *Development,* 126**,** 5073-84.

PAMPALONI, F., REYNAUD, E. G. & STELZER, E. H. K. 2007. The third dimension bridges the gap between cell culture and live tissue. *Nature Reviews Molecular Cell Biology,* 8**,** 839-845.

PAN, J.-B., HU, S.-C., SHI, D., CAI, M.-C., LI, Y.-B., ZOU, Q. & JI, Z.-L. 2013. PaGenBase: A Pattern Gene Database for the Global and Dynamic Understanding of Gene Function. *PLOS ONE,* 8**,** e80747.

PAPASSOTIROPOULOS, A., FOUNTOULAKIS, M., DUNCKLEY, T., STEPHAN, D. A. & REIMAN, E. M. 2006. Genetics, transcriptomics, and proteomics of Alzheimer's disease. *J Clin Psychiatry,* 67**,** 652-70.

PAPILI GAO, N., UD-DEAN, S. M. M., GANDRILLON, O. & GUNAWAN, R. 2017. SINCERITIES: inferring gene regulatory networks from time-stamped single cell transcriptional expression profiles. *Bioinformatics,* 34**,** 258-266.

PATIR, A., FRASER, A. M., BARNETT, M. W., MCTEIR, L., RAINGER, J., DAVEY, M. G. & FREEMAN, T. C. 2020. The transcriptional signature associated with human motile cilia. *Scientific Reports,* 10**,** 10814.

PATIR, A., SHIH, B., MCCOLL, B. W. & FREEMAN, T. C. 2019. A core transcriptional signature of human microglia: Derivation and utility in describing region-dependent alterations associated with Alzheimer's disease. *Glia,* 67**,** 1240-1253.

PAVELKO, K. D., BELL, M. P., HARRINGTON, S. M. & DONG, H. 2017. B7-H1 Influences the Accumulation of Virus-Specific Tissue Resident Memory T Cells in the Central Nervous System. *Front Immunol,* 8**,** 1532.

PEEL, A. L., ZOLOTUKHIN, S., SCHRIMSHER, G. W., MUZYCZKA, N. & REIER, P. J. 1997. Efficient transduction of green fluorescent protein in spinal cord neurons using adeno-associated virus vectors containing cell type-specific promoters. *Gene Therapy,* 4**,** 16-24.

PELLIN, D., LOPERFIDO, M., BARICORDI, C., WOLOCK, S. L., MONTEPELOSO, A., WEINBERG, O. K., BIFFI, A., KLEIN, A. M. & BIASCO, L. 2019. A comprehensive single cell transcriptional landscape of human hematopoietic progenitors. *Nature Communications,* 10**,** 2395.

PENDE, D., FALCO, M., VITALE, M., CANTONI, C., VITALE, C., MUNARI, E., BERTAINA, A., MORETTA, F., DEL ZOTTO, G., PIETRA, G., MINGARI, M. C., LOCATELLI, F. & MORETTA, L. 2019. Killer Ig-Like Receptors (KIRs): Their Role in NK Cell Modulation and Developments Leading to Their Clinical Exploitation. *Frontiers in Immunology,* 10.

PERRY, V. H. & HOLMES, C. 2014. Microglial priming in neurodegenerative disease. *Nature Reviews Neurology,* 10**,** 217-224.

PICCIOLI, D., TAVARINI, S., BORGOGNI, E., STERI, V., NUTI, S., SAMMICHELI, C., BARDELLI, M., MONTAGNA, D., LOCATELLI, F. & WACK, A. 2007. Functional specialization of human circulating CD16 and CD1c myeloid dendritic-cell subsets. *Blood,* 109**,** 5371-9.

PIZZOLLA, A., NGUYEN, T. H. O., SMITH, J. M., BROOKS, A. G., KEDZIESKA, K., HEATH, W. R., READING, P. C. & WAKIM, L. M. 2017. Resident memory CD8(+) T cells in the upper respiratory tract prevent pulmonary influenza virus infection. *Sci Immunol,* 2.

PRATAPA, A., JALIHAL, A. P., LAW, J. N., BHARADWAJ, A. & MURALI, T. M. 2020. Benchmarking algorithms for gene regulatory network inference from single-cell transcriptomic data. *Nature Methods,* 17**,** 147-154.

PRICE, J. G., IDOYAGA, J., SALMON, H., HOGSTAD, B., BIGARELLA, C. L., GHAFFARI, S., LEBOEUF, M. & MERAD, M. 2015. CDKN1A regulates Langerhans cell survival and promotes Treg cell generation upon exposure to ionizing irradiation. *Nature Immunology,* 16**,** 1060.

PUELLMANN, K., KAMINSKI, W. E., VOGEL, M., NEBE, C. T., SCHROEDER, J., WOLF, H. & BEHAM, A. W. 2006. A variable immunoreceptor in a subpopulation

of human neutrophils. *Proceedings of the National Academy of Sciences,* 103**,** 14441-14446.

PUGA, I., COLS, M., BARRA, C. M., HE, B., CASSIS, L., GENTILE, M., COMERMA, L., CHORNY, A., SHAN, M., XU, W., MAGRI, G., KNOWLES, D. M., TAM, W., CHIU, A., BUSSEL, J. B., SERRANO, S., LORENTE, J. A., BELLOSILLO, B., LLORETA, J., JUANPERE, N., ALAMEDA, F., BARO, T., DE HEREDIA, C. D., TORAN, N., CATALA, A., TORREBADELL, M., FORTUNY, C., CUSI, V., CARRERAS, C., DIAZ, G. A., BLANDER, J. M., FARBER, C. M., SILVESTRI, G., CUNNINGHAM-RUNDLES, C., CALVILLO, M., DUFOUR, C., NOTARANGELO, L. D., LOUGARIS, V., PLEBANI, A., CASANOVA, J. L., GANAL, S. C., DIEFENBACH, A., AROSTEGUI, J. I., JUAN, M., YAGUE, J., MAHLAOUI, N., DONADIEU, J., CHEN, K. & CERUTTI, A. 2011. B cell-helper neutrophils stimulate the diversification and production of immunoglobulin in the marginal zone of the spleen. *Nat Immunol,* 13**,** 170-80.

QUAKE, S. R., WYSS-CORAY, T. & DARMANIS, S. 2017. Transcriptomic characterization of 20 organs and tissues from mouse at single cell resolution creates a Tabula Muris. *bioRxiv*.

RACLE, J., DE JONGE, K., BAUMGAERTNER, P., SPEISER, D. E. & GFELLER, D. 2017. Simultaneous enumeration of cancer and immune cell types from bulk tumor gene expression data. *Elife,* 6.

RAJ, A., PESKIN, C. S., TRANCHINA, D., VARGAS, D. Y. & TYAGI, S. 2006. Stochastic mRNA Synthesis in Mammalian Cells. *PLOS Biology,* 4**,** e309.

RAMACHANDRAN, P., DOBIE, R., WILSON-KANAMORI, J. R., DORA, E. F., HENDERSON, B. E. P., LUU, N. T., PORTMAN, J. R., MATCHETT, K. P., BRICE, M., MARWICK, J. A., TAYLOR, R. S., EFREMOVA, M., VENTO-TORMO, R., CARRAGHER, N. O., KENDALL, T. J., FALLOWFIELD, J. A., HARRISON, E. M., MOLE, D. J., WIGMORE, S. J., NEWSOME, P. N., WESTON, C. J., IREDALE, J. P., TACKE, F., POLLARD, J. W., PONTING, C. P., MARIONI, J. C., TEICHMANN, S. A. & HENDERSON, N. C. 2019. Resolving the fibrotic niche of human liver cirrhosis at single-cell level. *Nature,* 575**,** 512-518.

