Transcriptional regulation and DNA damage response of extrachromosomal DNA in human glioblastoma stem cells

Karin Rita Purshouse

PhD

The University of Edinburgh

2023

Declaration

I declare that this thesis is of my own composition. The work presented in it is my own, unless stated otherwise by reference or acknowledgement. Any information derived from the published or unpublished work of others has been acknowledged in the text and references are listed in the bibliography. This work has not been submitted, in whole or in part, in any previous application for a degree.

6th April 2023 _

Date Signature

Abstract

Glioblastoma is a cancer characterised by limited treatment options and poor prognosis. Glioblastoma is driven by neural stem cell-like cells and is characterised by intratumoral heterogeneity. Extrachromosomal DNA (ecDNA) are circular regions of DNA that are seen in many cancers and are particularly frequent in glioblastoma. They are an important means of oncogene amplification, and correlate with treatment resistance and poor prognosis. Due to their frequency and association with accessible chromatin, it has been proposed that ecDNA oncogene transcription is amplified by their clustering with each other and key components of the transcriptional machinery. Such a mechanism could lead to greater levels of oncogene transcription than expected from copy number amplification alone. A major mechanism of ecDNA generation is massive DNA damage (chromothripsis), but the impact of DNA damage on existing ecDNA is yet to be fully characterised.

This thesis characterises ecDNA in five glioblastoma cell lines derived from patients using WGS and DNA FISH analysis. Super-resolution imaging and quantitative image analysis are used to evaluate the spatial organisation of ecDNA-resident oncogenes in glioblastoma cell lines. EcDNA are widely distributed throughout the nucleus, but a novel cluster analysis method demonstrates that ecDNA do not cluster closely with each other, nor do they closely engage with large transcriptional hubs. Focusing on the *EGFR* oncogene, transcriptional output is increased in cells harbouring *ecEGFR*. A combination of RNA:DNA FISH and RNAseq:WGS analysis is used to demonstrate that transcription per gene copy number is similar between chromosomal and ecDNA *EGFR* loci. This suggests increased ecDNA-resident oncogene transcription is primarily driven by copy number amplification rather than synergistic ecDNA regulatory processes and interactions.

Glioblastoma cells have many active DNA damage sites not closely related to ecDNA. To explore the impact of random and targeted DNA damage on ecDNA characteristics and dynamics, ionising radiation (IR) and CRISPR/Cas9 are utilised respectively. Treatment of *ecEGFR*-harbouring cells with IR results in a reduction in *EGFR* foci copy number and EGFR expression. Preliminary experiments suggest EGFR expression reduces further with co-treatment with the PARP inhibitor olaparib, although *ecEGFR* copy number appears protected. In a glioblastoma cell line with *ecEGFR*, CRISPR/Cas9 targeting *EGFR* generates cells with reduced EGFR expression. Cells can be FACS-sorted into EGFR High, Low and Null pools, in which Low and Null pools lose *ecEGFR* by DNA FISH and WGS analysis. Repeating this in a glioblastoma cell line with *EGFR* as a chromosomal amplification and *MYC* ecDNA demonstrates minimal effect on *EGFR*, but marked rearrangement of *ecMYC* loci.

Overall, these findings suggest that ecDNA are an important mechanism of oncogene amplification, the transcriptional effect of which is primarily driven by copy number. Understanding the impact of DNA damage on ecDNA and cell selection is important for development of new therapeutic strategies that successfully target genomic vulnerabilities.

Lay Summary

Glioblastoma is the most common type of brain cancer and is particularly aggressive with limited treatment options. Extrachromosomal DNA (ecDNA) are small circles of genetic information that exist separately to the rest of our genetic code, and can contain cancer information, known as oncogenes. EcDNA are an important way for oncogenes to make cancer cells more powerful, mainly because they drive messages that make cancer active. Whereas most of our genetic code is very organised with only one copy of each piece of information from each parent, ecDNA can exist as many copies in each cancer cell. They are seen in many cancers, but particularly glioblastoma. It has been suggested that ecDNA group together, or with other important machinery in cancer cells, to make the information carried on their oncogenes more powerful. It is also uncertain what happens to ecDNA when they are damaged when DNA is cut.

This thesis uses cells donated from patients with glioblastoma to explore these questions. EcDNA were examined with a powerful microscope. Mathematical analysis of these images shows that ecDNA don't cluster closely together, nor are they close to other important machinery in the cell. Oncogenes hold the code for cancer messages which can be converted into structures called proteins. Cells with more oncogenes have more cancer messages and more proteins. Analysing the genetic code in ecDNA and their cancer messages shows that this is mainly because there are more oncogenes, rather than the oncogenes on ecDNA behaving differently to those on normal DNA.

Radiotherapy, which creates damage in DNA, is used to treat most glioblastoma. My work has shown that ecDNA are affected by DNA damage. Radiotherapy creates random DNA cuts. My research show radiotherapy affects the number of ecDNA, even a short time after the radiotherapy dose is received. A drug that can stop DNA being repaired after cuts, called a PARP inhibitor, seems to limit this effect. I have also shown, using molecular scissors to make more precise DNA cuts, that making a cut in an oncogene present on ecDNA completely removes ecDNA. However, when cuts were created elsewhere - not on ecDNA - ecDNA were also unexpectedly affected. Understanding what happens when ecDNA are damaged is important for current and future cancer treatments.

Acknowledgements

I would like to thank Steve Pollard and Wendy Bickmore for supervising my PhD across their two labs and supporting me through two major unexpected interruptions to my studies. Our conversations and discussions have always been stimulating and inspiring, and I have learned a great deal from you both. Thank you also for encouraging me to believe in my ideas, whilst also sharing your knowledge and challenging me to think outside the box. I am grateful for your ongoing guidance and mentorship as I take my next steps in science.

I would like to thank the Wellcome Trust and the Edinburgh Clinical Academic Track (ECAT) programme for supporting my PhD; in particular, Jo Ness and Nick Gilbert. Thank you to my ECAT cohort who have provided laughs and friendship along our various PhD/pandemic experiences. I am also grateful to Lesley Dawson, Medical Oncology Consultant and trainee lead for NHS Lothian, for her support, kindness and enthusiasm. I am very fortunate to have such a supportive clinical ‘home’.

In the Pollard lab, I am especially grateful to Vivien Grant, who helped me with many of the practicalities of cell lines and transportation between two labs. Thanks also to Gillian Morrison and Alhafidz Hamdan who provided important insights and ideas. In the Bickmore lab, I am particularly indebted to Shelagh Boyle and Elias Friman, both for sharing their extensive expertise and skills, and more generally for their support and friendship. Thanks to Sjoerd Beentjes, Chancellor’s Fellow in the School of Mathematics, with whom I collaborated for the cluster analysis, for making maths achievable and fun! To the many informal discussions and friendships made in both labs, particularly in the Bickmore lab where I was based for the latter part of my PhD, I am very grateful.

Beyond the individual labs, I am grateful to the facilities and communities provided by both the MRC Human Genetics Unit (HGU) and Centre for Regenerative Medicine (CRM). In particular, I would like to thank the Advanced Imaging Resource (AIR) and flow cytometry facilities in the HGU, and the flow cytometry and imaging facility in the CRM.

Finally, it feels important to reflect that my PhD has formed the backdrop to the most personally challenging period of my life. From four months of clinical redeployment in the oncology department due to COVID19, to the short illness and loss of my father, and other major personal challenges in between, the end didn’t always seem possible. Getting there in one piece, still smiling and still enjoying science, has been largely due to the immense support of everyone listed above, to whom it is difficult to express my gratitude in strong enough terms. I am grateful to my family and friends who have walked the journey with me. Most importantly, thank you to my husband Dan who has been my constant supporter. Thank you for being the best team-mate.

Dedicated to my dad, Michael Purshouse, who would have been proud to see the end of my PhD journey.

Contents

List of figures

List of tables

Abbreviations

AA	AmpliconArchitect
AC-like	Astrocyte-like
ACTH	Adrenocorticotropic hormone
ADPr	ADPribose
AFP	Alpha-fetoprotein
AKT	Akt, also known as Protein Kinase B
AMPK	5' AMP-activated protein kinase
ARH3	(ADPr)hydrolase 3
ASR	Age–standardised incidence rate
ATAC-seq	Assay for Transposase-Accessible Chromatin using sequencing
ATM	Ataxia-telangiectasia mutated
ATR	ATM and Rad3 Related
BER	Base excision repair
BFB	Breakage-Fusion-Bridge
Bis-Tris	Bis(2-hydroxyethyl)iminotris(2-hydroxymethyl)methane
BRD4	Bromodomain-containing protein 4
BSA	Bovine Serum albumin
CDK4	Cyclin-dependent kinase 4
CDKN2A	Cyclin-dependent kinase inhibitor 2A
CEA	Carcinoembryonic antigen
CH13L1	Chitinase-3 like-protein-1
ChIA-PET	Chromatin interaction analysis by paired-end tag sequencing
ChIP	Chromatin immunoprecipitation
CHO	Chinese Ovarian Hamster
chr	Chromosome
CM	Complete Media
CMLE	Classic Maximum Likelihood Estimate
CNV	Copy number variation
CRE	Cis-regulatory element
CRISPR	Clustered Regularly Interspaced Short Palindromic Repeats
CSC	Cancer stem cell
CTCF	CCCTC-Binding factor
CTLA4	Cytotoxic T-lymphocyte-associated protein 4
DAPI	4′,6-diamidino-2-phenylindole
DDR	DNA Damage Response
DF	Deionised formamide
DHFR	Dihydrofolate reductase
DMs	Double minutes
DMSO	Dimethyl sulfoxide
DNA-PK	DNA Protein Kinase
DSB	Double stranded breaks
DSBR	DSB repair
DTT	Dithiothreitol
eccDNA	Extrachromosomal circular DNA
ecDNA	Extrachromosomal DNA
EDTA	Ethylenediaminetetraacetic acid
EGF(R)	Epidermal growth factor (receptor)
ERK	Extracellular signal–regulated kinase
EtOH	Ethanol
FACS	Fluorescence-activated cell sorting
FANC	Fanconi Anaemia
FCS	Fetal calf serum
FGF(R2)	Fibroblast growth factor (receptor 2)
FISH	Fluorescence in situ hybridisation
GCGR	Glioma Cellular Genetics Resource
GDP	Guanosine diphosphate
GFP	Green fluorescent protein
GRB2	Growth factor receptor-bound protein 2

GSCs	glioblastoma stem cells
GTP	Guanosine triphosphate
HCl	Hydrochloric acid
HER2	Receptor tyrosine-protein kinase erbB-2
HNSPCC	Hereditary Non-Polyposis Colorectal Cancer
HPF1	Histone PARylation Factor 1
HR	Homologous recombination
HSRs	Homogeneously staining regions
hTERT	Human telomerase catalytic component
IDH	Isocitrate dehydrogenase
IF	Immunofluorescence
iPSCs	Induced pluripotent stem cells
IR	Ionising Radiation
JAK	Janus Kinase
KCl	Potassium chloride
LDS	Lithium dodecyl sulfate
LLPS	Liquid-liquid phase separation
LOH	Loss of heterozygosity
MAPK	Mitogen activated protein kinase
MDM2	Mouse double minute 2 homolog
MDR1	Multidrug Resistance 1
Me:Ac	Methanol:acetic acid
MEK	MAPK/ERK kinase
MERTK	MER Proto-Oncogene, Tyrosine Kinase
MES-like	Mesenchymal-like
MET	Mesenchymal–epithelial transition factor
MGMT	Methyl-guanine methyl transferase
$MgSO_4$	Magnesium sulphate
MLH	MutL homologs
MMEJ	Microhomology-mediated end joining
MMR	Mismatch repair
mOS	Median overall survival
MRE	MRE11–RAD50–NBS1
MSH	MutS homologs
mTOR(c)	Mammalian target of rapamycin (complex)
MTX	Methotrexate
NaCit	Sodium citrate dihydrate
NaCl	Sodium chloride
NAD+	Nicotinamide-adenine-dinucleotide
NaOH	Sodium hydroxide
NER	Nucleotide excision repair
NES	Nestin
NF1	Neurofibromatosis 1
NHEJ	Non-homologous end joining
NKX2-2	NK2 Homeobox 2
NP-40	Nonidet P-40
NPC-like	Neural-progenitor-like
NSC	Neural stem cell
NSCs	Neural stem cells
ODG	Oligodendrogliomas
OLIG2	Oligodendrocyte transcription factor
OM	Optical mapping
OPC-like	Oligodendrocyte-progenitor-like
PARG	PolyADPr glycohydrolase
PARP	Poly(ADP-ribose) polymerase
PARPi	PARP inhibitor
PBS	Phosphate buffered saline
PD1	Programmed cell death protein 1
PDGFRA	Platelet derived growth factor receptor A
PDK1	Pyruvate Dehydrogenase Kinase 1

PDL1	Programmed cell death protein ligand 1
pFA	Paraformaldehyde
pH3	Phospho-Histone H3
PI3K	Phosphatidylinositol 3-kinase
PIC	Preinitiation Complex
PIK3CA	p110α catalytic subunit of PI3K
PKC	Protein kinase C
PLC	Phosphoinositide-specific phospholipase C
PML	Promyelocytic leukemia
PPV	Positive predictive value
PSF	Point spread function
PTEN	Phosphatase and tensin homolog
PTH	Parathyroid hormone
Rb	Retinoblastoma
RNA Pol II	RNA polymerase II
RNP	Ribonucleoprotein
ROS	Reactive oxygen species
RPA	Replication Protein A
RTK	Receptor tyrosine kinase
SDS	Sodium dodecyl sulfate
SDSA	Synthesis-dependent strand annealing
SE	Super enhancer
Ser-ADPr	Serine-ADPr
Shh	Sonic hedgehog
SNPs	Single nucleotide polymorphisms
SNR	Signal-to-noise ratio
SOS	Son of sevenless homolog
SSB	Single stranded breaks
SSC	Trisodium citrate and sodium chloride
STAT	Signal transducer and activator of transcription
TAD	Topologically associating domain
TARG	Terminal ADP-ribose protein glycohydrolase
TBS	Tris-buffered saline
TBST	Tris-buffered saline with 0.1% Tween 20
TE	Tris/EDTA
TF	Transcription factor
TMZ	Temozolomide
TNF	Tumour necrosis factor
TOP	Topoisomerase
TP53	Tumour protein p53
TREX1	Three prime repair exonuclease 1
Tris	Trisaminomethane
TSC	Tuberous Sclerosis Complex
TSG	Tumour suppressor gene
TSS	Transcription start site
TTF	Tumour Treating Fields
VEGF-A	Vascular endothelial growth factor A
WGS	Whole genome sequencing
WM	Wash Media
wt	Wild type
XLF	XRCC4-like factor
XP	Xeroderma Pigmentosa
XRCC	X-Ray Repair Cross Complementing

Chapter 1 Introduction

1.1 Glioblastoma

1.1.1 Clinical context of Glioblastoma

Over 12,000 people are diagnosed with brain tumours every year in the UK, with incidence rates increasing by 11% in the last decade (CRUK, 2021). Gliomas represent over half of these diagnoses. Adult-type diffuse gliomas are classified, based on histological features (microscopic morphology, grade and immunohistochemistry) and molecular markers, as astrocytomas, oligodendrogliomas (ODG) or glioblastomas (Lin et al., 2021; Louis et al., 2021). They are histologically diffuse, meaning that the tumours are infiltrative and the boundary between tumour and normal brain tissue poorly demarcated, rendering complete resection surgically challenging. Molecular markers are now fully integrated into the most recent (5th) edition of the World Health Organisation diagnostic classification system (Louis et al., 2021). Prior to this glioblastoma was described as a primary (i.e. *de novo*) tumour (90% of cases), or secondary low-grade astrocytoma that subsequently progressed (~10% of cases) (Louis et al., 2016). Secondary glioblastoma can, however, no longer be used diagnostically as these tumours have hallmark mutations of isocitrate dehydrogenase (IDH) genes. Glioblastoma is therefore now considered diagnostically distinct from other diffuse gliomas, by combining the wild-type (wt) IDH status with histological features (microvascular proliferation, nuclear pleomorphism and regions of necrosis) of a high grade glioma. Within this revised classification, glioblastoma is defined as IDH-wt and Grade IV, the highest grading available to describe the undifferentiated, abnormal microscopic pattern (Louis et al., 2021).

Glioblastoma represents the most common primary brain tumour, with an age–standardised incidence rate of 3.2-5.0 per 100,000 person years (Philips et al., 2018; Ostrom et al., 2019). Glioblastoma is a devastating diagnosis, with limited treatment options, almost universal recurrence and a median overall survival (mOS) from initial diagnosis of only 12-18 months (Stupp et al., 2005, 2014). Clinical presentation of glioblastoma is dependent on the site and size of the tumour, with symptoms including headache, nausea, motor/visual/speech deficit or seizures. Most tumours arise in the supra-tentorial space, with around a quarter arising in the frontal lobe (Gilard et al., 2021). Diagnosis is via magnetic resonance imaging, where tumours are classically visualised as a poorly circumscribed non-enhancing central region surrounded by contrast enhancement, which reflects disruption to the blood-brain barrier (Yan et al., 2019; Gilard et al., 2021). Imaging guides the strategy by which tissue for histological and molecular testing is obtained, either by surgical resection or biopsy.

Patients generally undergo maximum surgical resection, followed by radiotherapy and, if eligible, temozolomide (TMZ) chemotherapy (Stupp et al., 2014). TMZ, an alkylating agent, results in greater treatment response in patients with *methyl-guanine methyl transferase* (*MGMT*) promoter methylation by inhibiting repair of TMZ-induced DNA damage (Hegi et al., 2005, 2008; Stupp et al., 2014). Tumours virtually always recur, and do so locally in about 80% of cases (Piper et al., 2018). Options at recurrence are limited to retreatment (surgery or radiotherapy) or alternative chemotherapy.

Since the introduction of TMZ ~20 years ago, there have been no major treatment developments, despite trials involving highly diverse therapeutic strategies. Targeting of dysregulated angiogenesis and upregulated vascular endothelial growth factor A (VEGF-A) signalling seen in glioblastoma by the VEGF inhibitor bevacizumab showed promise in early-phase trials, but ultimately showed no survival benefit in a phase 3 clinical trial (Jain et al., 2007; Kreisl et al., 2009; Gilbert et al., 2014). Targeting the immune system via cytotoxic T-lymphocyte-associated protein 4 (CTLA-4), programmed cell death protein 1 (PD1) and PD1 ligand (PDL1) inhibitors, termed checkpoint inhibitors, has revolutionised the treatment of other cancers, but has shown no benefit in glioblastoma. This is thought to be due to the presence of an immunosuppressive tumour microenvironment (Reardon et al., 2020; Chuntova et al., 2021; Gangoso et al., 2021; Omuro et al., 2023). Combining immunotherapy with the activation of a CD8-T cell response via gene therapy is currently being explored in early phase clinical trials in glioblastoma (Lang et al., 2018; Chiocca et al., 2022). Non-invasive innovations such as Tumour Treating Fields have shown promise with a randomised non-sham device-controlled trial suggesting an overall survival advantage of 4.9 months (Stupp et al., 2017).

Challenges remain at every level of innovation development in glioblastoma, including drug delivery, the blood-brain barrier, markers of efficacy and clinical trial design (Aldape et al., 2019). Ultimately, none of these innovations has proven successful enough to integrate into clinical practice, either in primary or recurrent glioblastoma.

1.1.2 Genomic landscape of glioblastoma

Glioblastoma is characterised by its intra- and inter-tumoural heterogeneity, featuring alterations in major oncogenic signalling pathways, most prominently receptor tyrosine kinase (RTK)/Ras/phosphatidylinositol 3-kinase (PI3K) (altered in 88-90% of glioblastoma), tumour protein p53 (TP53) (86-87%) and Retinoblastoma (Rb) (78-79%), with 74% harbouring alterations in all three (Cancer Genome Atlas Research Network, 2008; Brennan et al., 2013) (Figure 1.1). These pathways involve key oncogenes, such as *epidermal growth factor receptor (EGFR)* and *cyclin-dependent kinase 4* (*CDK4)*, and tumour suppressor genes (TSGs), such as *TP53* and *phosphatase and tensin homolog (PTEN)*. Alterations in these pathways, which upregulate downstream pathways as varied as mitogen activated protein kinase (MAPK) pathway signalling, senescence, apoptosis and cell cycle regulation, are essential in glioblastoma tumorigenesis (Brennan et al., 2013).

Genomic characterisation by consensus clustering led to the proposal of four molecular subclasses of glioblastoma, each enriched in different genomic aberrations and survival trends (Phillips et al., 2006; Verhaak et al., 2010). Reanalysis due to greater stringency in glioblastoma-specific mRNA classification during transcriptome analysis has since proposed that the neural class may have represented a subtype artifact due to neuronal contamination. Classical, Mesenchymal and Proneural emerged as the three main transcriptional molecular subtypes (IDH-wt only) (Wang et al., 2017) (Table 1.1). Glioblastoma subtypes have also been proposed based on DNA methylation, with 6 distinct glioblastoma methylation clusters, M1-M6, where M5 is the G-CIMP phenotype and has been linked to the improved survival advantage seen in the proneural subtype (Brennan et al., 2013). Reanalysis of the datasets from which these subtypes were derived may yet lead to an evolution in glioblastoma subtype categorisation, given that the presence of IDH-mutant (previously secondary glioblastoma) has been associated with more favourable clinical outcomes (Brennan et al., 2013).

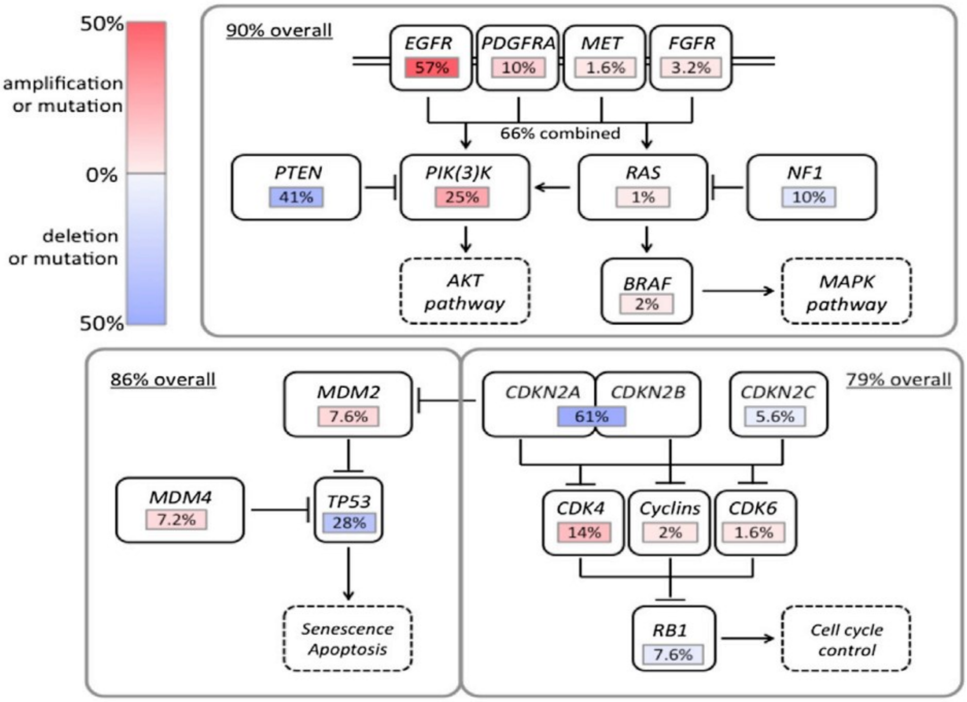


Figure 1.1 | Pathway alterations in glioblastoma

From TCGA datasets (Brennan et al., 2013 (Figure 4A)) showing shared genetic aberrations affecting RTK/PI3K/MAPK, p53 and Rb pathways (see large boxes with percentage of tumours in TCGA glioblastoma dataset with a mutation in the corresponding pathway). Red = activating mutations, Blue = inactivating mutations, with colour intensity corresponding with frequency of mutation in that gene, shown as a percentage of tumours in the TCGA glioblastoma dataset.

Molecular subtype	Genomic profile	Impact of intensive treatment on survival
Classical	Chr10 loss, chr7 gain EGFR amp CDKN2A del No TP53 mutations Highly expressed pathways: NES, Notch, Shh	Highest sensitivity to intensive treatment; significantly reduced mortality
Mesenchymal	NF1 del CH13L1, MET, CD44 MERTK Highly expressed pathways: TNF, NFκB	Significantly reduced mortality
Proneural	PDGFRA mut and amp IDH1 mut TP53 mut and LOH NKX2-2, OLIG2 PIK3CA amp Highly expressed pathways : Sox	No impact, note markedly better baseline prognosis than other subtypes

Table 1.1 | Molecular classification of glioblastoma

Key genomic alterations and other expressed pathways, with impact on clinical outcomes, summarised from (Verhaak et al., 2010; Wang et al., 2017). Chr – chromosome, amp – amplification, del – deletion, mut – mutation.

The signalling pathways mutated in glioblastoma are typically also seen in other carcinomas. RTKs initiate signalling pathways activated by ligand-binding to a corresponding extracellular kinase domain, which in turn leads to kinase homo- or heterodimerisation and activation of downstream pathways including the Ras/Raf/MAPK/extracellular signal–regulated kinase (ERK) pathway and PI3K/Akt/mTOR pathways (Figure 1.2). Examples of RTKs that are commonly amplified in glioblastoma include EGFR, platelet derived growth factor receptor A (PDGFRA) and mesenchymal–epithelial transition factor (MET). EGFR is the most commonly amplified RTK in glioblastoma and *EGFR* is altered in 57.4% of glioblastoma (Brennan et al., 2013). Aberrations in *EGFR*, its protein and downstream effector pathways are observed and indeed successfully targeted in many cancers, such as non-small cell lung and colorectal cancers (Furnari et al., 2015; An et al., 2018). A unique set of *EGFR* mutations is seen in glioblastoma, with *EGFR variant III* (EGFRvIII - deletion of exons 2-7) the most common (Brennan et al., 2013; Koga et al., 2018). *EGFRvIII* results in a truncated protein that is constitutively activated independent of ligand binding. Despite the wide range of available EGFR-targeting agents, EGFR inhibitors have shown no impact on patient outcomes (Reardon, Wen and Mellinghoff, 2014; Lee et al., 2020), and no other RTK-targeting agents have shown efficacy in glioblastoma. Contributory factors likely include the presence of other driver mutations and intra-tumoural heterogeneity in the signalling pathway mutations that make resistance easy to achieve (Verhaak et al., 2010; Snuderl et al., 2011; Szerlip et al., 2012; Brennan et al., 2013). Such redundancy has been addressed in other cancers by targeting more than one pathway component, such as dual Raf and MAPK/ERK kinase (MEK) inhibitor therapy (e.g. dabrafenib and trametinib) in BRAF-mutant melanoma which now represents standard-of-care (Sanchez, Wang and Cohen, 2018). This is a promising approach that many groups are exploring now for glioblastoma.

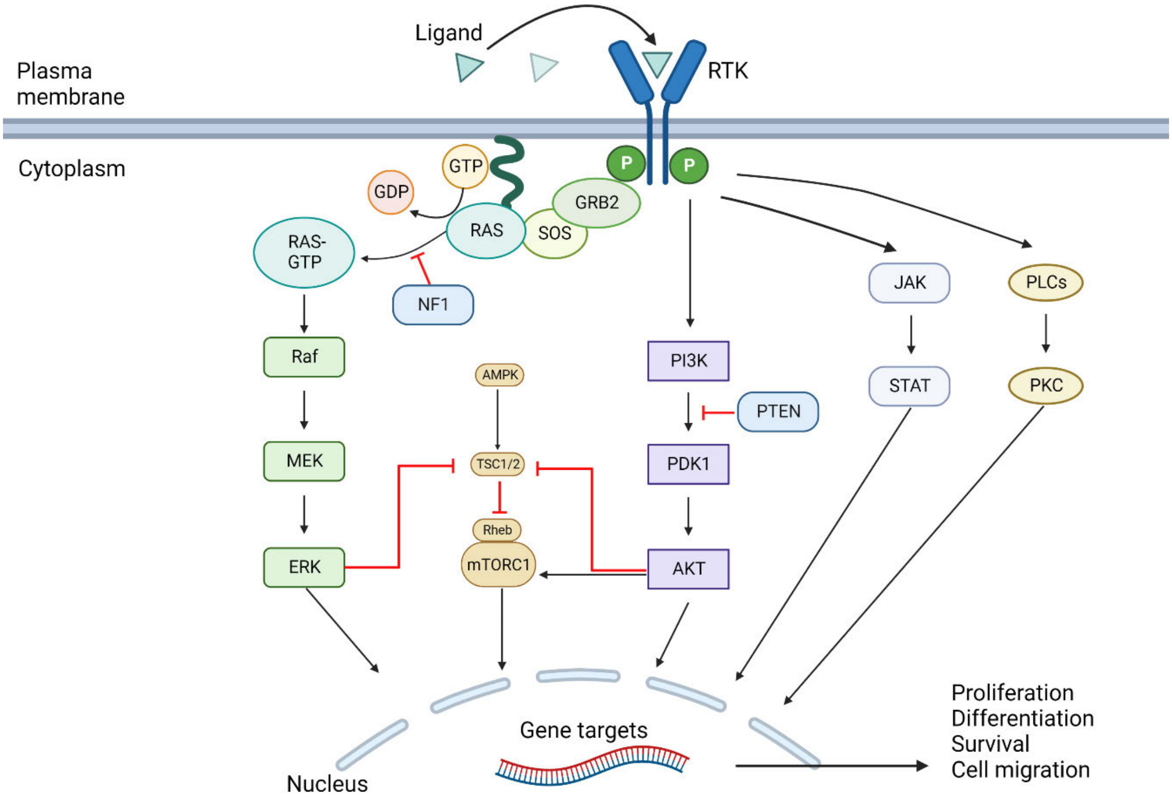


Figure 1.2 | Overview of RTK pathways

RTKs facilitate signal transduction across the plasma membrane. The ligand binds to the extra-cellular receptor, initiating intracellular signalling pathways via dimerisation-triggered kinase phosphorylation (P). These include the Ras/Raf/MAPK/ERK pathway, the PI3K/AKT pathway, the Janus Kinase (JAK)/ signal transducer and activator of transcription (STAT) pathway and the protein kinase C (PKC) pathway. These pathways interact with a wide range of gene targets, driving proliferation, differentiation, survival and cell migration. In addition, they link with cell cycle, angiogenesis, autophagy and lipid synthesis pathways. EGFRvIII lacks the extracellular ligand binding domain yet displays constitutive signalling. Figure created with Biorender.com.

Glioblastoma may be categorised by a genomic profile defined at diagnosis, but this profile remains highly heterogeneous over time and space, both within and between patients. It has been shown that tumours from an individual patient can harbour multiple genomically distinct clonal populations (Sottoriva et al., 2013; Patel et al., 2014; Meyer et al., 2015), which have been anatomically mapped by the Ivy Glioblastoma Atlas Project to link transcriptional signatures to their anatomical location within a tumour (Puchalski et al., 2018). The genomic profile of tumours is likely to evolve between initial diagnosis and recurrence, with only 55% of tumours retaining their subtype, along with other changes in the immune microenvironment (Wang et al., 2017). The clonal diversity and evolution of glioblastoma is supported by processes including functional subclone cooperativity, whereby driver mutations such as *EGFRvIII* may represent a minority population but promote paracrine activation of *EGFR-wt*-harbouring clones (Inda et al., 2010; McGranahan and Swanton, 2017).

Transcriptome-defined glioblastoma subtypes represent a dynamic spectrum rather than distinct entities, with functional redundancy and plasticity driving ongoing evolution, including in response to cancer treatment (Brennan et al., 2013). The failure of any therapeutic agents to successfully target these pathways highlights the challenges posed by genomic heterogeneity, clonal evolution and the dynamic tumour microenvironment. The limited benefit across all cancer types of genome-focussed targeted therapies over the last decade (Marquart, Chen and Prasad, 2018) has been frustrating in many instances for common solid tumours. This suggests a need for greater understanding of cancer biology (beyond genome sequencing) in order to identify cancer-acquired vulnerabilities and their corresponding pathways.

1.1.3 Cellular hierarchy

Attempts to understand the origins of glioblastoma heterogeneity have led to the key discovery that stem cell-like cells underlie glioblastoma initiation and progression. These glioblastoma stem cells (GSCs) are thought to hijack neurodevelopment mechanisms to achieve their characteristic heterogeneity and plasticity (Aldape et al., 2019; Neftel et al., 2019).

Stem cells represent the apex of the developmental hierarchy. They often exist in both quiescent and proliferative states, and are regulated by local niches. (Barker, Bartfeld and Clevers, 2010; Lathia et al., 2015). The cancer stem cell (CSC) hypothesis suggests that tumour cells acquire (or arise from) tissue stem cells and are the drivers of cancer and their cellular heterogeneity either by stochastic clonal evolution influenced by intrinsic or extrinsic factors, a more hierarchical model driven by CSCs and its differentiated derivatives, or a ‘unified’ dynamic model (Kreso and Dick, 2014). After extensive study in hematopoietic systems, CSCs were first isolated in solid tumours in breast cancer and shown to exist in many solid-organ cancers (Al-Hajj et al., 2003; Tan et al., 2006). CSCs are thought to become more restricted in their tumour initiating potential by the process of differentiation. Despite sharing the same cancer genome, not all tumour cells are functionally equivalent and must therefore have altered transcriptional and epigenetic programs that restrict their malignant phenotype.

Glioblastomas have been shown to harbour tumorigenic neural stem cell-like cells (Ignatova et al., 2002; Galli et al., 2004). CD133+ CSCs isolated from glioblastoma tumours, have the capacity to self-renew, proliferate and differentiate in culture (Singh et al., 2003). These cells can also generate brain tumours when injected into mice, in contrast with CD133- cells (Singh et al., 2004). This has been further developed in models using other definitions of GSCs, given that CD133 is neither a sensitive nor specific GSC marker (Beier et al., 2007; Pollard et al., 2009). Using a genetically engineered mouse model to follow glioblastoma tumorigenesis, neural cell types with different differentiated states revealed an inverse relationship between malignant transformation and lineage restriction, suggesting that stem cells may often arise through transformation of the endogenous neural stem and progenitor cells (Alcantara Llaguno et al., 2019).

Neural stem- and progenitor cell fates are dictated by master regulatory transcription factors (TFs) (Takahashi and Yamanaka, 2006; Orkin and Hochedlinger, 2011), and the same is true of GSCs. GSCs are driven and sustained by four core neurodevelopmental TFs - POU3F2 (BRN2), SOX2, SALL2 and OLIG2 (Suvà et al., 2014). These TFs operate downstream of recognised signalling pathways, including RTK signalling. Silencing *SOX2* has been shown to inhibit proliferation and tumorigenicity in glioblastoma tumour initiating cells (Gangemi et al., 2009). These TFs operate downstream of recognised signalling pathways, including RTK signalling.

Combining lineage tracing with transcriptome data has led to a more integrated perspective of the cellular states driving glioblastoma and how these correlate with genomic alterations. Glioblastoma cells have been proposed to exist in four cellular states - neural-progenitor-like (NPC-like), oligodendrocyte-progenitor-like (OPC-like), astrocyte-like (AC-like) and mesenchymal-like (MES-like) (Neftel et al., 2019) (Figure 1.3). These do not describe restricted subtypes, but a continuum of developmental cell states between which cells in an individual tumour can transition. A similar model of plasticity between cell states has since been described for immune evasion cell states (L.-B. Wang et al., 2021) and a developmental/inflammatory injury response (Richards et al., 2021) which has been functionally confirmed using mouse models (Gangoso et al., 2021).

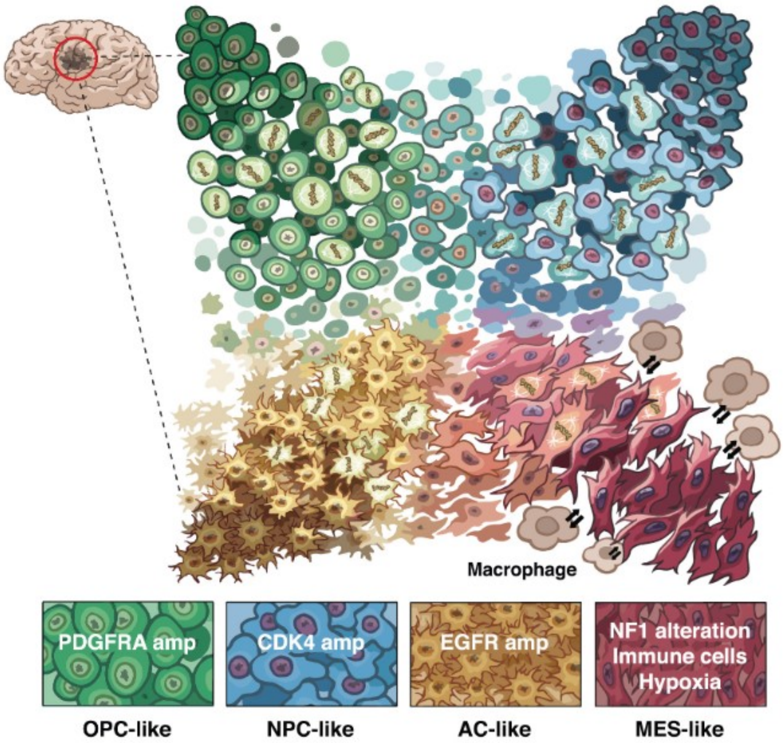


Figure 1.3 | Cellular states in glioblastoma

From (Neftel et al., 2019) Figure 7G. A model for cellular states, integrating lineage tracing and transcriptome data, and demonstrating intermediate and transition states between the four neural lineage-defined cellular states.

Taken together, although glioblastoma is highly heterogeneous and dynamic owing to genetic, epigenetic, and neurodevelopmental mechanisms, the commonality seems to be that a neural stem cell-like identity reminiscent of foetal forebrain development is always observed molecularly and histologically. These considerations have led to recent improved mouse and human glioblastoma models.

1.1.4 Models of glioblastoma

It is important to select an *in vitro* model that recapitulates the features of GSCs (Robertson et al., 2019). Human astrocytes can be immortalised via human telomerase catalytic component (hTERT) in cooperation with Ras-pathway activation, or inhibition of the TP53 pathway (Rich et al., 2001; Sonoda et al., 2001). Both of these models fail to truly recapitulate GSC biology, including genetic heterogeneity and stem cell pathways. Established glioma cell lines, such as U87 or U251, which are grown in serum-containing media, are often inadvertently driven away from a stem cell pathway towards a differentiated astrocytic state with the resulting tumours lacking the infiltrative and migratory histology seen in glioblastoma (Lee et al., 2006). Induced pluripotent stem cells (iPSCs) genetically altered by deletion of *PTEN/NF1* and *TP53/PDGFRA* (exon 8-9), result in mesenchymal and proneural progenitor and glioma cell lines (Koga et al., 2020). These create tumour forming cells that recapitulate many features of GSCs, including tumour heterogeneity and genomic instability. However, this only reflects a discrete group of genetic mutations and glioblastoma subtypes, and it is unclear if iPSCs reflect a relevant primary cell.

Patient-derived primary glioblastoma cell cultures permit the study of GSC pathobiology in the model format that most reflects the disease. These can be grown in neurospheres (Galli et al., 2004), or in an adherent 2D culture with laminin which facilitates a greater range of experimental opportunities (Pollard et al., 2009). The key culture conditions used here were those first optimised for human foetal neural stem cells (NSCs) (Conti et al., 2005; Pollard et al., 2009). A limitation of these genetically heterogeneous primary cell cultures is experimental standardisation. However, targeted gene editing via CRISPR/Cas9 is both feasible and efficient in NSCs (Bressan et al., 2017) and they can act as a control for stem cell pathways and normal copy number of genes that might be altered in GSCs. The Glioma Cellular Genetics Resource (GCGR) has generated a novel set of genetically characterised human glioblastoma and foetally-derived NSC lines with matched molecular profiling data (Figure 1.4).

Figure 1.4 | GCGR Cell lines

Summary of GCGR cell lines (x-axis), commonly mutated gene (left y-axis = gene name, right y axis = % of cell lines with mutation), type of mutation and ploidy-status (see legend). WGS data. Note bottom panel denotes SNV mutational signatures. The two cell lines with primarily C>T mutations (E51 and E62) are not further explored in this thesis.

1.2 Extrachromosomal DNA

Extrachromosomal DNA (ecDNA) were first described when karyotype analyses of human cancers revealed abnormal chromosomal and additional non-chromosomal structures. The medulloblastoma tumour of an 8-year-old girl was shown to have multiple small DNA fragments, and a glioblastoma resected from a 43-year-old man harboured very long, abnormal chromosomes (Lubs and Salmon, 1965). The long, abnormal chromosomes would go on to be characterised in drug-resistant cell lines as homogeneously staining regions (HSRs) (Biedler and Spengler, 1976). A larger case series (3 neuroblastoma, 1 medulloblastoma, 1 rhabdomyosarcoma, 1 bronchial carcinoma) identified minute centromere-free double chromatin bodies which were variable from cell to cell (Cox, Yuncken and Spriggs, 1965) and defined as double minutes (DMs) (Kaufman, Brown and Schimke, 1979). The existence of these chromatin bodies as doublets or singlets has resulted in the more global description of these chromosome-independent bodies as ecDNA (Hamkalo et al., 1985).

1.2.1 EcDNA composition

EcDNA structures were described with remarkable accuracy via simple karyotyping and molecular approaches, with novel imaging and bioinformatics tools confirming and further enhancing our understanding of ecDNA structures. EcDNA often exist as singlets, with only 30% shown to be paired doublets (Turner et al., 2017). EcDNA interact dynamically to integrate with chromosomes to form more stable chromosomal focal amplifications – HSRs (Balaban-Malenbaum and Gilbert, 1980; Hamkalo et al., 1985; Vogt et al., 2004; Storlazzi et al., 2010; Verhaak, Bafna and Mischel, 2019). Sequencing and associated analytical tools have enabled higher resolution structural characterisation of ecDNA. EcDNA in mammalian cancer cells had originally been hypothesised to be circular through comparison with similar structures in other cell types such as protozoa (Schimke, 1984). Combining sequencing tools with microscopy has confirmed ecDNA are indeed circular and, approximately 1-3Mb in size, although this may extend up to 5Mb (Turner et al., 2017; Deshpande et al., 2019; Verhaak, Bafna and Mischel, 2019; Wu et al., 2019). Their large size differentiates these elements from other circular extrachromosomal structures, such as extrachromosomal circular DNA (eccDNA), which are predominantly

<25kb in size, do not usually contain genes or regulatory sequences and are present in normal and cancer cells (Møller et al., 2016).

The first studies linking ecDNA to gene amplifications were in a range of murine cell types and Chinese Ovarian Hamster (CHO) cells treated with methotrexate (MTX) to generate MTX resistance. Amplifications in the gene encoding dihydrofolate reductase (DHFR) were observed, the copy number of which correlated with the number of emergent DMs (Alt et al., 1978; Kaufman, Brown and Schimke, 1979; Haber and Schimke, 1981; Haber et al., 1981). Subsequently, ecDNA and HSRs were confirmed as the location of oncogene amplifications in a range of glioma, neuroblastoma and colorectal cell lines (Alitalo et al., 1983; Kohl et al., 1983; Bigner et al., 1987). Studies across many cancers have shown that the most common focal oncogene amplifications are all located on ecDNA and/or HSRs, and indeed enable oncogene copy number to be amplified tens to hundreds of times, with significant copy number heterogeneity (Turner et al., 2017; Lange et al., 2022). Given that ecDNA replicate only once during the S phase of each cell cycle (Barker et al., 1980), it has been suggested that oncogene amplification occurs via random segregation at mitosis with subsequent cell selection favouring ecDNA-harbouring cells (Figure 1.5) (Lange et al., 2022).