RAMIREZ, G. A., YACOUB, M. R., RIPA, M., MANNINA, D., CARIDDI, A., SAPORITI, N., CICERI, F., CASTAGNA, A., COLOMBO, G. & DAGNA, L. 2018. Eosinophils from Physiology to Disease: A Comprehensive Review. *Biomed Res Int,* 2018**,** 9095275.

RAND, U., RINAS, M., SCHWERK, J., NÖHREN, G., LINNES, M., KRÖGER, A., FLOSSDORF, M., KÁLY-KULLAI, K., HAUSER, H., HÖFER, T. & KÖSTER, M. 2012. Multi-layered stochasticity and paracrine signal propagation shape the type-I interferon response. *Molecular Systems Biology,* 8**,** 584.

RANKIN, L. C., GIRARD-MADOUX, M. J. H., SEILLET, C., MIELKE, L. A., KERDILES, Y., FENIS, A., WIEDUWILD, E., PUTOCZKI, T., MONDOT, S., LANTZ, O., DEMON, D., PAPENFUSS, A. T., SMYTH, G. K., LAMKANFI, M., CAROTTA, S., RENAULD, J.-C., SHI, W., CARPENTIER, S., SOOS, T., ARENDT, C., UGOLINI, S., HUNTINGTON, N. D., BELZ, G. T. & VIVIER, E. 2016. Complementarity and

redundancy of IL-22-producing innate lymphoid cells. *Nature Immunology,* 17**,** 179-186.

RANSOHOFF, R. M. 2016. A polarizing question: do M1 and M2 microglia exist? *Nat Neurosci,* 19**,** 987-91.

RAZA, S., BARNETT, M. W., BARNETT-ITZHAKI, Z., AMIT, I., HUME, D. A. & FREEMAN, T. C. 2014. Analysis of the transcriptional networks underpinning the activation of murine macrophages by inflammatory mediators. *J Leukoc Biol,* 96**,** 167-83.

REGEV, A., TEICHMANN, S. A., LANDER, E. S., AMIT, I., BENOIST, C., BIRNEY, E., BODENMILLER, B., CAMPBELL, P., CARNINCI, P., CLATWORTHY, M., CLEVERS, H., DEPLANCKE, B., DUNHAM, I., EBERWINE, J., EILS, R., ENARD, W., FARMER, A., FUGGER, L., GÖTTGENS, B., HACOHEN, N., HANIFFA, M., HEMBERG, M., KIM, S., KLENERMAN, P., KRIEGSTEIN, A., LEIN, E., LINNARSSON, S., LUNDBERG, E., LUNDEBERG, J., MAJUMDER, P., MARIONI, J. C., MERAD, M., MHLANGA, M., NAWIJN, M., NETEA, M., NOLAN, G., PE'ER, D., PHILLIPAKIS, A., PONTING, C. P., QUAKE, S., REIK, W., ROZENBLATT-ROSEN, O., SANES, J., SATIJA, R., SCHUMACHER, T. N., SHALEK, A., SHAPIRO, E., SHARMA, P., SHIN, J. W., STEGLE, O., STRATTON, M., STUBBINGTON, M. J. T., THEIS, F. J., UHLEN, M., VAN OUDENAARDEN, A., WAGNER, A., WATT, F., WEISSMAN, J., WOLD, B., XAVIER, R., YOSEF, N. & HUMAN CELL ATLAS MEETING, P. 2017. The Human Cell Atlas. *eLife,* 6**,** e27041.

RHODES, J. W., TONG, O., HARMAN, A. N. & TURVILLE, S. G. 2019. Human Dendritic Cell Subsets, Ontogeny, and Impact on HIV Infection. *Frontiers in Immunology,* 10.

RICARDO-GONZALEZ, R. R., VAN DYKEN, S. J., SCHNEIDER, C., LEE, J., NUSSBAUM, J. C., LIANG, H.-E., VAKA, D., ECKALBAR, W. L., MOLOFSKY, A. B., ERLE, D. J. & LOCKSLEY, R. M. 2018. Tissue signals imprint ILC2 identity with anticipatory function. *Nature Immunology,* 19**,** 1093-1099.

ROMAGNOLI, P. A., SHERIDAN, B. S., PHAM, Q. M., LEFRANÇOIS, L. & KHANNA, K. M. 2016. IL-17A-producing resident memory γδ T cells orchestrate the innate immune response to secondary oral Listeria monocytogenes infection. *Proc Natl Acad Sci U S A,* 113**,** 8502-7.

ROSENBLUM, M. D., WAY, S. S. & ABBAS, A. K. 2016. Regulatory T cell memory. *Nat Rev Immunol,* 16**,** 90-101.

ROSSER, E. C. & MAURI, C. 2015. Regulatory B cells: origin, phenotype, and function. *Immunity,* 42**,** 607-12.

RŐSZER, T. 2015. Understanding the Mysterious M2 Macrophage through Activation Markers and Effector Mechanisms. *Mediators Inflamm,* 2015**,** 816460.

ROTH, A., FRANCESCHINI, A., SZKLARCZYK, D., MULLER, J., STARK, M., KUHN, M., SIMONOVIC, M., MINGUEZ, P., DOERKS, T., MERING, C. V., JENSEN, L. J. & BORK, P. 2010. The STRING database in 2011: functional interaction

networks of proteins, globally integrated and scored. *Nucleic Acids Research,* 39**,** D561-D568.

RUBIN, A. J., PARKER, K. R., SATPATHY, A. T., QI, Y., WU, B., ONG, A. J., MUMBACH, M. R., JI, A. L., KIM, D. S., CHO, S. W., ZARNEGAR, B. J., GREENLEAF, W. J., CHANG, H. Y. & KHAVARI, P. A. 2019. Coupled Single-Cell CRISPR Screening and Epigenomic Profiling Reveals Causal Gene Regulatory Networks. *Cell,* 176**,** 361-376.e17.

SACHS, U. J., ANDREI-SELMER, C. L., MANIAR, A., WEISS, T., PADDOCK, C., ORLOVA, V. V., CHOI, E. Y., NEWMAN, P. J., PREISSNER, K. T., CHAVAKIS, T. & SANTOSO, S. 2007. The neutrophil-specific antigen CD177 is a counter-receptor for platelet endothelial cell adhesion molecule-1 (CD31). *J Biol Chem,* 282**,** 23603-12.

SAFARI-ALIGHIARLOO, N., TAGHIZADEH, M., REZAEI-TAVIRANI, M., GOLIAEI, B. & PEYVANDI, A. A. 2014. Protein-protein interaction networks (PPI) and complex diseases. *Gastroenterol Hepatol Bed Bench,* 7**,** 17-31.

SAGEBIEL, A. F., STEINERT, F., LUNEMANN, S., KÖRNER, C., SCHREURS, R. R. C. E., ALTFELD, M., PEREZ, D., REINSHAGEN, K. & BUNDERS, M. J. 2019. Tissue-resident Eomes+ NK cells are the major innate lymphoid cell population in human infant intestine. *Nature Communications,* 10**,** 975.

SATHI, G. A., FARAHAT, M., HARA, E. S., TAKETA, H., NAGATSUKA, H., KUBOKI, T. & MATSUMOTO, T. 2017. MCSF orchestrates branching morphogenesis in developing submandibular gland tissue. *J Cell Sci,* 130**,** 1559-1569.

SATIJA, R., FARRELL, J. A., GENNERT, D., SCHIER, A. F. & REGEV, A. 2015. Spatial reconstruction of single-cell gene expression data. *Nat Biotechnol,* 33.

SATIJA, R. & SHALEK, A. K. 2014. Heterogeneity in immune responses: from populations to single cells. *Trends in Immunology,* 35**,** 219-229.

SAVAS, P., VIRASSAMY, B., YE, C., SALIM, A., MINTOFF, C. P., CARAMIA, F., SALGADO, R., BYRNE, D. J., TEO, Z. L., DUSHYANTHEN, S., BYRNE, A., WEIN, L., LUEN, S. J., POLINESS, C., NIGHTINGALE, S. S., SKANDARAJAH, A. S., GYORKI, D. E., THORNTON, C. M., BEAVIS, P. A., FOX, S. B., DARCY, P. K., SPEED, T. P., MACKAY, L. K., NEESON, P. J., LOI, S. & KATHLEEN CUNINGHAM FOUNDATION CONSORTIUM FOR RESEARCH INTO FAMILIAL BREAST, C. 2018. Single-cell profiling of breast cancer T cells reveals a tissue-resident memory subset associated with improved prognosis. *Nature Medicine,* 24**,** 986-993.

SCHAUM, N., KARKANIAS, J., NEFF, N. F., MAY, A. P., QUAKE, S. R., WYSS-CORAY, T., DARMANIS, S., BATSON, J., BOTVINNIK, O., CHEN, M. B., CHEN, S., GREEN, F., JONES, R. C., MAYNARD, A., PENLAND, L., PISCO, A. O., SIT, R. V., STANLEY, G. M., WEBBER, J. T., ZANINI, F., BAGHEL, A. S., BAKERMAN, I., BANSAL, I., BERDNIK, D., BILEN, B., BROWNFIELD, D., CAIN, C., CHEN, M. B., CHEN, S., CHO, M., CIROLIA, G., CONLEY, S. D., DARMANIS, S., DEMERS, A., DEMIR, K., DE MORREE, A., DIVITA, T., DU BOIS, H., DULGEROFF, L. B. T., EBADI, H., ESPINOZA, F. H., FISH, M., GAN, Q.,

GEORGE, B. M., GILLICH, A., GREEN, F., GENETIANO, G., GU, X., GULATI, G. S., HANG, Y., HOSSEINZADEH, S., HUANG, A., IRAM, T., ISOBE, T., IVES, F., JONES, R. C., KAO, K. S., KARNAM, G., KERSHNER, A. M., KISS, B. M., KONG, W., KUMAR, M. E., LAM, J. Y., LEE, D. P., LEE, S. E., LI, G., LI, Q., LIU, L., LO, A., LU, W.-J., MANJUNATH, A., MAY, A. P., MAY, K. L., MAY, O. L., MAYNARD, A., MCKAY, M., METZGER, R. J., MIGNARDI, M., MIN, D., NABHAN, A. N., NEFF, N. F., NG, K. M., NOH, J., PATKAR, R., PENG, W. C., PENLAND, L., PUCCINELLI, R., RULIFSON, E. J., SCHAUM, N., SIKANDAR, S. S., SINHA, R., SIT, R. V., SZADE, K., TAN, W., TATO, C., TELLEZ, K., TRAVAGLINI, K. J., TROPINI, C., WALDBURGER, L., VAN WEELE, L. J., et al. 2018. Single-cell transcriptomics of 20 mouse organs creates a Tabula Muris. *Nature,* 562**,** 367-372.

SCHMIEDER, A., MICHEL, J., SCHÖNHAAR, K., GOERDT, S. & SCHLEDZEWSKI, K. 2012. Differentiation and gene expression profile of tumor-associated macrophages. *Seminars in Cancer Biology,* 22**,** 289-297.

SCHROEDER, J. T. 2009. Basophils beyond effector cells of allergic inflammation. *Adv Immunol,* 101**,** 123-61.

SCHYNS, J., BAI, Q., RUSCITTI, C., RADERMECKER, C., DE SCHEPPER, S., CHAKAROV, S., FARNIR, F., PIROTTIN, D., GINHOUX, F., BOECKXSTAENS, G., BUREAU, F. & MARICHAL, T. 2019. Non-classical tissue monocytes and two functionally distinct populations of interstitial macrophages populate the mouse lung. *Nature Communications,* 10**,** 3964.

SCOTT, C. L., ZHENG, F., DE BAETSELIER, P., MARTENS, L., SAEYS, Y., DE PRIJCK, S., LIPPENS, S., ABELS, C., SCHOONOOGHE, S., RAES, G., DEVOOGDT, N., LAMBRECHT, B. N., BESCHIN, A. & GUILLIAMS, M. 2016. Bone marrow-derived monocytes give rise to self-renewing and fully differentiated Kupffer cells. *Nature Communications,* 7**,** 10321.

SERE, K., BAEK, J. H., OBER-BLOBAUM, J., MULLER-NEWEN, G., TACKE, F., YOKOTA, Y., ZENKE, M. & HIERONYMUS, T. 2012. Two distinct types of Langerhans cells populate the skin during steady state and inflammation. *Immunity,* 37**,** 905-16.

SERIN, E. A. R., NIJVEEN, H., HILHORST, H. W. M. & LIGTERINK, W. 2016. Learning from Co-expression Networks: Possibilities and Challenges. *Frontiers in Plant Science,* 7.

SHARMA, A. & RUDRA, D. 2018. Emerging Functions of Regulatory T Cells in Tissue Homeostasis. *Frontiers in Immunology,* 9.

SHAW, T. N., HOUSTON, S. A., WEMYSS, K., BRIDGEMAN, H. M., BARBERA, T. A., ZANGERLE-MURRAY, T., STRANGWARD, P., RIDLEY, A. J. L., WANG, P., TAMOUTOUNOUR, S., ALLEN, J. E., KONKEL, J. E. & GRAINGER, J. R. 2018. Tissue-resident macrophages in the intestine are long lived and defined by Tim-4 and CD4 expression. *Journal of Experimental Medicine,* 215**,** 1507-1518.

SHEN, E. H., OVERLY, C. C. & JONES, A. R. 2012. The Allen Human Brain Atlas: comprehensive gene expression mapping of the human brain. *Trends in neurosciences,* 35**,** 711-714.

SHEN-ORR, S. S. & GAUJOUX, R. 2013. Computational deconvolution: extracting cell type-specific information from heterogeneous samples. *Curr Opin Immunol,* 25**,** 571-8.

SHESHACHALAM, A., SRIVASTAVA, N., MITCHELL, T., LACY, P. & EITZEN, G. 2014. Granule protein processing and regulated secretion in neutrophils. *Front Immunol,* 5**,** 448.

SHISSLER, S. C., SINGH, N. J. & WEBB, T. J. 2020. Thymic resident NKT cell subsets show differential requirements for CD28 co-stimulation during antigenic activation. *Scientific Reports,* 10**,** 8218.

SHOEMAKER, J. E., LOPES, T. J., GHOSH, S., MATSUOKA, Y., KAWAOKA, Y. & KITANO, H. 2012. CTen: a web-based platform for identifying enriched cell types from heterogeneous microarray data. *BMC Genomics,* 13**,** 460.

SILVESTRE-ROIG, C., FRIDLENDER, Z. G., GLOGAUER, M. & SCAPINI, P. 2019. Neutrophil Diversity in Health and Disease. *Trends in Immunology,* 40**,** 565-583.

SILVESTRE-ROIG, C., HIDALGO, A. & SOEHNLEIN, O. 2016. Neutrophil heterogeneity: implications for homeostasis and pathogenesis. *Blood,* 127**,** 2173-81.

SIMONETTA, F., PRADIER, A. & ROOSNEK, E. 2016. T-bet and Eomesodermin in NK Cell Development, Maturation, and Function. *Front Immunol,* 7**,** 241.