As well as their resident oncogenes, ecDNA also harbour regulatory elements (enhancers) required to drive oncogene expression (Morton et al., 2019; Helmsauer et al., 2020; Zhu et al., 2021). EcDNA-resident enhancers have been proposed to interact with oncogenes in *cis* (i.e. same ecDNA) and *trans* (structurally separate oncogene and ecDNA-resident enhancer) (Helmsauer et al., 2020; Zhu et al., 2021). EcDNA may harbour regulatory elements independent of their relevant oncogenes, which has been proposed to facilitate *trans*-activation between enhancers and other ecDNA-inhabiting oncogenes (Hung et al., 2022).

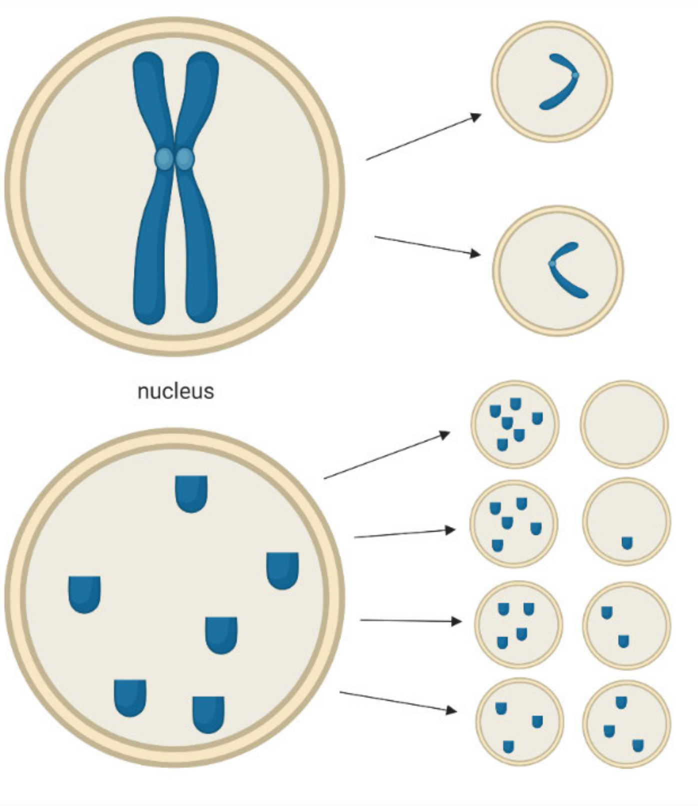


Figure 1.5 | EcDNA segregation as a source of copy number heterogeneity

Schematic of chromosomal segregation (top) during cell division, where daughter cells contain the product of equally divided sister chromatids following DNA replication owing to microtubule attachment to centromeres. EcDNA (bottom) lack centromeres, so ecDNA are randomly distributed between daughter cells, leading to copy number heterogeneity. Created in Biorender.

1.2.2 EcDNA and cancer

EcDNA are a frequent feature of many cancer types but are very rare in normal tissue (Benner, Wahl and Von Hoff, 1991; Turner et al., 2017; Kim et al., 2020). Although analysis of the Mitelman database initially suggested ecDNA were present in only 1.4% of cancers (Fan et al., 2011), an integrated study combining whole genome sequencing (WGS) and imaging across 17 cancer types using primarily cancer cell lines identified ecDNA in nearly half of cancers (Turner et al., 2017). A subsequent WGS study from 3,212 patients with cancer and 1810 non-cancer samples showed 14.3% of tumour samples harboured ecDNA and in 25 of 29 cancer types (Kim et al., 2020). Linking clinical and WGS data has shown that ecDNA amplification is associated with worse 5-year survival outcomes, although interestingly ecDNA level was not associated with metastatic status or previous cancer treatment (either systemic or radiation) (Turner et al., 2017; Kim et al., 2020).

EcDNA are particularly common in glioblastoma. The first karyotype analyses suggested 20-30% of glioblastoma harboured DMs (Bigner et al., 1988), increasing to 73.3% when considering only glioblastoma samples containing an oncogene amplification (Bigner et al., 1987). Large-scale analysis of WGS data has shown that ~50-60% glioblastoma cells carry ecDNA, rising to 90% of patient-derived glioblastoma tumour models (Turner et al., 2017; Kim et al., 2020). Not only are these preserved in neurosphere and xenograft models, but analysis of primary and recurrent glioblastoma cells also revealed high preservation of ecDNA (deCarvalho et al., 2018), likely due to the dynamic relationship between ecDNA and HSRs, in the face of treatment-based selection pressures (Nathanson et al., 2014). EcDNA are now considered the location of the key oncogene amplifications observed in glioblastoma (Vogt et al., 2004; Brennan et al., 2013).

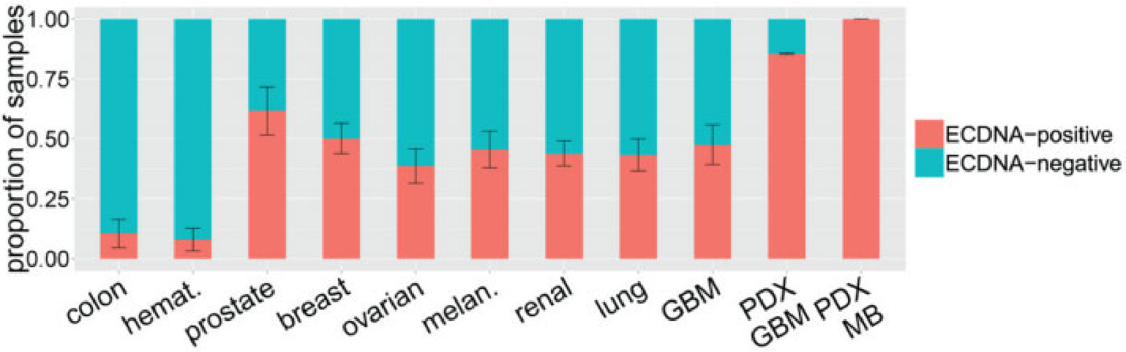


Figure 1.6 | EcDNA frequency across tumour types

From (Turner et al., 2017) Figure 2D. Proportion of cell cultures harbouring ecDNA by tumour type. PDX = Patient derived xenograft; MB = Medulloblastoma

The connection between ecDNA-resident oncogenes and cancer treatment resistance built on the earliest cellular models of ecDNA in murine cancer cells (Alt et al., 1978; Kaufman, Brown and Schimke, 1979). EcDNA behave dynamically in response to the selection pressure created by cancer treatment. A range of cancer cell lines, including a glioblastoma cell line, harbouring *MYC* amplifications on DMs, were treated with hydroxyurea. This resulted in a marked reduction in *MYC* copy number that was not observed in a cell line harbouring *MYC* on an HSR (Colo320HSR) (Von Hoff et al., 1992). The localisation of oncogenes on ecDNA have also been shown to play a role in RTK-targeted treatment resistance. For example, GBM39 cells (a patient-derived *EGFRvIII* expressing xenograft cell line) with high levels of *EGFRvIII*-containing ecDNA have increased RTK signalling and cell proliferation, increased metabolism, reduced apoptosis and enhanced cell death in the presence of the EGFR inhibitor, erlotinib (Nathanson *et al.*, 2014). Furthermore, cells that developed erlotinib resistance through chronic erlotinib exposure had a significant reduction in *EGFRvIII* ecDNA, an increase in *MDM2* ecDNA and maintenance of *EGFR* HSRs. *EGFRvIII* ecDNA were restored when erlotinib was removed (Nathanson et al., 2014). As might be expected given early studies linking ecDNA and chromosomal integration in the form of HSRs, ecDNA-resident oncogenes have previously been shown to integrate with chromosomes in this manner (Storlazzi et al., 2010; L'Abbate et al., 2014). The ability of single copy *EGFRvIII*-harbouring ecDNA to reintegrate with chromosomes to form multi-copy *EGFRvIII*-HSRs highlights this dynamic reintegration/release process as means of RTK treatment resistance (Turner et al., 2017). A similar mechanism for DM to HSR reintegration has been proposed following dual Raf/MEK-inhibitor treatment in *BRAF*-mutant ecDNA-null melanoma cells, where *BRAF* amplification, primarily via ecDNA formation, developed following Raf/MEK-inhibitor treatment. Interestingly, this study observed a preference from ecDNA to HSR conversion, initially via integration, during stable dual drug dosing but also showed that rare ecDNA could re-emerge from a predominantly HSR-*BRAF* population (Song et al., 2021).

Overall, these studies reveal that ecDNA are the primary source of oncogene amplification and heterogeneity in cancer, they integrate/re-emerge dynamically in response to selection pressures, and they are directly linked to treatment resistance and poor prognosis. Understanding the function and regulation of ecDNA is particularly important in glioblastoma as this might reveal new ways to target these tumours.

1.2.3 Formation and structure of ecDNA

A number of hypotheses for the formation of ecDNA have been proposed and the varied genomic composition of ecDNA has led to the proposal that multiple mechanisms may all contribute to ecDNA formation in different scenarios (T. Wang et al., 2021) (Figure 1.7).

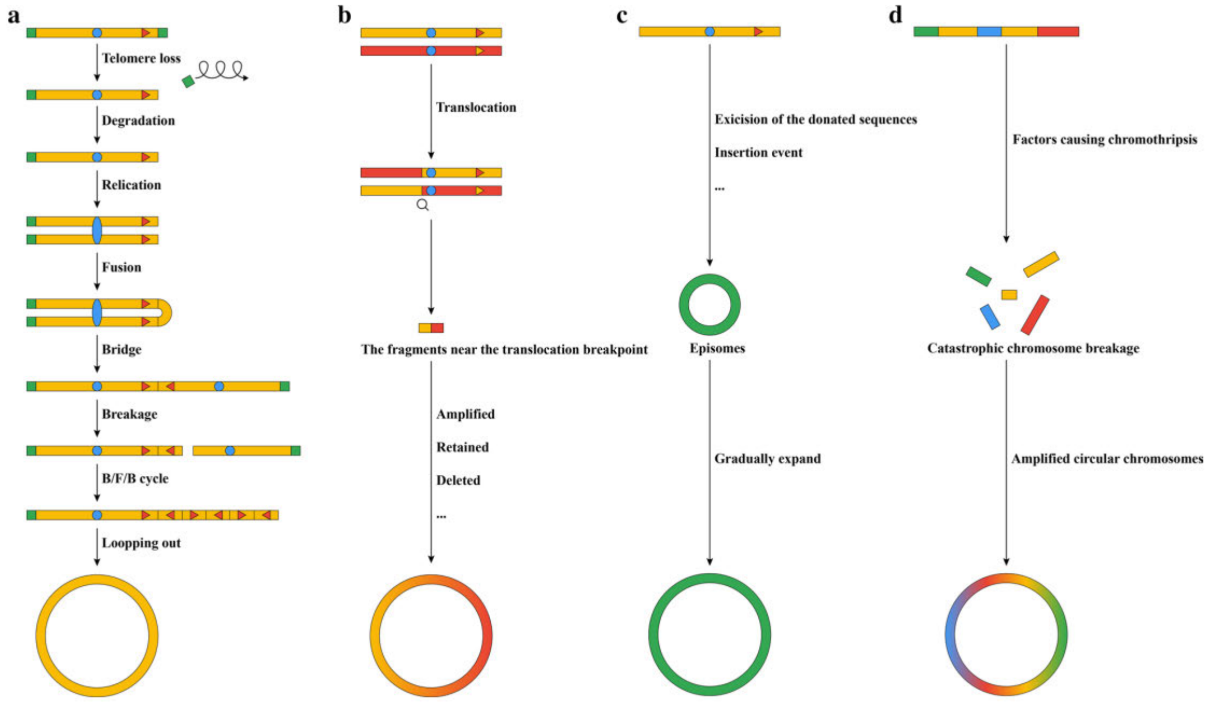


Figure 1.7 | Models for ecDNA formation

From (*T. Wang et al., 2021*)*, Figure 1. Schematic of the four proposed models of ecDNA formation. A) The Breakage-Fusion-Bridge (BFB) Cycle B) Translocation-(excision)-deletion-amplification C) Episome model D) Chromothripsis.*

1.2.3.1 Breakage-Fusion-Bridge (BFB) Cycle

BFB as a mechanism of chromosomal instability was first described by Barbara McClintock for formation of ring chromosomes in maize (McClintock, 1938, 1941). BFB cycles have been shown to generate genetic heterogeneity in a range of solid tumours (Gisselsson et al., 2000). The cycle describes a single chromatid break repaired via breakpoint fusion to form a dicentric chromosome, which leads to an anaphase bridge, which in turn can re-enter the BFB cycle (McClintock, 1938, 1941; Ly and Cleveland, 2017). Dicentric chromosomes form as a result of telomere fusion or fusion of two chromosome fragments both harbouring a centromere, with the formation of a chromatin bridge during mitosis that remains intact potentially throughout cytokinesis and into early interphase (Garsed et al., 2014; Ly and Cleveland, 2017). It is proposed that following nuclear envelope rupture, bridge breakage occurs via three prime repair exonuclease 1 (TREX1) (Maciejowski et al., 2015). The manner of breakage in the dicentric bridge remains debated, with the bridge, and therefore the BFB cycle as a whole, proposed as the site where chromothripsis occurs (Ly and Cleveland, 2017; Guérin and Marcand, 2022).

1.2.3.2 Translocation-(excision)-deletion-amplification

In this model, exogenous stimuli result in chromosomal translocations. The deleted or amplified DNA segments excised at translocation breakpoints are the result of DNA damage repair and can lead to ecDNA formation (Röijer et al., 2002; Van Roy et al., 2006). This offers a mechanism for complex translocations across different chromosomes to form on single ecDNA or HSR structures.

1.2.3.3 Episome model

The episome model suggests that ecDNA are formed during DNA synthesis as a result of DNA slippage and R-loop formation causing excised DNA segments, which then circularise, replicate and enlarge to form ecDNA (Carroll et al., 1988; Stark et al., 1989) EcDNA in seven gliomas were shown to be formed from a single excision event, supporting a post-replication event in forming these initial amplicons (Vogt et al., 2004). A study of 34 patients with haematological malignancies identified a defined, *MYC*-containing amplicon where the clustering of breakpoints appeared consistent with an excision/amplification episome model (Storlazzi et al., 2006). The chromosomal origin of *EGFRvIII*-harbouring ecDNA in the GBM39 cell line also points to the episome model, and, as might be expected given the mechanism, indicates these episomes are derived from a single parental allele (Hung et al., 2022). The episome model may explain simple continuous fragments of amplified chromosomes, but is more challenging as a model for multi-fragment or other complex ecDNA structures.

1.2.3.4 Chromothripsis

The term ‘chromothripsis’ is derived from the Greek ‘chromos’ (chromosome) and ‘thripsis’ (shattering) (Stephens et al., 2011). The chromothripsis model is based on the principle that the acquisition of multiple genomic arrangements occurs as a single catastrophic event rather than gradual stepwise accumulation of mutations. Chromothripsis, manifesting as complex rearrangements in single chromosomes, occurs across many cancer types and can involve multiple chromosomes. While initially described as occurring in 2-3% of cancer cells, a more recent analysis proposes a rate of 55% (Stephens et al., 2011; Cortés-

Ciriano et al., 2020). Rates vary by cancer type, with glioblastoma harbouring among the highest rates of chromothripsis of all cancers (Malhotra et al., 2013; Zack et al., 2013; Cortés-Ciriano et al., 2020). The absence of an ever-increasing number of breakpoints, the retention of heterozygosity and the clustering of breakpoints across a single chromosome all support the single catastrophic event hypothesis for genomic rearrangement acquisition (Stephens et al., 2011). More recently it has been proposed that further chromothriptic events can occur via micronuclei formation to drive evolution of ecDNA with ever-increasing copies of ecDNA-resident oncogenes (C.-Z. Zhang et al., 2015; Shoshani et al., 2021).

Two mechanisms for chromothripsis have been proposed: BFB chromosome fusion (see above) or micronuclei formation during cell division leading to premature chromosome condensation and shattering (Hatch et al., 2013; C.-Z. Zhang et al., 2015; Ly and Cleveland, 2017; Shoshani et al., 2021). Spectral karyotyping suggests that such rearrangements occur in a single parental copy of each chromosome pair (Stephens et al., 2011).

Chromothripsis can drive tumorigenesis if chromosome shattering leads to formation of key oncogene(s) or loss/disruption of a TSG, or even generate oncogenic fusions (Stephens et al., 2011). Fragment reassembly into extrachromosomal structures results in formation of oncogene-resident ecDNA, suggesting these cells survive such a catastrophic event owing to a selective advantage (Stephens et al., 2011; Rausch et al., 2012; Nones et al., 2014). It has been shown that chromothripsis and consequent oncogene-resident ecDNA formation represents the central source of seismic oncogene amplification in cancer (Rosswog et al., 2021; Shoshani et al., 2021).

The available evidence suggests that the majority of ecDNA derive their origins from chromothripsis; however the structure of ecDNA indicate cases where other, or indeed multiple, models can explain their origin. What remains unclear is what initiates such events in the first place, particularly a cellular disaster as major as chromothripsis. Given that glioblastoma has no clear environmental cause other than age-related risks (e.g. associations or risk factors) and the available evidence places chromothripsis as the source of most ecDNA in glioblastoma, it will be important to elucidate the causes of chromothripsis.

1.2.4 Cellular Models for studying ecDNA

EcDNA dynamics are increasingly studied in established cell lines. For example, the cell lines Colo320DM and Colo320HSR were originally derived from a patient with cancer of the sigmoid colon, and were among the first cells where oncogene amplification, *c-MYC*, was localised to DM or HSR sites, respectively (Quinn et al., 1979; Alitalo et al., 1983). The resident ecDNA have more recently been characterised as 4.328Mb in size carrying multiple copies of *c-MYC* (Hung et al., 2021). Histopathology described moderately undifferentiated adenocarcinoma, atypical pseudo-glands with an area of poorly differentiated carcinoid. Hormone and polypeptide levels were atypical for colorectal cancer, with raised parathyroid hormone (PTH), adrenocorticotropic hormone (ACTH), serotonin, norepinephrine and epinephrine, and low alpha-fetoprotein (AFP) and carcinoembryonic antigen (CEA) – characteristics of a neuroendocrine carcinoma. Growth of the cells differed significantly from other colorectal carcinoma cell lines, specifically regarding their shape, lack of CEA and rapid growth (Quinn et al., 1979). Overall, these paired cell lines may be highly atypical. Other examples of established cell lines with resident ecDNA that are widely studied include PC3 (prostate cancer – *c-MYC* ecDNA), SNU16 (gastric cancer – *c-MYC* and *fibroblast growth factor receptor 2* (*FGFR2)* ecDNA) and TR14 (Neuroblastoma - *c-MYC* ecDNA) (Kaighn et al., 1979; Cowell and Rupniak, 1983; Park et al., 1990).

Cell selection methodologies have been used to study the origins and evolution of ecDNA, building on strategies which were developed in the initial studies of ecDNA (Alt et al., 1978; Kaufman, Brown and Schimke, 1979). Low-dose, continuous, MTX drives the formation of MTX-resistant, *DHFR*-harbouring ecDNA and HSRs in HeLa (cervical) and H29T (colorectal) cancer cell lines (Shoshani et al., 2021). A similar strategy was deployed using a BRAF/MEK inhibitor combination resulting in *BRAF*-harbouring ecDNA in a previously ecDNA-null melanoma cell line (Song et al., 2021). Many such studies include development and subsequent study of single cell clones, and are reliant on continuous drug exposure. This may not be directly applicable to the heterogeneity observed in glioblastoma, where ecDNA exist de novo rather than in response to drug-based selection.

EcDNA were originally discovered in patient samples, and these likely represent the best cellular model for their study. EcDNA are markedly less frequent in established cell lines in comparison with primary human tumours, with the former also harbouring a higher proportion of HSRs (Benner, Wahl and Von Hoff, 1991). This observation was corroborated more recently in a study that identified ecDNA in ~40% tumour cell lines and ~90% patient-derived brain tumour models (Turner et al., 2017). Examples of patient-derived models include the GBM39 cell line, which is an *ecEGFRvIII*-expressing xenograft cell line comprising cells originally resected from a patient with primary glioblastoma which were transduced with lentiviral firefly luciferase prior to mouse injection (Sarkaria et al., 2006, 2007; Nathanson et al., 2014). The HK359 glioblastoma cell line, which also harbour *ecEGFRvIII,* was derived directly into neurosphere culture from a heavily pre-treated patient with recurrent glioblastoma (Laks et al., 2016). Other examples of primary cells used in recent studies include primary neuroblastoma cells (Helmsauer et al., 2020; Hung et al., 2021; Lange et al., 2022; Stöber et al., 2023) and primary (and one matched recurrent) glioblastoma cell culture (deCarvalho et al., 2018). Comparing the genomic profile from tumours, their derived cell cultures and subsequent xenografts indicates good molecular synergy between these entities (deCarvalho et al., 2018). Patient-derived glioblastoma cell cultures likely represent the best tool to study ecDNA dynamics in glioblastoma, and best recapitulate the complexities of glioblastoma and ecDNA biology.

1.2.5 Tools for the study of ecDNA

EcDNA can be directly visualised on metaphase spreads via 4′,6-diamidino-2-phenylindole (DAPI) staining, and DNA fluorescence in situ hybridisation (FISH) allows visualisation of key genomic loci including oncogenes and regulatory elements (van der Hout et al., 1989; Shapiro et al., 1993). These classic cytogenetic analysis tools remain the most robust way of characterising individual ecDNA within a cell, describing ecDNA/HSR dynamics and represent the only means of differentiating between ecDNA and HSRs. However, they are low throughput and it can be difficult to generate metaphases in individual cell lines. For example, in one study, metaphase spreads could only be obtained from 72/117 (61.5%) cancer cell lines (Turner et al., 2017). Interpretation relies on high quality epifluorescence or confocal-based imaging, particularly when combined with quantitative analysis.

DNA FISH relies on hybridisation onto fixed metaphase spreads or nuclei. Live-cell imaging strategies are in development, but single-copy locus detection is technically challenging. The Casilio system uses dead-Cas9 and gRNAs targeting ecDNA breakpoints to enable binding of fluorescent reporter molecules (Yi et al., 2021; Clow et al., 2022). Other systems, such as the ANCHOR system, whereby a ParB/green fluorescent protein (GFP) construct binds to a naturally occurring plant sequence and 'cages' further ParB/GFP fluorescent proteins, could be adapted for live cell imaging of ecDNA (Meschichi et al., 2021). Live-cell imaging tools require further validation to avoid fluorophore binding artefacts, as well as addressing uncertainties around binding affinity and effects on function.

The need to reconstruct highly complex ecDNA structures using genomic data has led to the development of novel sequencing-based analysis tools. AmpliconArchitect (AA) is a tool for ecDNA amplicon reconstruction with paired-end WGS data (Turner et al., 2017; Deshpande et al., 2019). This has since been developed further to the AmpliconSuite analysis pipeline (Luebeck *et al.*, 2022), which incorporates AmpliconClassifer (for output classification) (Kim et al., 2020) and CNVkit (for calling copy number variation (CNV) and alterations across the genome) (Talevich et al., 2016). AA uses WGS data to link copy number variant (CNV) regions of increased copy number, and identify discordantly mapped regions which are then defined as linked segments. However, short read WGS data inherently may not be long enough to span these long repeats, which limits their ability to differentiate between multiple possible structures. This in turn can result in multiple possible candidate amplicons from which the user can then choose as their 'amplicon of interest' (Deshpande et al., 2019). This caveat means that if AA 'predicts' a number of different possible amplicons, the user might select the amplicons of greatest interest, resulting in reporting bias. In addition, correlation of AA with DNA FISH of metaphase spreads suggested an 85% positive predictive value (PPV) of amplicons characterised as 'circular' by AA corresponding to extrachromosomal FISH signal. The sensitivity of AA to identify circular ecDNA was 83% (i.e. 83% of signals identified via DNA FISH as being extrachromosomal were also classified as ‘circular’ by AA). EcDNA may be more accurately characterised by long-read sequencing and optical mapping (OM), recently integrated as a computational tool (AmpliconReconstructor) (Wu et al., 2019; Luebeck et al., 2020; Hung et al., 2022). Future strategies may focus on single-cell or even single-read characterisation of ecDNA structures and breakpoints to truly capture the heterogeneity observed in recent studies of ecDNA (Hung et al., 2022; Stöber et al., 2023). However, computational tools alone are insufficient to accurately capture all ecDNA features and must be complemented by other approaches. A recent study combining FISH and genomic data at single cell resolution in a glioblastoma cohort (Walentynowicz et al., 2023) offers some insight into the potential of spatial transcriptomics in exploring spatial, as well as temporal, ecDNA dynamics.

1.3 Nuclear Organisation of ecDNA

1.3.1 Chromosome regions and territories

Nuclear organisation is not random. Eukaryotic chromosomes have been known to exist in chromosome territories for over 100 years (Rabl, 1885; Boveri, 1909; Cremer and Cremer, 2010; Dixon, Gorkin and Ren, 2016). It has been shown that central parts of the nucleus are preferentially occupied by active chromosome regions (Croft et al., 1999; Boyle et al., 2001). Other features, such as centromeric heterochromatin, also predict peripheral localisation (Carvalho et al., 2001). Indeed, chromosome 7 (chr7), the resident chromosome for the peri-centromeric gene *EGFR*, is among the most peripherally located chromosomes (Boyle et al., 2001). The organisation of the nucleus can be described via A and B compartments, with the inner A compartment ring hosting active chromatin and the most highly expressed genes (Lieberman-Aiden et al., 2009; Stevens et al., 2017).

Chromosome regions can be reorganised via loop extrusion. This process is mediated by cohesin, a ring-shaped protein complex loaded onto chromatin through which chromatin is extruded, and CCCTC-Binding factor (CTCF), a boundary element that restricts further extrusion (Sanborn et al., 2015; Fudenberg et al., 2016). This leads to the formation of topologically associating domains (TADs), which represent genomic regions containing genes and cis-regulatory elements (CREs - e.g. promoter, enhancer(s)) contained within CTCF/cohesin boundaries (Dixon et al., 2012; Rao et al., 2014). The manner by which different chromatin regions interact has been studied through the development of genome-wide chromosome conformation capture ('C' e.g. Hi-C (Lieberman-Aiden et al., 2009) or Micro-C (Hsieh et al., 2015)) techniques (de Wit and de Laat, 2012; Dixon et al., 2012; Dixon, Gorkin and Ren, 2016)), which can be combined with high-resolution imaging techniques such as DNA FISH to characterise TAD interactions and the regulatory environment of genes (e.g.(Williamson et al., 2019; Boyle et al., 2020)). Chromosome territories and the 3D nuclear environment they inhabit play a central role in how genes are regulated.

1.3.2 Enhancers and regulatory elements

Gene expression is regulated by CREs, which comprise a proximal promoter element (≤1kb from the transcription start site, TSS) and more distal regulatory elements (≥1kb from the TSS) comprising enhancers and other elements such as insulators or silencers (Vernimmen and Bickmore, 2015).

The initiation, elongation and termination of transcription requires the coordinated assembly and interaction of TFs, co-activators and key transcription proteins. Sequence-specific TFs bind to enhancers and recruit co-activators, such as p300, to modify chromatin structure to increase the accessibility of chromatin, facilitating the further recruitment of co-activators to aid formation of the Preinitiation Complex (PIC) (Esnault et al., 2008; Soutourina et al., 2011; Ortega et al., 2018). The PIC (which includes RNA polymerase II (RNA Pol II)) and the RNA Pol II-regulating protein Mediator form a 'bridge' linking the enhancer and promoter (Soutourina et al., 2011; Petrenko et al., 2016). Once these initiation steps are complete, and the first few RNA nucleotides have been synthesised, RNA Pol II is released from the promoter via Mediator-regulated phosphorylation (Soutourina 2018; Kim et al. 1994). RNA synthesis (elongation) proceeds until the transcript is complete, at which point RNA Pol II and the RNA transcript are released from the template (termination) (Kuehner, Pearson and Moore, 2011).

How enhancers activate gene expression in 3D space is an ongoing area of research. Initial models of looping fail to explain directionality and specificity. Cohesin-mediated loop extrusion compacts TADs, bringing CREs close to their target genes, providing a plausible mechanism (Kane et al., 2022). In addition, so-called ‘super-enhancers’ (SEs), dense regions of enhancers characterised by high TF, co-activator and Mediator occupancy, have been shown to play a key role in stem cell and cancer gene expression (Lovén et al., 2013; Whyte et al., 2013). Overall, these studies highlight the dynamic role of the non-coding genome in bringing together distant genomic regulatory regions to control transcription expression.

1.3.3 Transcriptional hubs

The discovery of SEs led to the hypothesis that co-activators bind in high density at these CREs (Lovén et al., 2013). A parallel hypothesis suggests that liquid-liquid phase separation (LLPS) allows membrane-less sub-compartmentalisation of the nucleus in order for high concentrations of the transcriptional machinery to drive gene transcription (Rai et al., 2018; Strom and Brangwynne, 2019) . It was subsequently shown that many TFs, co-activators (e.g. Bromodomain-containing protein 4 (BRD4), Mediator) and RNA Pol II densely co-bind and form condensates or ‘hubs’ (Cho et al., 2018; Chong et al., 2018; Sabari et al., 2018). Subsequent studies suggest that LLPS may not be required to define these highly concentrated regulatory hubs (Mir et al., 2018, 2019; McSwiggen et al., 2019). Indeed, the necessity of such regulatory hubs in driving gene expression remains a topic for discussion.

1.3.4 EcDNA nuclear localisation

EcDNA are chromatinised regions of DNA that lack centromeres. This raises interesting questions about their nuclear location. Early studies in Colo320DM cells reported that ecDNA were preferentially localised at the nuclear periphery, a region associated with heterochromatin and transcriptional repression. Specifically, e*cMYC* was reported to be localised at the nuclear periphery during G1 and moved centrally during S phase, during which ecDNA could also be expelled via micronucleus formation (Itoh and Shimizu, 1998; Shimizu et al., 1998). In contrast, a study of neuroblastoma cell lines with *MYC* ecDNA observed even nuclear distribution (rather than a peripheral preference) in >85% of interphase nuclei, moving peripherally during the interphase to prophase transition point (Lundberg et al., 2008). In a pan-cancer study, ecDNA have been found to have regions of largely accessible chromatin (assayed by Assay for Transposase-Accessible Chromatin using sequencing (ATAC-seq)), indicative of nucleosome displacement by bound TFs, and to be decorated with histone modifications associated with active chromatin (Wu et al., 2019). Widespread chromatin connectivity suggestive of an active transcriptional function of ecDNA was also observed via Hi-C and chromatin interaction analysis by paired-end tag sequencing (ChIA-PET) with RNA Pol II chromatin immunoprecipitation (ChIP) in ecDNA-harbouring glioblastoma patient-derived cell lines (Zhu et al., 2021). These are features more associated with active, central nuclear regions. These contradictory findings suggest that further study into ecDNA nuclear localisation and its relationship with gene regulation is warranted.

1.3.5 Transcription in the context of ecDNA

EcDNA harbour oncogenes and regulatory elements, mirroring the central components of a TAD, and indeed chromatin conformation analysis and CTCF and cohesin binding suggests that ecDNA chromatin is organised like a TAD (Wu et al., 2019). In glioblastoma, *EGFR*-ecDNA have been shown to be co-amplified with two functional enhancer elements, suggesting a novel regulatory environment on ecDNA that facilitates an SE environment (Figure 1.8) (Morton et al., 2019). EcDNA have also been proposed to act as mobile enhancers for both chromosomal and extrachromosomal genes (Helmsauer et al., 2020; Zhu et al., 2021). Small subclones of ecDNA may indeed exist that only harbour enhancers (or oncogenes), as a further source of dynamic enhancer interactions (Hung et al., 2022).

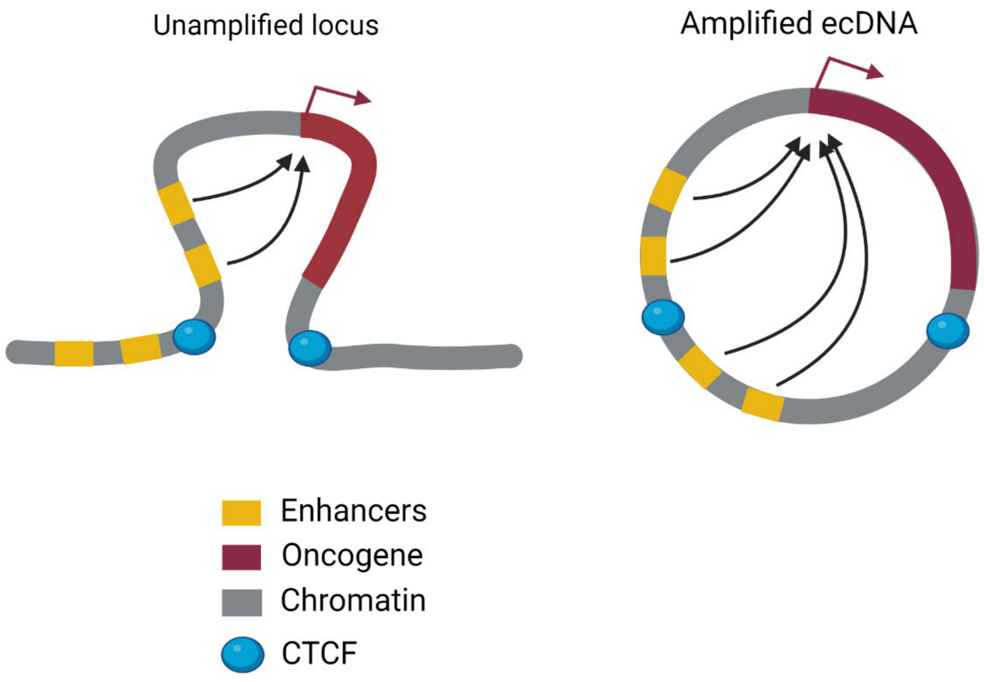


Figure 1.8 | Enhancer rewiring on ecDNA

Adapted from (Morton et al., 2019, Graphical Abstract) showing the regulatory environment on an unamplified locus (left), where loop extrusion within an insulated region (CTCF sites) enables enhancer-promoter contact to drive oncogene expression. Enhancers outside the insulated region do not drive gene transcription. In ecDNA (right), it is proposed that novel enhancers are able to drive oncogene expression via a process of enhancer rewiring in spite of insulated regions. The further addition of active enhancers might occur via chromothripsis (Morton et al., 2019).

Given the co-localisation of enhancers and driver oncogenes on ecDNA, it has been suggested that ecDNA cluster together in the nucleus, driving the recruitment of a high concentration of RNA Pol II in a condensate, thereby creating ecDNA-driven nuclear hubs (Hung et al., 2021; Yi et al., 2021; Zhu et al., 2021). A series of established tumour cell lines and patient neuroblastoma cells were evaluated using an autocorrelation approach of hybridisation signals on fixed cells labelled by DNA FISH (Figure 1.9). This analysis suggested that ecDNA were more clustered than expected. RNA/DNA FISH combined with an autocorrelation tool suggested that clustering was a greater predictor of transcription than copy number (Hung et al., 2021). EcDNA clustering was further suggested from live-cell imaging in glioblastoma neurospheres (Figure 1.10A) (Yi et al., 2021). It was proposed that ecDNA colocalise with RNA Pol II (Figure 1.10B), in addition to other markers of transcription (Cajal bodies and promyelocytic leukemia (PML) nuclear bodies).

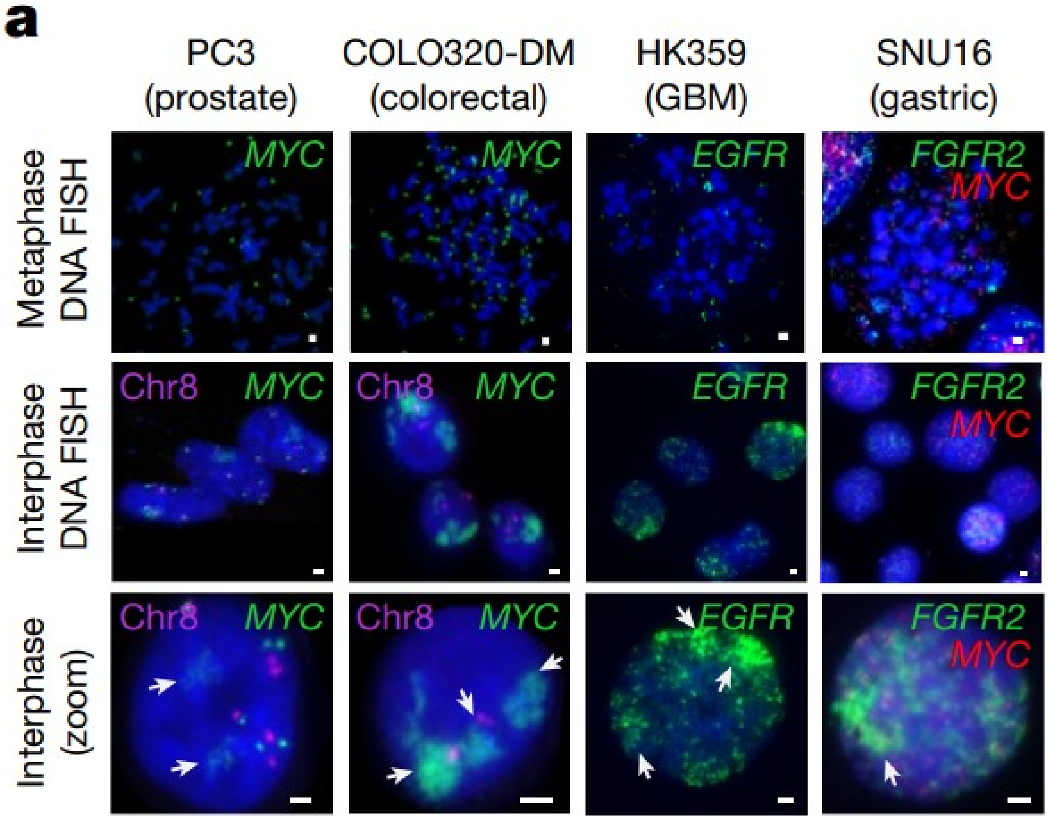


Figure 1.9 | Clustering images from (Hung *et al.*, 2021)

Images from (Hung et al., 2021), Figure 1A, and are described as representative DNA FISH images of the indicated cell lines and their ecDNA-resident oncogenes (colour of visualised gene/chromosome as indicated). Arrows = regions of clustering. Scale bar = 2µm.

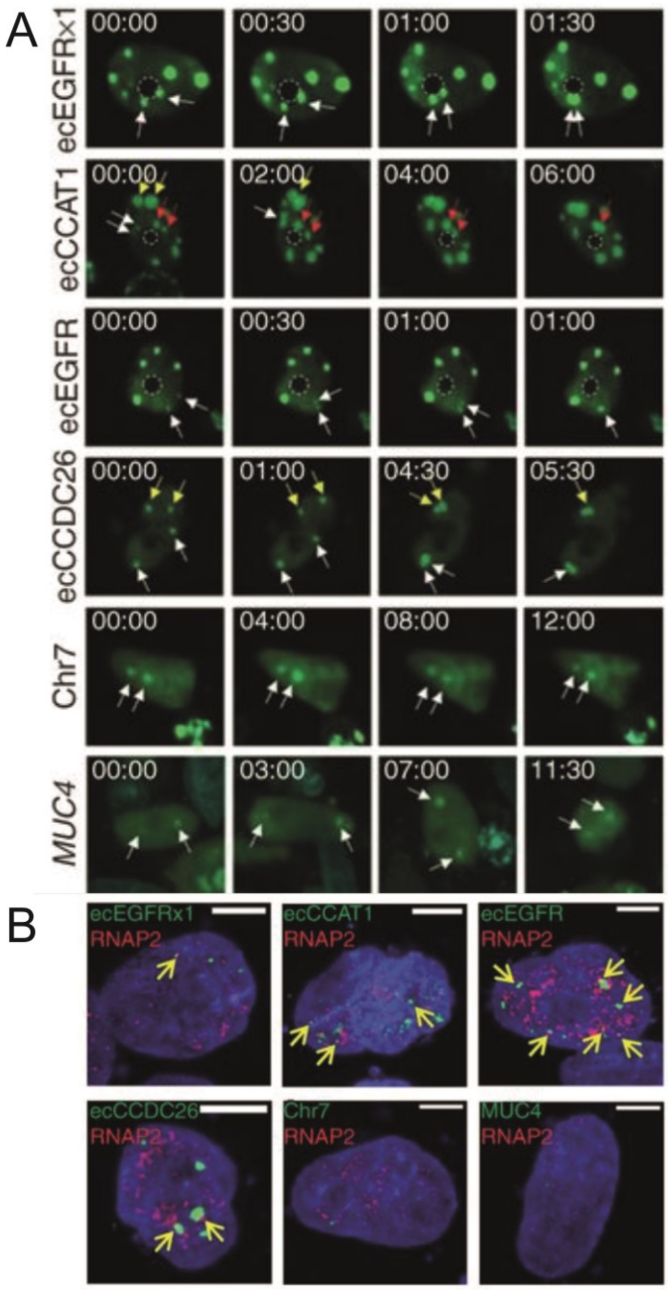


Figure 1.10 | Clustering images from (Yi et al., 2021)

Images from (Yi et al., 2021). A) Figure 3E. HF3016 neurosphere cells (primary glioblastoma). Live cell imaging using ecTag - sgRNAs targeting ecDNA breakpoint loci (ecEGFRx1, ecCCAT1, ecEGFR, and ecCCDC26), an intronic chr7 region (Chr7) and MUC4 as linear DNA amplicon controls. SgRNAs were co-transfected with other components of the Casilio system (Clover, dCas9) so sgRNA targets are visualised in green. Dashed circle = nucleolus. Arrows indicate ecDNA hub formation. Captured time-lapse images where 00:00 = hour:minute. B) Figure 4A. Representative images of RNA Pol II (RNAP2) immunofluorescence (red) on ecTag (green)-transfected cells, DNA = DAPI (blue). Scale bar = 10μm. Colocalisation of sgRNAs with RNAP2 by visual calling of partially or completely overlapping of fluorescence signals (arrows).

Further, it has been suggested that ecDNA-resident oncogene expression is amplified beyond that expected from copy number alone. RNA-seq data comparing gene copy number between circular and non-circular amplifications from panels of cell lines showed significantly higher RNA copy number from circular amplicons (Wu et al., 2019; Kim et al., 2020). This was also shown by identifying ecDNA-resident single nucleotide polymorphisms (SNPs) (Wu et al., 2019). These analyses confirm ecDNA-resident oncogenes result in massively amplified gene expression, but cannot determine whether this is copy number driven. Copy number-normalised gene expression has been shown to be higher in ecDNA-resident genes, although only in 3/11 genes evaluated, of which only one was a key oncogene (*CDK4*) (Wu et al., 2019). A recent study of neuroblastoma cell lines and primary cells showed a linear relationship between ecDNA copy number and gene expression, using single cell analysis to further suggest significant intra-tumoural heterogeneity in ecDNA transcription (Stöber et al., 2023). Taken together, authors of these studies concluded that ecDNA copy number has a significant role in driving the amplified gene expression from ecDNA-resident oncogenes. It remains uncertain as to whether there are ecDNA specific mechanisms of gene expression that are not seen in normal chromosomal context that explains this increase, or whether it is solely due to copy number. This question, alongside the potential physical clustering of ecDNA, is explored in this thesis.