SIRACUSA, M. C., SAENZ, S. A., HILL, D. A., KIM, B. S., HEADLEY, M. B., DOERING, T. A., WHERRY, E. J., JESSUP, H. K., SIEGEL, L. A., KAMBAYASHI, T., DUDEK, E. C., KUBO, M., CIANFERONI, A., SPERGEL, J. M., ZIEGLER, S. F., COMEAU, M. R. & ARTIS, D. 2011. TSLP promotes interleukin-3-independent basophil haematopoiesis and type 2 inflammation. *Nature,* 477**,** 229-33.

SMITH-GARVIN, J. E., KORETZKY, G. A. & JORDAN, M. S. 2009. T cell activation. *Annu Rev Immunol,* 27**,** 591-619.

SOJKA, D. K., PLOUGASTEL-DOUGLAS, B., YANG, L., PAK-WITTEL, M. A., ARTYOMOV, M. N., IVANOVA, Y., ZHONG, C., CHASE, J. M., ROTHMAN, P. B., YU, J., RILEY, J. K., ZHU, J., TIAN, Z. & YOKOYAMA, W. M. 2014a. Tissue-resident natural killer (NK) cells are cell lineages distinct from thymic and conventional splenic NK cells. *eLife,* 3**,** e01659.

SOJKA, D. K., TIAN, Z. & YOKOYAMA, W. M. 2014b. Tissue-resident natural killer cells and their potential diversity. *Semin Immunol,* 26**,** 127-31.

SOJKA, D. K., YANG, L., PLOUGASTEL-DOUGLAS, B., HIGUCHI, D. A., CROY, B. A. & YOKOYAMA, W. M. 2018. Cutting Edge: Local Proliferation of Uterine Tissue-Resident NK Cells during Decidualization in Mice. *J Immunol,* 201**,** 2551-2556.

SOLER ARTIGAS, M., WAIN, L. V., REPAPI, E., OBEIDAT, M. E., SAYERS, I., BURTON, P. R., JOHNSON, T., ZHAO, J. H., ALBRECHT, E., DOMINICZAK, A. F., KERR, S. M., SMITH, B. H., CADBY, G., HUI, J., PALMER, L. J., HINGORANI, A. D., WANNAMETHEE, S. G., WHINCUP, P. H., EBRAHIM, S., SMITH, G. D., BARROSO, I., LOOS, R. J. F., WAREHAM, N. J., COOPER, C., DENNISON, E., SHAHEEN, S. O., LIU, J. Z., MARCHINI, J., MEDICAL RESEARCH COUNCIL NATIONAL SURVEY OF, H., DEVELOPMENT RESPIRATORY STUDY, T., DAHGAM, S., NALUAI, A. T., OLIN, A.-C., KARRASCH, S., HEINRICH, J., SCHULZ, H., MCKEEVER, T. M., PAVORD, I. D., HELIÖVAARA, M., RIPATTI, S., SURAKKA, I., BLAKEY, J. D., KÄHÖNEN, M., BRITTON, J. R., NYBERG, F., HOLLOWAY, J. W., LAWLOR, D. A., MORRIS, R. W., JAMES, A. L., JACKSON, C. M., HALL, I. P., TOBIN, M. D. & SPIROMETA, C. 2011. Effect of five genetic variants associated with lung function on the risk of chronic obstructive lung disease, and their joint effects on lung function. *American journal of respiratory and critical care medicine,* 184**,** 786-795.

SONAWANE, A. R., PLATIG, J., FAGNY, M., CHEN, C.-Y., PAULSON, J. N., LOPES-RAMOS, C. M., DEMEO, D. L., QUACKENBUSH, J., GLASS, K. & KUIJJER, M. L. 2017. Understanding Tissue-Specific Gene Regulation. *Cell Reports,* 21**,** 1077-1088.

SPITS, H., ARTIS, D., COLONNA, M., DIEFENBACH, A., DI SANTO, J. P., EBERL, G., KOYASU, S., LOCKSLEY, R. M., MCKENZIE, A. N. J., MEBIUS, R. E., POWRIE, F. & VIVIER, E. 2013. Innate lymphoid cells — a proposal for uniform nomenclature. *Nature Reviews Immunology,* 13**,** 145-149.

SPITS, H., BERNINK, J. H. & LANIER, L. 2016. NK cells and type 1 innate lymphoid cells: partners in host defense. *Nat Immunol,* 17**,** 758-64.

STEINMAN, R. M. & COHN, Z. A. 1973. Identification of a novel cell type in peripheral lymphoid organs of mice. I. Morphology, quantitation, tissue distribution. *J Exp Med,* 137**,** 1142-62.

STOECKIUS, M., HAFEMEISTER, C., STEPHENSON, W., HOUCK-LOOMIS, B., CHATTOPADHYAY, P. K., SWERDLOW, H., SATIJA, R. & SMIBERT, P. 2017. Simultaneous epitope and transcriptome measurement in single cells. *Nat Methods,* 14**,** 865-868.

STREET, K., RISSO, D., FLETCHER, R. B., DAS, D., NGAI, J., YOSEF, N., PURDOM, E. & DUDOIT, S. 2018. Slingshot: cell lineage and pseudotime inference for single-cell transcriptomics. *BMC Genomics,* 19**,** 477.

STUBBINGTON, M. J. T., LÖNNBERG, T., PROSERPIO, V., CLARE, S., SPEAK, A. O., DOUGAN, G. & TEICHMANN, S. A. 2016. T cell fate and clonality inference from single-cell transcriptomes. *Nature Methods,* 13**,** 329.

SUKHBAATAR, N. & WEICHHART, T. 2018. Iron Regulation: Macrophages in Control. *Pharmaceuticals (Basel),* 11.

SUN, B. & ZHANG, Y. 2014. Overview of Orchestration of CD4+ T Cell Subsets in Immune Responses. *In:* SUN, B. (ed.) *T Helper Cell Differentiation and Their Function.* Dordrecht: Springer Netherlands.

SUN, S., ZHU, J., MA, Y. & ZHOU, X. 2019. Accuracy, robustness and scalability of dimensionality reduction methods for single-cell RNA-seq analysis. *Genome Biology,* 20**,** 269.

SUNKIN, S. M., NG, L., LAU, C., DOLBEARE, T., GILBERT, T. L., THOMPSON, C. L., HAWRYLYCZ, M. & DANG, C. 2013. Allen Brain Atlas: an integrated spatio-temporal portal for exploring the central nervous system. *Nucleic Acids Res,* 41**,** D996-d1008.

SUPEK, F., BOŠNJAK, M., ŠKUNCA, N. & ŠMUC, T. 2011. REVIGO Summarizes and Visualizes Long Lists of Gene Ontology Terms. *PLOS ONE,* 6**,** e21800.

SVENSSON, V., VENTO-TORMO, R. & TEICHMANN, S. A. 2018. Exponential scaling of single-cell RNA-seq in the past decade. *Nat Protoc,* 13**,** 599-604.

SWAIN, P. S., ELOWITZ, M. B. & SIGGIA, E. D. 2002. Intrinsic and extrinsic contributions to stochasticity in gene expression. *Proceedings of the National Academy of Sciences,* 99**,** 12795.

SZABO, P. A., LEVITIN, H. M., MIRON, M., SNYDER, M. E., SENDA, T., YUAN, J., CHENG, Y. L., BUSH, E. C., DOGRA, P., THAPA, P., FARBER, D. L. & SIMS, P. A. 2019. Single-cell transcriptomics of human T cells reveals tissue and activation signatures in health and disease. *Nature Communications,* 10**,** 4706.

T'JONCK, W., GUILLIAMS, M. & BONNARDEL, J. 2018. Niche signals and transcription factors involved in tissue-resident macrophage development. *Cellular Immunology,* 330**,** 43-53.