1.4 EcDNA and DNA repair pathways

1.4.1 The DNA Damage Response (DDR)

The ability to respond to DNA damage is essential for cell survival. The DNA in every human cell will undergo potentially thousands of damaging events every day (Lindahl and Barnes, 2000; Jackson and Bartek, 2009). This damage can be secondary to physiological processes such as DNA replication or epigenetic modification, inflammation giving rise to reactive oxygen species (ROS) or the impact of cellular reactions causing DNA adduct formation. In addition, the human body deals with a range of environmental DNA-damaging agents, such as ultraviolet light, tobacco and aflatoxins (Jackson and Bartek, 2009; Ciccia and Elledge, 2010). Exogenous DNA damage is now a major tool in treating disease, including glomerulonephritides (Ponticelli, Escoli and Moroni, 2018), psoriasis

(Bishop and Abel, 1985) and, most commonly, cancer (Weber, 2015). The DNA Damage Response (DDR) encompasses a cell's ability to identify a DNA lesion, initiate a signalling response and repair the lesion (Rouse and Jackson, 2002; Jackson and Bartek, 2009).

DNA damage can take the form of DNA modifications, such as base methylation, oxidation or adducts, interstrand crosslinks, base mismatch or breaks to the DNA itself (Ciccia and Elledge, 2010). A summary of the major DDR mechanisms for different types of lesions is outlined in Table 1.2. Other forms of DDR mechanisms include trans-lesion bypass, the Fanconi Anaemia (FANC) pathway and Ataxia-telangiectasia mutated (ATM) and ATM and Rad3 Related (ATR)-mediated DDR.

DDR Mechanism	Type of lesion	Key proteins
Direct lesion reversal	Base alkylation e.g. O6-alkylguanine	O6-methylguanine-methyltransferase
Mismatch repair (MMR)	DNA mismatch from DNA replication	MSH and MLH proteins e.g. MSH2-MSH6, MLH1-PMS2
Base excision repair (BER)	Incorrect DNA base	DNA glycosylases, APE1 endonuclease, DNA polymerases (β, δ, ε)
Single strand break (SSB) repair	SSB created de novo, or by BER or TOP I activity	As BER plus XRCC1, PARP1, PARP2,
Nucleotide excision repair (NER)	Nucleotide adducts e.g. UV photo products, base adducts	RNA Polymerase, Xeroderma Pigmentosa (XP) C, A, E, F, G, DNA polymerases
Non-homologous recombination (NHEJ)	DSBs	Ku, DNA-dependent Protein Kinase (DNA PKs), XRCC4, XLF/Cernunnos, Ligase IV
Homologous recombination (HR)	DSBs, stalled replication forks, inter-strand DNA crosslinks, meiotic recombination, TOP II activity	RAD51 (and related proteins incl XRCC2/3), BRCA1, BRCA2, RPA, DNA polymerase, ATM CHK1/2

Table 1.2 | Major DDR Mechanisms

Adapted from (Jackson and Bartek, 2009, Table 1), summary of the key mechanisms by which DNA damage is repaired, which is dependent on the type of DNA damage lesion.

The repair of single stranded (SSB) or double stranded (DSB) DNA breaks is intrinsically linked. SSBs are the commonest lesions, can be initiated directly or indirectly via base excision repair (BER) and are most often caused by ROS (Bradley and Kohn, 1979; Caldecott, 2008). Indirect BER leads to SSBs by the excision of oxidised bases, a process which can lead to DSBs (Cannan et al., 2014). Stalled replication forks, which can form as a result of SSBs, can also trigger DSBs (Saleh-Gohari et al., 2005). Other causes of SSBs include abortive topoisomerase (TOP) I activity, a process whereby a DNA 'nick' is created to relax supercoiling during transcription and replication (Wang, 2002). Exogenous causes of SSBs include chemotherapy, such as TOP1 and TOP2 inhibitors which cause SSB and DSBs respectively, and ionising radiation (IR), which generates roughly ten times more SSBs than DSBs (Koster et al., 2007; Ma et al., 2012; Mehta and Haber, 2014).

SSBs are bound by poly(ADP-ribose) polymerase 1 (PARP1) thereby initiating recruitment of end-processing repair proteins, although detection via PARP1 recruitment is not required for all causes of SSBs, e.g. TOP1-induced (Durkacz et al., 1980; D'Amours et al., 1999; Caldecott, 2008). Failure to repair SSBs most commonly results in collapsed DNA replication forks (and subsequent DSBs), stalled transcription and ultimately cell death (Kuzminov, 2001; Kathe, Shen and Wallace, 2004; Saleh-Gohari et al., 2005; Caldecott, 2008).

DSBs are much less common than SSBs, but extremely toxic to cells. In addition to SSBs as a source of DSBs, other endogenous causes include ROS, abortive TOP II activity and meiotic crossovers (Spell and Holm, 1994; Lam and Keeney, 2014; Mehta and Haber, 2014). DSBs can be repaired by two major repair pathways: homologous recombination (HR) or non-homologous end joining (NHEJ). HR can faithfully repair the genome by using the homologous sister chromatid as a template and occurs during the late S to G2 phase of the cell cycle (Rothkamm et al., 2003; Moynahan and Jasin, 2010). The DSB is detected and bound by the MRE11–RAD50–NBS1 (MRE) complex which recruits replication protein A (RPA) to generate single stranded (ss) DNA (Buis et al., 2008). This allows RAD51 and key accessory factors (e.g. BRCA2) to bind and initiate strand synthesis, either via the synthesis-dependent strand annealing (SDSA) pathway, or the canonical DSB repair (DSBR) pathway (Haaf et al., 1995; Sung and Klein, 2006; Moynahan and Jasin, 2010). In contrast, NHEJ occurs at all phases of the cell cycle (Rothkamm et al., 2003). DSBs are detected and bound by the Ku80/Ku70 heterodimer, which in turn recruits a DNA Protein Kinase (DNA-PK) complex (Yaneva, Kowalewski and Lieber, 1997; Downs and Jackson, 2004). This leads to the recruitment of DNA polymerases, and subsequent binding of X-Ray Repair Cross Complementing (XRCC) 4, XRCC4-like factor (XLF) and ligase IV leads to direct religation of the DSB ends (Ahnesorg, Smith and Jackson, 2006; Lieber, 2010b; Lord and Ashworth, 2012). An alternative NHEJ (alt-NHEJ) pathway has been proposed (also known as microhomology-mediated end joining (MMEJ)), whereby PARP1 recognises DSBs and initiates a repair process that concludes by direct end ligation via DNA Ligase I or III (Lu et al., 2016; Yang et al., 2018). Alt-NHEJ is thought to have roles as a ‘backup’ pathway for NHEJ and HR, but other roles have been proposed such as in IR-induced DSB repair (Lieber, 2010a; Della-Maria et al., 2011; Dutta et al., 2017). Owing to the lack of a repair template, NHEJ is a far more error prone pathway than HR and can lead to the introduction of mutational insertions or deletions.

1.4.2 DNA Damage in Cancer

Genomic instability is considered a key characteristic of cancer (Hanahan and Weinberg, 2011). Induction of the DDR by oncogenes has been proposed as an early event in cancer initiation, whereby DSBs form and multiple checkpoint proteins, such as ATM, ATR and p53, fail to activate (Bartkova et al., 2006; Halazonetis, Gorgoulis and Bartek, 2008). In most cancers, the specific DDR dysfunction is unknown (Lord and Ashworth, 2012), however, in some cancers there is a specific link between a particular DDR mechanism and a cancer phenotype. Examples include Hereditary Non-Polyposis Colorectal Cancer (HNPCC), a form of familial colorectal cancer caused by loss-of function mutations in mismatch repair (MMR) genes, and breast, ovarian or pancreatic cancers caused by mutations in *BRCA1*, *BRCA2* or other genes involved with HR repair (Lord and Ashworth, 2012). These genetic aberrations are also seen in sporadic cancers e.g. microsatellite instability in colorectal cancer (Boland and Goel, 2010), and defects in HR in ovarian cancer (Cancer Genome Atlas Research Network, 2011).

Upregulation of the DDR has been shown in a number of CSC models (Eyler and Rich, 2008; Ogawa et al., 2013; Desai, Webb and Gerson, 2014). The DDR has also been shown to be constitutively activated in glioblastoma, and has been attributed as a major cause of resistance to cancer treatment (Bartkova et al., 2010). In GSCs, the upregulated DDR leads to chemo-radioresistance and subsequent tumour recurrence (Bao et al., 2006; Chen et al., 2012). It has been proposed that this is driven by replication stress, the process whereby replication forks stall and result in DSBs (Carruthers et al., 2018). This suggests that further interrogation of the DDR is warranted, with a view to identifying potential therapeutic vulnerabilities.

1.4.3 The role of PARP

PARPs are a group of enzymes that catalyse the post-translational modification of proteins, of which PARP1 is the most extensively studied and most active (D'Amours et al., 1999). The *PARP1* gene comprises a double zinc-finger DNA binding domain, a central automodification domain and an nicotinamide-adenine-dinucleotide (NAD+) binding catalytic domain (Kraus and Lis, 2003). PARP1 catalyses the addition of ADPribose (ADPr) from NAD+ (Nicotinamide + ADPr) to target proteins, thereby generating a branched polyADPribosylated protein in a process known as PARylation (Figure 1.11) (Hayashi et al., 1983; D'Amours et al., 1999). These polymers are broken down by the enzyme polyADPr glycohydrolase (PARG), which hydrolyses the glycosidic bonds between ADPr polymers, and (ADPr)hydrolase 3 (ARH3), which cleaves the final serine-ADPr bond (Miwa et al., 1974; D'Amours et al., 1999; Kim et al., 2004; Slade et al., 2011; Palazzo, Suskiewicz and Ahel, 2021; Schützenhofer, Rack and Ahel, 2021). MonoADPr bonds can also be hydrolysed by PARG, in addition to other enzymes such as terminal ADP-ribose protein glycohydrolase (TARG1) or macrodomain hydrolases (MacroD1, MacroD2) which are responsible for removing acidic residues (glutamate and aspartamate) (Rack, Perina and Ahel, 2016; Munnur and Ahel, 2017).

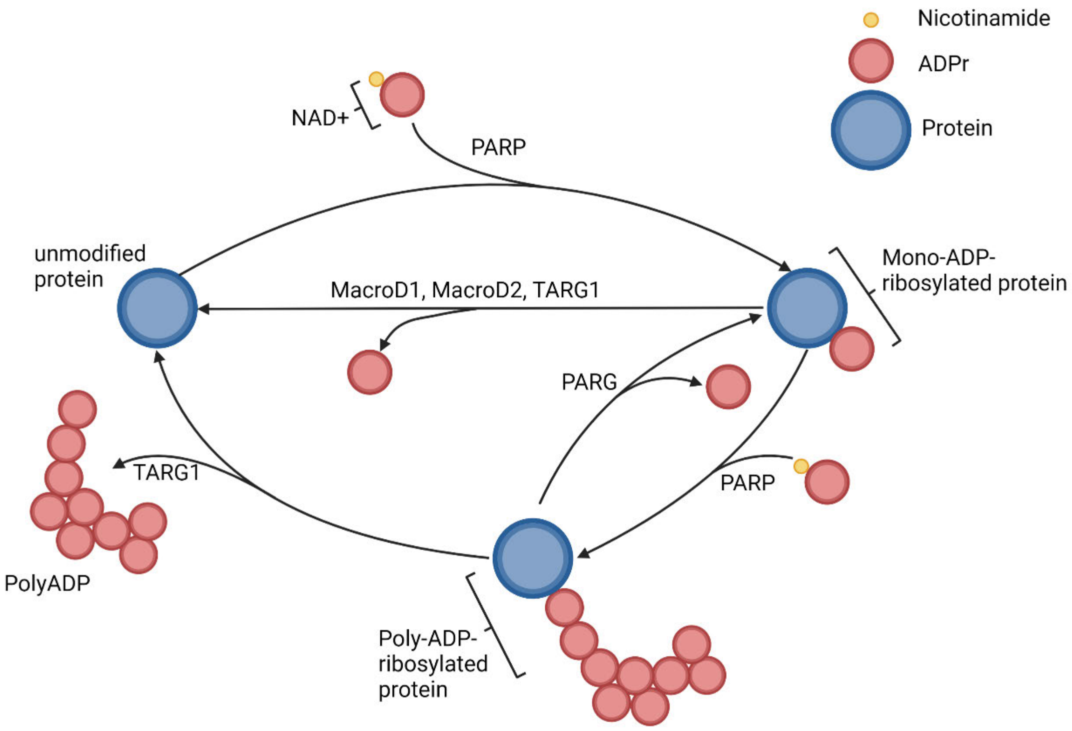


Figure 1.11 | PARylation

PARP catalyses the addition of an ADPr to a target protein, with further polymerisation leading to the generation of a poly-ADPr-protein. These polymers are broken down by PARG, which breaks down the glycosidic bonds between ADPr monomers. MonoADPr bonds are hydrolysed by PARG and the acidic residue-cleaving enzymes TARG1 and MacroD1 and D2. The final serine-ADPr bonds, which represent the initial PARylation event following DNA damage, are cleaved by ARH3 (not shown). Created with Biorender, modified from Ahel lab.

PARylation without DNA damage generally results in mono- or oligo-ADPribosylation and has many important cellular functions (Ferro and Oppenheimer, 1978; Benjamin and Gill, 1980). For example, PARylation of histones, primarily histone H1, leads to chromatin decondensation (Poirier et al., 1982). PARylation also regulates transcription by contributing to enhancer/promoter complex formation (D'Amours et al., 1999; Kraus and Lis, 2003; Kim et al., 2004) and appears to be required for altered physical enhancer-promoter proximity (Benabdallah et al., 2019). PARylation also has a role in DNA replication (Bryant et al., 2009; Ray Chaudhuri et al., 2012), mitosis (Halappanavar and Shah, 2004) and apoptosis (Yu et al., 2002).

In the case of DNA damage, the target protein for PARylation is PARP1 itself, with PARP1 recruited within seconds to SSBs (Benjamin and Gill, 1980; Ludwig et al., 1988; Satoh and Lindahl, 1992). Compared with basal scenarios, PARylation driven by DNA strand breaks results in much longer branched chains, with 500-1000 times greater PARP activity (Simonin et al., 1993; D'Amours et al., 1999; Langelier et al., 2018). Serine residues are the main target for ADPr (Ser-ADPr) following DNA damage (Palazzo et al., 2018). PARP is involved in the repair of primarily SSBs, but also some DSBs (see above), and is a key coordinator of the DDR (Ljungman and Lane, 2004). PARP1 binds to DNA breaks via its DNA binding domain and is involved in both the assembly of DNA repair complexes and PAR binding domains in proteins (Ljungman and Lane, 2004). PARP binding leads to the generation of a stalled replication fork (Bryant et al., 2009; Ying et al., 2016). The accessory protein, Histone PARylation Factor 1 (HPF1), is needed for PARP1 activation, and a combined HPF1/PARP complex represents an active DNA damage repair complex (Suskiewicz et al., 2020), enabling binding of key proteins such as XRCC1, DNA polymerase beta and Ligase III (Pandey and Black, 2021). Failure to repair the damage leads to DSB formation and ultimately cell death.

1.4.4 The DDR and ecDNA

DNA damage and repair are thought to be central to ecDNA formation (see Section 1.2.3), but also to ecDNA-directed drug resistance and are a driver of chromosomal/extra-chromosomal rearrangements. EcDNA formation driven by MTX resistance is dependent on NHEJ, HR and PARP (Meng et al., 2015; Cai et al., 2019; Shoshani et al., 2021). Fewer ecDNA and MTX-resistant colonies were formed in the evolution of MTX-resistant cells when NHEJ was inhibited by a DNA-PK inhibitor or the PARP inhibitor (PARPi) veliparib (Shoshani et al., 2021).

EcDNA also evolve in the face of DNA strand breaks. Epidermoid cells with ecDNA harbouring the drug resistance gene *Multidrug Resistance 1 (MDR1)* and Colo320 cells harbouring *ecMYC* were both shown to lose ecDNA copy number following IR, even at relatively low doses, with *ecMDR1* relocated to micronuclei (Sanchez, Barrett and Schoenlein, 1998; Schoenlein et al., 2003). Increased chromosomal reintegration of *MDR1* was not observed following IR (Schoenlein et al., 2003). A more recent study in established cell lines harbouring ecDNA (Colo320) or driven to form ecDNA by MTX resistance showed, in contrast, that both random DNA damage caused by IR or doxorubicin, and targeted nuclease-induced DNA damage near the amplified ecDNA gene *DHFR*, drove ecDNA to form ectopic chromosome integrations. This was increased three-fold in the presence of veliparib, but not a DNA-PK inhibitor (Shoshani et al., 2021), suggesting PARylation is a key pathway in subsequent ecDNA rearrangements, although not for the maintenance of existing ecDNA. What is unclear is whether these studies are identifying new cell populations generated as a result of the DDR, or whether this represents cell selection. No studies have explored the effect of ecDNA strand breaks in primary patient-derived cell cultures where ecDNA are already established.

1.4.5 PARP inhibitors

The potential of PARP as a therapeutic target, particularly to enhance the action of known DNA damaging agents, has long been identified (Satoh and Lindahl, 1992) and PARPi have transformed the management of HR-deficient cancers, notably those harbouring BRCA mutations.

Inhibition of PARP1 makes cells reliant on NHEJ to repair DNA breaks (Schultz et al., 2003), thus tumours deficient in HR proteins caused by *BRCA1/2* mutations are particularly vulnerable to PARPi via synthetic lethality (Bryant et al., 2005; Farmer et al., 2005; Fong et al., 2009). PARP inhibitors block PARylation and some covalently trap PARP on DNA at a SSB, preventing its dissociation from a replication fork (Farmer et al., 2005). The resulting DSB cannot be repaired by HR due to *BRCA1/2* mutation, and the cell dies. In addition, the ‘trapping’ of PARP itself blocks its catalytic activity (Pommier, O’Connor and de Bono, 2016; Lord and Ashworth, 2017). PARP inhibitors are considered trapping (in order of potency: talozaparib, niraparib, rucaparib, olaparib) or non-trapping (veliparib) (Lord and Ashworth, 2017).

1.4.6 PARP inhibitors in glioblastoma

All existing treatment options for glioblastoma induce DNA damage such as base methylation (TMZ), reactive oxygen species (TMZ and IR) and DNA strand breaks (IR) (Erasimus et al., 2016), with essentially all patients likely to receive IR as part of their treatment journey. While *BRCA1/2* mutations are rare in glioblastoma (Cancer Genome Atlas Research Network, 2008), IR generates DNA strand breaks (primarily SSBs), leading to the hypothesis that PARPi may potentiate IR via impaired DNA repair. *In vitro* studies suggest that the radiosensitivity of glioma cell lines is increased with drug-induced PARP inhibition (Dungey, Löser and Chalmers, 2008). Consequently, PARPi are being trialled in both recurrent and newly diagnosed glioblastoma in combination with IR. For example, the PARADIGM-2 study is a phase 1 trial evaluating the PARPi olaparib in combination with IR +- TMZ in patients with newly diagnosed glioblastoma trial (Fulton et al., 2018). The OPARATIC phase 1 study is assessing this in the recurrent glioblastoma setting, also using olaparib (Hanna et al., 2020).

These studies suggest that, while glioblastoma stem cells have a highly activated DDR, the impact of this on ecDNA dynamics remains unclear. Uncertainties remain concerning the response of ecDNA in glioblastoma stem cells to DSBs and impaired PARylation. Further study exploring the impact of the DDR on ecDNA dynamics is therefore highly warranted, both to elucidate the mechanism of existing therapeutic strategies, and to potentially identify novel tools to target ecDNA.

1.5 Aims of PhD Thesis

The overarching aim of my PhD was to characterise ecDNA in patient-derived glioblastoma stem cells with a particular focus on their transcriptional regulation and dynamic response to DNA damage. I addressed two main hypotheses. First, I hypothesised that the 3D organisation between individual ecDNA and between ecDNA and transcriptional hubs plays a central role in ecDNA transcriptional regulation. Second, I hypothesised that ecDNA characteristics evolve in response to DNA damage caused by DNA strand breaks, and that the mechanism of DNA damage and the relationship with PARylation both impact the evolution of ecDNA in glioblastoma cells.

I had three primary aims:

1. To determine the quality, quantity and stability of ecDNA in patient-derived glioblastoma stem cell lines for their use as an experimental model.

2. To characterise the spatial organisation and transcriptional regulation of ecDNA in glioblastoma stem cells to determine if clustering of ecDNA underpins their activity.

3. To determine the effect of DNA damage and PARylation on ecDNA dynamics in glioblastoma stem cells in search for targetable vulnerabilities.

To achieve Aims 1 and 2 I used super-resolution imaging and devised quantitative image analysis tools in combination with classic cytogenetic methods such as FISH and a range of bioinformatics analysis strategies. For Aim 3, I used both clinical and molecular biology strategies to evaluate the effect of different DNA damage modalities on ecDNA dynamics.

Chapter 2 Methods and Materials

Additional source, reference and identifier data for Methods and Materials is provided in the Appendix, with specific sections referenced in the text below.

2.1 Cell culture

2.1.1 Cell lines

Patient-derived glioblastoma stem cells (GSCs) and foetal neural stem cell (NSC) lines were selected from the Glioma Cellular Genetics Resource (GCGR) (http:/gcgr.org.uk) based on copy number variation (CNV) data (See Figure 1.4). The GSC lines selected are shown with their corresponding oncogene amplifications in Table 2.1. When the study was commenced, oncogene amplification was only predicted from CNV profile of WGS data, and it was not known whether these oncogenes were resident on ecDNA or how these related to glioblastoma subtypes. The NSC cell line GCGR-NS9FB_B (aka NS9) was derived from the forebrain of a 9-week-of-gestation foetus. All GSC lines were derived from patients with treatment-naive glioblastoma, with the exception of E37, which is derived from recurrent glioblastoma. This was originally thought to be from a patient with treatment-naïve glioblastoma, but was subsequently reclassified. There is no cell line available from this patient from their original primary tumour. Cell lines generated for the GCGR were obtained via informed consent and ethical approval (East of Scotland Research Ethics service, REC ref 15/ES/0094; South East Scotland Research Ethics Committee, REC ref 08/S1101/1).

	Oncogene amplifications				
Cell line	EGFR	MDM2	CDK4	PDGFRA	MYC
E20			X	X	
E25		X	X	X	
E26	X				
E28	X				
E37	X				X
NS9 (NSC)					

Table 2.1 | Cell lines with predicted oncogene amplifications by CNV profile

CNV data derived from WGS, provided by Dr Gillian Morrison.

Colo320DM cells were kindly provided by Professor Paul Mischel (Stanford University, California, USA).

2.1.2 Cell culture conditions

Cells from the GCGR were cultured in pre-warmed serum-free basal Dulbecco's Modified Eagle's Medium/Nutrient Mixture F-12 Ham (DMEM/F12) media (Sigma) supplemented with N2 and B27 (Life Technologies), 2 µg/mL Laminin-1 (Cultrex),1% (v/v) Penicillin-Streptomycin (Pen-Strep) (Gibco) and 10 ng/mL growth factors EGF and FGF-2 (Peprotech) (Pollard *et al.*, 2009), hereafter referred to as Complete Media (CM). In addition, E20, E25 and E26 cells were cultured on plates pre-laminated (Laminin-1 10 µg/mL in phosphate buffered saline (PBS)) for at least 1 hour (h) at 37°C. Where cells from any cell line were grown on glass coverslips or slides, the same prelamination protocol was used.

Cells were maintained as an adherent monolayer at 37 °C with 5% CO_2 with CM changed at least weekly. For passaging, cells were split with Accutase solution (Sigma) approximately weekly, or whenever 70-90% confluence was observed; cell growth rates varied by cell line, so passaging frequency varied between ~7-10 days. CM was removed, and cells were incubated in 1ml (T25 flask) or 2ml (T75 flask) Accutase until cells dissociated. Cells were lifted into a 50ml falcon with Wash Media (WM) (DMEM/F12 supplemented with 1% (v/v) Pen Strep and Bovine Serum Albumin (BSA)) and centrifuged at 300*g* for 4 minutes (mins) in a bench-top centrifuge. WM was aspirated and cells were resuspended in CM and replated at between 1:3 and 1:5 dilutions (end volume of CM - T25 - 8ml, T75 - 16ml).

For long-term storage, approximately 0.5-1 x 10^6 cells were pelleted as described above and resuspended in 1ml of freezing media (10% dimethyl sulfoxide (DMSO), 90% WM), transferred to -80°C in a polystyrene box and then to liquid nitrogen storage. To resuspend cells from frozen, cells were transferred to dry ice and swiftly thawed in a 37°C water bath. Cells were immediately resuspended in pre-warmed WM and centrifuged as above. Cells were resuspended in 3ml CM and plated in a 6-well plate, with CM replaced the following day.

Colo320DM cells were cultured in DMEM (Thermo Fisher Scientific) supplemented with 10% foetal calf serum (FCS). For passaging, adherent cells were first washed in PBS prior to dissociation with 0.25% (v/v) Trypsin-EDTA solution (Gibco), using the same volumes outlined above. This was then neutralised with the FCS-containing growth media prior to centrifugation and resuspension of cells as previously described.

Where required, cells were counted manually with a haemocytometer counting chamber (Marienfeld Neubauer-improved #0640030).

2.1.3 Preparation of cells for experimental work

Cells were grown on Poly-Prep (poly-L-lysine) coated glass slides (Sigma, P0425) in a four-well dish or on 22x22mm glass cover slips in a six-well plate and pre-laminated as described above. Cells were washed with PBS, fixed for 10 minutes in 4% paraformaldehyde (pFA) and permeabilised for 15 mins in 0.5% Triton X-100 (Merck) in 1x PBS (4ml per slide). Slides were washed three times with PBS between each step and incubated overnight (o/n) at 4°C in PBS prior to ensure the pFA was completely removed. Slides for DNA FISH could be dried and stored at -80°C prior to use.

2.2 Immunofluorescence (IF)

IF was performed in a humidified chamber, where, in all cases, 50-100ul of blocking buffer+-antibody was applied to a cover slip which was laid onto the cell-covered slide. Where IF was performed on cells plated on a coverslip, 50ul of blocking buffer +- antibody was applied directly to parafilm sealing film (Parafilm, HS234526B) and the cover slip laid cell-side down onto the parafilm. Slides and coverslips were washed 3x after primary and secondary antibody steps by being completely submerged in PBS in a Coplin jar.

Slides or cover slips were blocked in blocking buffer A (1% BSA and 0.1%Triton X-100 in 1x PBS) for 30 min at 37°C before incubation o/n at 4°C with the primary antibody diluted in blocking buffer A (Rpb1 NTD (D8L4Y), Cell Signalling Technology, 1 in 1000; mCherry (Rabbit), abcam, 1 in 500; mCherry (Rat), 1 in 500; phospho-Histone H2A.X (Ser139), Sigma-Aldrich, 1 in 2000; PolyADP-ribose, Sigma-Aldrich, 1 in 250). The following day, slides were washed 3x in

PBS before incubation for 1 h at 37°C with an appropriate corresponding secondary antibody (1 in 1000 AlexaFluor). After further PBS washes and DAPI staining (50ng/ml for 3 mins), slides were mounted with Vectashield anti-fade mounting medium (Vector laboratories).

The only exception to this protocol was IF for phospho-Histone H3 (pH3) and Ki67 which was performed in the Pollard lab, and imaged using the Nikon Eclipse Ti inverted microscope (CRM Imaging facility, UoE). The protocol was as follows:

For pH3: 50,000 NSCs per well were plated in a pre-laminated 48 well plate (3 wells per condition) suspended in a total of 450µl CM per well and were left to adhere o/n. The following concentrations of each drug were then added in a volume of 150µl to each well, to make a total final volume per well of 600ul: colcemid 0.1, 0.05, 0.0125, 0.01µg/ml; paclitaxel: 100, 50, 10, 5nm; nocodazole: 0.5, 0.25, 0.1, 0.05µg/ml. The volume of colcemid was normalised with PBS, and for paclitaxel and nocodazole, with DMSO, so the volumes of PBS and DMSO respectively were the same across all drug concentrations. The duration of drug exposure in order of the concentrations shown for all three drugs was: 30 mins, 1 h, 6 h, 6 h. Cells were then washed with PBS, fixed in 4% pFA for 10 mins, incubated in blocking buffer B (10% goat serum (Sigma), 0.3% Triton X-100 in 1x PBS) for 20 mins and then incubated o/n at 4°C in fluorophore-tagged primary antibody (pH3 (ser28) eFluor 660, eBioscience, 1 in 50) diluted in blocking buffer. After three thorough washes in PBS, cells were incubated in DAPI (50ng/ml for 3 mins), and washed and retained in PBS.

For Ki67: 3000 cells of each cell pool (E37 EGFR High, Low and Null) were plated in a pre-laminated 8-well chamber slide (Ibidi, 80826) up to a CM volume of 300µl and left to adhere o/n. The following day the CM was removed and was replaced with fresh CM (as a control) or 5% FCS in CM where no EGF or FGF had been added up to a total volume of 400µl. Cell were grown in their respective media for 7 days, washed twice with PBS, fixed for 10 mins in 4% pFA and washed again thoroughly in PBS. After a 20 min incubation in 200µl of blocking buffer B, cells were incubated o/n at 4°C in 200µl Ki67 (Thermo Fisher) primary antibody diluted in blocking buffer B. After three PBS washes, cells were intubated in anti-rabbit secondary antibody (1:1000) for 1 h at 4°C. Cells were washed and incubated in DAPI as above, then washed and retained in PBS.

2.3 Metaphase spreads

Cell lines were optimised to generate metaphase spreads. Cells were grown until near-confluence in a T75 flask and incubated for 30 mins with colcemid (Thermo Fisher) added directly to the media to 0.1ug/ml. Media and cells dissociated with Accutase were centrifuged as above, washed in PBS (cell pellet gently resuspended in 10ml PBS in a 14ml Falcon tube and centrifuged as above) and resuspended in 10ml 0.56% KCl added drop-wise while gently vortexing. Cells were incubated for 10 mins in a 37°C water bath, centrifuged (300 *g*, 5 mins) and the pellet resuspended in 10ml methanol:acetic acid (Me:Ac – dilution ratio 3:1) into a 14ml Falcon tube drop-wise while gently vortexing. Cells in Me:Ac were either allowed to stand at room temperature for an hour, or stored at -20°C before proceeding. Following two further Me:Ac washes, cells were dropped onto slides. Cells in Me:Ac could be stored long term at -20°C suspended in 10ml Me:Ac. To resuspend a stored sample, cells were centrifuged at 300 *g* for 5 mins, the stored Me:Ac poured off and the cell pellet resuspended in fresh Me:Ac.

To generate a metaphase spread, cells were resuspended in a small amount of fresh Me:Ac until the suspension appeared slightly milky. Using a thin-tipped plastic dropping pipette, a small amount of the cell suspension was aspirated and a single drop expressed from ~30cm onto a glass slide, and spread by rotating and blowing on the slide. Once completely dry, slides were stained with 50ng/ml DAPI diluted in PBS for 3 mins, and mounted in Vectashield mounting medium.

This was effective in generating metaphase spreads in E37 cells. All other GCGR cell lines generated few spreads, and those present were bunched, short and not adequate for analysis despite optimising the duration and concentration of colcemid (Table 2.2).

Cell line	Response to colcemid
E20	Very few metaphases, very deformed nuclei
E25	Very few metaphases and those present are very bunched
E26	Very few metaphases and remain very bunched
E28	Very few and bunched metaphases
E37	Appropriate number, length and spread with colcemid
NS9	No metaphases seen

Table 2.2 | Summary of cell line response to colcemid

2.3.1 Optimisation of metaphase spreads

2.3.1.1 Mitotic arrest agent

To gain a preliminary assessment of agents that might be more successful at inducing mitotic arrest, immunocytochemistry (ICC) was performed for phospho-Histone 3 (pH3) in NS9 cells using two agents predicted to achieve mitotic arrest, nocodazole (Cambridge BioScience) and paclitaxel (Sigma-Aldrich) at a range of doses and durations (See Section 2.2 – Immunofluorescence and Section 2.10.3.1 – Intensity Analysis). This suggested that while the proportion of mitotic cells remained small, nocodazole and paclitaxel may be more successful at inducing mitotic arrest than colcemid in neural stem cell-like cells (Figure 2.1A and B). To further test this, and estimate an appropriate dose for metaphase spreads, I performed a growth curve analysis using the Incucyte Live Cell Imager (See Section 2.9.1 – Cell growth assay). E25 and E26 cells were plated on a pre-laminated 48 well plate with 10,000 cells/well in 500µl CM. The following day, CM was replaced with 0, 0.5, 1 and 2µM of nocodazole and paclitaxel in CM on E25 and E26 cells, volume-standardised DMSO as the control, and plotted growth curves using the Incucyte. This suggested growth arrest occurred with both agents within the range of drug doses used (Figure 2.1C).

I generated metaphase spreads in E25 and E26 cells using 1µM nocodazole and 100nm paclitaxel for 16 hours. The protocol was otherwise as described in Section 2.3. The best metaphase spreads in terms of number, length of chromosomes and spread were generated using paclitaxel for both cell lines (Figure 2.2). For other cell lines that subjectively grew more quickly (E28, NS9), I found a combination of paclitaxel and nocodazole at shorter incubation times generated a good number of metaphase spreads featuring chromosomes of appropriate length (summarised in Table 2.3).

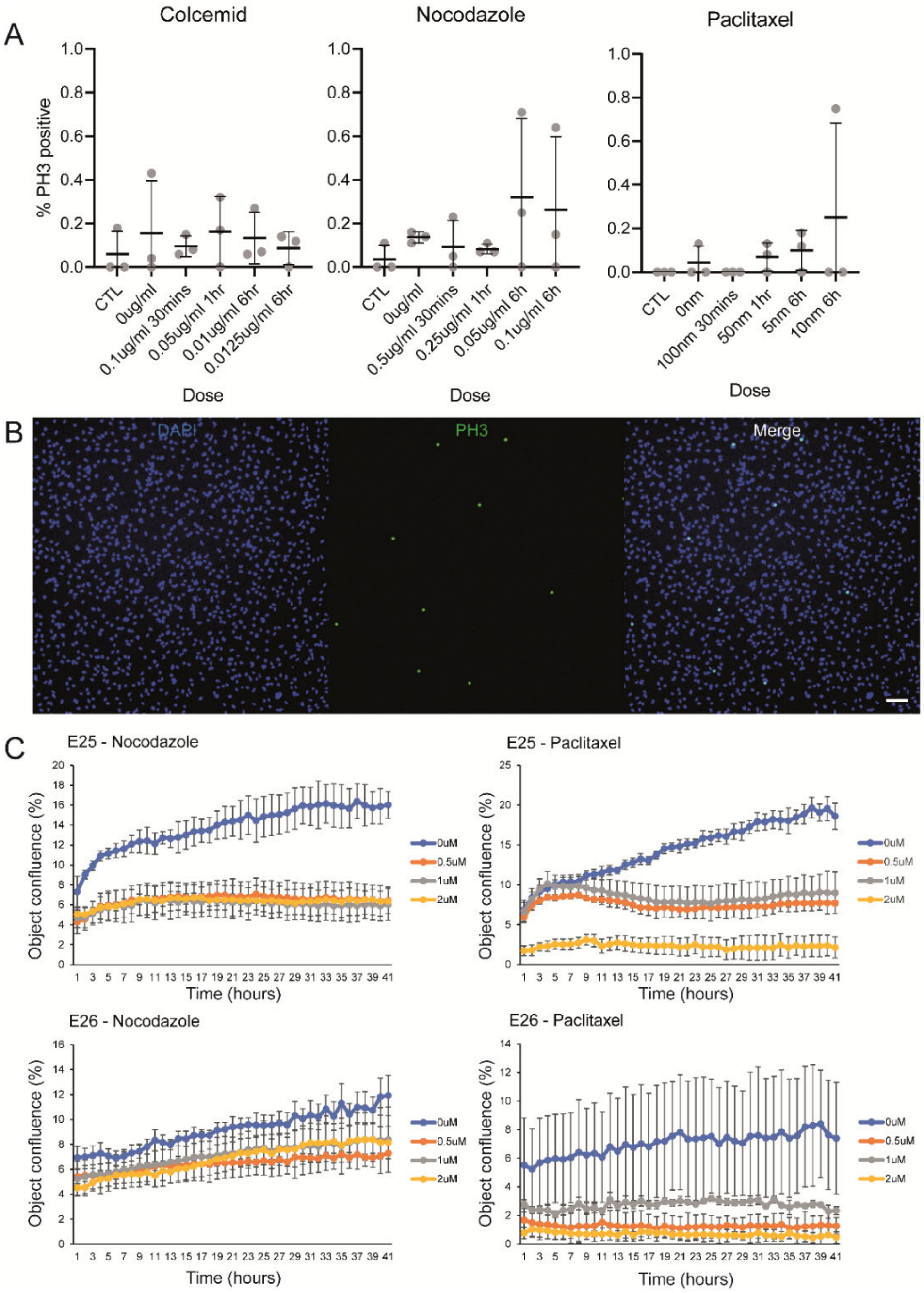


Figure 2.1 | Induction of mitotic arrest with alternative agents

A) NS9 cells treated with colcemid, nocodazole or paclitaxel, doses/durations shown, 3 fields of view (technical replicates), nuclei positive for pH3 as % of all nuclei (DAPI) per field of view, mean +- standard deviation (SD) of 3 technical replicates. B) Representative image of NS9 cells treated with nocodazole 0.05µg/ml over 6 hours, IF for PH3 (Green), nuclei (DAPI – blue). Scale bar = 100µm. C) Growth curves for E25 and E26 nuclei incubated with nocodazole and paclitaxel (concentration/duration shown), mean +- SD of 3 technical replicates. CTL/0µM = DMSO.

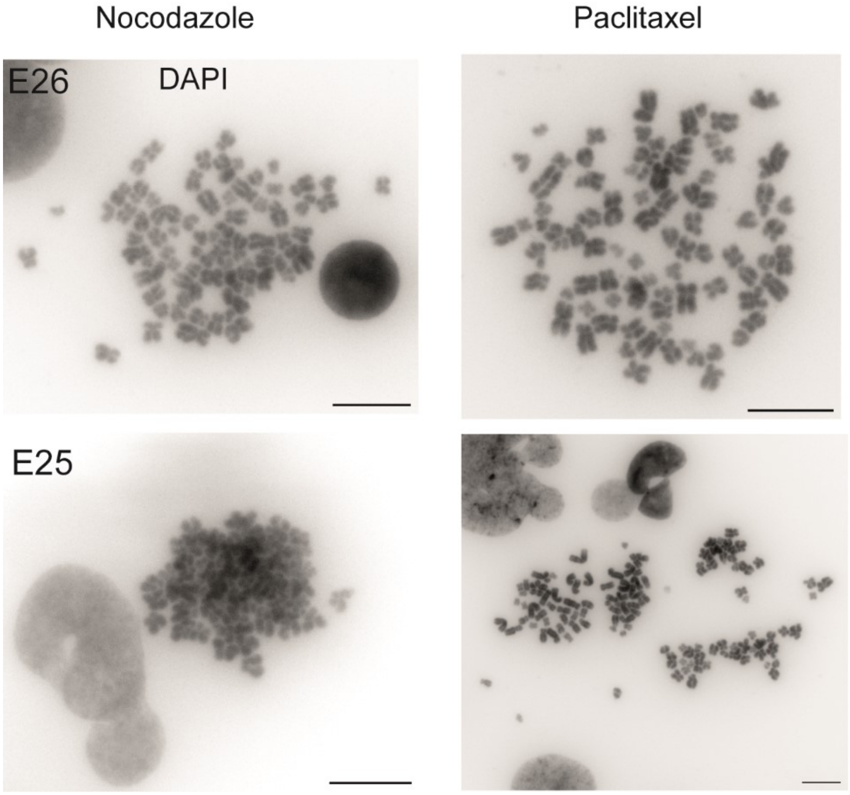


Figure 2.2 | Metaphase spreads with nocodazole and paclitaxel

Representative metaphase spreads, DNA = DAPI (greyscale). Scale bar = 10μm. 1μM nocodazole overnight, 100nm paclitaxel overnight.

2.3.1.2 Hypotonic solution

Incubation in a hypotonic solution results in cell lysis, with 0.56% KCl the most widely used solution. This resulted in metaphase spreads where the chromosomes were clumped together rather than being identifiable as separate structures, so I trialled sodium citrate dihydrate 0.9% (NaCit) as an alternative hypotonic solution, both alone and in combination with 0.56% KCl (MacLeod, Kaufmann and Drexler, 2011). The surface area of the metaphase spreads was measured by an imaging analysis script written in ImageJ to establish which hypotonic solution generated well separated chromosomes suitable for karyotyping (See 2.10.1 – Metaphase spread area analysis), as shown for E37 EGFR CTL/Low/KO cells (Figure 2.3). For E28 cells, I found a 50:50 mix of NaCit 0.9%:KCl 0.56% gave the best spreads (Figure 2.4).

Humidity is another factor which may affect the ‘spread’ of metaphase spreads, in addition to duration of hypotonic solution incubation (MacLeod, Kaufmann and Drexler, 2011). Humidity can be increased by drying slides in a metal tray floated in a 50°C water bath. Cells in Me:Ac fixative were dropped in different humidity conditions, measuring the area of the resulting spreads by the same ImageJ Macro script as above. Slides dried in humid conditions resulted in a larger spread area, allowing for easier analysis of metaphase spreads (Figure 2.5). The optimised mitotic arrest agent and hypotonic solution for each cell line where metaphase spreads are presented in this thesis are shown in Table 2.3.

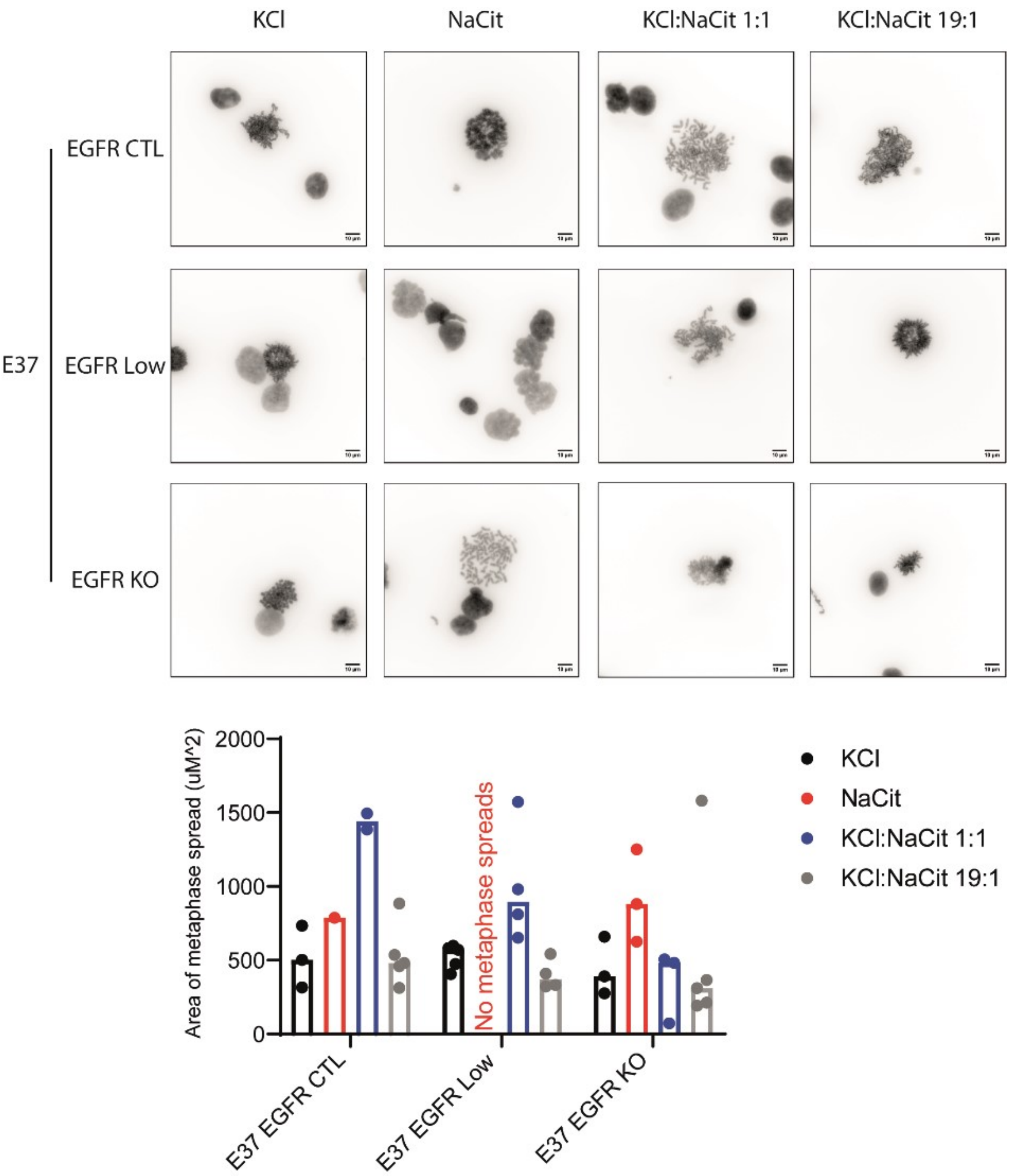


Figure 2.3 | Metaphase spread – optimisation of hypotonic solution.

E37 EGFR CTL/Low/KO cell lines (see Chapter 5) incubated in 0.56% (KCl), 0.9% NaCit or combinations thereof in the ratios shown. Representative metaphase spreads shown, DNA = DAPI (greyscale), scale bar = 10µm. Graph plotting area of individual metaphase spreads per condition, calculated by ImageJ macro.

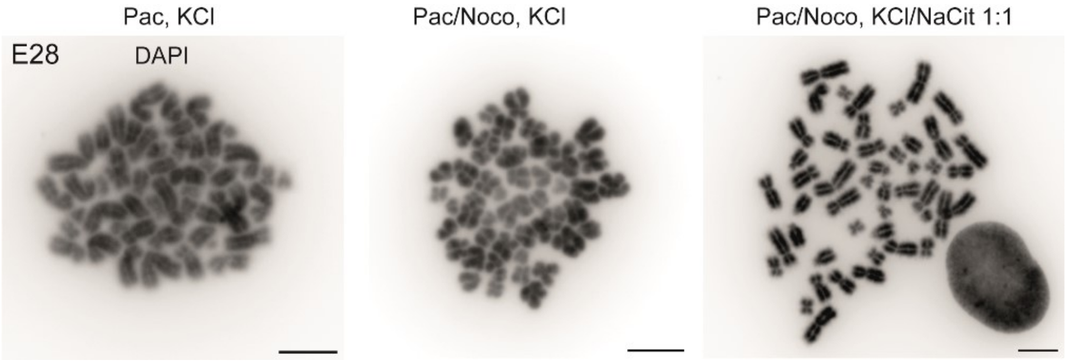


Figure 2.4 | E28 metaphase spread – optimisation of hypotonic solution

Pac = paclitaxel 10nm 6hr; Noco = nocodazole 0.1µg/ml 4hr. KCl = Potassium chloride 0.56%; NaCit = Sodium Citrate 0.9%. DNA = DAPI (greyscale). Scale bar = 10µm.

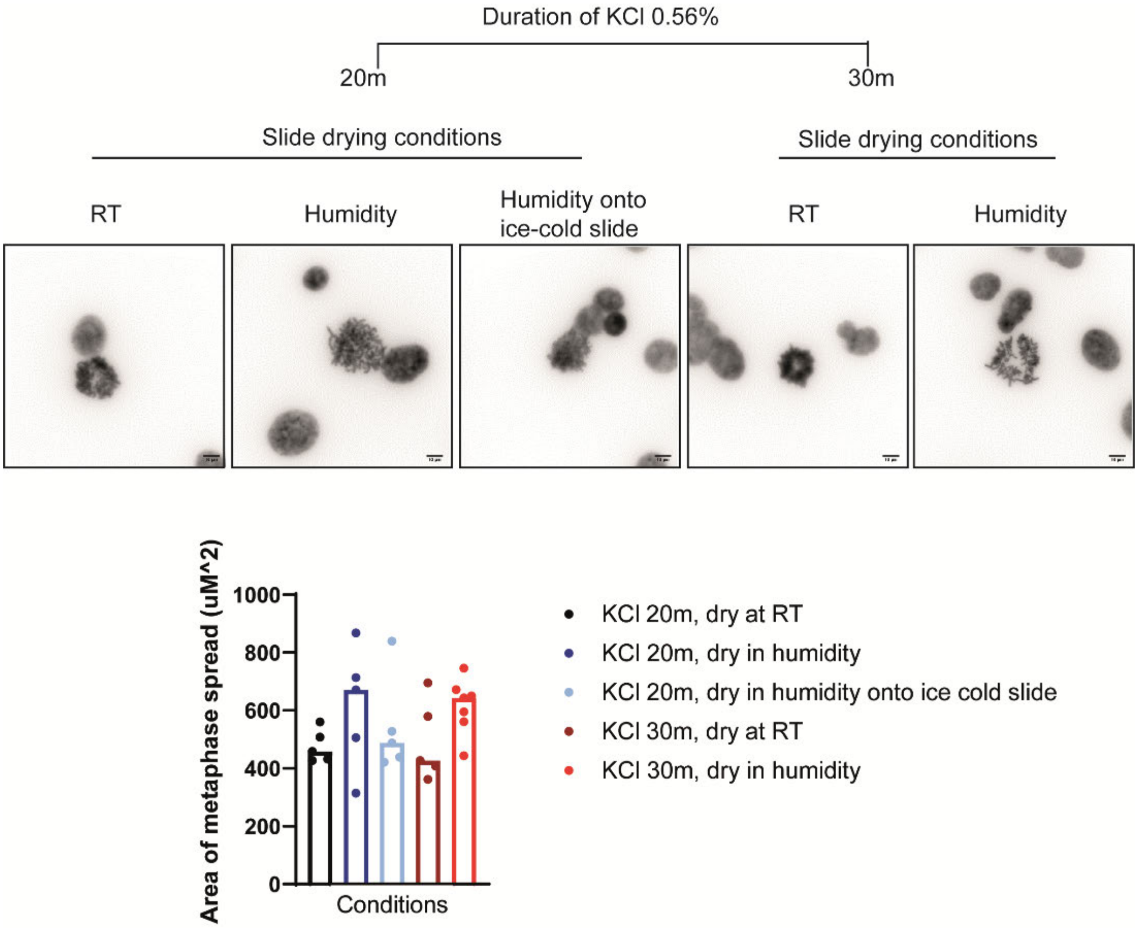


Figure 2.5 | Effect of humidity on metaphase spread area

Conditions as indicated, representative images shown. DNA = DAPI (greyscale). Scale bar = 10µm. Cells – E37 EGFR Low (See Chapter 5). Graph shows area of metaphase spread, with each dot representing one metaphase spread.

Cell line	Mitotic arrest agent	Hypotonic solution
E37	Colcemid 0.1µg/ml, 30 mins	KCl 0.56% 10 mins
E20	Paclitaxel 100nm 16 hours	KCl 0.56% 20 mins
E25	Paclitaxel 100nm 16 hours	KCl 0.56% 20 mins
E26	Paclitaxel 100nm 16 hours	KCl 0.56% 20 mins
E28	Paclitaxel 10nm 6 hours with Nocodazole 0.1µg/ml 4 hours	KCl 0.56% 20 mins
NS9	Paclitaxel 10nm 4 hours with Nocodazole 0.05µg/ml 4 hours	NaCit 0.9%:KCl 0.56% 1:1, 20 mins
E26 EGFR CTL/Low/KO	Paclitaxel 10nm 6 hours with Nocodazole 25ng/ml 6 hours	NaCit 0.9% 20 mins (25 mins for EGFR KO)
E37 EGFR CTL/Low/KO	Colcemid 0.01µg/ml 16 hours	CTL/Low - NaCit 0.9%:KCl 0.56% 1:1, 10 mins KO - NaCit 0.9% 10 mins
Colo320DM	Colcemid 0.025µg/ml 2 hours	KCl 0.56% 10 mins

Table 2.3 | Summary of optimised metaphase spread conditions for all cell lines

2.4 Fluorescence in situ hybridisation (FISH)

2.4.1 DNA FISH

2.4.1.1 Fosmid probe preparation

Fosmid probes were ordered from the WIBR2 Library (BACPAC Genomics Inc, Emeryville, California). A bacterial stab from the probe was streaked onto a plate of LB-Agar with 12.5ng/µL chloramphenicol and incubated o/n at 37°C. A single colony was picked and cultured o/n at 37°C in a shaking incubator, in a 50ml Falcon tube containing 5ml L-broth with 12.5ng/µL chloramphenicol, including a stab-free culture tube as a negative control. A glycerol stock for each fosmid was prepared using 800µl of the o/n culture suspension with 400µl 100% glycerol, which was stored long term at -80°C.

An alkaline miniprep was performed to generate fosmid DNA. An o/n culture of bacteria containing the relevant fosmid probe was centrifuged (5 mins at 4000 x *g*), the supernatant discarded and the pellet resuspended vigorously in 200µl GTE buffer (50 mM Glucose, 25 mM Tris pH 8, 10 mM EDTA) with 50mg lysozyme. After incubating for 5 mins at room temperature (r.t), 400 µL lysis buffer (0.2 M NaOH, 1% SDS) was added, the solution mixed by inversion and incubated on ice for 5 mins. Then 300 µL acetate buffer (3M potassium acetate, 11.5% glacial acetic acid prepared in dH2O) was added and mixed gently but thoroughly by vortexing to ensure flocculation could be observed. After incubating on ice for 5 mins, samples were centrifuged (16,000 x *g*, 5 mins, 4°C) and the supernatant transferred to a fresh 1.5ml Eppendorf tube. DNA was purified by phenol:chloroform extraction, 500μL phenol:chloroform were added, mixed by inversion and centrifuged (16,000 *g*, 4 mins, 4°C). The top layer was transferred to a fresh tube and this process was repeated with chloroform. DNA was precipitated from the aqueous phase with 500μl isopropanol. After >1 h at -20°C DNA was precipitated by centrifugation (16,000 *g*, 15 mins, 4°C). The pellet was washed in 70% EtOH and dried before resuspension in 25μl Tris/EDTA pH 8.0 (TE) supplemented with 2μl 20mg/ml RNase A (Invitrogen). This was incubated for 5 mins at 37°C before long term storage at -20°C. Fosmid DNA was quantified with the Qubit dsDNA broad range assay (Thermo Fisher) using the Qubit 4 fluorometer.

Fosmids used in this thesis are listed in Appendix 8.1. All fosmids were validated to ensure they bound specifically and sensitively to the intended locus prior to use. Validation was either by performing DNA FISH on an existing Me:ac preparation of Neo3 cells, a human parental lymphoblastoid cell line with a normal karyotype, or by performing FISH on NS9 cells with a previously validated fosmid binding to the same chromosome.

2.4.1.2 Nick Translation

Fosmid DNAs (6μl - (0.5 – 1 μg)) were directly labelled by nick translation to incorporate a fluorescent dUTP (2.5μl) (Green496-dUTP, ENZO life sciences; ChromaTide AlexaFluor 594-5-dUTP, Thermo Fisher Scientific) by incubation with 2.5μl each of unlabelled 0.5mM dATP, dCTP and dGTP, 1ul ice cold DNase I (Roche; 1:5 dilution in ice cold water), 2 μL nick translation salts (0.5 M Tris pH 7.5, 0.1 M $MgSO_4$, 1 mM dithiothreitol (DTT) and 0.5 mg/ml BSA) and 1μl DNA polymerase I (Invitrogen) for 90 min at 16°C. The reaction was quenched with 3μl 0.5M EDTA and 2μl 20% SDS, TE buffer added to a total volume of 90μl and the reaction mix purified using a Quick Spin Sephadex G50 column (Roche). Once labelled, fosmid probes were stored at -20°C.

2.4.1.3 Probe Hybridisation

Cells on slides or cover-slips for 3D FISH were prepared by briefly rinsing them in 2x Trisodium citrate and sodium chloride (SSC). Cells on slides for 2D FISH (metaphase spreads dropped in Me:Ac fixative) were additionally dehydrated for 1 h at 70°C if they had been dropped the same day.

Slides were incubated for 1 h in 2xSSC with 100 µg/ml RNaseA (Invitrogen) at 37°C in Coplin jars, then dehydrated in 70%, 90% and 100% EtOH for 2 minutes each. After air drying, slides were incubated for 5 minutes at 70°C and immersed in a denaturing solution (2xSSC/70% formamide, pH 7.5) heated to 70°C (Me:ac-fixed cells) or 80°C (pFA-fixed cells) for 1-2 mins for 2D DNA FISH, depending on the timeframe since the cells were dropped, and 40 mins for 3D DNA FISH. Slides were immediately immersed in ice-cold 70% ethanol, followed by 90% and 100% EtOH at r.t., for 2 mins each before air drying.

The only deviation from this protocol was for slides where RNA FISH had previously been performed on a slide. Here, slides were transferred directly from PBS into a denaturing solution at 80°C for 15-30 mins without the preceding 2xSSC or EtOH washes. Following denaturing, slides were immediately washed in 2xSSC and minimally dried.

FISH probes were prepared by combining approximately 80-100ng of each directly labelled fosmid probe (per slide), 6 µg Human Cot-1 DNA (per probe; Invitrogen), 5 µg sonicated salmon sperm (per slide; Invitrogen) plus two volumes of 100% ethanol. The probe mix was pelleted and dried. The pellet was suspended in hybridisation mix (50% deionised formamide (DF), 2× SSC, 10% dextran sulphate (Sigma), 1% Tween 20 in ddH2O) by vortexing and briefly centrifuging, and then incubated for 1 h at r.t. in the dark. Alternatively, in order to perform chromosome territory analysis (Chapter 4), FISH probes were instead suspended in 10µl of Chromosome 7 paint (XCP 7 Orange, Metasystems) and incubated in the same way. Probes were then denatured for 5 mins at >70°C and annealed at 37°C for 15 min. Probes were then hybridised to the denatured slides under a sealed coverslip o/n at 37°C.

The following day, slides were washed four times for 3 mins each in 2xSSC (45°C) then 4x 3 mins in 0.1% SSC (60°C), then immersed in 4xSSC/0.1% Tween 20 with 50ng/ml DAPI for 3 mins. Slides were mounted with 25-50µl of Vectashield onto an appropriately sized coverslip and sealed with clear nail varnish.

Where noted, an additional Centromere 7 (CEN7 - CHR07-Dig Control) FISH probe (Pisces Scientific) was prepared by combining 2µl CEN7 probe with 8µl of CEN7 hybridisation mix. This was denatured for 5 min at 80°C and snap frozen on crushed ice. This CEN7 hybridisation mix was combined with the relevant DNA FISH probe/hybridisation mix which was prepared as above except in a slightly greater volume (20µl). This was applied directly to denatured slides and hybridised overnight as above. Following washing (see above), slides were incubated at 37°C in a humidified chamber with 50µl blocking buffer (4x SSC/5% Marvel) for 5 mins. This was followed by a 1 h incubation with 50µl anti-digoxigenin antibody (Roche; 1 in 10) and a 1 h incubation with 50µl anti-sheep Alexa Fluor 647 secondary antibody (Thermo Fisher Scientific; 1 in 10) in the same conditions with 4xSSC/0.1% Tween 20 washes (3 x 2 mins) in between. After the final wash, slides were stained with DAPI and mounted with Vectashield as above.

2.4.2 RNA FISH

Custom RNA FISH probes labelled with Quasar® 570 targeting the first intron (pool of 48 22-mer probes) of *EGFR* were designed and ordered via the Stellaris probe designer (Appendix 0; Biosearch Technologies, Inc., Petaluma, CA) (https://www.biosearchtech.com/support/tools/design-software/stellaris-probe-designer, version 4.2). Slide preparation was as described above, with RNAse-free PBS used for the o/n wash, with all PBS and ddH2O treated with DEPC and autoclaved. Solutions were pipetted with filter pipette tips and all surfaces and equipment were treated with RNaseZAP (Invitrogen AM9780).

Slides were immersed in RNA FISH wash buffer (2xSSC, 10% DF in ddH2O) for 2-5 min. The RNA FISH probe was diluted to 125nM and 1µl added to 100µl RNA FISH hybridisation mix (10% DF in Stellaris® RNA FISH Hybridization Buffer). Once mixed by pipetting, 25µl of the probe/hybridisation mix was applied to a coverslip and incubated on the slide for 4-16 h at 37°C. Thereafter, slides were washed at 37°C for 30 min in RNA FISH wash buffer, followed by 30 min with DAPI (5ng/ml) in RNA FISH wash buffer. Slides were washed with PBS for 2-5 mins before mounting with Vectashield.

For combined RNA:DNA FISH, RNA FISH was performed as described. Once imaging had been completed, coverslips were removed using acetone and the slides washed in PBS. A modified version of DNA FISH was then performed, as outlined above.

2.4.3 Immuno-FISH

For immuno-FISH (DNA), IF was performed as outlined above. Following the final washes after secondary antibody incubation, slides were incubated with 4% pFA for 30 mins at r.t. to fix the signal. Following thorough PBS washes (three times), the DNA FISH protocol was then followed as above.

For immuno-FISH (RNA), the antibodies were added at the same concentration as described above directly to the hybridisation mix (primary antibody) and 2xSSC/10% DF washes (secondary antibody), without further deviation from the RNA-FISH protocol outlined.

2.5 Flow Cytometry and fluorescence-activated cell sorting (FACS)

2.5.1 Flow Cytometry

Cells were grown and treated as required then prepared for flow cytometry by adding WM (which does not contain EGF) for 30 mins. Cells were dissociated as described above, and the pellet resuspended in wash solution (WS) (0.1% BSA/PBS). Cells were counted and 50,000 cells added to a 1.5ml Eppendorf. These were centrifuged (2300 *g* for 4 mins) and resuspended in 200µl WS. This was supplemented with 100ng/ml EGF-647 (Thermo Fisher Scientific - reconstituted in WS), with an equivalent volume/number of cells incubated in WS as a negative control. After incubation for 25 mins at 37°C, cells were washed 3x in WS, centrifuging (600 *g* for 1 min) and discarding the WS each time, before resuspending in 400µl WS. Samples were analysed on the BD

FACSAria III FUSION or Fortessa (CRM Flow Facility, UoE), or the BD LSRFortessa X-20 (IGC Flow Facility, UoE).

2.5.2 Fluorescence-activated cell sorting (FACS)

The protocol was overall the same as outlined under flow cytometry. Where a cell population was intended for sorting, more cells were prepared and incubated in EGF-647, with the volume of WS and concentration of EGF-647 increased proportionally. Cells were sorted on the BD FACSAria III FUSION by EGF-647 fluorescence signal, with gating defined according to the relevant positive and negative controls. A sort check was performed to verify these were true populations prior to expanding cells in pre-warmed CM on pre-laminated 48-well plates.

For the data presented in Chapter 4, sorted cells were immediately expanded onto pre-laminated 22x22mm coverslips containing pre-warmed CM. Fifteen days after the cells were sorted, the coverslips were fixed, permeabilised and DNA FISH performed.

For data presented in Chapter 5, cells sorted by FACS were reanalysed by flow cytometry to confirm the EGFR-status once cells of all sorted populations were near confluence. This varied between 13 days (and 13 days following EGFR Low repeat FACS) (E37– replicate 1), 31 days (and 10 days following EGFR Null repeat FACS) (E37– replicate 2), and 74 days (E26).

For both flow cytometry and FACS, and across both facilities, BD FACSDiva Software (BD Biosciences) was used for flow cytometer and application set up, and data acquisition. Where further analysis was performed, this was performed with FCS Express software (v7, https://denovosoftware.com/). For flow cytometry analysis following IR +- olaparib, the mean EGF-647 signal of the experimental (EGF-647+) samples was normalised to their corresponding negative (EGF-647-) control samples.

2.6 DNA Damage

2.6.1 Ionising Radiation (IR)

2.6.1.1 Adapted Colony Survival Assay

An adapted colony survival assay was performed to validate the selected dose of IR, and to confirm that the glioblastoma cells selected were radiosensitive. Cells were plated in pre-laminated 6-well plates with two technical samples for each cell density (2,000, 4,000, 8,000 and 16,000 cells per well) and by limited serial dilution for consistent cell density. The following day, plates were irradiated with or without 4Gy using the Faxitron X-Ray machine, a clinically relevant dose used in previously published *in vitro* studies of IR in glioblastoma (Ahmed *et al.*, 2015; Carruthers *et al.*, 2018). Cells were then cultured, changing media carefully after 5 days, until the untreated cells became confluent or day 12 was reached, whichever occurred first. Plates were washed with PBS and incubated at r.t. in methylene blue for 45 mins, washed 3x with PBS and dried at r.t. Once dry, plates were imaged using an Epson Scanner in 24 bit colour, 2400dpi resolution. Analysis was performed with Ilastik (v1.3.3) with settings dependent on the cell line to automate counting of stained cells in each pair of technical replicates.

Other cell viability assays were considered, such as fluorescent cell viability assays (e.g. RealTime-Glo MT Cell Viability Assay) or fluorescence assays that measure other markers of cell survival such as ATP (e.g. CellTiter-Glo Luminescent Cell Viability Assay) or protease activity (e.g. CellTiter-Fluor Cell Viability Assay). I chose methylene blue as this assay was previously optimised by others using the radiation facility, and this overall methodology is the most widely used assay in order to assess cell viability following IR (e.g. Ahmed *et al.*, 2015; Carruthers *et al.*, 2018 where crystal violet was used as an alternative dye).

2.6.1.2 IR and analysis of ecDNA

E26 and E28 cells were plated and grown until near confluence on 150mm tissue culture plates. These plates were used because culture flasks were not compatible with the Faxitron X-ray machine. Cells were treated with or without 4Gy IR. They were then treated with their optimised mitotic arrest agent for metaphase spreads (Table 2.3) so that cell dissociation would be performed 24h after IR. Cells were then processed as per metaphase spread and DNA FISH protocols. Metaphase spreads and nuclei were imaged as single slices using the Zeiss AxioImager 2. Nuclear signal intensity was analysed using an ImageJ macro (see Signal Intensity Analysis: 2.10.3.2)).

2.6.2 Treatment of cells with olaparib

Cells were plated at a density of 400,000 cells per Poly-Prep slide in a four-well dish and left to settle overnight in 4ml CM. The following day, the media was changed for 4ml CM containing 1µM olaparib (Selleck Chemicals) or the same volume of DMSO as a control. After 1 hour incubation, cells were either left untreated, or treated with 1 in 5000 hydrogen peroxide (H_2O_2) 30% (w/w) solution (Merck) for 10 minutes. Cells were then washed in PBS, and fixed, permeabilised and stained by IF.

To assess the effect of olaparib treatment over 24h and 2 weeks, E26 cells were plated in CM on Poly-Prep slides (2 slides at 400,000 cells/slide for 24h treatment, 2 slides at 100,000 cells/slide for 2 week treatment). Once settled, the media was changed for 4ml CM containing 1µM olaparib or the same volume of DMSO as a control. For the 2 week treatment, media was replaced with fresh olaparib/DMSO every 4 days to ensure continued inhibition of PARP (Prasad *et al.*, 2017; Cahuzac *et al.*, 2022). Cells were then washed, fixed, permeabilised and analysed by DNA FISH.

2.6.3 Combined IR and olaparib

Cells were plated in 10cm plates ($1x10^6$ cells) or pre-laminated cover slips in a 6-well plate ($1x10^5$ cells) in CM. Once cells had settled, the media was replaced with 1µM olaparib or the same volume of DMSO as a control. Plates were then immediately treated with or without 2Gy of IR. Both steps were repeated each day for 4 days (See Figure 5.6). The experiment was concluded on the fifth day by flow cytometry and DNA FISH.

IR doses of 4Gy (single fraction) and 8Gy (2Gy per fraction across 4 fractions) were used in experiments as described above. These doses are within the range of many previously published studies of IR in glioblastoma cells (i.e. single fractions of 1-5Gy at 1Gy increments (Ahmed et al., 2015; Carruthers et al., 2018) and so can both be considered clinically relevant. The higher single dose was chosen to ensure radiosensitivity at a relatively high dose, whilst remaining within previously published ranges. When combining IR and olaparib, a longer and more aggressive treatment approach was selected to ensure an effect would be observed, as I was concerned that any treatment effect observed with a single 4Gy fraction at a short time point might be too subtle. As such, a higher overall dose with a corresponding lower dose per fraction was selected to achieve a balance between sufficient IR exposure and limited acute cell death.

2.7 CRISPR/Cas9 genome editing

2.7.1 Assembly of the Cas9/crRNA complex

CRISPR-Cas9 crRNA (crRNA, 100 pmoles, Integrated DNA Technologies (IDT), USA) and universal tracrRNA (100 pmoles, IDT, USA) were combined in a 0.2ml PCR tube and assembled to a cr:tracrRNA complex by annealing at the following settings on a PCR block: 95°C for 5 min, step down cooling from 95°C to 85 °C at 0.5°C/sec, step down cooling from 85°C to 20°C at 0.1°C /sec, store at 4°C. A tracrRNA-only PCR tube was prepared and processed in the same way as a negative control. Recombinant Cas9 protein (10µg, purified in house - see (Dewari *et al.*, 2018)) was thawed on ice and added to the cr:trRNA complex and corresponding negative control, thereby forming the ribonucleoprotein (RNP) complex, by incubating at r.t. for 10 mins. The cr:tracrRNA/RNP complex was stored on ice until ready for electroporation.

For EGFR CRISPR/Cas9 knock-out: Two Alt-R crRNA were designed to target the 3' and 5' ends of *EGFR*. For POLR2G mCherry knock-in: A single Alt-R crRNA was used targeting the 3' end of *POLR2G* (Appendix 8.2).

2.7.2 Cell nucleofection

A 6-well plate was prepared by prelamination and adding 2ml prewarmed CM to each well. Cells were dissociated and prepared as above and 2-3x10^5 cells were aliquoted into a 1.5ml Eppendorf. Cells were centrifuged, all CM removed by pipetting and cells resuspended in 20µl of Lonza SG cell line buffer (Lonza). This cell suspension was immediately mixed by pipetting with the pre-prepared cr:tracrRNA/RNP complex, transferred to a microcuvette and electroporated into cells using the 4D Amaxa X Unit (programme DN-100). Pre-warmed media (100µl) was added immediately to the microcuvette chamber after nucleofection to aid cell recovery, before pipetting the cell suspension directly into the previously prepared 6-well plate. Media was changed for fresh CM the following day. Cells were allowed to recover until at least near-confluence on a 6-well-plate, before analysis by flow cytometry +- FACS as outlined above.

Where FACS was performed after nucleofection, this occurred 9-11 days later to allow cells to recover.

2.7.1 mCherry_PolR2G knock in cell line

Generation of this E28 cell line was performed by Dr Pooran Singh Dewari who designed the crRNA (Appendix 8.2) and donor DNA (Appendix 8.3) using the TAG-IN tool incorporating the fluorescent reporter gene sequence for mCherry to create a donor double-stranded DNA (dsDNA - Twist Bioscience) (Dewari *et al.*, 2018). CrRNA, tracrRNA and Cas9 protein were assembled to form the cr:tracrRNA/RNP complex as outlined above. Donor dsDNA (300ng) was prepared in 30% DMSO, denatured by incubating at 95 °C for 5 min, and immediately placed on ice to stop the reaction. Nucleofection was performed as described above, mixing the donor dsDNA with the RNP complex prior to mixing with the cells, using the same 4D Amaxa X Unit (programme DN-100). Cells were grown by serial expansion over two weeks. Knock-in efficiency was assessed by flow cytometry for mCherry, with tracrRNA:Cas9 only cells as a negative control. 5-7x 10^5 cells were suspended in 0.2% BSA/PBS and analysed on BD LSRFortessa Cell Analyzer, followed by FACS into a pure mCherry-knock-in population.

I performed validation of this cell line by IF, the methods for which are outlined in Methods: Immunofluorescence Section 2.2.

2.8 Western Blotting

Cells were dissociated with Accutase and centrifuged into a cell pellet as described above. Pellets were resuspended in RIPA buffer (50 mM Tris-HCl pH 8.0, 150 mM NaCl, 1% NP-40, 0.5% deoxicholate, 0.1% SDS) and protease inhibitors (Complete, Roche) on ice and protein lysate extracted following centrifugation (16,000 x *g*, 10 mins, 4°C). Protein quantification was performed (Pierce BCA Protein Kit) to allow concentration standardisation, and samples were prepared with 4X lithium dodecyl sulfate (LDS) Sample Buffer (Thermo Fisher Scientific) and 20mM DTT (Thermo Fisher Scientific) prior to being boiled at 95°C for 2 mins. Samples were run on a 4-12% Bis-Tris Gel (prepared by Dr Carla Blin, Pollard lab) in MES buffer (Thermo Fisher Scientific) with BioRad Precision Plus Protein Dual Colour standards (5µl). The gel was transferred via electroblotting to a methanol-activated Immobilon-P PVDF membrane (Millipore) in transfer buffer (Tris/Glycine, BioRad) with 20% methanol. Having verified successful transfer with ponceau stain, membranes were blocked for 30 mins in Tris-buffered saline (TBS) with 0.1% Tween 20 detergent (TBST) with 5% milk prior to incubation overnight in primary antibody diluted in TBS-5% milk at 4°C (EGFR (D38B1), #4267, Cell Signalling, 1 in 1000; c-Myc (9E10): sc-40, Santa Cruz Biotechnology, 1 in 500; GAPDH, Invitrogen, 1 in 10,000). The following day, membranes were washed 3x for 5 mins each in TBST prior to incubation in the appropriate secondary antibody diluted in TBS-5% milk at r.t. for 1 h (Goat anti-rabbit Blue 700, BioRad, 1 in 10,000; Goat anti-mouse Blue 520, BioRad, 1 in 10,000). After 3 further washes in TBST, membranes were imaged using the Bio-Rad ChemiDoc Imager. A complete list of antibodies used is provided in Appendix 8.5.

2.9 Imaging

2.9.1 Cell growth assay

Growth curves and confluence analysis were performed using the Incucyte Live Cell Imaging System (Essen Bioscience). Cells were plated in triplicate wells in a 12-well plate at a density of 1-5x10^4 cells per well and left to adhere o/n. Where indicated, cell media was replaced with CM (+EGF/FGF), CM without EGF/FGF, CM without EGF/FGF supplemented with bone morphogenetic protein (BMP, 20ng/ml) or CM without EGF/FGF supplemented with 5% FCS. Wells were imaged hourly until confluence was reached or the experiment was complete.

See section 2.3.1.1 of the Methods where the Incucyte was used during optimisation of metaphase spreads.

2.9.2 Fluorescence Microscopy

Slides were imaged on epifluorescence microscopes (Zeiss AxioImager 2 and Zeiss AxioImager.A1) and the SoRa spinning disk confocal microscope (Nikon CSU-W1 SoRa) (IGC Advanced Imaging Facility, UoE). The only exception were images (Chapter 3.1 and 3.2 and associated Methods Section 2.3) taken with the Nikon Eclipse Ti inverted microscope (CRM Imaging Facility, UoE).

For 3D image analysis in Chapters 4 and 5, images were taken with the SoRa microscope and a 3µm section across each nucleus was imaged in 0.1 µm z steps. Images were denoised and deconvolved using NIS deconvolution software (blind preset or Lucy-Richardson, 10 iterations) (Nikon). 3D images are shown in the figures as maximum intensity projections (MIP) prepared using ImageJ.

2.10 Image analysis

2.10.1 Metaphase spread area

This was devised to optimise the area of metaphase spreads by altering the hypotonic solution and humidity (See Section 2.3.1 – Optimisation of metaphase spreads). An imaging macro was written in ImageJ which uses the DAPI channel to mask and calculate the area of each metaphase spread.

2.10.2 Automated ecDNA counting in metaphase spreads

I used the E37 cell line to perform this analysis as this was the first cell line where the metaphase spread protocol was fully optimised and I was able to generate sufficient metaphase spreads across three passages (P6, P17, P26). A sample size of 50 metaphases per passage number was selected based on conventions from other studies (Turner et al., 2017). E37 metaphase spreads stained with DAPI were imaged at a magnification of 1000x using the DAPI channel, Z stack (0.2µm step size, 17 steps over a range of 3µm) with a single metaphase spread per field of view, with the Nikon Eclipse Ti inverted microscope. An analysis pipeline was devised (Figure 2.6). Deconvolution was performed using Huygens (SVI) Professional with the following settings – Classic Maximum Likelihood Estimate (CMLE) algorithm; point spread function (PSF): auto; signal-to-noise ratio (SNR): 1 in 3; imaging parameters (i.e. excitation wavelength, depth of acquisition as per imaging).

To ensure the macro was accurate, a sample of 6 metaphases per slide were counted both manually and using the macro (Figure 2.7). This suggested that the macro was consistent with manual ecDNA copy number counting.

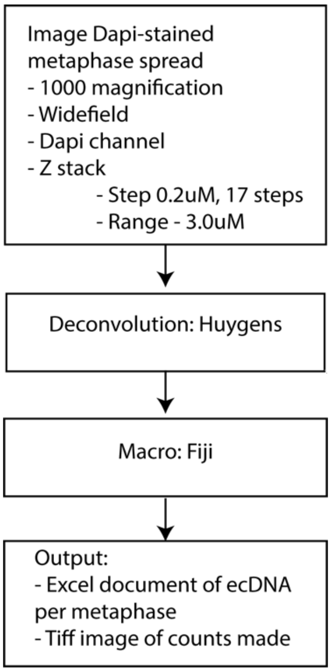


Figure 2.6 | Workflow of ecDNA count automation

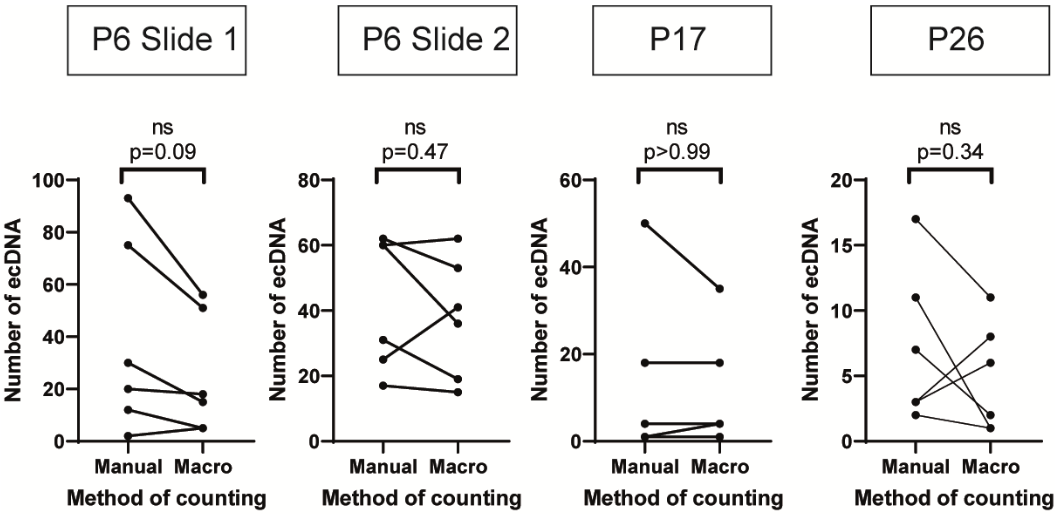


Figure 2.7 | Manual vs macro count of ecDNA.

Statistical test - Wilcoxon matched pairs signed rank test, p = not significant between manual vs macro ecDNA counts.

2.10.3 Signal Intensity Analysis

2.10.3.1 Intensity Analysis

Single slice images were taken with a 10x lens using the Nikon Eclipse Ti inverted microscope, with three fields of view taken per condition per cell line. Scripts were written and performed in ImageJ. The DAPI signal was used to segment and count the number of individual nuclei. The same was then performed for the other relevant channel(s), providing a ratio of number of positive nuclei in a given channel to all DAPI-stained nuclei in the field of view.

2.10.3.2 Nuclear Territory analysis

At least 50 nuclei were imaged per cell line as single-slice images with a 20x lens using the Zeiss AxioImager 2. Nuclear territory analysis scripts were adapted from those previously published and were performed in ImageJ (Croft *et al.*, 1999; Boyle *et al.*, 2001). The script first segments individual nuclei and defines the nuclear area as that occupied by the DAPI signal. This nuclear area was then segmented into a user-defined number of equally sized concentric shells from the centre to the periphery of each nucleus. In the case of the nuclear territory analysis performed here, 5 concentric shells were used. The signal intensity (DAPI, FISH probe or chromosome paint) was calculated for each shell. The FISH probe or chromosome paint signal was normalised to the DAPI signal in each shell.

This was adapted to performed signal intensity analysis of DNA FISH foci following IR (Section 2.6.1.2) by only segmenting the nucleus into one shell. The signal intensity of the FISH signal was normalised to the DAPI signal and size for each individual nucleus.

2.10.4 EcDNA and large RNA Pol II foci

For 3D analysis of nuclei, images were deconvolved as previously described, and all analysis performed on 3D images using Imaris (Oxford Instruments, v9.7). The Spots function in Imaris was used to define foci (RNA FISH, DNA FISH, large RNA Pol II foci), converting a 3D input image into a 3D image of Spots. Using representative deconvolved image(s) as the input tool, the user defines the characteristics of the focus of interest, including the imaging channel and the size (XY diameter) by measuring across a single plane across representative foci. The PSF is derived from the diameter to allow for 3D analysis. Imaris then identifies all foci fulfilling these definitions in the image and plots them as spots. These settings can be run as a batch to define spots across a set of images. The user can adjust the thresholds for individual images to control for varying signal intensity between nuclei, which is a feature of DNA FISH. An example is shown in Figure 2.8.

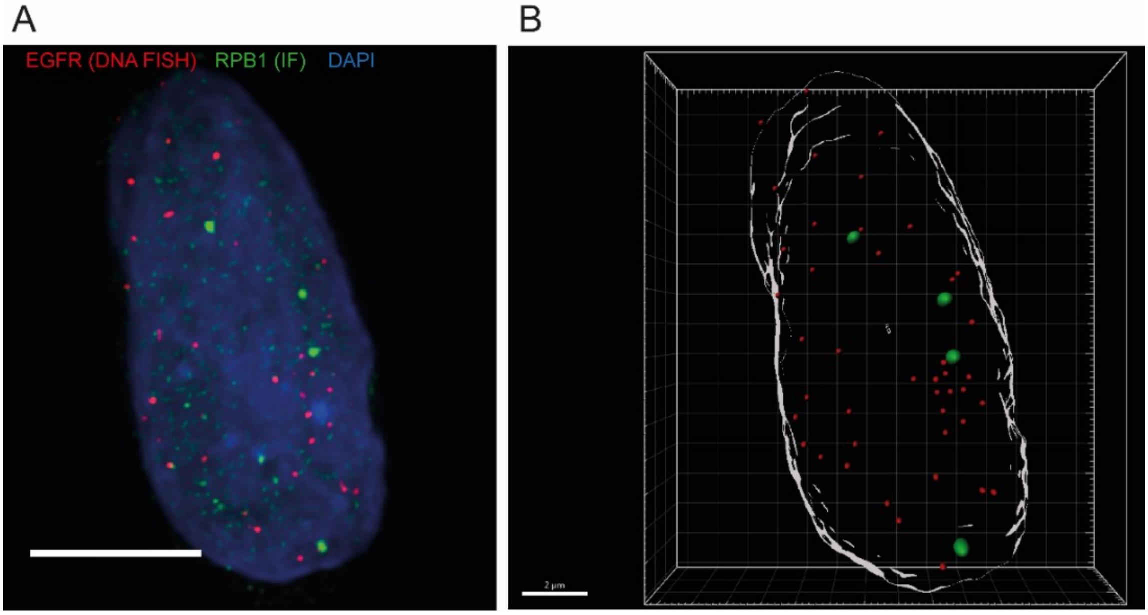


Figure 2.8 | Definition of foci using Imaris

A) DNA immunoFISH of an E26 nucleus: IF for Rpb1 protein (green) and EGFR DNA FISH (red). Scale bar = 5 µm, MIP. B) The same nucleus with foci shown as spots defined using Imaris Spots function. Large Rpb1 Spots $\geq$500nm diameter (green), EGFR DNA FISH Spots = 300nm diameter (red). Edge of nuclear surface = white. Scale bar = 2 µm.

For DNA foci (Chapter 4), the diameter of foci in cell lines E25, E26 and E28 were measured at 300nm, and were also reanalysed with a diameter of 150nm to ensure small foci were not being missed. For E20 and all RNA FISH foci, the diameter was measured at 200nm. Large RNA Pol II foci (RPB1 and POLR2G) were defined as spots of $\geq$500nm diameter (Cho *et al.*, 2018; Sabari *et al.*, 2018). All E26, E28 and NSC foci were defined as being 200nm in diameter in the analysis presented in Chapter 5.

The Spots function was used to count and measure distances between foci. The Spots function also provides the x,y,z coordinates for each Spot which are necessary for 3D Cluster analysis.

2.10.5 Ripley's K Cluster Analysis

The design and code for Ripley's K analysis was conducted in collaboration with Dr Sjoerd Beentjes, from the Department of Mathematics at the University of Edinburgh.

The method for 3D cluster analysis by Ripley's K function is described in Chapter 4 (Section 4.3.2). The script was written in Python (v3.9), using an existing Python Package for Ripley's K calculation (ripleyk 0.0.3). Additional notes on the mathematical basis of Ripley's K function are outlined here.

$$K(r) = \frac{1}{\lambda} \sum_{i \neq j} \frac{\mathbb{I}\{d(i,j) \leq r\}}{n}$$

Ripley's K function compares the number of points at a distance smaller than a given radius r, relative to the average number of points in the volume. This average is the density lambda, in this case the number of foci, n, divided by the volume. In the above equation, II is the indicator function which equals one if the distance between points i and j is no larger than r, and zero otherwise. A high value of Ripley's K function represents clustering at the given radius r, whereas a low value represents dispersion. Consequently, a high Ripley's K function at a given radius is indicative of clustering at this radius. By comparing the observed value of Ripley's K function at a given radius with that computed on the same number of foci and with the same volume but drawn from a uniform null distribution, the presence of significant clustering in the given cluster at the given radius can be detected. (Purshouse *et al.*, 2022b, paragraph 2, Methods 'Image analysis of ecDNA and large PolII foci')

2.10.6 Combined RNA:DNA FISH

For combined RNA:DNA FISH, RNA FISH was performed as described and nuclei imaged with the SoRa microscope and the x,y,z coordinates for each nucleus recorded using NIS. DNA FISH was then performed as described above. The stored x,y,z coordinates were then used to relocate and reimage each nucleus. The nuclei were variable in their morphology, and so it was possible to be confident that the same nucleus was being imaged between RNA and DNA FISH. Nuclei were excluded where this was not the case, or in the case where a nucleus could not be relocated, both of which were rare. Spot counting was performed separately for RNA and DNA FISH foci, with cross correlation only performed with NSC cells to confirm that RNA loci corresponded with genomic loci visualised by DNA FISH.