TAKEUCHI, O. & AKIRA, S. 2010. Pattern Recognition Receptors and Inflammation. *Cell,* 140**,** 805-820.

TANG, F., BARBACIORU, C., WANG, Y., NORDMAN, E., LEE, C., XU, N., WANG, X., BODEAU, J., TUCH, B. B., SIDDIQUI, A., LAO, K. & SURANI, M. A. 2009. mRNA-Seq whole-transcriptome analysis of a single cell. *Nat Methods,* 6**,** 377-82.

TARASHANSKY, A. J., XUE, Y., QUAKE, S. R. & WANG, B. 2018. Self-assembling Manifolds in Single-cell RNA Sequencing Data. *bioRxiv***,** 364166.

TEIJARO, J. R., TURNER, D., PHAM, Q., WHERRY, E. J., LEFRANÇOIS, L. & FARBER, D. L. 2011. Cutting edge: Tissue-retentive lung memory CD4 T cells mediate optimal protection to respiratory virus infection. *J Immunol,* 187**,** 5510-4.

TENTLER, J. J., TAN, A. C., WEEKES, C. D., JIMENO, A., LEONG, S., PITTS, T. M., ARCAROLI, J. J., MESSERSMITH, W. A. & ECKHARDT, S. G. 2012. Patient-derived tumour xenografts as models for oncology drug development. *Nat Rev Clin Oncol,* 9**,** 338-50.

THE IMMGEN, C. & BENOIST, C. 2016. Open-source ImmGen: mononuclear phagocytes. *Nature Immunology,* 17**,** 741.

THEISEN, D. & MURPHY, K. 2017. The role of cDC1s in vivo: CD8 T cell priming through cross-presentation. *F1000Res,* 6**,** 98.

THRUPP, N., FRIGERIO, S., WOLFS, L., SKENE, N. G., POOVATHINGAL, S., FOURNE, Y., MATTHEWS, P. M., THEYS, T., MANCUSO, R., DE STROOPER, B. & FIERS, M. 2020. Single nucleus sequencing fails to detect microglial activation in human tissue. bioRxiv.

TOBER, J., KONISKI, A., MCGRATH, K. E., VEMISHETTI, R., EMERSON, R., DE MESY-BENTLEY, K. K., WAUGH, R. & PALIS, J. 2007. The megakaryocyte lineage originates from hemangioblast precursors and is an integral component both of primitive and of definitive hematopoiesis. *Blood,* 109**,** 1433-41.

TRAEGER, T., MIKULCAK, M., EIPEL, C., ABSHAGEN, K., DIEDRICH, S., HEIDECKE, C. D., MAIER, S. & VOLLMAR, B. 2010. Kupffer cell depletion reduces hepatic inflammation and apoptosis but decreases survival in abdominal sepsis. *Eur J Gastroenterol Hepatol,* 22**,** 1039-49.

TRAPNELL, C., CACCHIARELLI, D., GRIMSBY, J., POKHAREL, P., LI, S. & MORSE, M. 2014. The dynamics and regulators of cell fate decisions are revealed by pseudotemporal ordering of single cells. *Nat Biotechnol,* 32.

TRITSCHLER, S., BÜTTNER, M., FISCHER, D. S., LANGE, M., BERGEN, V., LICKERT, H. & THEIS, F. J. 2019. Concepts and limitations for learning developmental trajectories from single cell genomics. *Development,* 146**,** dev170506.

TSOU, Y.-A., TUNG, M.-C., ALEXANDER, K. A., CHANG, W.-D., TSAI, M.-H., CHEN, H.-L. & CHEN, C.-M. 2018. The Role of BPIFA1 in Upper Airway Microbial Infections and Correlated Diseases. *BioMed research international,* 2018**,** 2021890-2021890.

TURVEY, S. E. & BROIDE, D. H. 2010. Innate immunity. *Journal of Allergy and Clinical Immunology,* 125**,** S24-S32.

UJIHARA, Y., KANAGAWA, M., MOHRI, S., TAKATSU, S., KOBAYASHI, K., TODA, T., NARUSE, K. & KATANOSAKA, Y. 2019. Elimination of fukutin reveals cellular and molecular pathomechanisms in muscular dystrophy-associated heart failure. *Nature Communications,* 10**,** 1-17.

ULIRSCH, J. C., LAREAU, C. A., BAO, E. L., LUDWIG, L. S., GUO, M. H., BENNER, C., SATPATHY, A. T., KARTHA, V. K., SALEM, R. M., HIRSCHHORN, J. N., FINUCANE, H. K., ARYEE, M. J., BUENROSTRO, J. D. & SANKARAN, V. G. 2019. Interrogation of human hematopoiesis at single-cell and single-variant resolution. *Nature Genetics*.

VALE, A. M., KEARNEY, J. F., NOBREGA, A. & SCHROEDER, H. W. 2015. Chapter 7 - Development and Function of B Cell Subsets. *In:* ALT, F. W., HONJO, T., RADBRUCH, A. & RETH, M. (eds.) *Molecular Biology of B Cells (Second Edition).* London: Academic Press.

VAN DE LAAR, L., SAELENS, W., DE PRIJCK, S., MARTENS, L., SCOTT, C. L., VAN ISTERDAEL, G., HOFFMANN, E., BEYAERT, R., SAEYS, Y., LAMBRECHT, B. N. & GUILLIAMS, M. 2016. Yolk Sac Macrophages, Fetal Liver, and Adult Monocytes Can Colonize an Empty Niche and Develop into Functional Tissue-Resident Macrophages. *Immunity,* 44**,** 755-68.

VAN DER WIJST, M. G. P., BRUGGE, H., DE VRIES, D. H., DEELEN, P., SWERTZ, M. A. & FRANKE, L. 2018. Single-cell RNA sequencing identifies celltype-specific cis-eQTLs and co-expression QTLs. *Nat Genet,* 50**,** 493-497.

VAN DIJK, D., SHARMA, R., NAINYS, J., YIM, K., KATHAIL, P., CARR, A. J., BURDZIAK, C., MOON, K. R., CHAFFER, C. L., PATTABIRAMAN, D., BIERIE, B., MAZUTIS, L., WOLF, G., KRISHNASWAMY, S. & PE'ER, D. 2018. Recovering Gene Interactions from Single-Cell Data Using Data Diffusion. *Cell,* 174**,** 716-729.e27.

VAN GALEN, P., HOVESTADT, V., WADSWORTH II, M. H., HUGHES, T. K., GRIFFIN, G. K., BATTAGLIA, S., VERGA, J. A., STEPHANSKY, J., PASTIKA, T. J., LOMBARDI STORY, J., PINKUS, G. S., POZDNYAKOVA, O., GALINSKY, I., STONE, R. M., GRAUBERT, T. A., SHALEK, A. K., ASTER, J. C., LANE, A. A. & BERNSTEIN, B. E. 2019. Single-Cell RNA-Seq Reveals AML Hierarchies Relevant to Disease Progression and Immunity. *Cell,* 176**,** 1265-1281.e24.

VAN UNEN, V., LI, N., MOLENDIJK, I., TEMURHAN, M., HÖLLT, T., VAN DER MEULEN-DE JONG, ANDREA E., VERSPAGET, HEIN W., MEARIN, M. L., MULDER, CHRIS J., VAN BERGEN, J., LELIEVELDT, BOUDEWIJN P. F. & KONING, F. 2016. Mass Cytometry of the Human Mucosal Immune System Identifies Tissue- and Disease-Associated Immune Subsets. *Immunity,* 44**,** 1227-1239.

VAROL, C., MILDNER, A. & JUNG, S. 2015. Macrophages: Development and Tissue Specialization. *Annual Review of Immunology,* 33**,** 643-675.