As the CEN7 probe targets repetitive alpha-satellite regions, occasionally it was difficult to clearly identify, and therefore count, the number of CEN7 foci due to the signal being dispersed. Where this occurred, these nuclei were excluded from analysis.

A complete list of reagents and solutions used is provided in Appendix 8.6.

2.11 RNA and WGS sequencing sample preparation, analysis and processing

2.11.1 Sample preparation

2.11.1.1 GCGR RNA-seq and WGS

Cell lines from the GCGR were prepared and sent for RNA-seq and WGS by the GCGR, and the raw data made available to me for analysis.

RNA was extracted from cells using the QIAGEN RNeasy kit, and prepared with KAPA mRNA Hyper prep kit with KAPA SeqCap Adapters. RIN values were determined by processing RNA samples with Bioanalyser RNA 6000 nano chips, and, following quantification with Qubit RNA Broad Range (BR) Assay, 200ng of RNA was used per sample. Fragmentation was performed at 94°C for 6 mins with a library amplification of 12 cycles. Each library was quantified with the Qubit dsDNA High Sensitivity (HS) assay, and the average fragment size calculated with Bioanalyser DNA 1000 or DNA HS. The molarity of each library was calculated and normalised to approximately 10nM. Library pools were generated, containing 24 libraries per pool, and quantified/fragment sizes calculated as above. Sequencing was performed on the Illumina HiSeq 2500 by UCL Genomics, with “Dilute and Denature Libraries for cBot Clustering” performed as per the manufacturer’s instructions. A standard normalisation method was used. Each 24-library pool was run on two lanes of a HiSeq High Output flow cell.

WGS of blood and tumour samples was performed by Beijing Genomics (BGI) Tech Solutions with PE100 sequencing and normal library construction. At least 180Gb of paired-end data was generated per sample. For E26 and E28 analysis in Chapter 4, sequences from repeat samples were merged as follows: E26 tumour (11 samples), E28 tumour (18 samples), E26 blood (14 samples), E28 blood (10 samples).

2.11.1.2 WGS of CRISPR-Cas9 samples

Shallow WGS of all cell lines generated in Chapter 5 was performed by the Tumour Genome Analysis Core (TGAC), Cancer Center Amsterdam (CCA), Amsterdam UMC with kind assistance from Prof Bauke Ylstra. Samples were prepared for WGS using the QIAGEN DNEasy kit (Qiagen, 69504) and quantified using the Qubit dsDNA High Sensitivity (HS) assay. 250ng of DNA from each cell line was diluted in 35µl of elution buffer (10 mM Tris-HCl, pH 8.5). The library construction protocol included Roche KAPA Hyper Plus library prep with Roche dual adaptors. Sequencing was performed on the NovaSeq S4 300, with at least 50Gb and 167M reads per samples. Copy number profiles were prepared by the TGAC and transferred, along with raw data (paired-end WGS fastq files).

2.11.1.3 Other sequencing data

Data (WGS, RNA-seq, AmpliconArchitect) for the cell line GBM39 (Chapter 4) was available via publication and the NCBI Sequence Read Archive (BioProject: PRJNA506071) (Wu *et al.*, 2019).

2.11.2 Sequence alignment and visualisation

WGS and RNA-seq sequences were aligned to human genome assembly 38 (hg38) with STAR 2.7.1a with the following settings '--outFilterMultimapNmax 1'. The additional settings '--alignMatesGapMax 2000 --alignIntronMax 1 --alignEndsType EndToEnd' were used for WGS data (Dobin *et al.*, 2013). Picard (Broad Institute) was used to remove duplicate reads. Aligned sequences were converted to bigWig format with deepTools bamCoverage and '--normalizeUsingRPKM' within the settings (Ramírez *et al.*, 2016). BigWig files were visualised within the UCSC genome browser (Kent *et al.*, 2002).

For data presented in Chapter 4 (E26, E28), raw WGS sequencing files were merged using the 'cat' command prior to processing and analysis.

For GBM39 WGS and RNA-seq data (Chapter 4), files were downloaded from the NCBI BioProject using the NCBI SRA Toolkit v2.10.8 and sequence alignment performed as outlined above. AA data for GBM39 (Wu *et al.*, 2019) was aligned to hg19, so this was converted to hg38 using UCSC LiftOver (Lift Genome Annotations (ucsc.edu)).

2.11.3 AmpliconArchitect (AA)

AA (Deshpande *et al.*, 2019) was used to predict ecDNA regions (and any breakpoint coordinates), classify amplicons and, where required, classify exons as being primarily on ecDNA or only on chromosomal DNA. This was conducted via the AmpliconSuite pipeline, which also required AmpliconClassifier (Kim *et al.*, 2020; Luebeck *et al.*, 2022), CNVkit (Talevich *et al.*, 2016), BWA MEM (Li and Durbin, 2009), samtools (Li *et al.*, 2009) and additional python packages (pandas, Bio, reportlab, pomegranate, DNAcopy [via R]). The input to this pipeline was paired-end WGS fastq files.

Note that the AA coordinates in Chapter 4 were generated by Alhafidz Hamdan (CRM) on merged WGS data (E26 tumour (11 samples), E28 tumour (18 samples)) aligned using BWA (Li and Durbin, 2009) prior to analysis via AA and AmpliconClassifier (Deshpande *et al.*, 2019; Kim *et al.*, 2020). I performed all other AA analysis performed in this thesis, including all AA analysis in chapters 3 and 5, where all predicted amplicons are shown unless stated otherwise.

In Chapter 5, CNV regions as called by AA/CNVkit were plotted using scripts written in Python (v3.9), with copy number per CNV region visualised using seaborn (Waskom, 2021), and copy number of amplicon regions visualised in Graphpad Prism v9.0.

2.11.4 EcDNA EGFR RNA-seq/DNA transcription analysis

The coordinates for EGFR (isoform:EGFR-201) exons were extracted from Ensembl (Ensembl Transcript ID: ENST00000275493.7). The coordinates of E26 ecDNA amplicon break points were obtained via AA. Read counts of WGS and RNA in EGFR exons, or read counts of WGS in AA-defined blocks, were performed using HOMER2 (Heinz *et al.*, 2010) makeTagDirectory and annotation of this Directory was performed with AnnotatePeaks.pl (settings '-len 0 -size given').

For WGS read counts defined by AA blocks, WGS read counts were normalised by AA block size. For comparison with RNA-seq *EGFR* exon read counts, the RNA-seq read count for each *EGFR* exon was also normalised for exon size. This exon-size normalised RNA-seq read count was then normalised to the AA-block-size normalised WGS read count.

Direct comparison of *EGFR* exon read counts negated the need to normalise for exon or AA-block size. The RNA-seq read count was normalised to the WGS read count for each exon and plotted according to whether the exon was chromosomal (exons 2-7) or predominantly extrachromosomal (exons 1, 8-28).

Analysis of both size-normalised and direct RNA-seq counts per copy number was performed using scripts written in Python (v3.9). Statistical testing was performed in Graphpad Prism v9.0.

2.11.5 SNP analysis

SNP calling and associated analysis was performed by Dr Elias Friman (MRC HGU) using strelka v2.9.10 (Kim *et al.*, 2018). The configureStrelkaGermlineWorkflow.py command was separately used on all samples for both E26 and E28 cell lines (merged WGS blood, merged WGS tumour, and RNA-seq tumour). The bcftools (Danecek *et al.*, 2021) package was used to extract SNPs. SNPs were selected if they had an allele frequency between 40-60%, indicating that they were derived from a single parental homolog. Exon-overlapping SNPs were identified and the ratio of allele frequencies between the RNA-seq and WGS tumour samples determined (for SNPs with at least 20 RNA-seq reads). The ratio of RNA-seq to WGS reads for each SNP was calculated to give a ratio of allele frequencies for each SNP.

Analysis was performed using scripts written in Python (v3.9), with SNP allele frequency visualised using seaborn (Waskom, 2021), and the ratio of allele frequencies visualised in Graphpad Prism v9.0.

2.12 Statistical Analysis

Unless otherwise stated, statistical analysis was performed with Graphpad Prism v9.0 and the specific statistical tests used are described in the relevant sections of the thesis. For figures, p-value significance is denoted as follows: * < 0.05, ** < 0.01, *** < 0.001, **** < 0.0001. To avoid significant results due to multiple hypothesis testing, Bonferroni correction was performed where appropriate and the corrected p values used.

A list of all software and bioinformatics programmes used is provided in Appendix 8.7.

Chapter 3 Characterising ecDNA in glioblastoma stem cell lines

3.1 Introduction

Glioblastoma is a cancer characterised by highly heterogeneous stem cell-like cells which can host a range of dominant oncogene amplifications. Glioblastoma harbours transcriptionally-distinct subtypes, defined not only by their dominant oncogenes but also by developmental cell state, immune profile and injury response (Verhaak *et al.*, 2010; L.-B. Wang *et al.*, 2021; Richards *et al.*, 2021). *EGFR* (chr7), the most commonly altered oncogene, is altered or amplified in around two-thirds of glioblastoma, with other commonly amplified genes including those derived from chr4 (*PDGFRA*), chr7 (*MET, CDK6*) and chr12 (*CDK4, MDM2*) (Snuderl *et al.*, 2011; Szerlip *et al.*, 2012; Brennan *et al.*, 2013; Kim *et al.*, 2020). A major mechanism for this oncogene amplification is ecDNA (Brennan *et al.*, 2013).

EcDNA are particularly common in glioblastoma, and are able to contribute to glioblastoma heterogeneity via their highly variable and dynamic copy number (Nathanson *et al.*, 2014; Turner *et al.*, 2017). The frequency of ecDNA in glioblastoma is also reflective of the high frequency of chromothripsis, a major initiating event in ecDNA formation, in glioblastoma cells (Malhotra *et al.*, 2013; Zack *et al.*, 2013; Cortés-Ciriano *et al.*, 2020). EcDNA have been observed to be most frequent in patient-derived cells in comparison with established cell lines, and therefore patient-derived cells likely represent the most biologically relevant model for the study of ecDNA (Benner, Wahl and Von Hoff, 1991; Turner *et al.*, 2017). Patient-derived glioblastoma xenograft (GBM39) and recurrent, pre-treated glioblastoma (HK359) cell lines both harbour *ecEGFRvIII* and have been used to study ecDNA dynamics *in vitro* (Nathanson *et al.*, 2014; Turner *et al.*, 2017; Hung *et al.*, 2021). Further, primary glioblastoma cells have been shown to retain their molecular profile when taken into cell culture (deCarvalho *et al.*, 2018). Patient glioblastoma stem cells are therefore most likely to recapitulate glioblastoma pathobiology and genetic heterogeneity (Pollard *et al.*, 2009).

The GCGR comprises a library of glioblastoma and NSC lines that have been characterised by copy number profile, indicating that these cells harbour a range of key oncogene amplifications previously described in the glioblastoma genomic landscape ((Verhaak *et al.*, 2010; Brennan *et al.*, 2013), See Introduction, Figure 1.1). Given that amplified oncogenes in glioblastoma are routinely found on ecDNA, this suggests that cells within this collection will harbour oncogene-resident ecDNA.

This chapter describes the characterisation of ecDNA in five glioblastoma stem cell lines using both imaging and sequencing analysis tools. I use metaphase spread analysis to describe fundamental characteristics of ecDNA copy number. I then use DNA FISH, WGS and AA analysis to describe the oncogene and ecDNA landscape of the five glioblastoma cell lines which will form the central tools for onward study.

3.2 EcDNA copy number variation in glioblastoma cells

First, I generated metaphase spreads to characterise ecDNA copy number variation in glioblastoma cells. These proved very challenging to generate in many of the selected cell lines as optimised protocols were needed for each independent line due to their distinct cell biological features – see Methods Section 2.3.1 for further details.

E37 cells generated sufficiently good metaphase spreads for analysis, with clearly defined chromosome morphology. EcDNA could be visualised around chromosomes as distinct small DAPI-stained spots (Figure 3.1A). EcDNA copy numbers per metaphase spreads showed a wide variation and non-Gaussian distribution of ecDNA copy number, as would be expected (Figure 3.1B).

It has previously been reported in a study of GBM39 cells that ecDNA copy number decreases with increasing cell passage, possibly through chromosomal integration (Turner et al., 2017). I therefore aimed to characterise ecDNA copy number across increasing passages (P6, P17, P26) of E37 cells as I wanted subsequent experiments to be at a passage number where ecDNA copy number is maintained.

I developed an ecDNA-counting macro (see Methods 2.10.2) to enable counting of ecDNA across >50 metaphase spreads per passage number. This demonstrated two key features of ecDNA. First, the distribution of ecDNA copy number is highly variable and non-Gaussian between metaphase spreads across all passages. Second, ecDNA number decreases with increasing passage number, although further experiments are needed to determine whether this represents chromosomal integration (Figure 3.2).

Next, I characterised ecDNA-resident oncogenes. A subset of glioblastoma stem cell lines were selected for further study based on previously generated CNV data which predicted particular genomic amplifications of interest (Figure 1.4).

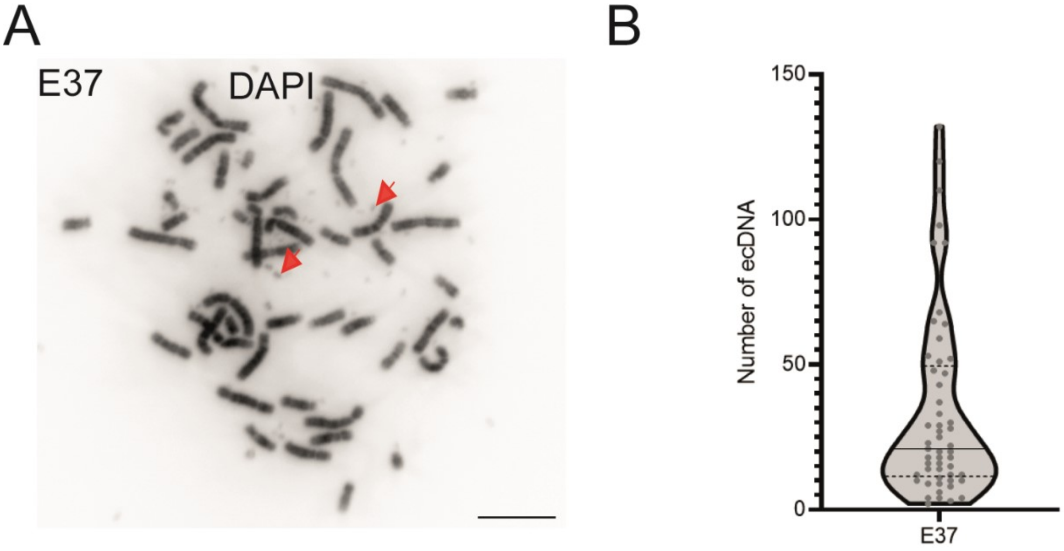


Figure 3.1 | E37 metaphase spreads and ecDNA copy number

A) Representative image of E37 metaphase spread with red arrows indicating ecDNA. Colcemid 0.1ug/ml for 30 mins. DNA = DAPI (greyscale). Scale bar = 10 µm. B) Violin plot of number of ecDNA per metaphase spread in E37 cells, median and quartiles shown. Number of metaphase spreads = 53. Biological replicates = 1.

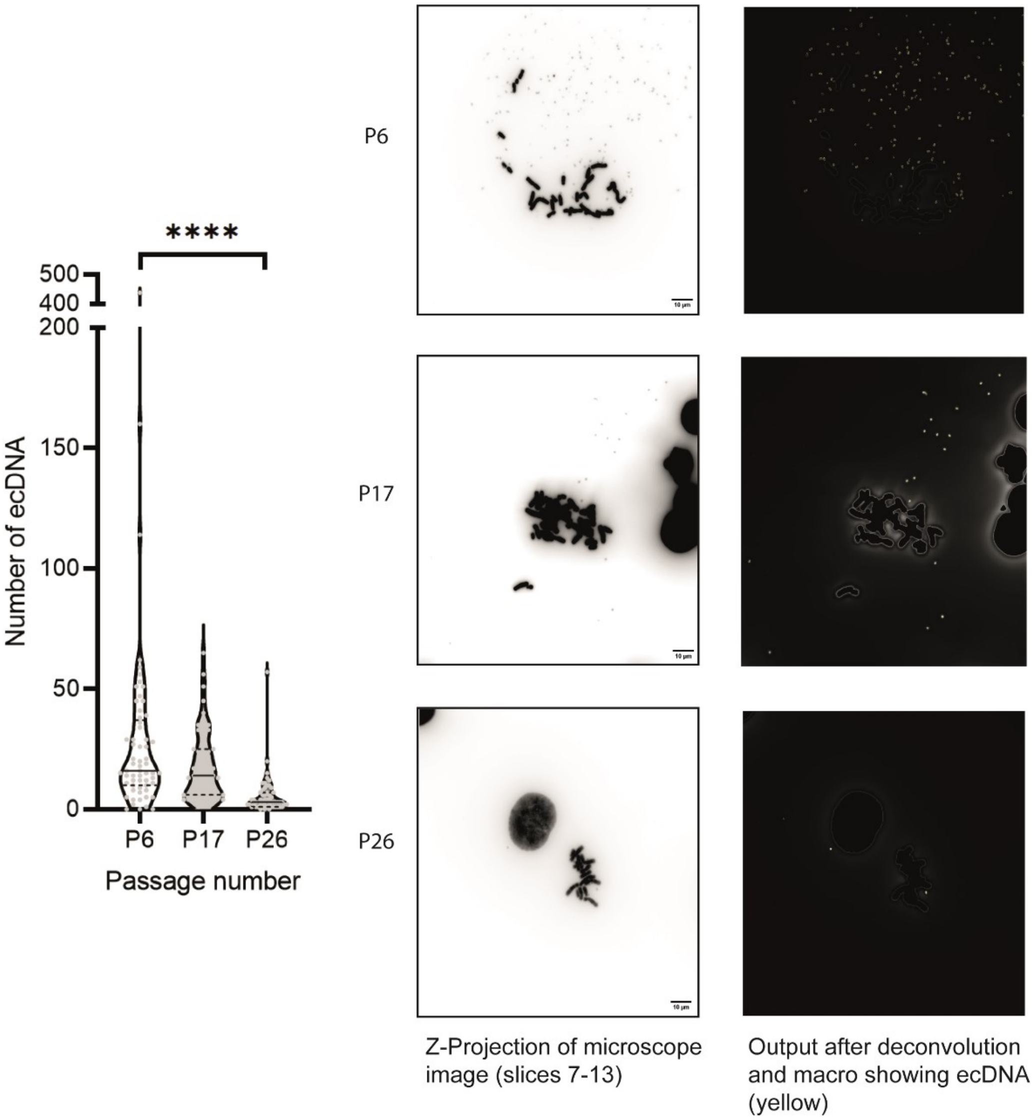


Figure 3.2 | Dynamics of ecDNA copy number changes with passage.

Violin plot of number of ecDNA per metaphase spread, showing median and quartiles over increasing passage (P) number (left panel). Median number of ecDNA per metaphase spread: P6 = 16, P17 = 14, P26 = 3. Statistical test – Kruskall-Wallis, p<0.0001. Number of metaphases counted – P6 = 61, P17 = 59, P26 = 55. Shown with representative images of metaphase spreads with the projection (left) and output (right) of imaged metaphase spreads by the ecDNA counting macro. Scale bar = 10 μm. Biological replicates = 1.

3.3 Primary glioblastoma cell lines harbouring *EGFR* ecDNA

CNV profiles suggested two primary glioblastoma cell lines harbour *EGFR* ecDNA: E26 and E28. Bulk WGS data, available for these cell lines, were analysed using AA, which predicted the presence of a single circular *EGFR*-containing amplicon in each cell line. In E26, ecDNA are 1,558,246bp (~1.5Mb) in size (Figure 3.3A) and in E28, 846,968bp (~850kb) (Figure 3.3B). E26 harbours an *EGFR* exon 2-7 deletion indicative of *EGFRvIII* and both amplicons harbour other genes from adjoining regions of chromosome 7 in addition to the oncogene *EGFR.* E28 also appears to have a subpopulation harbouring a deletion in *EGFR* exons 7–14, as indicated by a trough in the WGS plot (Figure 3.3C – see arrow). Similar to *EGFRvIII*, this would be expected to lead to shortening of the extracellular ligand-binding domain of EGFR, with possible activation of constitutive EGFR signalling.

EcDNA in these cell lines were further characterised by DNA FISH on metaphase spreads. This verified the presence of *EGFR* on ecDNA in E26 (Figure 3.4A) and E28 (Figure 3.4B) cells. In addition, E26 harboured *EGFR* HSRs on approximately 10% of metaphase spreads (Figure 3.4A), suggesting that AA analysis is unable to differentiate between ecDNA and chromosomal (HSR) amplicons. E26 had a higher number of *EGFR* foci per metaphase spread than E28, and the distribution of ecDNA copy number was again widely distributed and positively skewed (Figure 3.4C). Given that *EGFR* is located on chromosome 7 – which is frequently amplified in glioblastoma – we characterised this using a chromosome 7 paint. E26 had 3-6 copies, and E28 frequently 3 copies, of chromosome 7 per metaphase spread (Figure 3.4D).

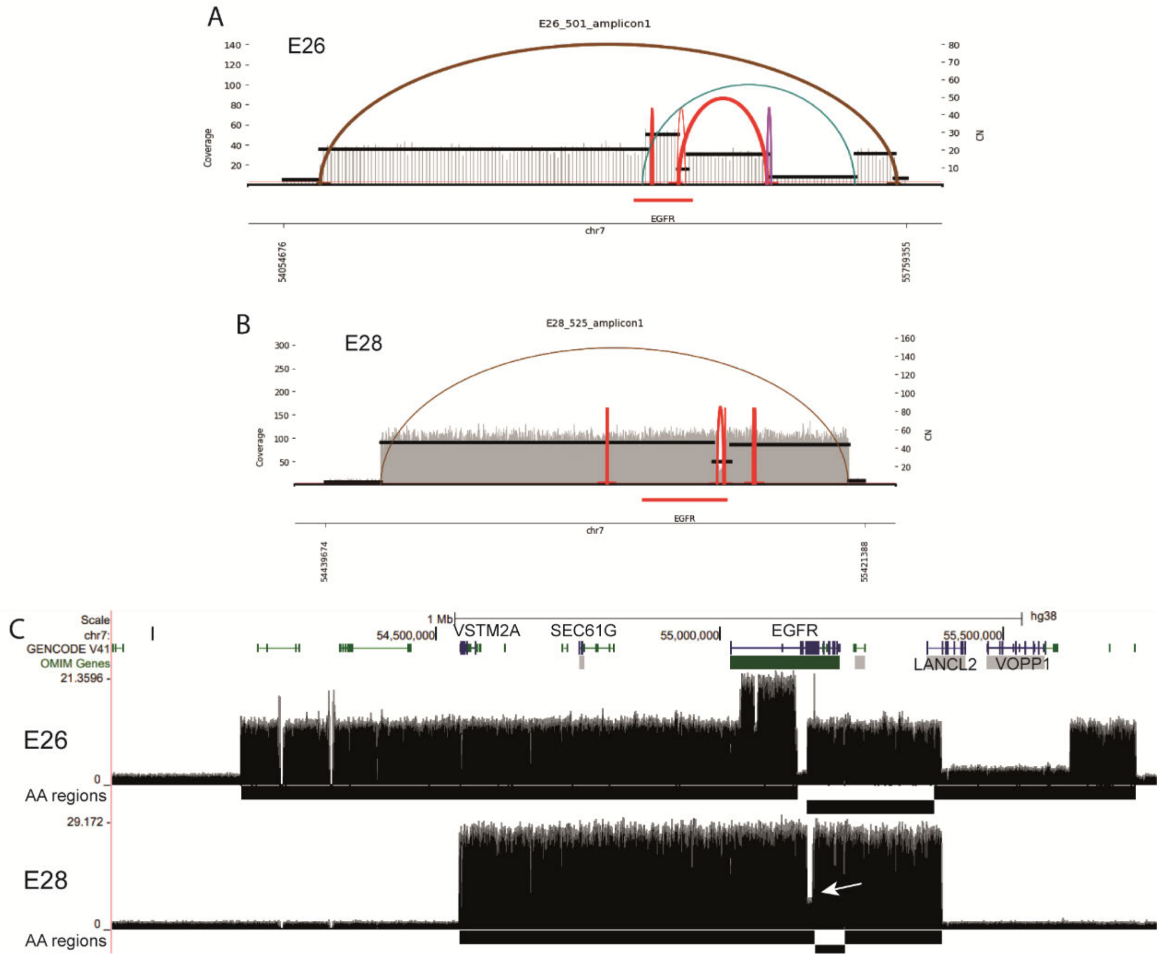


Figure 3.3 | AA and WGS analysis of E26 and E28 cell lines

A) AA predicted amplicon for E26, showing breakpoint edges and strands of the edge (arcs). EGFR gene is highlighted. B) As for A) but for E28. C) UCSC Genome Browser showing WGS and AA regions for E26 and E28. White arrow indicates region of exon 7-14 deletion in a proportion of E28 cells.

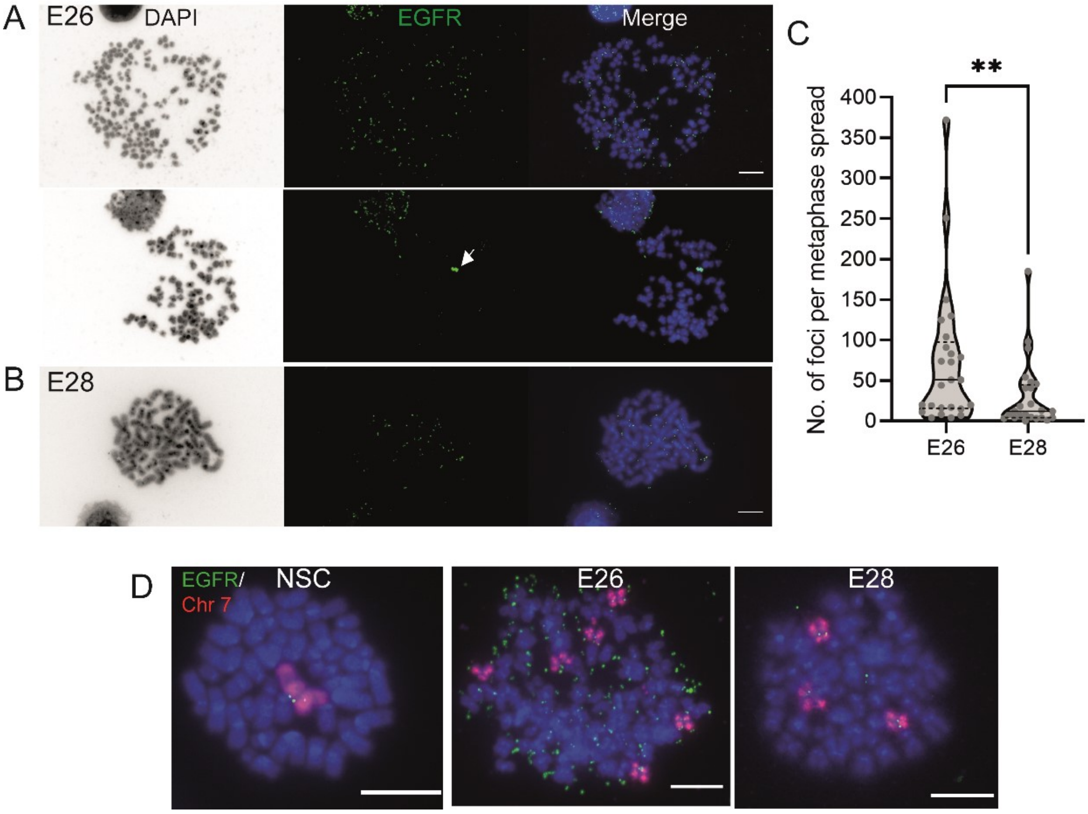


Figure 3.4 | *EGFR* DNA FISH of E26 and E28 metaphase spreads

*A) Representative DNA FISH of metaphase spread, E26 cells showing EGFR (green) on ecDNA, and on an HSR (white arrow). DNA = DAPI (greyscale, then blue). Scale bar = 10 µm. B) As for A) but for E28. C) Violin plot, number of EGFR DNA FISH signals per metaphase spread (median and quartiles shown). ** p=0.008 (Mann-Whitney test). D) Representative DNA FISH images of metaphase spreads for NSC, E26, and E28 cells showing signals for chromosome 7 (red) and EGFR (green). DNA = DAPI (blue). Scale bar = 10 µm. Adapted from Purshouse et al., 2022b. Biological replicates = 1.*

3.4 Primary glioblastoma cells harbouring *CDK4* and *PDGFRA* ecDNA

3.4.1 Analysis of E25 cells

Analysis of bulk WGS data for E25 with AA predicted a single ecDNA amplicon, although further analysis of the amplicon plot suggests that multiple different amplicons exist within this ecDNA region, which are not all interconnected. One distinct region (547,394bp (~500kb)) is comprised of six fragments from chromosome 4, including the *PDGFRA* oncogene. Another region (2,388,924bp (~2.4Mb)) is comprised of 10 fragments from chromosome 12, including the *CDK4* and *MDM2* oncogenes; however, closer evaluation of the segment arcs and copy number profiles suggests the majority of *CDK4* and *MDM2* oncogenes are located on separate ecDNA (Figure 3.5A).

Visualising the WGS copy number profiles against their predicted AA regions highlights the highly complex nature of ecDNA (Figure 3.5B). This suggests ecDNA contain fragments from multiple chromosomal regions rather than a single continuous fragment, supporting the chromothripsis hypothesis of ecDNA formation.

DNA FISH for *CDK4* and *PDGFRA* confirmed their localisation on separate ecDNA (Figure 3.6A), with ecDNA copy number widely distributed and more *CDK4* than *PDGFRA* foci (Figure 3.6B). It proved challenging to identify a DNA FISH probe targeting *MDM2* that could give a clean hybridisation signal. As the aim was to identify a cell line with ecDNA on separate populations in order to study ecDNA clustering, I did not pursue characterisation of *MDM2* ecDNA by DNA FISH as this was not essential for my hypothesis (Chapter 4).

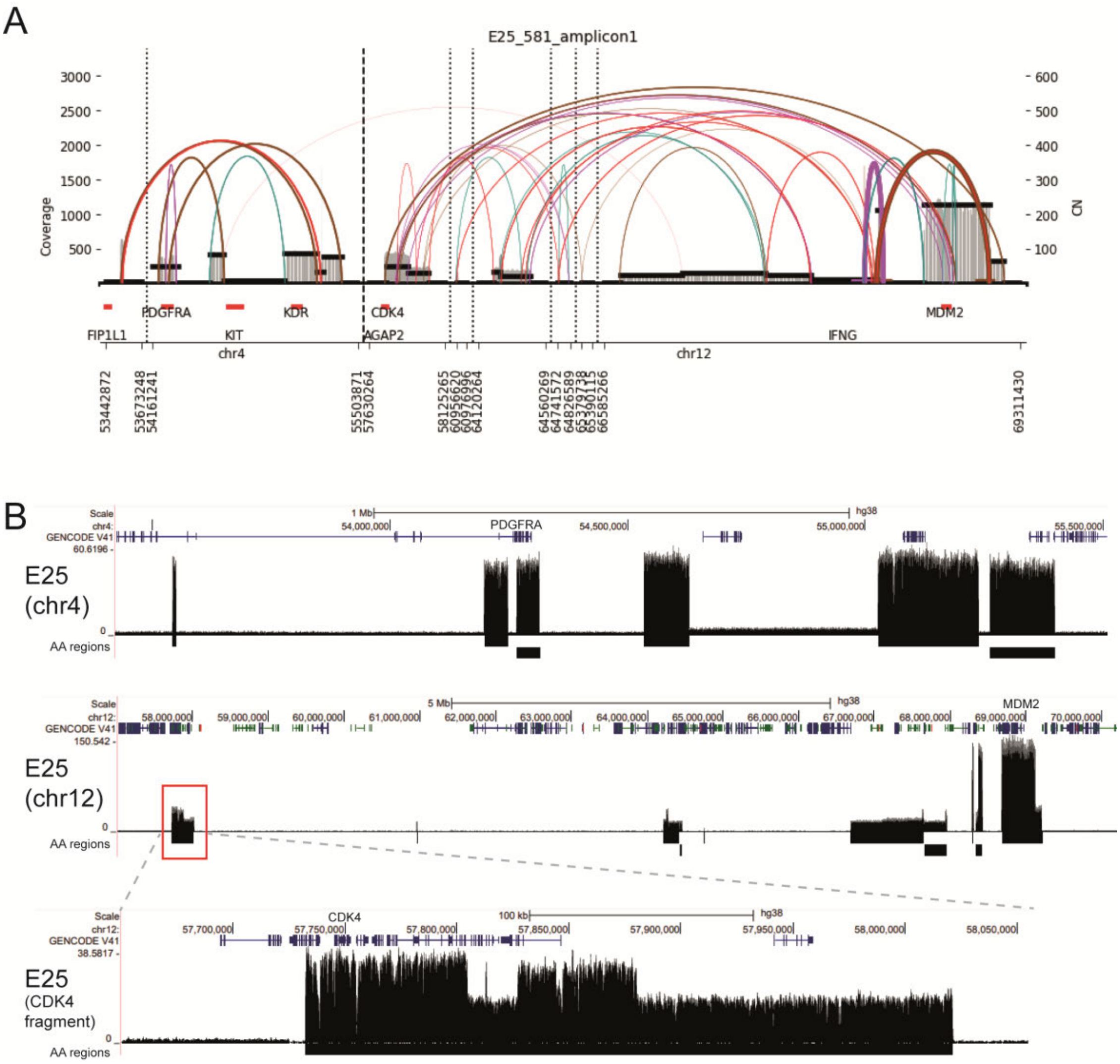


Figure 3.5 | AA and WGS analysis of E25 cell line

A) AA predicted amplicon for E25, showing breakpoint edges and strands of the edge (arcs). PDGFRA, CDK4 and MDM2 genes are highlighted. B) UCSC Genome Browser showing WGS and AA regions for E25.

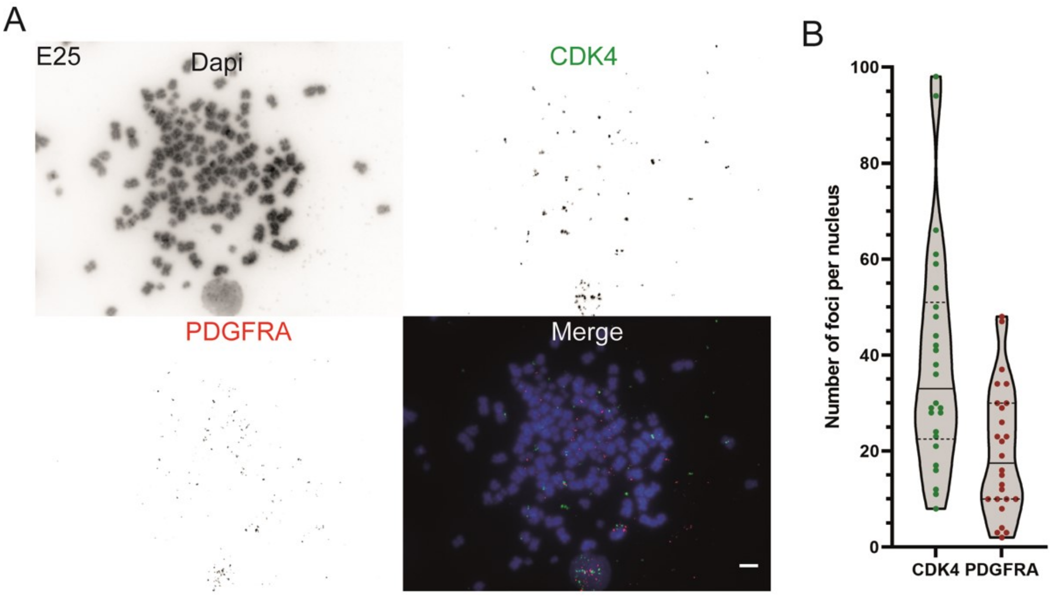


Figure 3.6 | DNA FISH of E25 metaphase spreads

A) Representative DNA FISH of metaphase spread, E25 cells showing CDK4 (greyscale, then green), and PDGFRA (greyscale, then red), on separate ecDNA. DNA = DAPI (greyscale, then blue). Scale bar = 5 µm. B) Violin plot, number of EGFR DNA FISH signals per nucleus (median and quartiles shown). Adapted from Purshouse et al., 2022b. Biological replicates = 1.

3.4.2 Analysis of E20 cells

I also explored an independent GSC line harbouring two separate oncogene-ecDNA populations to provide a second tool for visualising ecDNA colocalisation in nuclei. E20 is another cell line where CNV data suggested the presence of *CDK4* and *PDGFRA* oncogene amplifications. AA predicted two amplicons. The first (3,025,039bp (~3Mb)), derived from chromosome 4, harboured the *PDGFRA* oncogene, and this is called as a linear amplicon. The second is a more complex circular amplicon (1,269,074bp (~1.3Mb)) containing 7 chromosome 12 fragments, which includes *CDK4*. There are unclassified amplifications on chromosome 7 which may be involved with this second amplicon (Figure 3.7A). WGS profiles plotted alongside the predicted AA regions once again highlighted the complexity of chromosomal amplifications contained within each ecDNA (Figure 3.7B).

DNA FISH of E20 metaphase spreads using *CDK4* and *PDGFRA* probes confirmed that when two ecDNA populations were present, most ecDNA were located on separate ecDNA (Figure 3.8A). However, interestingly ~10% (4/38) metaphase spreads harboured colocalised *CDK4* and *PDGFRA* signals, suggesting these are located on the same structure (Figure 3.8B, insets a and b). This was unexpected, given that colocalisation was not predicted by bulk WGS and AA analysis, suggesting the presence of a subclone that could not be identified via these computational methods.

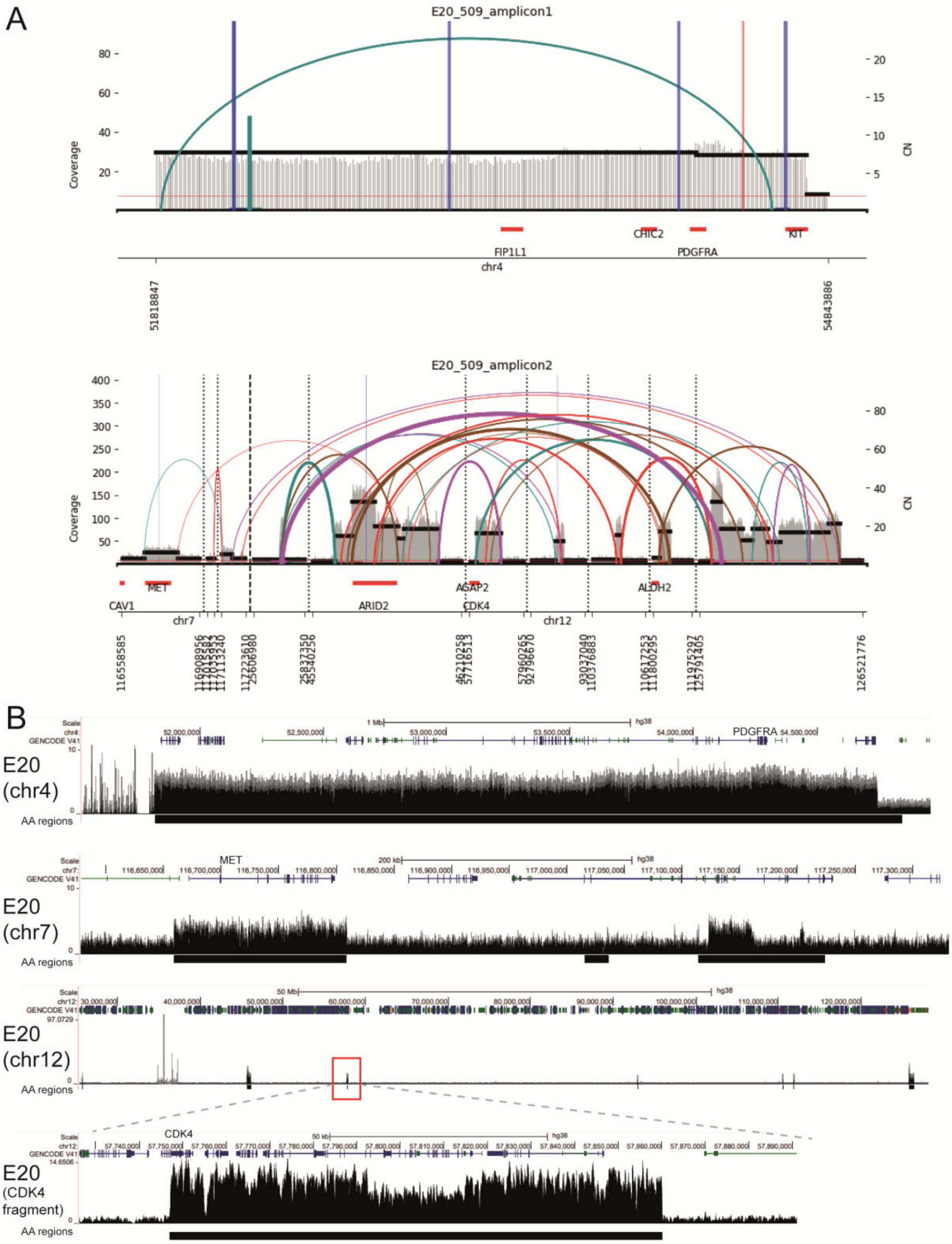


Figure 3.7 | AA and WGS analysis of E20 cell line

A) AA predicted amplicon for E20, showing breakpoint edges and strands of the edge (arcs). CDK4 (top) and PDGFRA (bottom) genes are highlighted. B) UCSC Genome Browser showing WGS and AA regions for E20. Note the peaked regions shown in E20 (chr4) and E20 (chr12) where there are no matched AA region are centromeric areas.

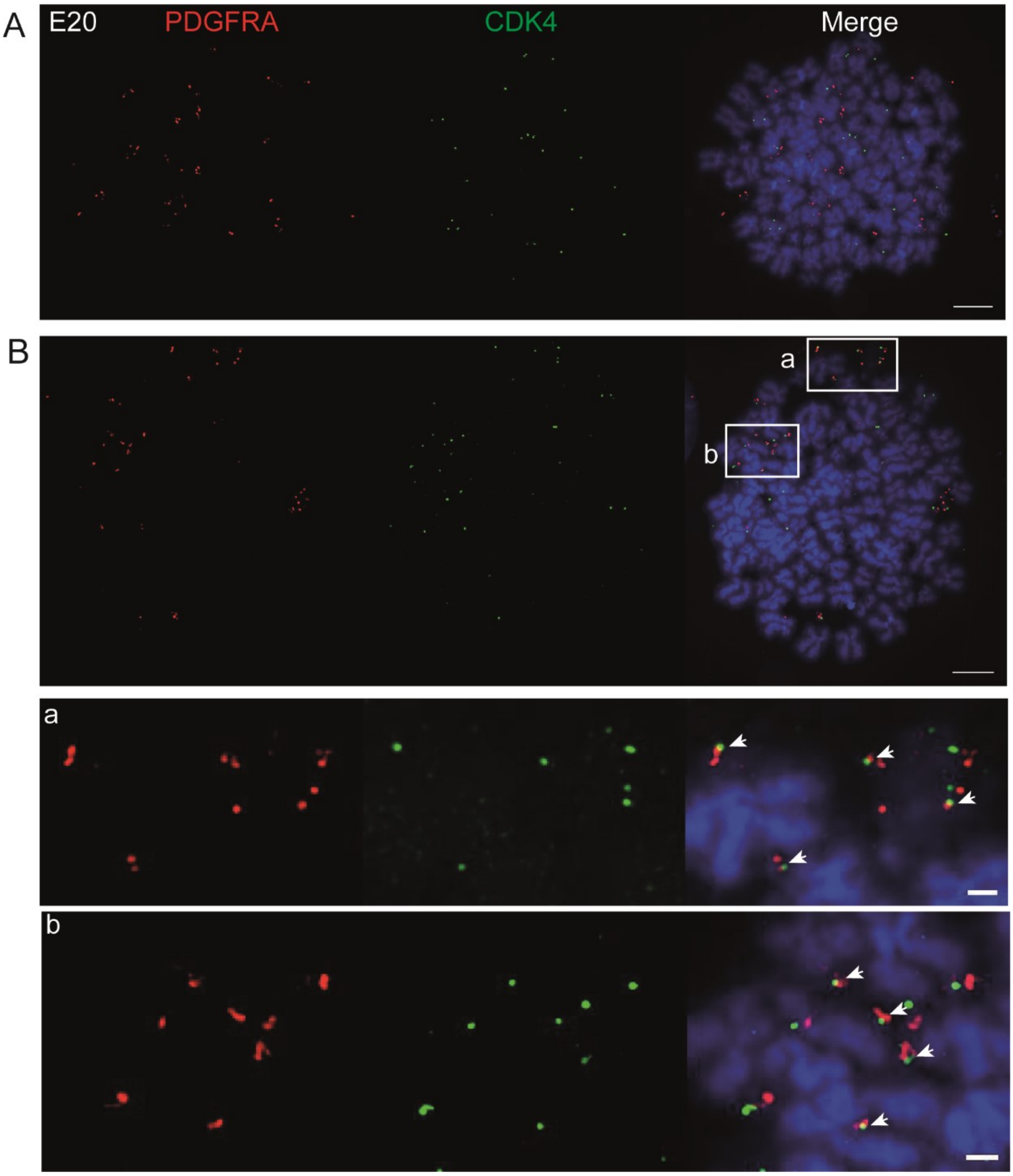


Figure 3.8 | DNA FISH of E20 metaphase spreads

A) Metaphase spread representative of most cells (34/38) with PDGFRA (red) and CDK4 (green) clearly on separate ecDNA. DNA = DAPI (blue) B) Metaphase spread showing an example of colocalised PDGFRA and CDK4 signals found in 4/38 metaphase spreads. Scale bar = 10 µm. Insets a) and b) are shown zoomed in (scale bar = 1 µm). Biological replicates = 1.

3.5 Recurrent glioblastoma cell line harbouring *EGFR* and *MYC* ecDNA

E37 is derived from a patient with recurrent glioblastoma and CNV predicts *MYC* and *EGFR* amplifications. AA analysis of bulk WGS from E37 DNA suggest two major amplicons. A 1,827,057bp (~1.8kb) circular amplicon derived from chromosome 8 harbours the *MYC* oncogene along with *PVT1*. A second amplicon (742,504bp (~750kb)) is a chr7/chr8 hybrid, which includes *EGFR* (Figure 3.9A). AA calls both of these amplicons as circular ecDNA. Note that the two chr8 regions are ~25Mb apart (Figure 3.9B).

DNA FISH on metaphase spreads confirm that *MYC* is located on ecDNA. In contrast, *EGFR* appears to be solely chromosomal, with an amplification in keeping with an HSR (Figure 3.10).

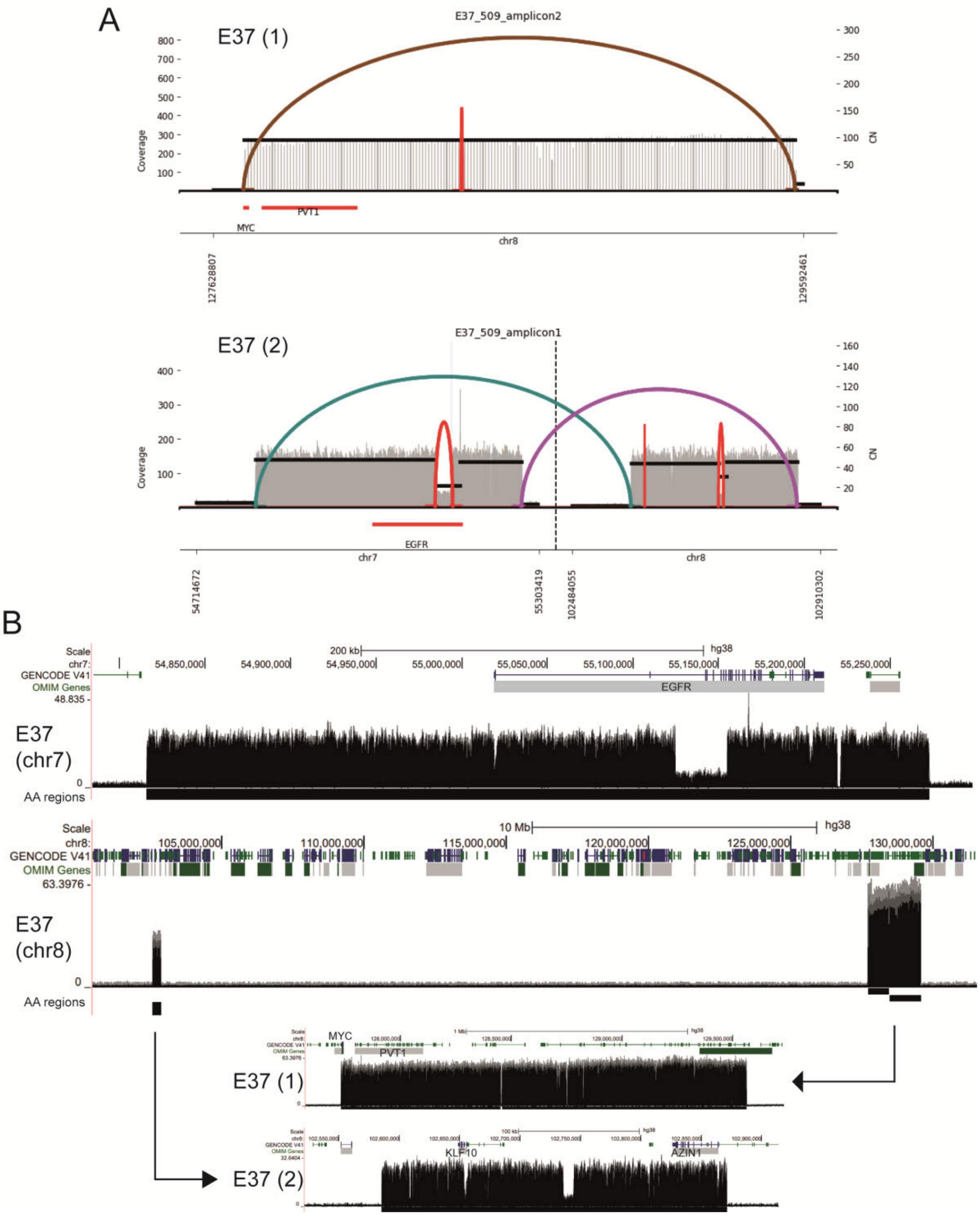


Figure 3.9 | AA and WGS analysis of E37 cell line

A) AA predicted amplicon for E37, showing breakpoint edges and strands of the edge (arcs). MYC and EGFR genes are highlighted. B) UCSC Genome Browser showing WGS and AA regions for E37.

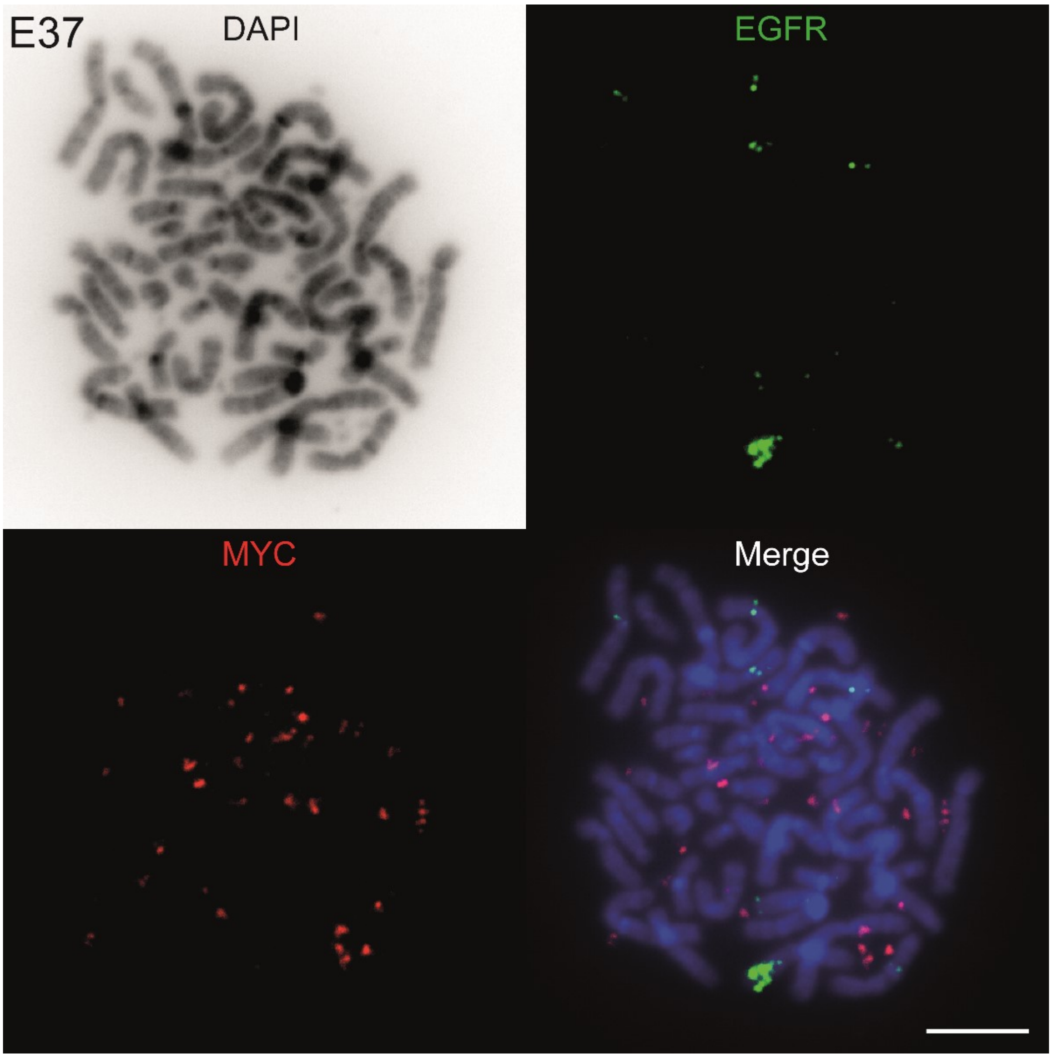


Figure 3.10 | DNA FISH of E37 metaphase spreads

A) Metaphase spread showing MYC (red) on ecDNA, and EGFR (green) on chromosomes, notably as an HSR amplification. DNA = DAPI (greyscale, then blue). Scale bar = 10 µm. Biological replicates = 1.

3.6 Discussion

EcDNA are a common feature of cancer, with significant variation in cell-to-cell ecDNA copy number (Turner *et al.*, 2017). Given that ecDNA harbour oncogenes and regulatory elements such as enhancers, this results in significant intra-tumoural copy number heterogeneity (Turner *et al.*, 2017; Kim *et al.*, 2020), which in turn evolves over the course of tumour evolution (Lange *et al.*, 2022). I observed a similar variation in the distribution of ecDNA copy number in the glioblastoma cell lines evaluated here, both within, and between, cell lines. EcDNA lack centromeres and therefore during cell division their segregation to daughter cells is random. The heterogeneity this facilitates renders ecDNA-harbouring cells highly adaptable to selection pressures (e.g. chemotherapy) (Lundberg *et al.*, 2008; Lange *et al.*, 2022). I showed that ecDNA copy number decreases with increasing passage number. This is perhaps an unexpected result given ecDNA are proposed to confer a survival advantage. This could be explained by ecDNA integration, which has been previously proposed in the GBM39 cell line where a loss of ecDNA across passage number was also observed (Turner *et al.*, 2017), or may be an artefact of 2D culture in the absence of selection pressures such as chemotherapy or metabolic environment (Lundberg *et al.*, 2008; Lange *et al.*, 2022). Although the methods utilised here cannot confirm this and were only performed in a single cell line, this does highlight the dynamic nature of ecDNA copy number.

Glioblastoma most commonly features oncogene amplifications including *EGFR*, *CDK4* and *PDGFRA* *(Brennan et al., 2013)*. Here, CNV data was available for primary and recurrent glioblastoma cell lines, allowing selection of cell lines with these key oncogene amplifications. WGS and DNA FISH analysis confirmed that all key oncogenes analysed here were located on ecDNA, with the exception of *EGFR* in E37 which was located on an HSR. This is in agreement with previous ecDNA-oncogene analysis, where all oncogene amplifications were shown to exist on ecDNA and/or HSRs (Turner *et al.*, 2017). Although tumour tissue samples for these cell lines were not available, previous studies comparing the genomic profile of tissue samples and subsequently derived cell cultures observed gross similarities between the two, suggesting these cell cultures are likely to reflect the originating tissue (deCarvalho *et al.*, 2018). In addition, WGS and AA of these cell lines demonstrates the multi-fragmented nature of some ecDNA, with ecDNA in some cell lines (E20, E25) comprising multiple fragments of the same, or different, chromosomes. This might be expected in the context of the chromothripsis hypothesis of ecDNA formation, which suggests that a chromosome shattering event is at the origins of ecDNA formation (Rosswog *et al.*, 2021; Shoshani *et al.*, 2021). In contrast, other cell lines harboured ecDNA that were simpler in structure, such as the E28 amplicon, which may be more indicative of the episome model (Vogt *et al.*, 2004; Storlazzi *et al.*, 2006; Hung *et al.*, 2022). These data suggest that more than one mechanism may be involved in ecDNA formation.

These results demonstrate some of the limitations of both bulk WGS and AA as analysis tools of ecDNA. Bulk WGS represents all the DNA in a given sample, and individual reads from short-read sequencing are 50-300bp in length, which naturally do not span the entire length of an ecDNA molecule. Consequently, it is difficult to be confident in reconstructing ecDNA structure with bulk WGS and AA alone, and these tools are unlikely to capture the full range of ecDNA present in a given cell population. As an example of this, in the E20 cell line, WGS/AA analysis predicted that *CDK4* and *PDGFRA* were present on different amplicons. However, direct visualisation of these oncogenes by DNA FISH on metaphase spreads confirmed that they were colocalised in only a small proportion of cells. This likely represents a subclone too rare to be identified by bulk WGS. Indeed, the gastric cancer cell line SNU16, which harbours *FGFR2* and *MYC* oncogenes on ecDNA, was analysed via a novel ecDNA isolation and analysis methodology, CRISPR-CATCH, which identified many subspecies of ecDNA (Hung *et al.*, 2022). This included *MYC* and *FGFR2* ecDNA hybrids validated by DNA FISH, where ~4% of *MYC* and chromosome 10 (region overlapping with *FGFR2)* DNA FISH probes signals colocalised. Long read sequencing is also a powerful tool to increase read lengths to many kilobases, which will also enable individual ecDNA to be more accurately defined (Y. Wang *et al.*, 2021).

In addition, AA is unable to differentiate between ecDNA and HSRs, as demonstrated here for both the E26 and E37 cell lines, which both harbour *EGFR*

HSRs (~10% in E26, 100% in E37). The creators of AA acknowledge that tandem duplicated reads, as can exist on an HSR, will be identified as circular (Deshpande *et al.*, 2019). In addition, circularity and linearity can be called incorrectly, such as in E20 cells where the *PDGFRA* amplicon was called as linear. DNA FISH of metaphase spreads here confirms that the linear prediction does not represent an HSR, and previous multimodality studies (electron microscopy and OM) have indicated that ecDNA are circular rather than linear structures evading nucleases (Wu *et al.*, 2019). AA has previously been shown to have a sensitivity of 83% and a PPV of 85% of defining circular amplicons by WGS (Kim *et al.*, 2020). As such, direct visualisation by DNA FISH remains the 'gold standard' modality for verifying whether a multicopy amplification identified via WGS and AA is either an ecDNA or a chromosomal amplification (HSR). Characterisation by DNA FISH is limited by the fact that only tens of metaphase spreads can feasibly be visualised, and probes are user-defined. Characterisation requires advanced fluorescence microscopy with high optical resolution and high-quality samples, and quantifying features such as ecDNA copy number, while indicative, are not accurate. There are some slight variations in DNA FISH signal intensity (e.g. Figure 3.10), which are expected due to differences in how a FISH probe accesses the nucleus and differences between cell lines. In addition, the definition of an HSR varies widely, with some appearing as large doublet foci on chromosomes (e.g. HeLa cells with drug-induced methotrexate DHFR HSR (Shoshani *et al.*, 2021)) and others forming an entire chromosome arm (e.g. neuroblastoma cell lines harbouring a *MYCN* HSR (Storlazzi *et al.*, 2010). Further tools are needed to differentiate ecDNA from HSRs, particularly as they represent dynamic entities, responding to selection pressures and DNA damage (Coquelle *et al.*, 2002; Nathanson *et al.*, 2014; Shoshani *et al.*, 2021). Until such tools exist, the data presented here highlight the importance of direct visualisation of genomic loci by DNA FISH of metaphase spreads to accurately characterise ecDNA and HSRs in combination with sequencing tools.

In summary, my data indicate that glioblastoma stem cells from the GCGR harbour ecDNA and represent an appropriate model in which to study ecDNA dynamics. By characterising the ecDNA and HSRs present on primary and recurrent glioblastoma cell lines, I have confirmed that driver oncogene amplifications previously reported in glioblastoma are found on these structures.

Chapter 4 Spatial organisation and gene regulation of ecDNA in primary glioblastoma stem cells

4.1 Introduction

EcDNA express their resident genes in a highly organised nucleus. EcDNA are proposed to be highly transcriptionally active, by virtue of their large regions of accessible chromatin, they harbour both oncogenes and their cognate enhancers, and can themselves appear to act as mobile enhancers (Morton *et al.*, 2019; Wu *et al.*, 2019; Zhu *et al.*, 2021). Enhancers with densely bound transcription factors have been shown to recruit high concentrations of the transcriptional machinery, such as Mediator and RNA Pol II, with the central theory that gene control and these dense transcriptional hubs are intrinsically linked (Cho *et al.*, 2018; Chong *et al.*, 2018; Sabari *et al.*, 2018).

Bringing together these concepts of ecDNA oncogenes, enhancers and transcriptional hubs has led to the proposal that ecDNA might cluster. In doing so, ecDNA-oncogene transcription might be amplified beyond that expected by copy number alone (Wu *et al.*, 2019; Kim *et al.*, 2020). These have led to the suggested role of ecDNA 'hubs', featuring clustering of ecDNA, recruitment of RNA Pol II and enhanced transcriptional output (Hung *et al.*, 2021; Yi *et al.*, 2021; Zhu *et al.*, 2021).

In Chapter 3, I characterised two glioblastoma cell lines (E26 and E28) harbouring *EGFR* ecDNA, and two (E25 and E20) harbouring *CDK4* and *PDGFRA* ecDNA. In this Chapter, I use these primary glioblastoma cell lines to characterise the spatial organisation and transcriptional regulation of ecDNA.

I used epifluorescence 2D imaging to characterise the overall nuclear organisation of ecDNA. I then used 3D super-resolution imaging and quantitative image analysis to quantify the spatial relationship between individual ecDNA, and between ecDNA and large RNA Pol II hubs. I characterised the relationship between ecDNA copy number and gene expression in cell lines harbouring *EGFR* ecDNA. Finally, I used a combination of super-resolution imaging and sequencing analysis to quantify the transcriptional efficiency (i.e., transcripts per copy number) of genes transcribed from ecDNA.

Note figures in this section are adapted from (Purshouse *et al.*, 2022b) and the corresponding pre-print (Purshouse *et al.*, 2022a) where stated. The publication can be found in Appendix 8.8.

4.2 EcDNA are widely distributed throughout the nucleus

First, I analysed the overall nuclear localisation of ecDNA using the two cell lines with *EGFR*-harbouring ecDNA (*ecEGFR*) - E26 and E28, and compared this with the genetically normal and euploid NSC line NS9. Chr7 territory was used as a control for chromosomal *EGFR*, given both the peripheral localisation of chr7 and increased chr7 copy number in E28 and particularly E26 cells. *EGFR* DNA FISH signal was compared with chr7 territory as probed using chr7 paint. 2D nuclei were analysed by signal intensity analysis for five equally sized bins eroded from the centre to the periphery of each nucleus (see Methods, Nuclear Territory Analysis) (Figure 4.1A). Chr7 and *EGFR* DNA FISH signal were more peripherally localised in nuclei in all three cell types (Figure 4.1B and C). Once chr7 signal, and therefore chromosomal *EGFR*, was accounted for, *EGFR* DNA FISH signal was still highest in the periphery in NSC nuclei ($p<0.0001$). In E28 nuclei, this radial organisation was still significant but less marked ($p=0.012$). In E26 nuclei, which harbour a higher number of ecDNA than E28, this peripheral localisation was lost ($p=0.0598$) (Figure 4.1D). These data suggest that *EGFR* loci on ecDNA are liberated from the constraints of chr7 localisation and can access more central regions of the nucleus.

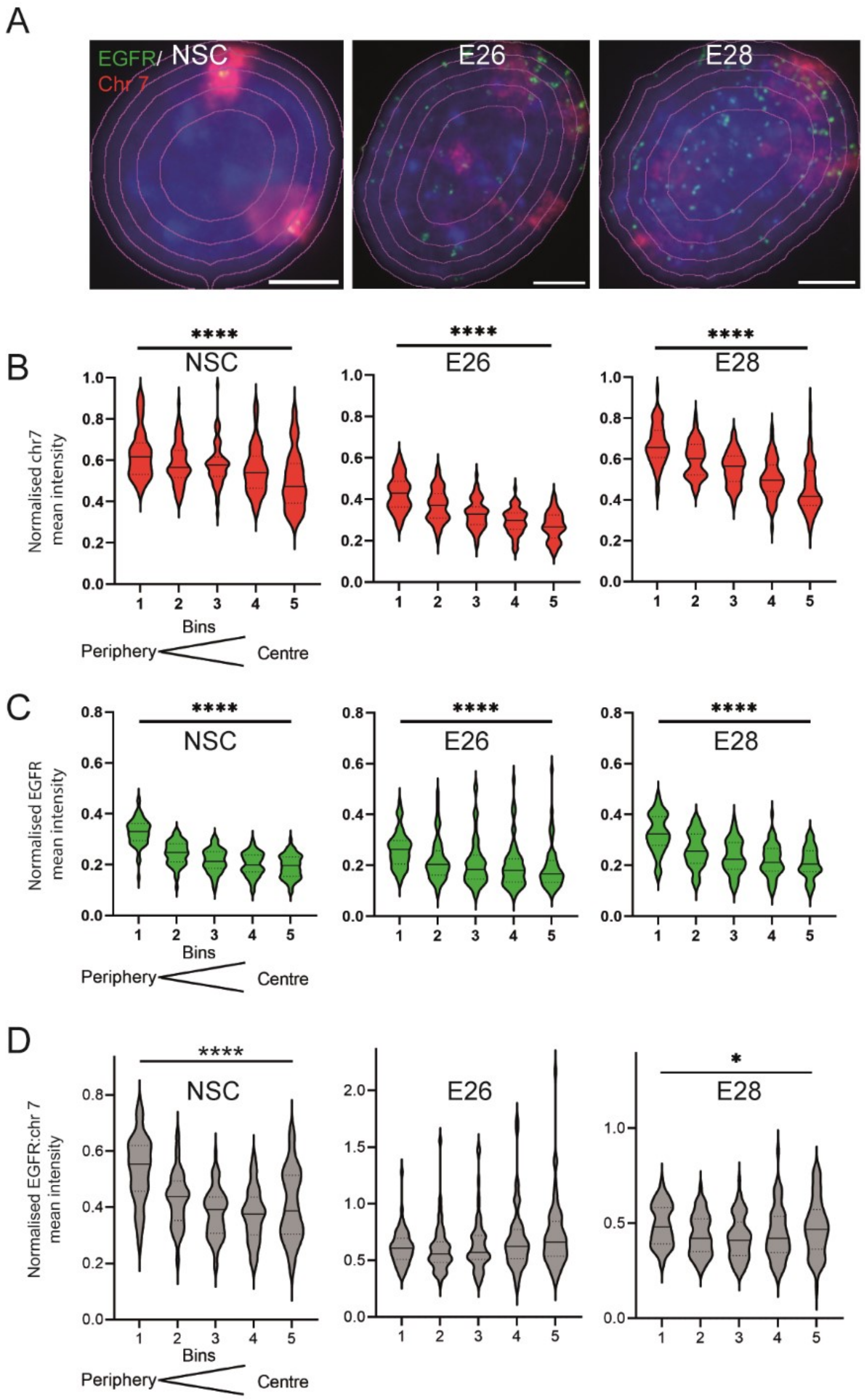


Figure 4.1 | Nuclear Territory Analysis

*A) Representative DNA FISH images of 2D nuclei for neural stem cell (NSC) (n=66 nuclei), E26 (n=59), and E28 (n=64) cells. Signals for chr7 (red), EGFR (green), DNA [DAPI] (blue). Erosion territories are shown. Scale bar = 10 µm. B) Chr7 signal intensity normalised to DAPI across five bins of equal area from the periphery (1) to the centre (5) of each nucleus. Median and quartiles shown. **** $p<0.0001$, * $p<0.05$. Kruskall-Wallis test. C) As for B) but for EGFR signal intensity. D) As for B) but for EGFR FISH signal intensity normalized to that for chr7 (EGFR:Chr7 Mean Intensity). Adapted from (Purshouse et al., 2022b). Biological replicates = 1.*

4.3 EcDNA do not closely cluster

4.3.1 EcDNA-ecDNA distances are not in keeping with close clustering

Previous work reported potential clustering of ecDNA in cancer nuclei via 2D analysis of fixed cells (Hung et al., 2021). I performed DNA FISH in cells harbouring *EGFR* ecDNA and took images in 3D (across 3μm of each nucleus, 100nm increments in z) using a Nikon SoRa microscope. The SoRA is a super-resolution spinning disc confocal microscope with an optic resolution of approximately 120nm in x,y, rendering it an ideal tool for capturing and analysing loci in close proximity. Subjectively, I observed a different, un-clustered pattern to that observed in published images (Figure 4.2A – see images from previous publications in the Introduction - Figure 1.9 and Figure 1.10).

Next, I used image analysis software (Imaris) to define *EGFR* foci as spots (see Methods 2.10.4). I used the spot-to-spot feature to record the shortest distance from each spot to each other spot in each nucleus.

Based on extensive studies of transcriptional co-localisation between genomic loci, I defined as 200nm or less, the spatial distance that might be compatible with a transcriptional hub or condensate, and therefore clustering (Williamson *et al.*, 2016, 2019; Hansen, Maeshima and Hendzel, 2021). Historically, this definition for genomic colocalisation was based on the optical resolution for conventional wide-field microscopy being ~200nm. The resolution of the SoRa super-resolution microscope used in our study is approximately 120nm. This definition is also in keeping with distances described in studies of super-enhancer proximity to BRD4/Med1 'condensates' (Sabari *et al.*, 2018).

Cumulative frequency distribution of these distances showed that distances were routinely >200nm (dotted line) (Figure 4.2B). This remained true when only nuclei with >10 *EGFR* foci (to ensure a reasonable number of ecDNA) were included (Figure 4.2C).

The mean shortest interprobe distances between spots were well above >200nm (dotted line), with no significant difference between E26 and E28 nuclei despite the higher number of ecDNA in E26 cells (Figure 4.2D). When only comparing nuclei with >10 *EGFR* foci, a difference in keeping with this higher frequency could be observed, but the mean distances themselves remained >500nm in all cases (Figure 4.2E). The single shortest interprobe distance in each nucleus was also routinely >200nm (overall shortest distance 0.24µm, E26; 0.25µm, E28) (Figure 4.2F and G). Taken together, these distances are not consistent with the hypothesis that ecDNA cluster in hubs.

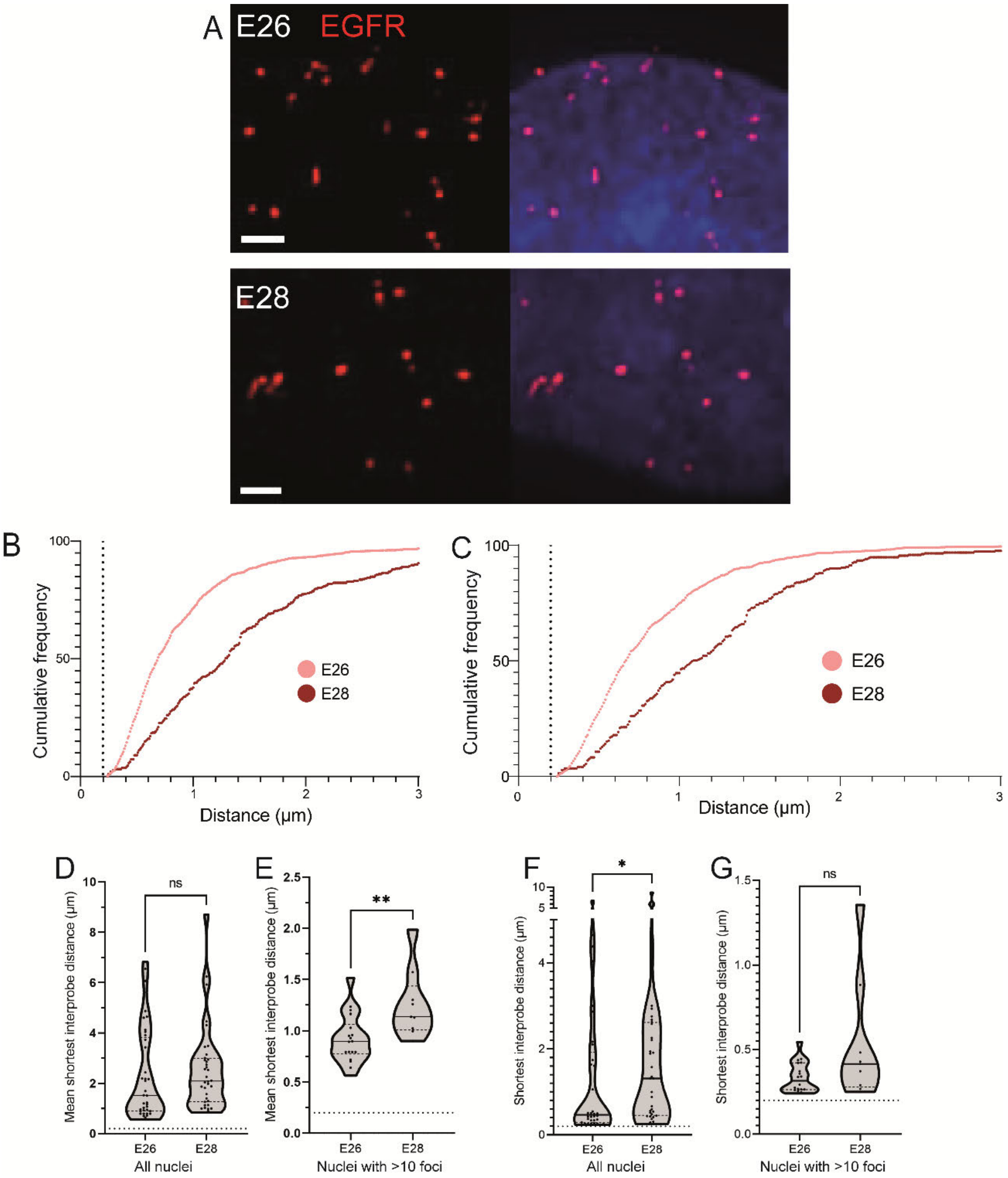


Figure 4.2 | Distances between *EGFR* foci in nuclei with ecEGFR

*A) Representative images of DNA FISH for EGFR (red) in the nuclei of E26 (top) and E28 (bottom) cells, scale bar = 1 µm, MIP image. B) Cumulative frequency distribution of shortest EGFR-EGFR distances between all foci in each nucleus across all E26 and E28 nuclei. Dotted line = 200nm. C) As for B) but only including nuclei with >10 EGFR foci. D) Violin plots showing the distribution of mean shortest interprobe distance between EGFR foci per nucleus in E26 and E28 cell lines. Dotted line denotes y=200 nm. Number of nuclei (n): E26 (n) = 37, E28 (n) = 36. (C) As for (B) but only showing nuclei with >10 EGFR foci. E26 (n) = 18, E28 (n) = 9. (D) As for (B) but for shortest single distance between two EGFR foci in any nucleus. (E) As for (D) but only showing nuclei with >10 EGFR foci. Statistical significance examined by Mann-Whitney test. ns = not significant, * p<0.05, ** p<0.01. Adapted from (Purshouse et al., 2022b). Biological replicates = 1.*

4.3.2 Ripley's K Function

Previous studies have utilised an autocorrelation approach that compares against the estimated probability of detecting another pixel (ecDNA) (Hung *et al.*, 2021) or subjective calling of colocalisation (Yi *et al.*, 2021). Neither strategy facilitates 3D analysis, controls for nuclear size or ecDNA copy number, or assessment of clearly defined distances inkeeping with transcriptional colocalisation ($\leq$200nm).

Ripley's K Function is an analytical approach that can quantify distances between ecDNA FISH hybridisation signals and to assess these distances against those expected from a random distribution. However, ecDNA copy number are highly variable between nuclei. In a nucleus with many ecDNA, these ecDNA will invariably be closer together than in a nucleus with few ecDNA. The same is true for small vs large nuclei. Existing analyses do not account for this random chance of colocalisation due to these factors. I aimed to devise an analysis tool to analyse 3D images of ecDNA-containing nuclei that would account for the varying number of ecDNA and nuclear size, in order to determine whether there is non-random distribution of ecDNA in glioblastoma stem cells at transcriptional distances.

4.3.2.1 Outline of Ripley's K Function

The design and code for Ripley's K analysis was conducted in collaboration with Dr Sjoerd Beentjes, from the Department of Mathematics at the University of Edinburgh, using an existing Python Package for Ripley's K calculation (ripleyk 0.0.3). See Methods for further details.

Ripley's K utilises a user-defined scale to describe a point pattern around increasing radii (Figure 4.3). The rationale for this method was that DNA foci were clearly resolved, each ecDNA only harboured a single oncogene targeted by DNA FISH hybridisation and the analysis could be performed in 3D. This allows for each nucleus to act as its own control in terms of number of foci and size.

Ripley's K plots two point patterns. The **expected** point pattern is a null distribution generated using the given number of foci in a given nucleus using their x, y, and z coordinates to define the size of the nucleus. The null distribution is generated by multiple simulations in order to have statistical confidence in the significance boundaries this defines and represents uniform distribution of spots with no clustering. This is then compared with the **observed** distribution of the x,y,z coordinates of the given nucleus. This allows a comparison, at each user-defined radius, of significant clustering (Figure 4.3).

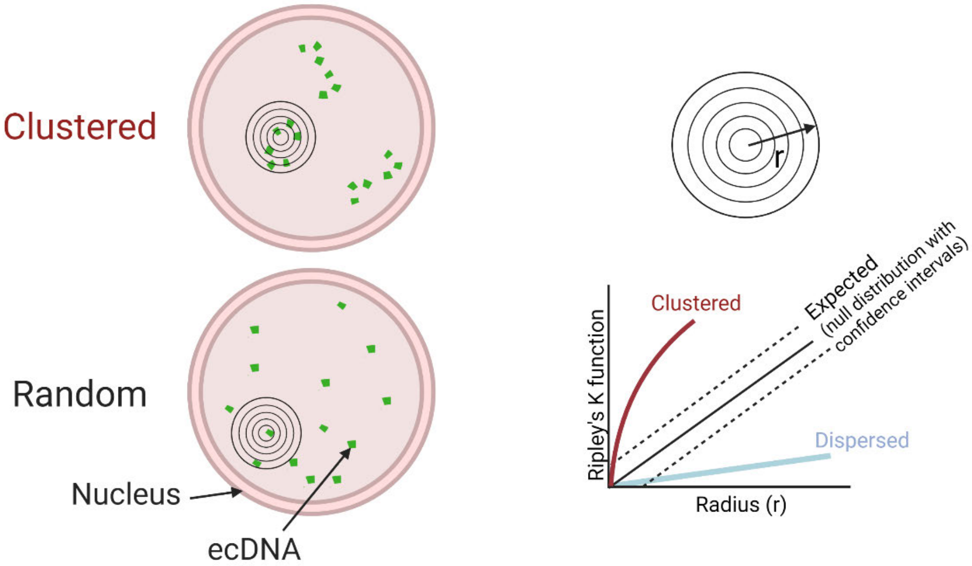


Figure 4.3 | Ripley's K Function

Graphical representation of nuclei showing clustered (top) and random (bottom) distribution of ecDNA, with overlay of increasing radii (r) to indicate Ripley's K. The graph shows how this is plotted per nucleus, with the expected values +- confidence interval shown to represent the null distribution i.e. random distribution. Observed values within this null distribution would be considered randomly distributed. The lines indicate observed values as they would plot if foci were clustered (burgundy) and dispersed (blue). Adapted from (Purshouse et al., 2022a)

4.3.2.2 Defining the Function

Ripley's K is a function of the following:

- df – dataframe of x,y,z coordinates of spots in a nucleus
- radii – list of radii
- n_data - number of draws from the null, i.e. the significance cut-off
- bounding_radius – the estimated radius of the sphere bounding the sample
- boundary – boundary correction, either 'True' or 'False'
- normalise – normalisation, either 'True' or 'False'

The x,y and z coordinates were defined using the Imaris software ‘spots’ function, with a single csv file containing the x,y and z coordinates for an individual nucleus. Within the function, these coordinates define the volume parameters as well as the number of foci within the volume, and this is used to derive the expected values of Ripley’s K that define the null distribution. Based on optimisation data, nuclei were only analysed if they harboured >20 foci, to ensure most foci were ecDNA, ensure sufficient granularity and to avoid false negative results if foci were too sparse. The diameter of spots was defined by measuring representative foci in Imaris (see Methods 2.10.4).

We selected a range of radii that were appropriate for transcriptional regulation and were feasibly detectable with the optic resolution of the SoRa microscope (0.1-1µm in 0.1µm increments).

The significance cut-off (n_data) represents how the *p*-value is defined for each observed value vs the expected value. We used the Neyman-Pearson lemma to establish the *p*-value. The null distribution is a list of expected Ripley’s K values at each radius in the list of radii. Having tested a sample of 1,000 and 50,000 expected values, we found generating 10,000 Ripley’s K values provided sufficient accuracy for the null distribution. If all expected values at each radius are ranked in size order, the 250th (2.5%) value at either end of the null distribution represents the significance cut-off of $p = 0.05$, which we chose for this analysis. On occasion, observed and expected Ripley’s K function at $p=0.05$ were the same because multiple expected Ripley’s K values at a given radius were the same. To generate a *p*-value in this scenario, a randomised binomial test was integrated into the function – a mathematical coin toss to establish if $p<0.05$ for the observed value. This was weighted as the ratio of the number of values $p<0.05$ and the total number of equal values. Using this approach tells us whether the value is significant or not, but not the exact *p*-value. Here there are 3 options – optimistic (i.e. the lowest *p*-value in the rank of expected Ripley’s K values), conservative (the highest) or the mean. We elected to use the most optimistic, i.e. lowest, *p*-value which would favour identification of a significant result. Whilst this represented a bias in favour of clustering, this would minimise the risk of a false negative result overall.

A bounding radius of 5μm was used for all calculations, having tested extremes of bounding radius of 1μm and 20μm and found a negligible effect on results. This was important given the nuclei themselves vary in size. Using a standardised bounding radius allowed batch analysis of nuclei. Boundary correction and normalisation were both set to ‘False’.

4.3.2.3 Defining the output

We defined the output of the function for each nucleus for each radius in the list of radii in the input:

- Minimum, median and maximum Ripley’s K values of the null distribution
- The significance cut-off for Ripley’s K to be clustered at that radius, as calculated from the null distribution
- The observed Ripley’s K value
- The p-value at that radius between observed and expected distribution

A Benjamini-Hochberg procedure was performed when combining the p-values from multiple nuclei at each given radius, to control for the false discovery rate (FDR = 0.05).

A summary of function components and how they were defined in the function and output are listed in Table 4.1. Further detail on the mathematical methodology is outlined in the Methods (Ripley’s K Cluster Analysis, Section 2.10.5).

Ripley’s K function therefore enables us to use 3D imaging data of ecDNA visualised by DNA FISH and quantify whether they are clustered at transcriptionally relevant distances by allowing each individual nucleus to act as its own ecDNA copy number control.

Component	**Rationale**	**Defined in function**
Range of radii	Interested in short distances where biological interactions may occur	Range **0.1-1µm in 0.1µm** increments. 10 radii
Definition of spots	Need to have observed spots, and these will form the basis of the expected random range	X,Y,Z coordinates from Spots defined in Imaris, >20 spots/nucleus included. Diameter in Imaris = 300nm
Bounding radius	Radius from the volume in which all spots reside. Challenges = spots contained in a range of nuclear volume and shape	**5µm** Based on trials of calculating per nucleus, or very high/low (20/1um)
Number of simulations	Need enough to have statistical confidence in significance, but balance with computational load	**10k** (having tested a sample with 1k and 50k simulations)
Defining the volume of the space and boundary correction	Null distributions for the simulations Boundary correction accounts for edge effect	**X,Y,Z** coordinates from a nucleus. Caveat – potential for underestimating size of the space and thus simulation too – will affect bigger radii more
Normalisation	Assumes the volume as a whole is spherical	Set normalisation to False
Significant clustering	How to define differences in the K functions at different nuclei	Neyman-Pearson lemma (*p*=0.05 cut-off with 'optimistic' *p*-value), Benjamini-Hochberg false detection rate (FDR) = 0.05.

Table 4.1 | Ripley's K function definitions

4.3.3 Ripley’s K in cells with single-ecDNA populations do not show close clustering

This 3D point-pattern analysis was applied to *EGFR* DNA FISH images from E26 and E28 nuclei to determine the extent of clustering. Examples of E26 and E28 nuclei and their corresponding Ripley’s K function are shown in Figure 4.4A and B.

The Ripley’s K analysis from several nuclei were combined. In E26, a proportion of nuclei had significant non-random distribution of foci at ≥400nm, but not at smaller distances. In E28, significant non-random distribution of foci was only seen in a proportion of nuclei at ≥700nm (Figure 4.4C).

To ensure that no small FISH foci were omitted, the analysis was repeated, reducing the spot size diameter in Imaris from 300nm to 150nm. Although this shifted the trend to the left, there was no significant clustering at <300nm (Figure 4.4D).