VASANTHAKUMAR, A., MORO, K., XIN, A., LIAO, Y., GLOURY, R., KAWAMOTO, S., FAGARASAN, S., MIELKE, L. A., AFSHAR-STERLE, S., MASTERS, S. L., NAKAE, S., SAITO, H., WENTWORTH, J. M., LI, P., LIAO, W., LEONARD, W. J., SMYTH, G. K., SHI, W., NUTT, S. L., KOYASU, S. & KALLIES, A. 2015. The transcriptional regulators IRF4, BATF and IL-33 orchestrate development and maintenance of adipose tissue–resident regulatory T cells. *Nature Immunology,* 16**,** 276-285.

VEIGA-FERNANDES, H., COLES, M. C., FOSTER, K. E., PATEL, A., WILLIAMS, A., NATARAJAN, D., BARLOW, A., PACHNIS, V. & KIOUSSIS, D. 2007. Tyrosine kinase receptor RET is a key regulator of Peyer's Patch organogenesis. *Nature,* 446**,** 547-551.

VERHAAK, R. G. W., HOADLEY, K. A., PURDOM, E., WANG, V., QI, Y., WILKERSON, M. D., MILLER, C. R., DING, L., GOLUB, T., MESIROV, J. P., ALEXE, G., LAWRENCE, M., O'KELLY, M., TAMAYO, P., WEIR, B. A., GABRIEL, S., WINCKLER, W., GUPTA, S., JAKKULA, L., FEILER, H. S., HODGSON, J. G., JAMES, C. D., SARKARIA, J. N., BRENNAN, C., KAHN, A., SPELLMAN, P. T., WILSON, R. K., SPEED, T. P., GRAY, J. W., MEYERSON, M., GETZ, G.,

PEROU, C. M. & HAYES, D. N. 2010. Integrated Genomic Analysis Identifies Clinically Relevant Subtypes of Glioblastoma Characterized by Abnormalities in PDGFRA, IDH1, EGFR, and NF1. *Cancer Cell,* 17**,** 98-110.

VILLANI, A. C., SATIJA, R., REYNOLDS, G., SARKIZOVA, S., SHEKHAR, K., FLETCHER, J., GRIESBECK, M., BUTLER, A., ZHENG, S., LAZO, S., JARDINE, L., DIXON, D., STEPHENSON, E., NILSSON, E., GRUNDBERG, I., MCDONALD, D., FILBY, A., LI, W., DE JAGER, P. L., ROZENBLATT-ROSEN, O., LANE, A. A., HANIFFA, M., REGEV, A. & HACOHEN, N. 2017. Single-cell RNA-seq reveals new types of human blood dendritic cells, monocytes, and progenitors. *Science,* 356.

VINCENTI, J. E., MURPHY, L., GRABERT, K., MCCOLL, B. W., CANCELLOTTI, E., FREEMAN, T. C. & MANSON, J. C. 2016. Defining the microglia response during the time course of chronic neurodegeneration. *Journal of virology,* 90**,** 3003-3017.

VON MOLTKE, J., JI, M., LIANG, H.-E. & LOCKSLEY, R. M. 2016. Tuft-cell-derived IL-25 regulates an intestinal ILC2–epithelial response circuit. *Nature,* 529**,** 221-225.

WAGNER, D. E. & KLEIN, A. M. 2020. Lineage tracing meets single-cell omics: opportunities and challenges. *Nature Reviews Genetics,* 21**,** 410-427.

WALKER, J. A., CLARK, P. A., CRISP, A., BARLOW, J. L., SZETO, A., FERREIRA, A. C. F., RANA, B. M. J., JOLIN, H. E., RODRIGUEZ-RODRIGUEZ, N., SIVASUBRAMANIAM, M., PANNELL, R., CRUICKSHANK, J., DALY, M., HAIM-VILMOVSKY, L., TEICHMANN, S. A. & MCKENZIE, A. N. J. 2019. Polychromic Reporter Mice Reveal Unappreciated Innate Lymphoid Cell Progenitor Heterogeneity and Elusive ILC3 Progenitors in Bone Marrow. *Immunity,* 51**,** 104-118.e7.

WALSH, K., MEGYESI, J. & HAMMOND, R. 2005. Human central nervous system tissue culture: a historical review and examination of recent advances. *Neurobiology of Disease,* 18**,** 2-18.

WANG, E. T., SANDBERG, R., LUO, S., KHREBTUKOVA, I., ZHANG, L., MAYR, C., KINGSMORE, S. F., SCHROTH, G. P. & BURGE, C. B. 2008. Alternative isoform regulation in human tissue transcriptomes. *Nature,* 456**,** 470.

WANG, L.-X., ZHANG, S.-X., WU, H.-J., RONG, X.-L. & GUO, J. 2019a. M2b macrophage polarization and its roles in diseases. *Journal of Leukocyte Biology,* 106**,** 345-358.

WANG, S. & COLONNA, M. 2019. Microglia in Alzheimer's disease: A target for immunotherapy. *Journal of Leukocyte Biology,* 106**,** 219-227.

WANG, X., ALLEN, W. E., WRIGHT, M. A., SYLWESTRAK, E. L., SAMUSIK, N., VESUNA, S., EVANS, K., LIU, C., RAMAKRISHNAN, C., LIU, J., NOLAN, G. P., BAVA, F.-A. & DEISSEROTH, K. 2018. Three-dimensional intact-tissue sequencing of single-cell transcriptional states. *Science,* 361**,** eaat5691.

WANG, X., HE, Y., ZHANG, Q., REN, X. & ZHANG, Z. 2019b. Direct Comparative Analysis of 10X Genomics Chromium and Smart-seq2. *bioRxiv*, 615013.

WANG, X., PARK, J., SUSZTAK, K., ZHANG, N. R. & LI, M. 2019c. Bulk tissue cell type deconvolution with multi-subject single-cell expression reference. *Nature Communications,* 10, 380.

WANG, Y., XU, Z., MAO, J.-H., HUNG, M.-S., HSIEH, D., AU, A., JABLONS, D. M. & YOU, L. 2015. Analysis of lung tumor initiation and progression in transgenic mice for Cre-inducible overexpression of Cul4A gene. *Thoracic cancer,* 6, 480-487.

WARDE-FARLEY, D., DONALDSON, S. L., COMES, O., ZUBERI, K., BADRAWI, R., CHAO, P., FRANZ, M., GROUIOS, C., KAZI, F., LOPES, C. T., MAITLAND, A., MOSTAFAVI, S., MONTOJO, J., SHAO, Q., WRIGHT, G., BADER, G. D. & MORRIS, Q. 2010. The GeneMANIA prediction server: biological network integration for gene prioritization and predicting gene function. *Nucleic Acids Research,* 38, W214-W220.

WATARAI, H., SEKINE-KONDO, E., SHIGEURA, T., MOTOMURA, Y., YASUDA, T., SATOH, R., YOSHIDA, H., KUBO, M., KAWAMOTO, H., KOSEKI, H. & TANIGUCHI, M. 2012. Development and Function of Invariant Natural Killer T Cells Producing TH2- and TH17-Cytokines. *PLOS Biology,* 10, e1001255.

WEGIEL, B. & OTTERBEIN, L. E. 2012. Go green: the anti-inflammatory effects of biliverdin reductase. *Front Pharmacol,* 3, 47.

WEINREB, C., RODRIGUEZ-FRATICELLI, A., CAMARGO, F. D. & KLEIN, A. M. 2020. Lineage tracing on transcriptional landscapes links state to fate during differentiation. *Science,* 367, eaaw3381.

WELIN, A., AMIRBEAGI, F., CHRISTENSON, K., BJÖRKMAN, L., BJÖRNSDOTTIR, H., FORSMAN, H., DAHLGREN, C., KARLSSON, A. & BYLUND, J. 2013. The Human Neutrophil Subsets Defined by the Presence or Absence of OLFM4 Both Transmigrate into Tissue In Vivo and Give Rise to Distinct NETs In Vitro. *PLOS ONE,* 8, e69575.

WELLER, P. F. & SPENCER, L. A. 2017. Functions of tissue-resident eosinophils. *Nature Reviews Immunology,* 17, 746-760.

WESCH, D., GLATZEL, A. & KABELITZ, D. 2001. Differentiation of resting human peripheral blood gamma delta T cells toward Th1- or Th2-phenotype. *Cell Immunol,* 212, 110-7.

WIESE, A. V., ENDER, F., QUELL, K. M., ANTONIOU, K., VOLLBRANDT, T., KÖNIG, P., KÖHL, J. & LAUMONNIER, Y. 2017. The C5a/C5aR1 axis controls the development of experimental allergic asthma independent of LysM-expressing pulmonary immune cells. *PLoS One,* 12, e0184956.

WIKLUND, P., ZHANG, X., PEKKALA, S., AUTIO, R., KONG, L., YANG, Y., KEINÄNEN-KIUKAANNIEMI, S., ALEN, M. & CHENG, S. 2016. Insulin resistance is associated with altered amino acid metabolism and adipose tissue dysfunction in normoglycemic women. *Scientific Reports,* 6, 24540.

WILLEKENS, F. L. A., WERRE, J. M., KRUIJT, J. K., ROERDINKHOLDER-STOELWINDER, B., GROENEN-DÖPP, Y. A. M., VAN DEN BOS, A. G., BOSMAN, G. J. C. G. M. & VAN BERKEL, T. J. C. 2005. Liver Kupffer cells rapidly remove red blood cell–derived vesicles from the circulation by scavenger receptors. *Blood,* 105**,** 2141-2145.

WOLF, F. A., HAMEY, F., PLASS, M., SOLANA, J., DAHLIN, J. S., GOTTGENS, B., RAJEWSKY, N., SIMON, L. & THEIS, F. J. 2018. Graph abstraction reconciles clustering with trajectory inference through a topology preserving map of single cells. *bioRxiv***,** 208819.

WOLF, F. A., HAMEY, F. K., PLASS, M., SOLANA, J., DAHLIN, J. S., GÖTTGENS, B., RAJEWSKY, N., SIMON, L. & THEIS, F. J. 2019. PAGA: graph abstraction reconciles clustering with trajectory inference through a topology preserving map of single cells. *Genome Biology,* 20**,** 59.

WOOD, G. S., WARNER, N. L. & WARNKE, R. A. 1983. Anti-Leu-3/T4 antibodies react with cells of monocyte/macrophage and Langerhans lineage. *The Journal of Immunology,* 131**,** 212-216.

WOODWARD DAVIS, A. S., ROOZEN, H. N., DUFORT, M. J., DEBERG, H. A., DELANEY, M. A., MAIR, F., ERICKSON, J. R., SLICHTER, C. K., BERKSON, J. D., KLOCK, A. M., MACK, M., LWO, Y., KO, A., BRAND, R. M., MCGOWAN, I., LINSLEY, P. S., DIXON, D. R. & PRLIC, M. 2019. The human tissue-resident CCR5⁺ T cell compartment maintains protective and functional properties during inflammation. *Science Translational Medicine,* 11**,** eaaw8718.

WU, C., OROZCO, C., BOYER, J., LEGLISE, M., GOODALE, J., BATALOV, S., HODGE, C. L., HAASE, J., JANES, J., HUSS, J. W., 3RD & SU, A. I. 2009. BioGPS: an extensible and customizable portal for querying and organizing gene annotation resources. *Genome biology,* 10**,** R130-R130.

WU, D., WU, P., QIU, F., WEI, Q. & HUANG, J. 2017. Human γδT-cell subsets and their involvement in tumor immunity. *Cellular & Molecular Immunology,* 14**,** 245-253.

WUCHTY, S. & UETZ, P. 2014. Protein-protein Interaction Networks of E. coli and S. cerevisiae are similar. *Scientific Reports,* 4**,** 7187.

XU, C. & SU, Z. 2015. Identification of cell types from single-cell transcriptomes using a novel clustering method. *Bioinformatics,* 31**,** 1974-1980.

XU, G., RAN, Y., FROMHOLT, S. E., FU, C., YACHNIS, A. T., GOLDE, T. E. & BORCHELT, D. R. 2015. Murine Aβ over-production produces diffuse and compact Alzheimer-type amyloid deposits. *Acta Neuropathol Commun,* 3**,** 72.

XUE, J., SCHMIDT, S. V., SANDER, J., DRAFFEHN, A., KREBS, W., QUESTER, I., DE NARDO, D., GOHEL, T. D., EMDE, M., SCHMIDLEITHNER, L., GANESAN, H., NINO-CASTRO, A., MALLMANN, M. R., LABZIN, L., THEIS, H., KRAUT, M., BEYER, M., LATZ, E., FREEMAN, T. C., ULAS, T. & SCHULTZE, J. L. 2014. Transcriptome-based network analysis reveals a spectrum model of human macrophage activation. *Immunity,* 40**,** 274-88.

YANG, J., HUANG, T., PETRALIA, F., LONG, Q., ZHANG, B., ARGMANN, C., ZHAO, Y., MOBBS, C. V., SCHADT, E. E., ZHU, J., TU, Z., THE, G. C., ARDLIE, K. G., DELUCA, D. S., SEGRÈ, A. V., SULLIVAN, T. J., YOUNG, T. R., GELFAND, E. T., TROWBRIDGE, C. A., MALLER, J. B., TUKIAINEN, T., LEK, M., WARD, L. D., KHERADPOUR, P., IRIARTE, B., MENG, Y., PALMER, C. D., WINCKLER, W., HIRSCHHORN, J., KELLIS, M., MACARTHUR, D. G., GETZ, G., SHABLIN, A. A., LI, G., ZHOU, Y.-H., NOBEL, A. B., RUSYN, I., WRIGHT, F. A., LAPPALAINEN, T., FERREIRA, P. G., ONGEN, H., RIVAS, M. A., BATTLE, A., MOSTAFAVI, S., MONLONG, J., SAMMETH, M., MELE, M., REVERTER, F., GOLDMANN, J., KOLLER, D., GUIGO, R., MCCARTHY, M. I., DERMITZAKIS, E. T., GAMAZON, E. R., KONKASHBAEV, A., NICOLAE, D. L., COX, N. J., FLUTRE, T., WEN, X., STEPHENS, M., PRITCHARD, J. K., LIN, L., LIU, J., BROWN, A., MESTICHELLI, B., TIDWELL, D., LO, E., SALVATORE, M., SHAD, S., THOMAS, J. A., LONSDALE, J. T., CHOI, C., KARASIK, E., RAMSEY, K., MOSER, M. T., FOSTER, B. A., GILLARD, B. M., SYRON, J., FLEMING, J., MAGAZINE, H., HASZ, R., WALTERS, G. D., BRIDGE, J. P., MIKLOS, M., SULLIVAN, S., BARKER, L. K., TRAINO, H., MOSAVEL, M., SIMINOFF, L. A., VALLEY, D. R., ROHRER, D. C., JEWEL, S., BRANTON, P., SOBIN, L. H., QI, L., HARIHARAN, P., WU, S., TABOR, D., SHIVE, C., SMITH, A. M., et al. 2015. Synchronized age-related gene expression changes across multiple tissues in human and the link to complex diseases. *Scientific Reports,* 5**,** 15145.