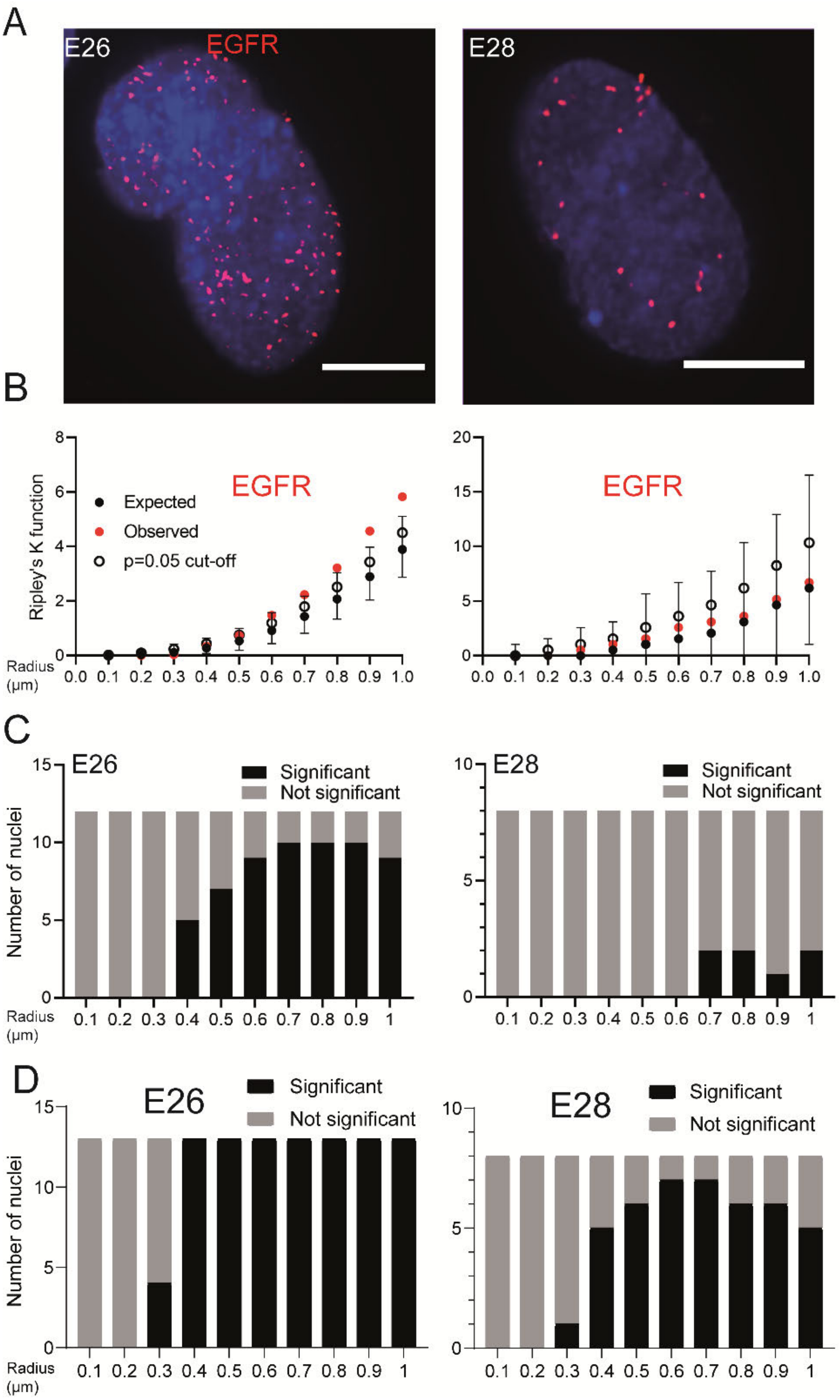


Figure 4.4 | Ripley's K analysis of E26 and E28 cell lines

A) Representative images of EGFR DNA FISH (red) in nuclei (blue=DNA) of E26 and E28 cells. Scale bar = 5 µm, MIP Z-projection. B) Ripley's K function for the nuclei in A) showing observed K function (red), max/min/median (black) of null distribution with p=0.05 significance cut-off (empty black circle). C) Ripley's K function for EGFR DNA FISH signals showing number of E26 (n=12) and E28 (n=8) nuclei with significant (black) and non-significant (grey) clustering at each given radius. D) As C) but with spot diameter of 150nm. Adapted from (Purshouse et al., 2022b). Biological replicates = 1.

4.4 EcDNA do not cluster closely in a two-ecDNA population cell line

4.4.1 EcDNA-ecDNA distances are not in keeping with coordinated close clustering in a two-ecDNA cell line

It is possible that foci hybridised by the same fluorophore are too tightly clustered to be resolved by FISH. To address this, I repeated the analysis with the E25 cell line, which harbours two oncogenes (*CDK4* and *PDGFRA*) on separate ecDNA populations, where the DNA FISH probes for each are directly labelled with different fluorophores. Again, subjectively there was no obvious clustering of foci in these nuclei (Figure 4.5A).

As previously, the shortest distance between each *CDK4-CDK4*, *PDGFRA-PDGFRA,* and *CDK4-PDGFRA* locus across all nuclei were measured and plotted by cumulative frequency. This revealed the frequency of distances ≤200nm was low (Figure 4.5B). Indeed, no *CDK4-CDK4* or *PDGFRA-PDGFRA* foci were <200nm apart, and <1% of *CDK4-PDGFRA* distances were <200nm (4/1011 (0.39%) *CDK4* foci, 4/518 (0.77%) *PDGFRA* foci). The mean shortest interprobe distance per nucleus was regularly >1µm (Figure 4.5C) across all combinations. The single shortest interprobe distances were shorter for *CDK4-CDK4* and *CDK-PDGFRA* foci than for *PDGFRA-PDGFRA* foci, noting that there are more *CDK4* than *PDGFRA* foci across nuclei, but there was no significant difference when comparing the shortest distance between *CDK4-CDK4* to *CDK4-PDGFRA* foci (Figure 4.5D). Taken together these data suggest that these two ecDNA populations do not appear to be routinely clustered at close distances.

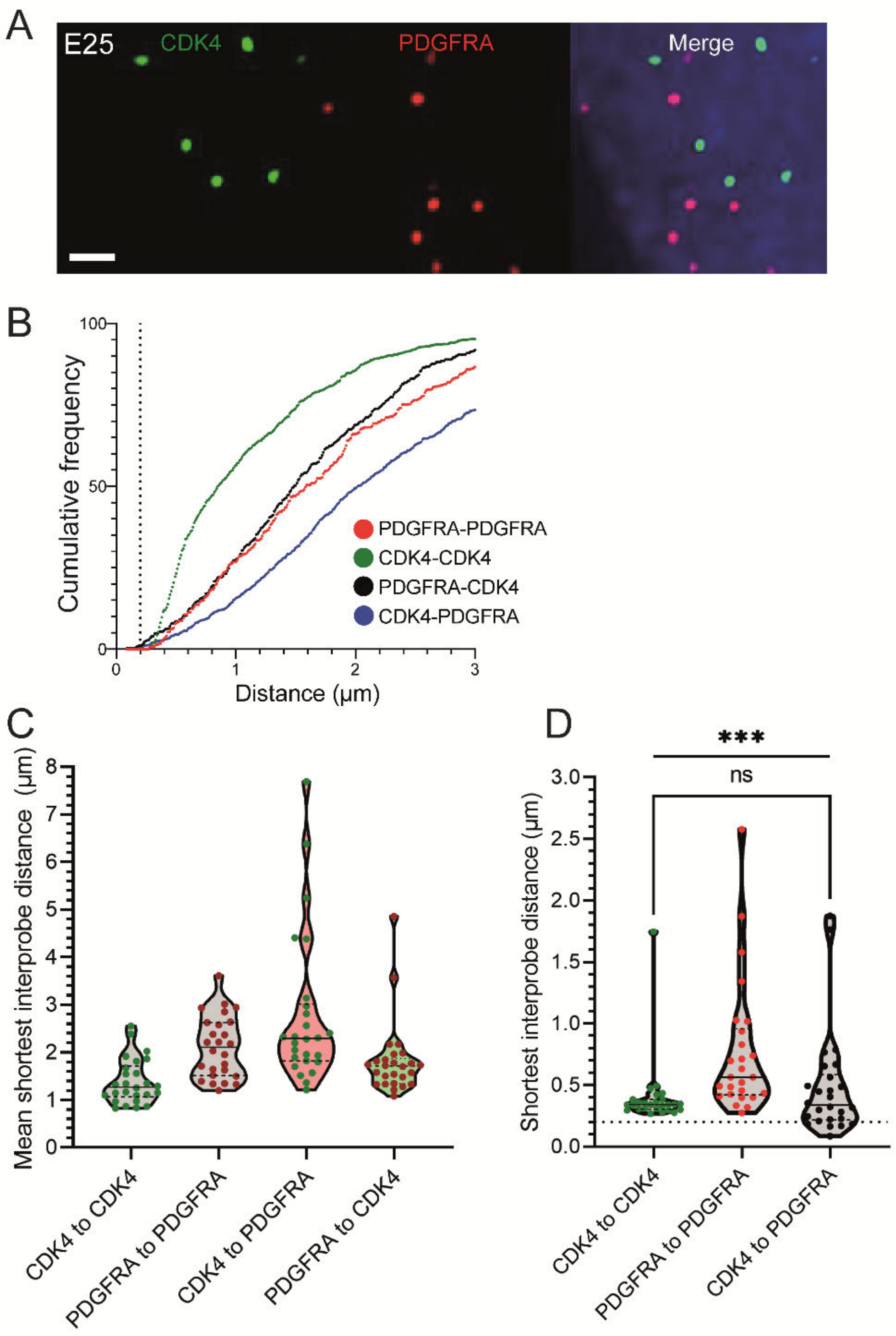


Figure 4.5 | Distances between CDK4 and PDGFRA ecDNA in E25 nuclei

*A) Representative image of DNA FISH for CDK4 (green) and PDGFRA (red) in an E25 nucleus (blue=DNA (DAPI)), Scale bar = 1 μm, MIP. B) Cumulative frequency distribution of shortest interprobe distances between all foci of each type as shown in each nucleus across all E25 nuclei (n=26). C) Violin plots showing distribution of mean shortest distance between CDK4 and PDGFRA foci per E25 nucleus. (D) As for (C) but for the shortest single interprobe distance measured in any nucleus. Dotted line denotes y=200 nm. Statistical significance examined by Mann-Whitney test (hooked line, ns = not significant) and Kruskall-Wallis (straight line, *** p<0.001). Adapted from (Purshouse et al., 2022b). Biological replicates = 1.*

4.4.2 Ripley’s K analysis in dual-ecDNA population do not show close clustering

Ripley’s K was then used to examine the 3D distribution of *CDK4* and *PDGFRA* ecDNA. Example of an E25 nucleus with the Ripley’s K function for *CDK4, PDGFRA* and *CDK4*/PDGFRA hybridisation signals combined are shown in Figure 4.6A and B.

When combining nuclei together, *CDK4* ecDNA had significant non-random distribution in many nuclei at $\geq$400nm, and *PDGFRA* ecDNA also had some significant non-random distribution of foci at $\geq$400nm. When *CDK4* and *PDGFRA* foci were combined, no significant non-random distribution was observed at <300nm. With increasing radial distance $\geq$300nm, the majority of nuclei had significant non-random distribution of foci (Figure 4.6C). Again, no significant clustering could be identified at <300nm when this analysis was repeated with a spot size of 150nm to avoid omission of small FISH foci (Figure 4.6D).

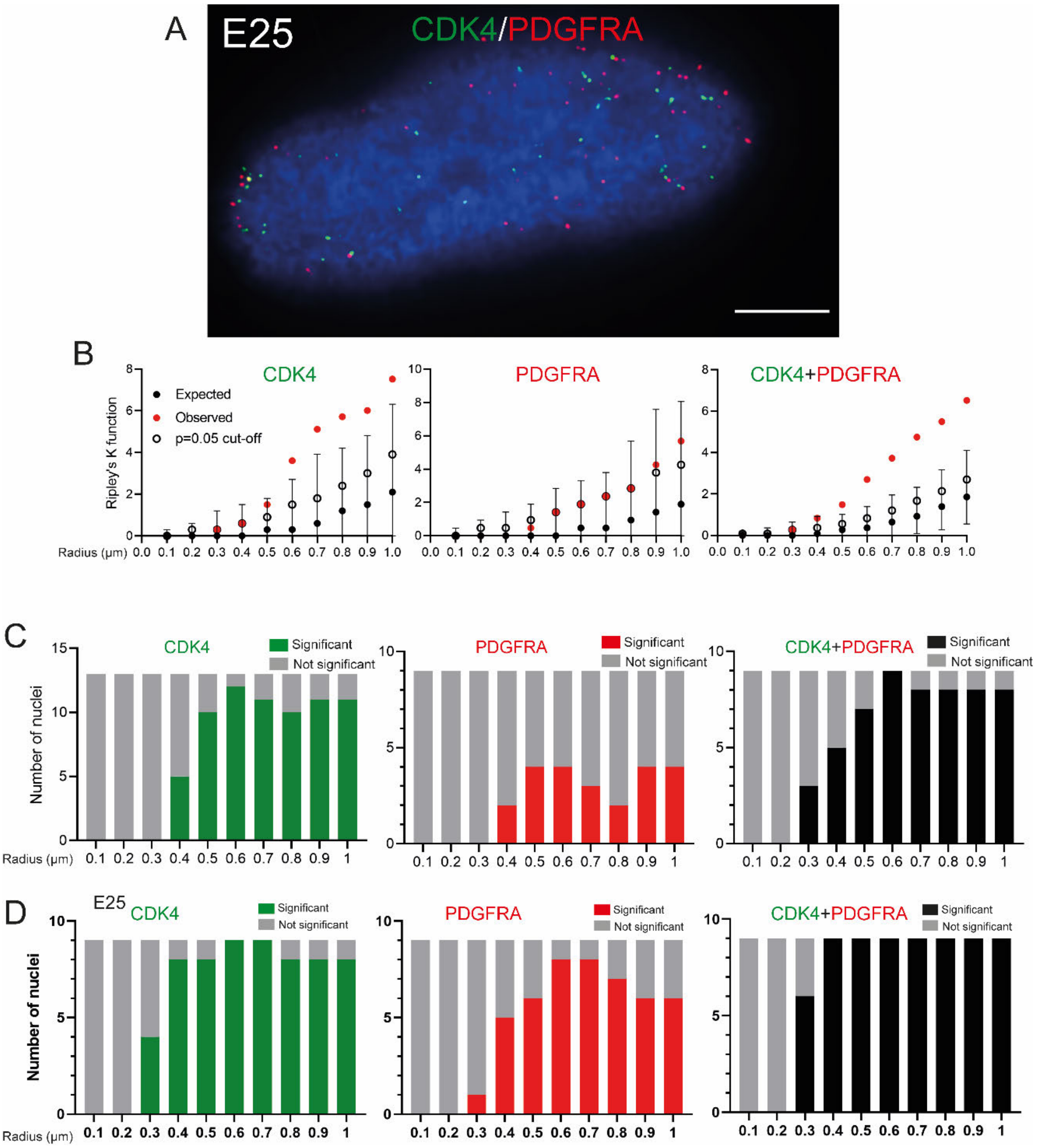


Figure 4.6 | Ripley's K analysis of E25 cell line

A) Representative image of DNA FISH shown of E25 nucleus (Blue=DNA (DAPI)), hybridized with probes for CDK4 (green) and PDGFRA (red). Scale bar = 5 μm, MIP. B) Ripley's K function for the nucleus in A) showing observed K function (red), max/min/median (black) of null distribution with p=0.05 significance cut-off shown (empty black circle) for CDK4, PDGFRA, and CDK4 and PDGFRA spots combined. C) Ripley's K function for E25 nuclei showing number of nuclei with significant and non-significant clustering at each radius for CDK4 (n=13 nuclei), PDGFRA (n=9 nuclei), and CDK4 and PDGFRA foci combined (n=9 nuclei). D) As C) but with spot diameter at 150nm. Adapted from (Purshouse et al., 2022b). Biological replicates = 1.

4.4.3 Clustering in dual-ecDNA cell line with some hybrid ecDNA did not show close clustering

I sought to confirm this finding in an independent cell line. The E20 cell line also harbours two oncogenes on separate ecDNA, and in addition, approximately 10% of ecDNA harboured *CDK4-PDGFRA* hybrid ecDNA (Chapter 3 – Section 3.4.2) as visualised by colocalised hybridisation signals on metaphase spreads. Therefore, this represented a positive control for Ripley's K, because we would expect a similar proportion of colocalised *CDK4-PDGFRA* signals in our 3D analysis. A greater number of nuclei were therefore imaged, and only imaged if they harboured >20 foci of each oncogene.

The relative frequency of interprobe distances of $\leq$200nm was, again, low in E20 nuclei (Figure 4.7A). Ripley's K analysis of *CDK4-PDGFRA* foci showed 22/24 did not have significant non-random distribution of foci at <300nm (Figure 4.7B and C). However 2/24 (8.3%) did have clustering at $\leq$200nm (Figure 4.7B and D), a similar proportion to hybridised *CDK4-PDGFRA* ecDNA observed in metaphase spreads. This suggests these two nuclei represent cells where *CDK4* and *PDGFRA* are on the same ecDNA. Doublets of *CDK4* foci could be identified by Ripley's K and verified by imaging in 4/24 (16.7%) nuclei (Figure 4.7B and E).

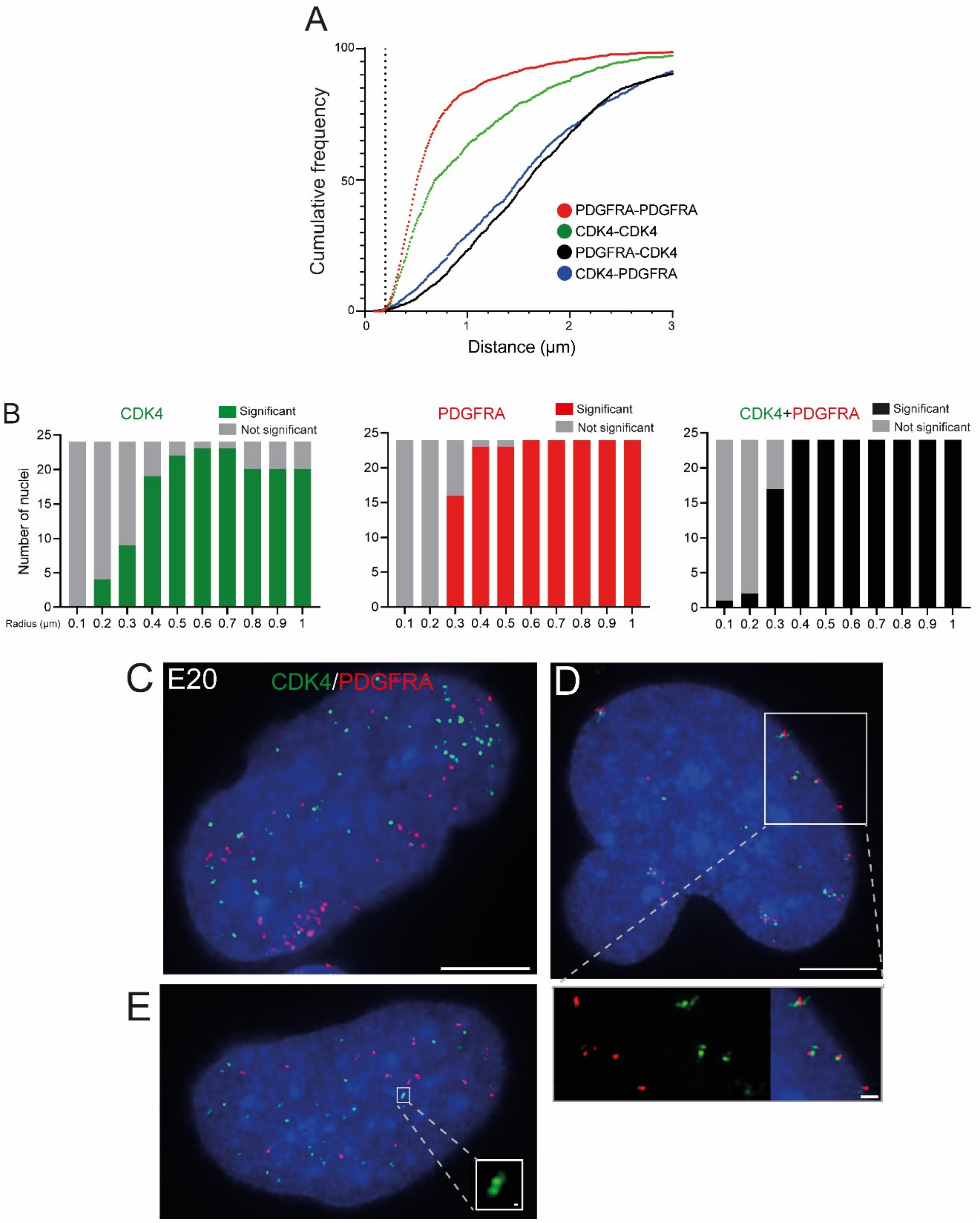


Figure 4.7 | Ripley's K analysis of E20 cell line

A) Cumulative frequency distribution of shortest interprobe distances between all foci (as shown) in each nucleus across all E20 nuclei (n=24).B) Ripley's K function for E20 nuclei showing number of nuclei with significant and non-significant clustering at each radius for CDK4, PDGFRA, and CDK4 and PDGFRA foci combined (n = 24 nuclei). C) Representative image of DNA FISH of E20 interphase nuclei (Blue=DNA (DAPI)), hybridized with probes for CDK4 (green) and PDGFRA (red). Scale bar = 5 µm, MIP. D) As for C) but for a nucleus where the close association of CDK4 and PDGFA signal suggests ecDNA hybrids harbouring both oncogenes. Scale bar = 1µm in inset. E) As for C) but inset showing CDK4 doublets in an E20 nucleus. Adapted from (Purshouse et al., 2022b). Biological replicates = 1.

4.4.4 EcDNA distribution in Colo320DM cells

Given the discrepant conclusions from my analysis of glioblastoma patient derived cells compared with other reports, I next analysed the colorectal cell line Colo320DM. This has been extensively studied to establish ecDNA dynamics, including as an exemplar of ecDNA clustering by DNA FISH of the ecDNA-resident oncogene, *MYC* (Hung *et al.*, 2021). I first sought to characterise Colo320DM ecDNA and establish whether I could use my Ripley's K analysis tool to quantify clustering of *MYC* ecDNA in this cell line.

DAPI staining and DNA FISH for *MYC* on metaphase spreads confirmed that Colo320DM harbours large numbers of ecDNA which are very large in size (approximately 1-2µm) and often appear as doublets (Figure 4.8A). In addition, where doublets can be visualised in isolation, each part of the doublet appears to harbour 3 copies of *MYC* (inset Figure 4.8A). 3D DNA FISH signal for *MYC* in Colo320DM nuclei have a swarm-like pattern in focussed regions of the nucleus (Figure 4.8B).

Considering the large size of the ecDNA (4.3Mb) in this cell line, and, more importantly, the fact that the ecDNA harbour multiple copies of *MYC*, it may be difficult to characterise clustering in Colo320DM cells using *MYC* as the visualisation tool. *MYC* foci may, by definition, appear clustered owing to their multi-copy presence on the same ecDNA molecule and this is a major caveat of previous studies using this cell line. This may also represent a limitation of using Ripley's K as a tool to assess clustering in cell lines with large ecDNA, and indeed a small pilot analysis (data not shown) suggested that the colocalization of *MYC* could not be confidently quantified using this tool, likely due to the large size of Colo320DM ecDNA.

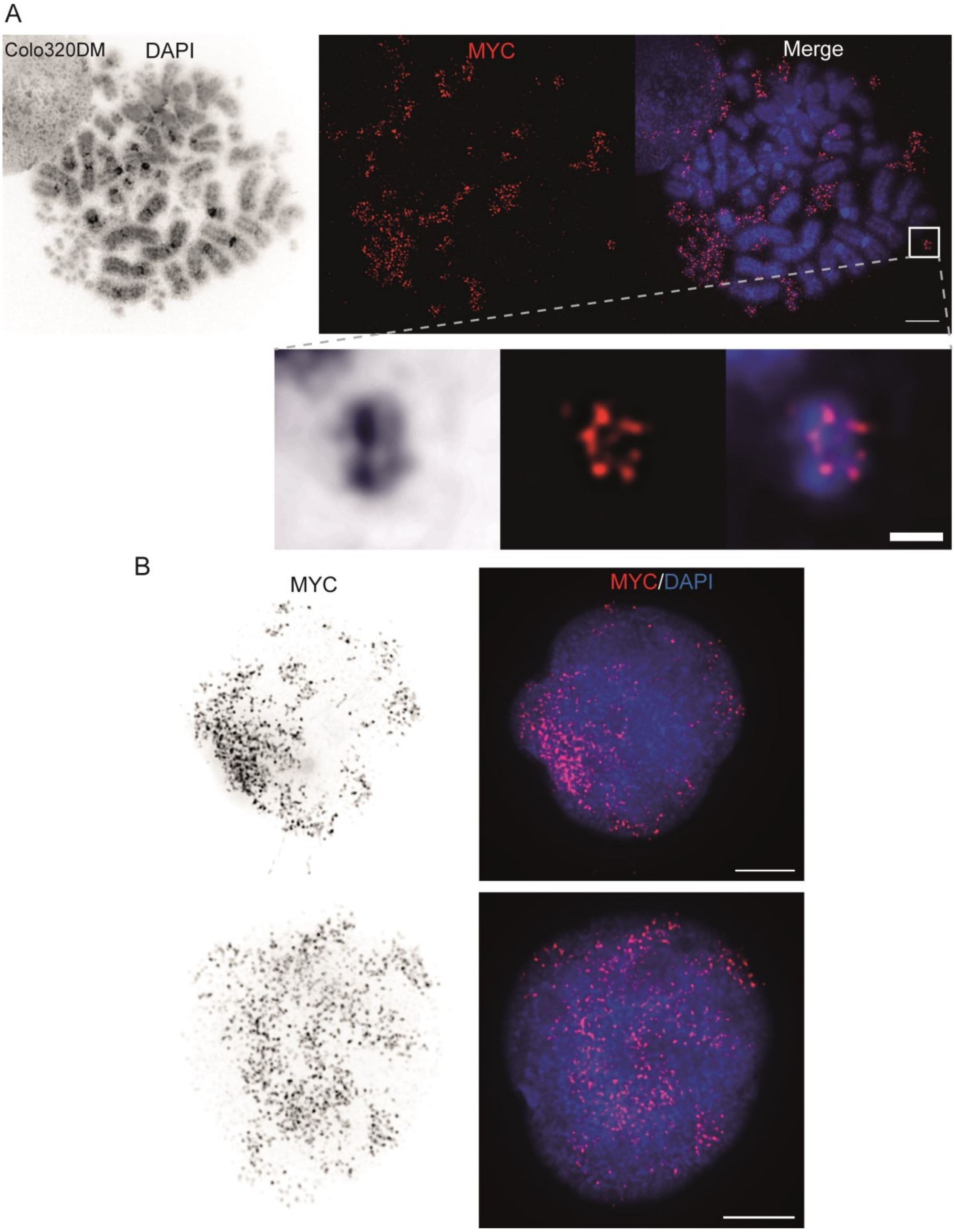


Figure 4.8 | Colo320DM metaphase spreads and nuclei

A) Representative DNA FISH of Colo320DM metaphase spread probing for MYC (red), DNA (greyscale, blue in merged image – DAPI (blue)). Inset highlighting doublet ecDNA with 3 MYC foci on each half of doublet. Scale bar = 5µm (inset – 1µm), single slice. B) Two representative Colo320DM nuclei probing for MYC (greyscale, red in merged image), DNA (DAPI (blue)). Scale bar = 5µm, MIP. Biological replicates = 1.

4.5 Transcribing ecDNA do not cluster closely

It is possible that only actively transcribing ecDNA are clustered. As such I probed for nascent *EGFR* RNA in E26 cells, using RNA FISH probes targeting the large first intron of *EGFR*. The E26 cell line was selected for analysis because there are more ecDNA in this cell line than in E28 cells so there would be an adequate number of cells with >20 foci/nucleus. An example of *EGFR* RNA FISH signal in a nucleus with its corresponding Ripley's K function is shown in Figure 4.9A.

Similar to the DNA-FISH results, there was only significant non-random distribution of nascent RNA FISH spots at $\geq$400nm (Figure 4.9B). This suggests actively transcribing *EGFR* loci do not cluster in the nucleus.

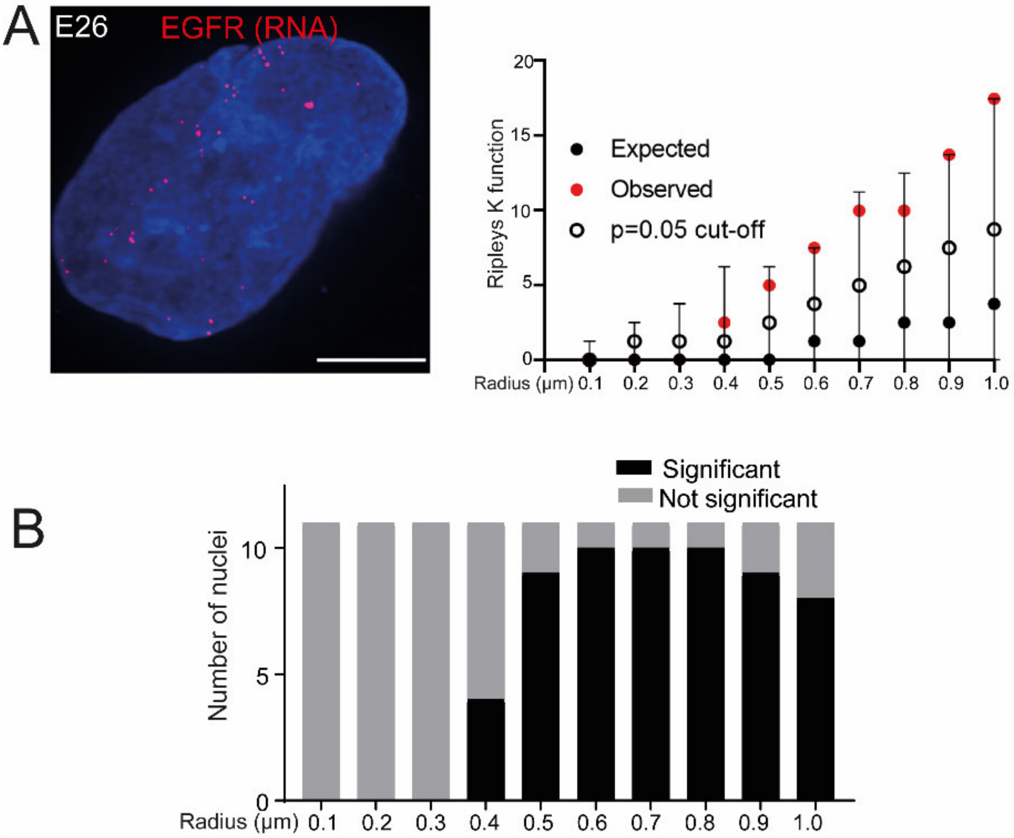


Figure 4.9 | Ripley's K of Nascent RNA FISH in E26 nuclei

A) Representative image of nascent EGFR RNA FISH (red) in E26 cell nucleus (Blue=DNA (DAPI)). Scale bar = 5 µm, MIP. Associated Ripley's K function for this nucleus showing observed K function (red), max/min/median (black) of null distribution with p=0.05 significance cut-off shown (empty black circle). B) Ripley's K function for E26 nuclei (n=11), EGFR nascent RNA FISH showing number of nuclei with significant and non-significant clustering at each given radius. Adapted from (Purshouse et al., 2022b). Biological replicates = 1.

4.6 Transcriptional hubs and transcription efficiency in glioblastoma and NSCs

The above results suggested ecDNA do not cluster with each other at short distances. Significant non-random distribution at higher radii do indicate a regional, rather than close, colocalisation, consistent with the results of nuclear erosion analysis (Section 4.2). As such, ecDNA may not cluster closely with other ecDNA, but rather colocalise with transcriptional hubs harbouring high concentrations of the transcriptional machinery.

4.6.1 Transcriptional hubs are sparse in glioblastoma stem cells

I chose to visualise the protein RPB1 (encoded by *POLR2A*) N-terminal domain (NTD) because RPB1 is the largest subunit of RNA Pol II and this antibody proved the most robust at identifying the large Pol II hubs previously reported by immunofluorescence (IF). Brd4 is also considered a key component of large transcriptional hubs. To confirm the RPB1-NTD antibody was an appropriate label of transcriptional hubs, IF for RPB1 and Brd4 was performed and colocalisation assessed via cross-section intensity plot analysis in ImageJ. Colocalisation of RPB1 and Brd4 large transcriptional hubs was seen in all NSC cells, and was also seen in E26 and E28 nuclei, although the Brd4 signal was weaker (Figure 4.10). These data also indicate, as shown in previous studies (Cho *et al.*, 2018; Sabari *et al.*, 2018), that transcriptional hubs are ≥500nm in diameter. This definition was therefore used for the analysis of subsequent experiments.

I noted that in NS9 (NSC) cells, transcriptional hubs were sparse (across 48 nuclei – mean number of hubs = 0.83, SD 1.098). I performed two biological replicates to count the number of transcriptional hubs in E26 and E28 glioblastoma cells (Figure 4.11A and B). They were also infrequent, suggesting transcriptional hubs may be less frequent in these cells than previously reported in other cell types, such as embryonic stem cells.

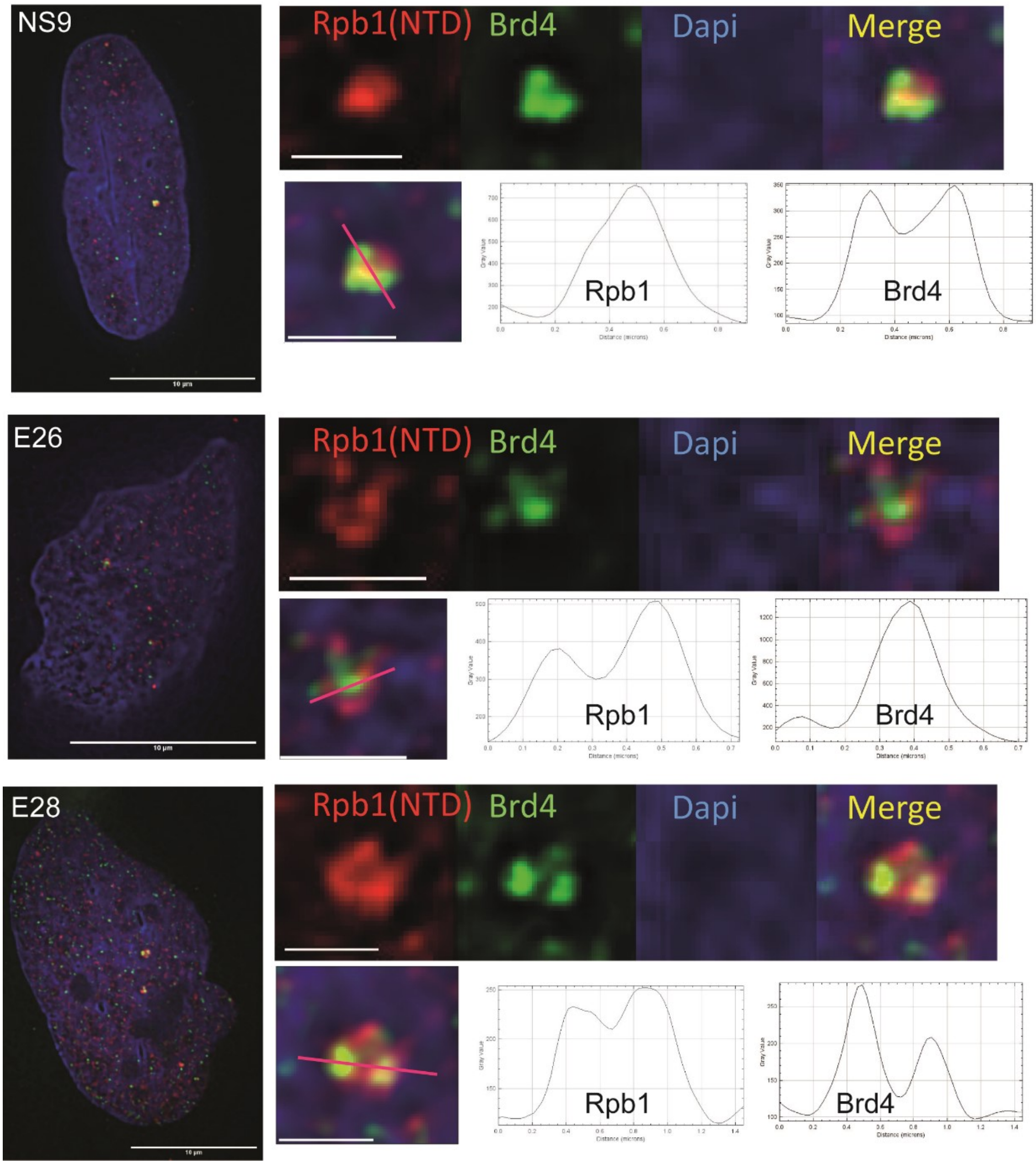


Figure 4.10 | Co-immunofluorescence for RPB1 (NTD) and Brd4

Example nuclei of NS9 (NSC), E26 and E28 cells. IF for Rpb1(NTD) protein (red), Brd4 protein (green), DNA (blue – DAPI). Scale bars – whole nucleus = 10µm, insets = 1µm. Cross-section of transcriptional hub shown in pink, with intensity plots for each signal. Analysis performed via ImageJ. Secondary antibody only performed for NS9. Biological replicates = 1.

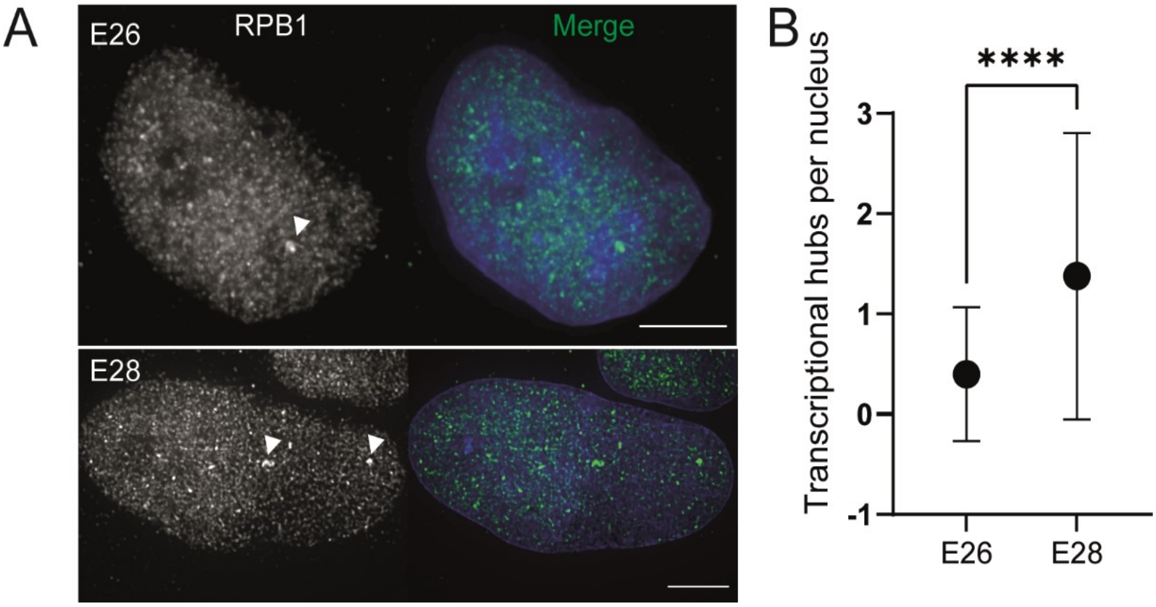


Figure 4.11 | Transcriptional hubs in glioblastoma stem cells

*A) Representative images of IF for RPB1 transcriptional hubs (arrows) in E26 and E28 nuclei (RPB1 – white on left, green on right; DNA – blue (DAPI). Scale bar – 5um. B) Number of RPB1 hubs (mean +-SD) detected per nucleus in E26 and E28 cell lines. Two biological replicates. Unpaired t-test **** p <0.0001. Adapted from (Purshouse et al., 2022b) and associated pre-print, (Purshouse et al., 2022a)*

4.6.2 Transcriptional hubs do not closely cluster with ecDNA by DNA immunoFISH

Next, I combined immunoFISH for Pol II (IF) and EGFR (DNA FISH) and performed 3D image analysis using the Spot function in Imaris to quantify the spatial relationship between large Pol II hubs (≥500nm) and *EGFR* in NSC cells compared with E26 and E28 cells which harbour *EGFR* ecDNA (Figure 4.12A).

There was no correlation between the number of *EGFR* foci and the number of RNA Pol II hubs (Figure 4.12B). The mean shortest distance between large Pol II hubs and EGFR was routinely >1µm in all three cell lines, regardless of the presence of ecDNA (Figure 4.12C). The single shortest distance between a large Pol II hub and an *EGFR* locus was also regularly >1µm, with no measurements of <200nm (Figure 4.12D).

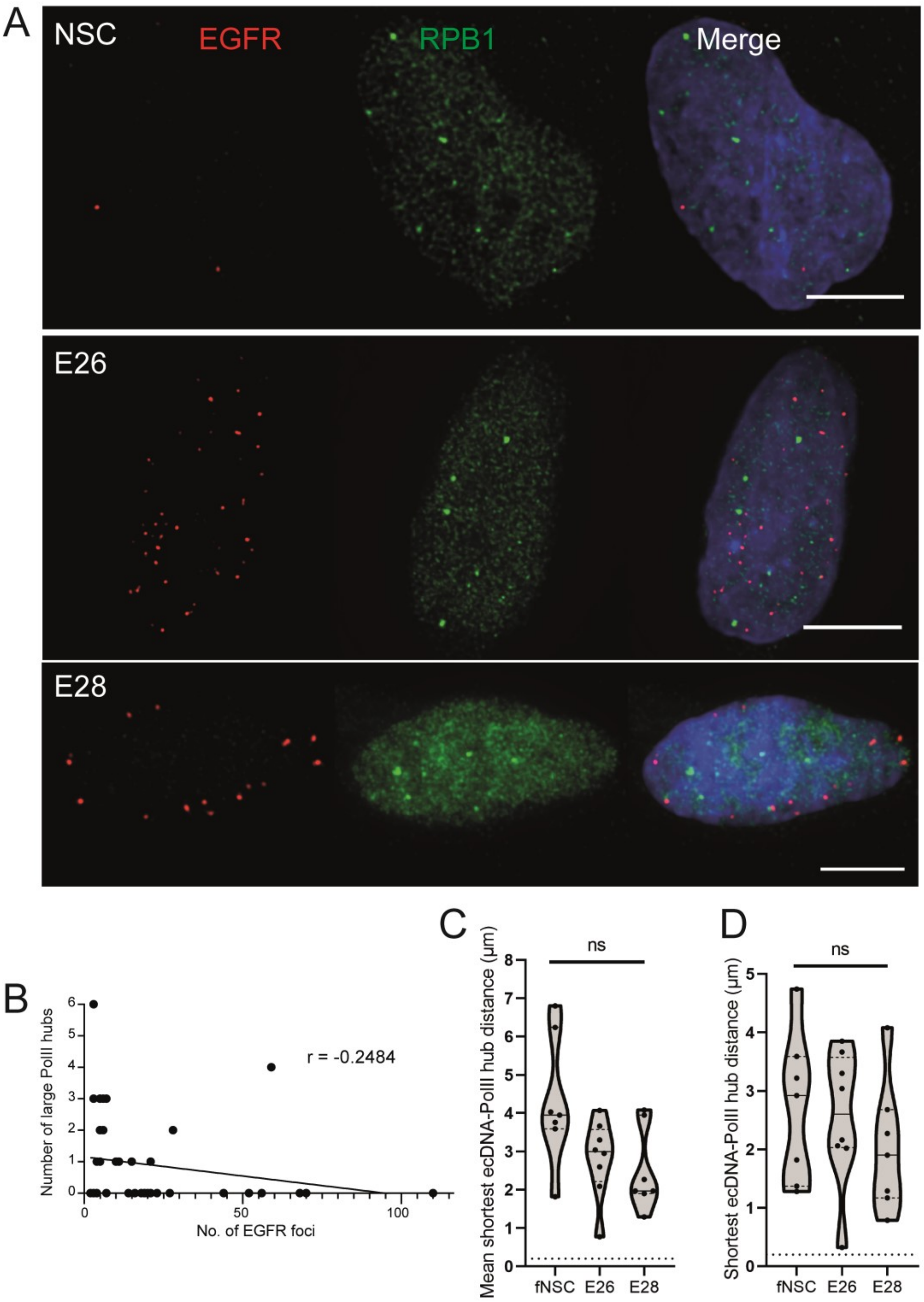


Figure 4.12 | DNA ImmunoFISH for EGFR and large Pol II hubs

A) Representative images of DNA immunoFISH in NSC, E26 and E28 cells: Immunofluorescence for RPB1 protein (green) and EGFR DNA FISH (red). DNA (blue – DAPI). Scale bar = 5 µm, MIP. (B) Spearman's correlation between number of EGFR foci and number of RPB1 foci, p = 0.13, E26 and E28 cell line data combined. (C) Violin plot of distribution of mean shortest interprobe distance per nucleus between EGFR foci and Pol II foci in NSC (n=7), E26 (n=8) and E28 (n=7) cell lines. (D) As for (C) but for shortest single distance in each nucleus. Median and quartiles plotted. Dotted line denotes y=200nm. Statistical significance examined by Kruskall-Wallis test. Adapted from (Purshouse et al., 2022b). Biological replicates = 1.