YAO, Y., XU, X. H. & JIN, L. 2019. Macrophage Polarization in Physiological and Pathological Pregnancy. *Front Immunol,* 10**,** 792.

YIN, X., YU, H., JIN, X., LI, J., GUO, H., SHI, Q., YIN, Z., XU, Y., WANG, X., LIU, R., WANG, S. & ZHANG, L. 2017. Human Blood CD1c+ Dendritic Cells Encompass CD5high and CD5low Subsets That Differ Significantly in Phenotype, Gene Expression, and Functions. *J Immunol,* 198**,** 1553-1564.

YOSHIDA, H., LAREAU, C. A., RAMIREZ, R. N., ROSE, S. A., MAIER, B., WROBLEWSKA, A., DESLAND, F., CHUDNOVSKIY, A., MORTHA, A., DOMINGUEZ, C., TELLIER, J., KIM, E., DWYER, D., SHINTON, S., NABEKURA, T., QI, Y., YU, B., ROBINETTE, M., KIM, K. W., WAGERS, A., RHOADS, A., NUTT, S. L., BROWN, B. D., MOSTAFAVI, S., BUENROSTRO, J. D. & BENOIST, C. 2019. The cis-Regulatory Atlas of the Mouse Immune System. *Cell,* 176**,** 897-912.e20.

YOUNG, K., BORIKAR, S., BELL, R., KUFFLER, L., PHILIP, V. & TROWBRIDGE, J. J. 2016. Progressive alterations in multipotent hematopoietic progenitors underlie lymphoid cell loss in aging. *J Exp Med,* 213**,** 2259-2267.

YOUNG, R., BUSH, S. J., LEFEVRE, L., MCCULLOCH, M. E. B., LISOWSKI, Z. M., MURIUKI, C., WADDELL, L. A., SAUTER, K. A., PRIDANS, C., CLARK, E. L. & HUME, D. A. 2018. Species-Specific Transcriptional Regulation of Genes Involved in Nitric Oxide Production and Arginine Metabolism in Macrophages. *Immunohorizons,* 2**,** 27-37.

YU, G., WANG, L. G., HAN, Y. & HE, Q. Y. 2012. clusterProfiler: an R package for comparing biological themes among gene clusters. *Omics,* 16**,** 284-7.

YUDANIN, N. A., SCHMITZ, F., FLAMAR, A. L., THOME, J. J. C., TAIT WOJNO, E., MOELLER, J. B., SCHIRMER, M., LATORRE, I. J., XAVIER, R. J., FARBER, D. L., MONTICELLI, L. A. & ARTIS, D. 2019. Spatial and Temporal Mapping of Human Innate Lymphoid Cells Reveals Elements of Tissue Specificity. *Immunity,* 50**,** 505-519.e4.

ZAID, A., MACKAY, L. K., RAHIMPOUR, A., BRAUN, A., VELDHOEN, M., CARBONE, F. R., MANTON, J. H., HEATH, W. R. & MUELLER, S. N. 2014. Persistence of skin-resident memory T cells within an epidermal niche. *Proc Natl Acad Sci U S A,* 111**,** 5307-12.

ZAYNAGETDINOV, R., SHERRILL, T. P., GLEAVES, L. A., MCLOED, A. G., SAXON, J. A., HABERMANN, A. C., CONNELLY, L., DULEK, D., PEEBLES, R. S., JR., FINGLETON, B., YULL, F. E., STATHOPOULOS, G. T. & BLACKWELL, T. S. 2015. Interleukin-5 facilitates lung metastasis by modulating the immune microenvironment. *Cancer Res,* 75**,** 1624-1634.

ZHANG, B., GAITERI, C., BODEA, L.-G., WANG, Z., MCELWEE, J., PODTELEZHNIKOV, A. A., ZHANG, C., XIE, T., TRAN, L. & DOBRIN, R. 2013. Integrated systems approach identifies genetic nodes and networks in late-onset Alzheimer's disease. *Cell,* 153**,** 707-720.

ZHANG, C., ZOU, Q., XIAO, S.-J. & JI, Z.-L. 2010. TiSGeD: a database for tissue-specific genes. *Bioinformatics,* 26**,** 1273-1275.

ZHANG, D.-L., QU, L.-W., MA, L., ZHOU, Y.-C., WANG, G.-Z., ZHAO, X.-C., ZHANG, C., ZHANG, Y.-F., WANG, M., ZHANG, M.-Y., YU, H., SUN, B.-B., GAO, S.-H., CHENG, X., GUO, M.-Z., HUANG, Y.-C. & ZHOU, G.-B. 2018a. Genome-wide identification of transcription factors that are critical to non-small cell lung cancer. *Cancer Letters,* 434**,** 132-143.

ZHANG, J., MAROTEL, M., FAUTEUX-DANIEL, S., MATHIEU, A. L., VIEL, S., MARCAIS, A. & WALZER, T. 2018b. T-bet and Eomes govern differentiation and function of mouse and human NK cells and ILC1. *Eur J Immunol,* 48**,** 738-750.

ZHANG, L. H., SHIN, J. H., HAGGADONE, M. D. & SUNWOO, J. B. 2016. The aryl hydrocarbon receptor is required for the maintenance of liver-resident natural killer cells. *J Exp Med,* 213**,** 2249-2257.

ZHANG, R., LAHENS, N. F., BALLANCE, H. I., HUGHES, M. E. & HOGENESCH, J. B. 2014. A circadian gene expression atlas in mammals: Implications for biology and medicine. *Proceedings of the National Academy of Sciences,* 111**,** 16219-16224.

ZHANG, S., WEINBERG, S., DEBERGE, M., GAINULLINA, A., SCHIPMA, M., KINCHEN, J. M., BEN-SAHRA, I., GIUS, D. R., YVAN-CHARVET, L., CHANDEL, N. S., SCHUMACKER, P. T. & THORP, E. B. 2019. Efferocytosis Fuels Requirements of Fatty Acid Oxidation and the Electron Transport Chain to Polarize Macrophages for Tissue Repair. *Cell Metabolism,* 29**,** 443-456.e5.

ZHANG, X., LAN, Y., XU, J., QUAN, F., ZHAO, E., DENG, C., LUO, T., XU, L., LIAO, G., YAN, M., PING, Y., LI, F., SHI, A., BAI, J., ZHAO, T., LI, X. & XIAO, Y. 2018c.

CellMarker: a manually curated resource of cell markers in human and mouse. *Nucleic Acids Research,* 47**,** D721-D728.

ZHAO, J., ZHANG, S., LIU, Y., HE, X., QU, M., XU, G., WANG, H., HUANG, M., PAN, J., LIU, Z., LI, Z., LIU, L. & ZHANG, Z. 2020. Single-cell RNA sequencing reveals the heterogeneity of liver-resident immune cells in human. *Cell Discovery,* 6**,** 22.

ZHOU, G., YU, L., FANG, L., YANG, W., YU, T., MIAO, Y., CHEN, M., WU, K., CHEN, F., CONG, Y. & LIU, Z. 2018. CD177⁺ neutrophils as functionally activated neutrophils negatively regulate IBD. *Gut,* 67**,** 1052-1063.

ZHOU, X., TANG, J., CAO, H., FAN, H. & LI, B. 2015. Tissue resident regulatory T cells: novel therapeutic targets for human disease. *Cell Mol Immunol,* 12**,** 543-52.

ZHU, L., LEI, J., DEVLIN, B. & ROEDER, K. 2018. A UNIFIED STATISTICAL FRAMEWORK FOR SINGLE CELL AND BULK RNA SEQUENCING DATA. *Ann Appl Stat,* 12**,** 609-632.