4.7 Transcriptional hubs do not closely cluster with ecDNA by RNA immunoFISH

To ascertain whether large Pol II hubs colocalise with only actively transcribed ecDNA, the analysis was repeated by performing immunoFISH but for nascent *EGFR* transcripts (Figure 4.13A). Again, large Pol II hubs were sparse and nascent *EGFR* transcripts were not close enough to these hubs to suggest colocalisation, examined either by mean shortest *EGFR*-Pol II hub distance per nucleus or the single shortest *EGFR*-Pol II hub distance per nucleus (Figure 4.13B and C).

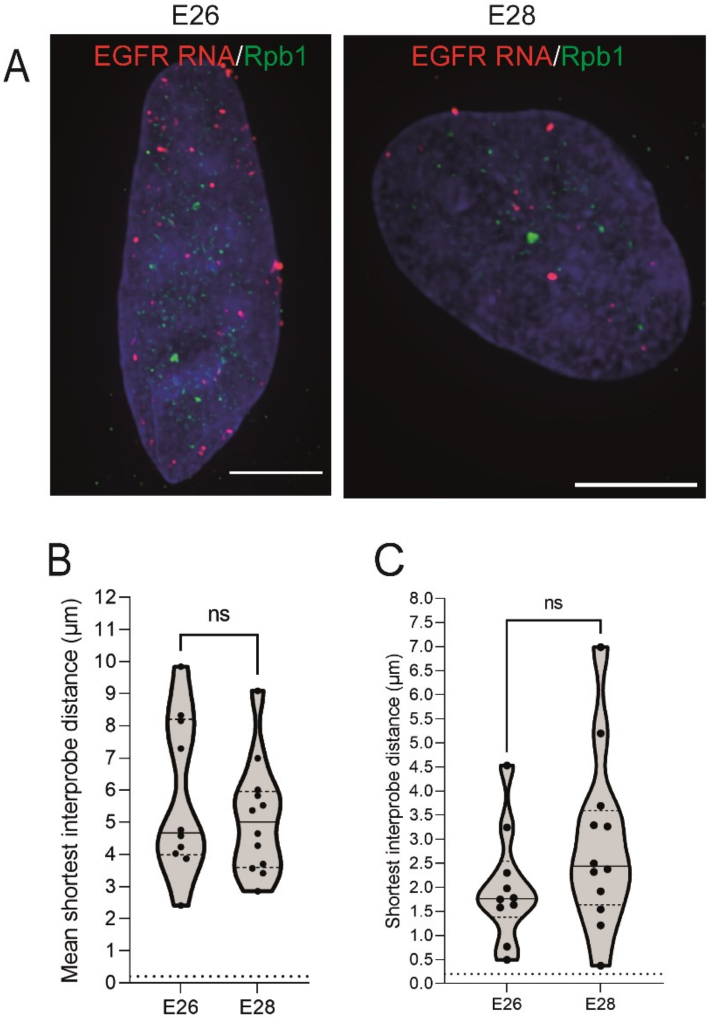


Figure 4.13 | RNA ImmunoFISH for EGFR and large Pol II hubs

A) Representative images of nascent RNA immunoFISH in E26 and E28 nuclei for EGFR (red) and Pol II (RPB1 – green). DNA (blue – DAPI). MIP, scale bar = 5 µm. B) Mean shortest distance between EGFR RNA and Pol II foci. (C) As for (B) but for single shortest distance per nucleus. Median and quartiles plotted. Dotted line denotes y = 200 nm. Statistical significance examined by Mann-Whitney test. Adapted from (Purshouse et al., 2022b). Biological replicates = 1.

4.7.1 Transcriptional hubs in an E28 *Pol2RG-mCherry* knock in cell line mirror these findings

Finally, I considered whether this could be an antibody effect by seeking an alternative means of visualising large RNA Pol II hubs. Previous work in the Pollard lab by Dr Pooran Dewari generated an E28 cell line where *mCherry* is fused by CRISPR-Cas9 knock-in to the key RNA Pol II subunit *POLR2G* at the stop codon, to create a C-terminal fusion product. These cells had been FACS-sorted by mCherry signal.

IF for mCherry confirmed the expression of RPB1-mCherry in the majority of these cells (Figure 4.14A). I validated the fusion of mCherry to RNA Pol II by co-IF for mCherry and RPB1, including visualisation of large mCherry/RNA Pol II hubs (Figure 4.14B).

I repeated DNA ImmunoFISH in the E28-*mCherry-PolR2G* cell line, probing for mCherry (IF) and *EGFR* (DNA FISH) (Figure 4.15A), and performed 3D spatial analysis as described for ecDNA-RNA Pol II foci. *EGFR*-mCherry foci distances observed were comparable to those observed by immunoFISH for *EGFR*-RPB1 (Figure 4.15B and C).

Together, these data suggest ecDNA do not colocalise with large RNA Pol II hubs to form ecDNA/RNA Pol II hubs.

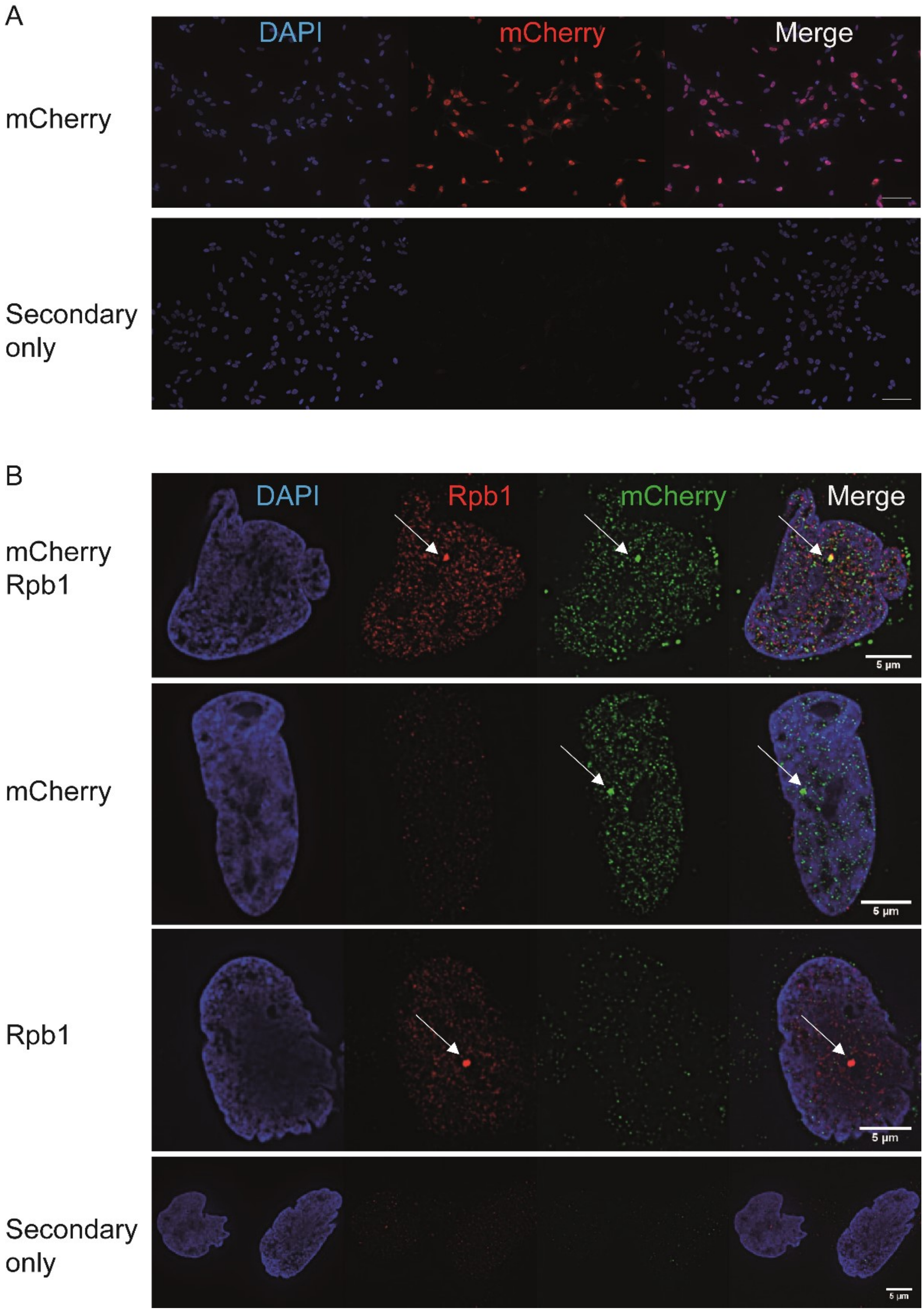


Figure 4.14 | Validation of E28-mCherry-PolR2G cell line

A) E28-mCherry-PolR2G nuclei stained by IF for mCherry (red), DNA (blue – DAPI). Scale bar – 100µm. Widefield microscopy. B) E28-mCherry-PolR2G nuclei stained by IF for RPB1 (Red) and mCherry (green), DNA (blue – DAPI). Arrows denote large Pol II hubs. Scale bar – 5µm. Biological replicates = 1.

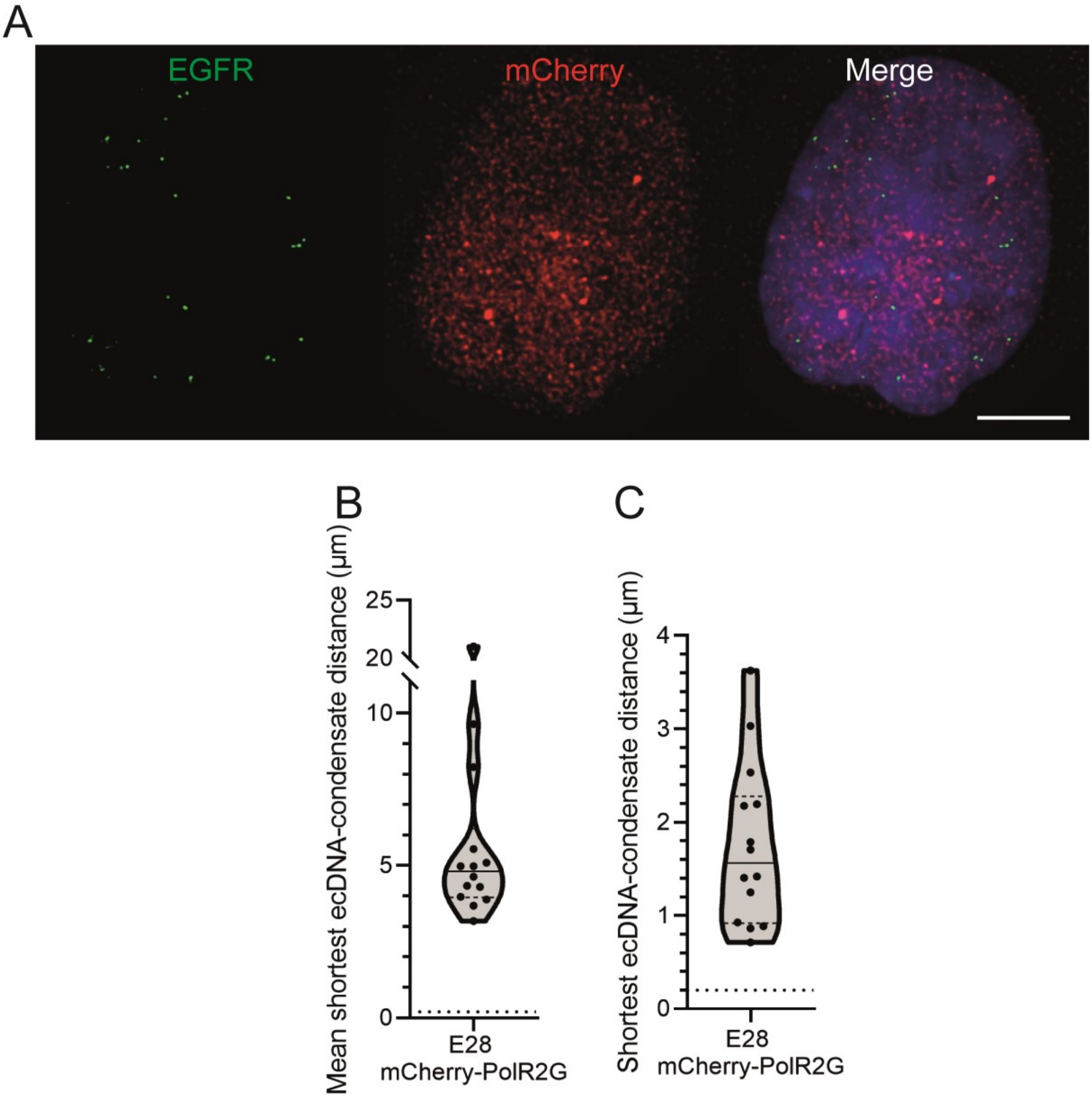


Figure 4.15 | DNA ImmunoFISH for EGFR and large Pol II hubs in E28-*mCherry-PolR2G* nuclei

A) Representative DNA immunoFISH in the E28 mCherry-POL2RG cell line: Immunofluorescence for mCherry (red) and EGFR (Green) DNA FISH. Scale bar = 5 µm. (B) Mean shortest distance per nucleus between EGFR foci and Pol II foci (visualised here by mCherry IF). Dotted line denotes y=200 nm. C) As for (B) but for shortest single distance in each nucleus. n=14 nuclei. Adapted from (Purshouse et al., 2022b). Biological replicates = 1.

4.8 Transcriptional efficiency in NSC, E26, E28 cells

4.8.1 Glioblastoma cells with ecEGFR have higher EGFR expression

It is expected that cells with more *EGFR* ecDNA would express higher levels of EGFR. To test this hypothesis, I used *EGFR* RNA FISH to quantify nascent

EGFR transcripts and compare these trends with the *EGFR* DNA copy number observed in *ecEGFR*-harbouring cell lines (Chapter 3 Section 3.3). As expected, E26 cells had the highest number of *EGFR* foci per nucleus, followed by E28. This mirrored the distribution of ecDNA copy number identified by DNA FISH in these cell lines (Chapter 3 Section 3.3).

Next, I evaluated EGFR protein expression by flow cytometry, using EGF conjugated to a fluorescent ligand (EGF-647). This enables monitoring of receptor levels on the cell surface. The median EGF-647 signal in live cells was as follows: NSC = 172.2; E28 = 985.64; E26 = 7191.81 (Figure 4.17). When live cells were gated by EGF-647 into negative (from EGF-647 negative signal), normal (from NSC signal) and elevated (from >NSC EGF-647 signal) gates, E26 harbours most cells in the elevated gate, followed by E28, with no NSCs in this gate (Figure 4.17B and C). Overall, this correlates with RNA and DNA FISH *EGFR* copy number data in these cell lines.

As a final validation that EGFR expression and ecDNA copy number are directly related, I sorted E28 and E26 cells by FACS into EGF-647 high and low fractions, aiming for EGF-647 High to represent the highest third and EGF-647 Low the lowest third of EGF-647 signal for each cell line and replated these fractions into separate cultures for expansion. A quality control sort confirmed the purity of these High and Low sorted populations (Figure 4.18A). 15 days after sorting, cells were fixed and hybridised with *EGFR* DNA FISH probes. This confirmed that the EGFR-High populations harboured more *EGFR* FISH foci than the EGFR-Low populations (Figure 4.18B and C), and confirming this is a useful method to separate cells with distinct ecDNA levels. Taken together, these data show, as expected, *ecEGFR* are associated with increased expression of EGFR.

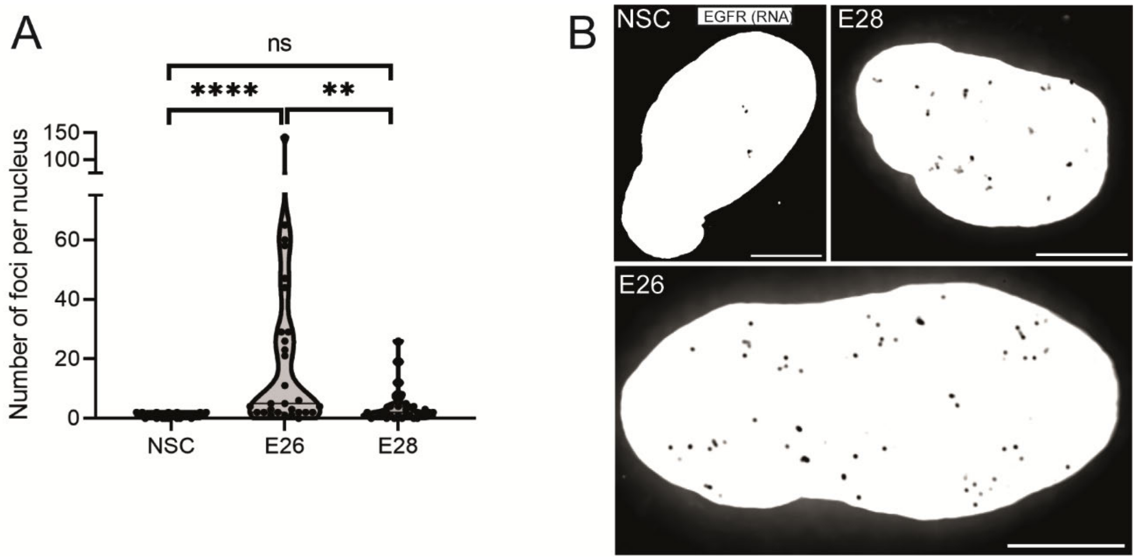


Figure 4.16 | Nascent RNA FISH in NSC, E26 and E28 nuclei

A) Number of EGFR RNA foci per cell line, at least 25 nuclei of each cell line imaged. Statistical significance examined by Mann-Whitney test. B) Representative images of nascent EGFR RNA FISH (shown in greyscale) in NSC, E26 and E28 cell lines. Scale bar = 5 µm, MIP. Adapted from (Purshouse et al., 2022b). Biological replicates = 1.

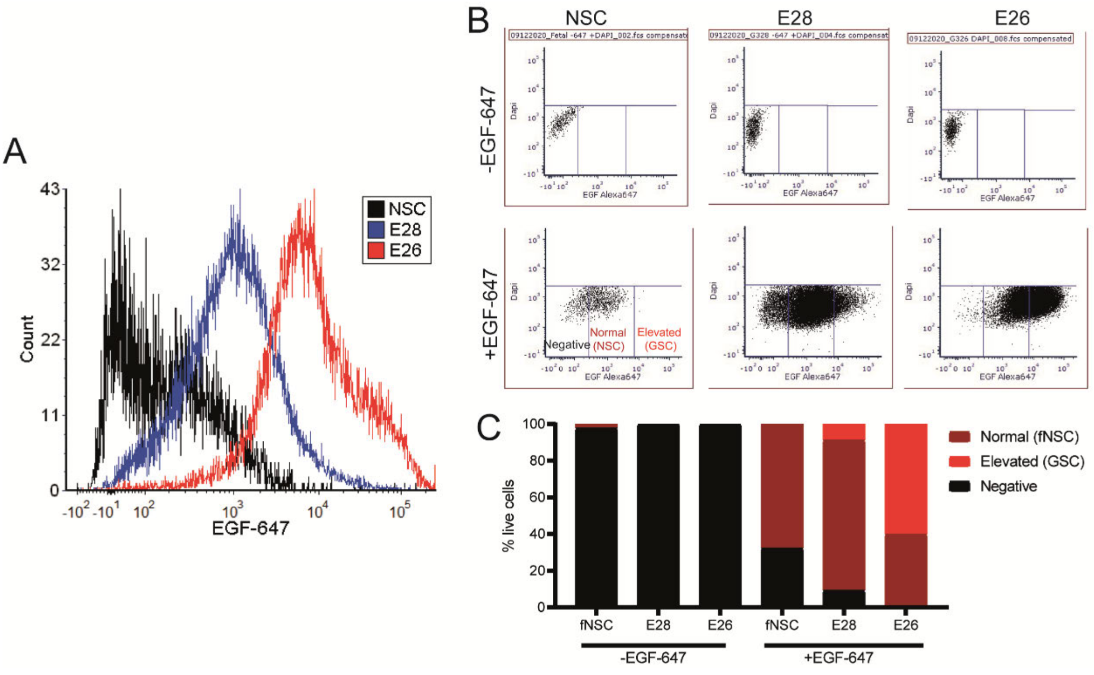


Figure 4.17 | Flow cytometry with EGF-647 in NSC, E26 and E28 cells

A) Histogram showing EGF-647 fluorescent signal (x-axis) in NSC, E28 and E26 cells from live cells, normalised to peak count (y axis). B) Flow cytometry with EGF-647; gated for negative, normal (NSC), and elevated (GSC) EGF-647 signal. C) Data in B) shown as bar chart, distribution of negative, normal and elevated EGF-647 signal gates across the three cell lines as a percentage of all live cells. Data shows one of two biological replicates. Adapted from (Purshouse et al., 2022b).

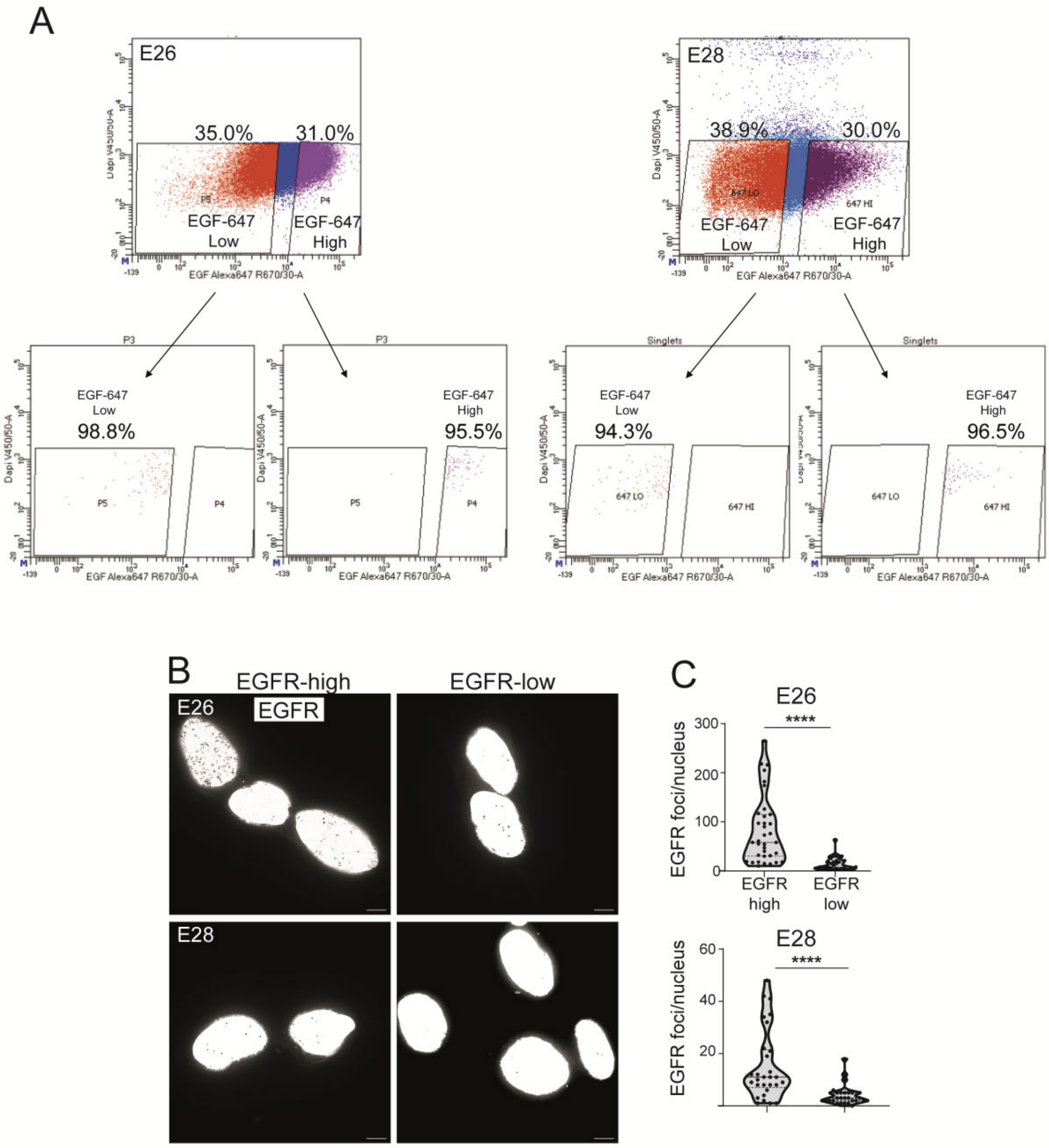


Figure 4.18 | EGF-647 FACS and DNA FISH on EGF-647 High and Low cells

A) E26 and E28 cells sorted by FACS into EGF-647 high and low populations (with % total live cell population in each sorting gate shown). Below – sort check of indicated gates for E26 and E28 confirming EGF-647 High and Low populations (arrows) and percentage of live cells. B) Representative EGFR DNA FISH images of E26 and E28 EGFR high and low cells. Scale bar = 5 µm, MIP, shown in greyscale. C) Number of EGFR DNA FISH per nucleus in E26 and E28 EGFR high and low cells. Statistical significance examined by Mann-Whitney. Adapted from (Purshouse et al., 2022b). Biological replicates = 1.

4.8.2 Copy-number normalised RNA foci number is the same between ec- and chromosomal EGFR

Next, I sought to characterise the relationship between ecDNA copy number and corresponding gene expression. Based on existing literature, I hypothesised that ecDNA gene expression would be higher than that expected by a simple linear relationship with gene copy number – indicative of an enhanced transcriptional mechanism. To test this, I performed sequential *EGFR* RNA and DNA FISH, co-hybridising for chr7 (centromeric region – CEN7) (Figure 4.19A) (n=3 biological replicates). *EGFR* RNA:DNA FISH ratio was highest in the E26 cell line, with E28 higher than NSCs but this was not significant (Figure 4.19B). This suggests differences in transcriptional efficiency between the two tumour lines.

As a first step to characterise possible differences between chromosomal and ecDNA *EGFR* transcription, I used a coarse method to categorise nuclei as harbouring mostly chromosomal *EGFR* (<10 *EGFR* DNA foci) compared with mostly *ecEGFR* (≥10 *EGFR* DNA foci). There was no difference in the RNA:DNA FISH ratio between most chromosomal and mostly ecDNA *EGFR* nuclei using these categories (Figure 4.19C). I repeated this using <6/≥6 *EGFR* DNA foci as the cut-off in E28 cells, owing to the smaller chr7 copy number in this cell line, and observed the same result (Figure 4.19D). This suggests that expression of *ecEGFR* is similar to that of chromosomal *EGFR*.

To characterise this further, I plotted the number of *EGFR* RNA foci against the number of *EGFR* DNA foci. The correlation between RNA and DNA FISH foci copy numbers demonstrated a linear relationship between *EGFR* copy number and gene expression in both cell lines (Figure 4.20A). To control more accurately for chr7 copy number, I probed for chr7 by DNA FISH (centromere 7 (CEN7) foci - Figure 4.19A) alongside *EGFR* RNA and DNA FISH. For each nucleus I calculated the RNA:DNA ratio as the number of RNA foci divided by the number of DNA foci. I plotted this against the proportion of ecDNA in the corresponding nucleus (EcDNA proportion = (number of *EGFR* DNA foci – number of CEN7 foci)/number of *EGFR* DNA foci). There was no correlation between the RNA:DNA ratio and the proportion of ecDNA in each nucleus (n=3 biological replicates) (Figure 4.20B).

These data suggest that the relationship between ecDNA copy number and gene expression is linear, with no increased transcriptional efficiency of genes present on ecDNA as compared with their chromosomal counterparts.

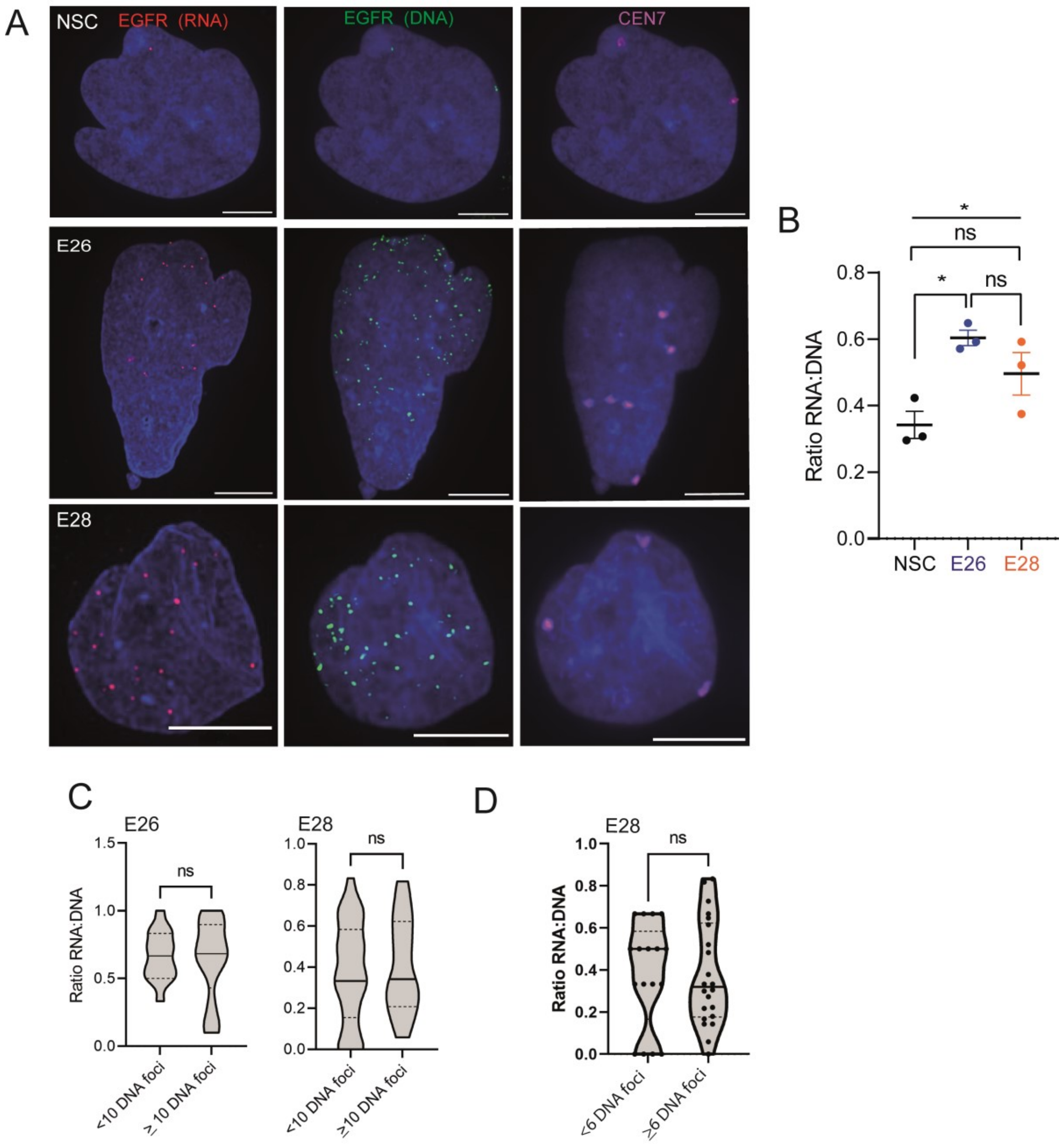


Figure 4.19 | Transcriptional efficiency visualised by RNA:DNA FISH

A) Representative images of nascent EGFR RNA, EGFR DNA, and centromere 7 (CEN7) DNA FISH in NSC, E26 and E28 cell lines. Scale bar = 5 µm, MIP. B) Ratio of RNA:DNA foci per nucleus in NSC, E26 and E28 cell lines. Flat line – one-way ANOVA, hooked lines – unpaired t-test. Mean and standard error of the mean (SEM) plotted, with 3 biological replicates for NSC (total n=67), E26 (98) and E28 (95) nuclei. C) Violin plot of ratio of RNA:DNA foci per nucleus, nuclei categorised to primarily chromosomal EGFR (<10 EGFR DNA foci) and primarily ecDNA EGFR (≥10 EGFR DNA foci). D) As for C) but using <6/≥6 cut-offs for the E28 cell line. Adapted from (Purshouse et al., 2022b) and associated pre-print, (Purshouse et al., 2022a)

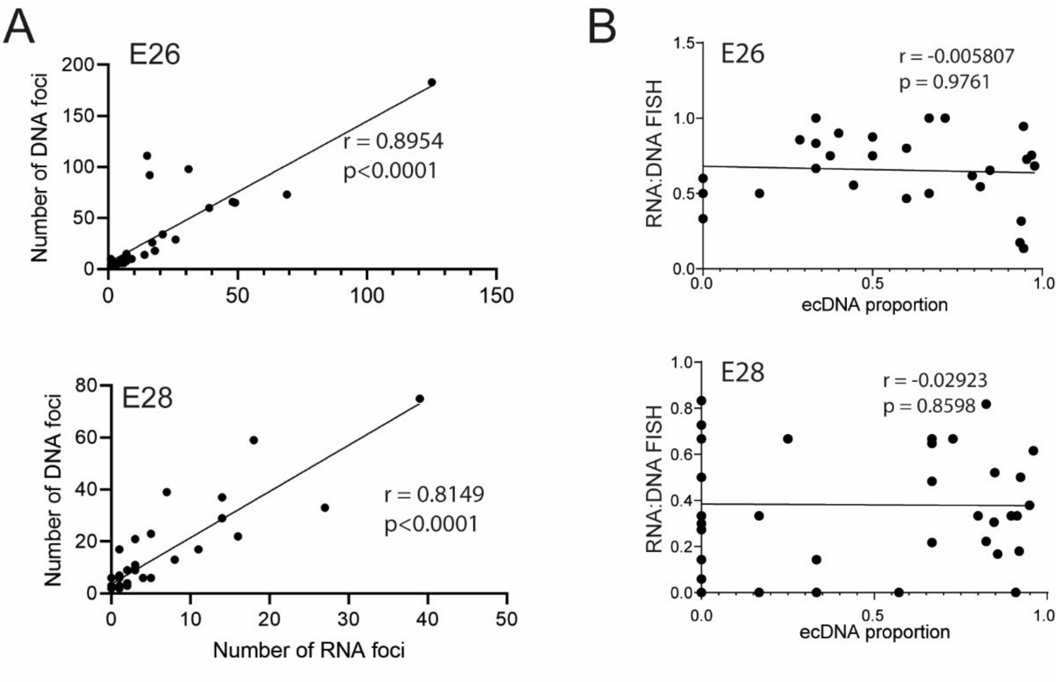


Figure 4.20 | Correlations between RNA and DNA FISH foci

A) Representative Spearman r correlation (ρ) and p-values between number of RNA foci (x axis) and number of DNA foci (y axis) in E26 and E28 cells. B) Representative Spearman r correlation (ρ) and p-values shown for E26 and E28 cells between RNA:DNA FISH and proportion of ecDNA. Three biological replicates performed, data from replicate 1 shown here. Adapted from (Purshouse et al., 2022b) and associated pre-print, (Purshouse et al., 2022a)

4.8.3 RNAseq/WGS analysis confirms no difference in EGFR transcription between ec- and chromosomal EGFR

Given the divergence of my findings from previously published data, I sought to evaluate this via an independent method. Bulk RNAseq (tumour) and WGS (tumour and blood) data were available for the cell lines studied in this thesis, including E26. As described in Chapter 3 Section 3.3, the ecDNA in E26 cells harbour *EGFRvIII*, which has an exon 2-7 deletion (Figure 4.21A). This can be utilised to compare the transcriptional efficiency of *EGFR* on ecDNA vs chromosomes, because the RNA transcripts from exons 2-7 can only be derived from chromosomal *EGFR*. By contrast, transcripts from *EGFR* exons 1, 8-28 are predominantly transcribed by extrachromosomal *EGFR* due to the copy number amplification generated by ecDNA. AA allows WGS regions to be defined by copy number variant (CNV) regions, including the region of exon 2-7. Further detail on AA methods can be found in Section 2.11.3.

I calculated the RNA count for each *EGFR* exon and normalised by exon size. These were in turn normalised by WGS counts normalised for each region size as defined by AA. These copy number and exon-size normalised RNA seq counts were compared between exons 2-7 (chromosomal exons) and exons 1,8-28 (predominantly ecDNA exons). They were not significantly different (Figure 4.21B).

I repeated this analysis using published RNAseq, WGS and AA data for another glioblastoma cell line - GBM39 - harbouring an *EGFR* exon 2-7 deletion (Figure 4.21A) (Wu et al., 2019). Normalised RNA counts were not significantly higher in ecDNA exons and were slightly lower compared to chromosomal exons (Figure 4.21C).

A simpler way of performing this analysis is to normalise the RNA read counts and DNA read counts in each *EGFR* exon, which automatically accounts for exon and AA region size for RNAseq counts and WGS read counts respectively. Normalised RNAseq counts in predominantly ecDNA exons were not significantly different from chromosomal exons in E26 (Figure 4.22A) and GBM39 cell lines (Figure 4.22B).

To confirm these new data were comparable, I performed correlation analysis on the normalised RNA counts for the two different strategies (AA regions - Figure 4.21B and C vs direct exon comparison - Figure 4.22A and B). This showed a positive correlation between normalised RNA counts in ecDNA or chromosomal exons calculated using the AA method and the direct exon comparison method (Figure 4.22C-F).

Together, these data show that the transcriptional efficiency of ec*EGFR* is not significantly different to that of the corresponding chromosomal *EGFR* oncogene.

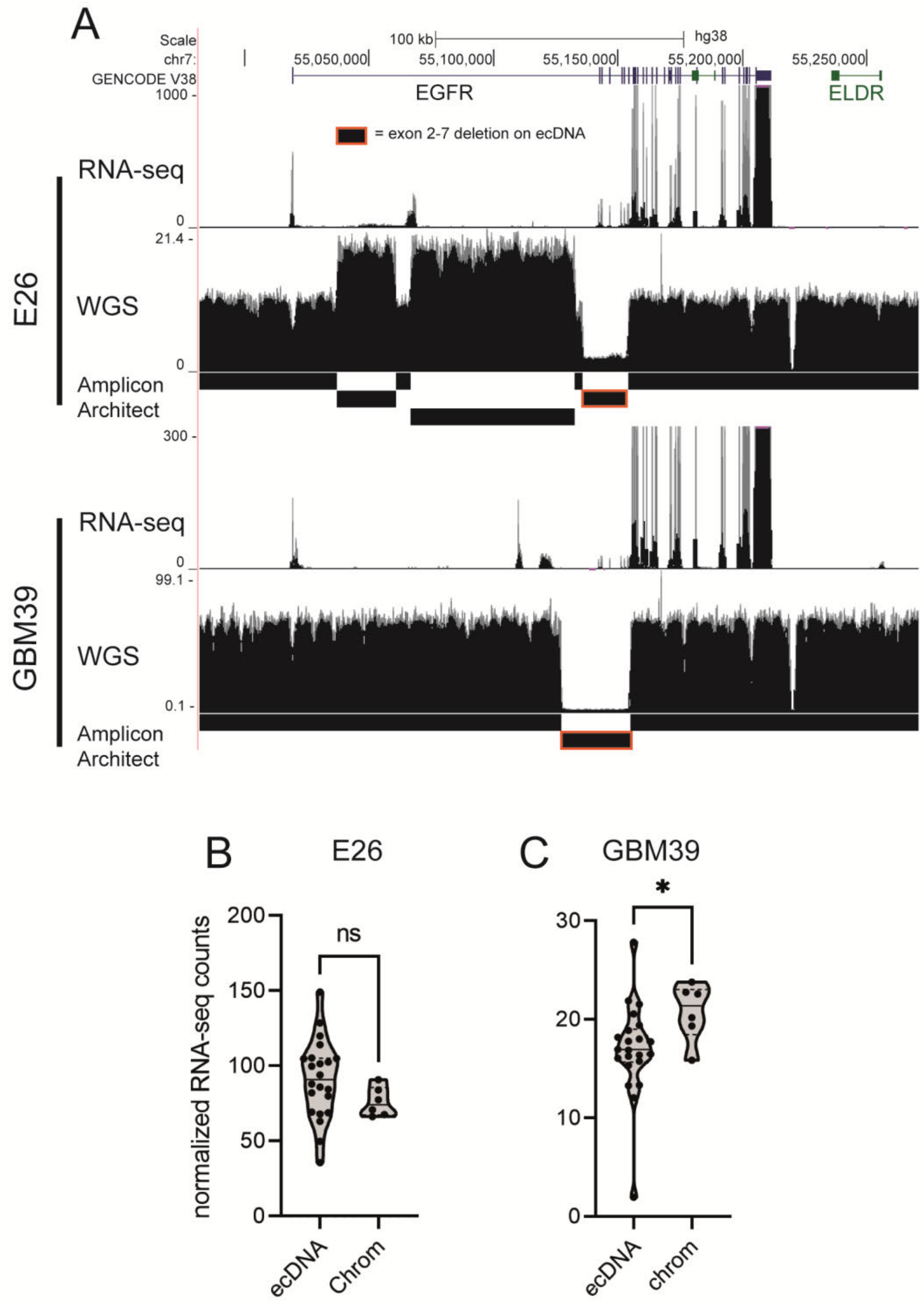


Figure 4.21 | Normalised RNA counts in EGFR exons (chromosomal vs predominantly ecDNA) via exons and AA regions

A) UCSC Genome browser tracks showing RNA-Seq, WGS and AA tracks for E26 and GBM39 cell lines, with region of AA-predicted exon 2-7 deletion highlighted in red. EGFR (GENCODE) exons shown. Note to aid visualisation of all exons, RNAseq counts in some ecDNA regions exceed that of the scale. B) Exon-size normalised RNA-seq counts normalised to WGS counts in AA regions (WGS counts per region, normalised by region size). EcDNA = exons 1, 2-28; chrom = exons 2-7. Statistical significance examined by Mann-Whitney test. Adapted from (Purshouse et al., 2022b) and associated pre-print, (Purshouse et al., 2022a)

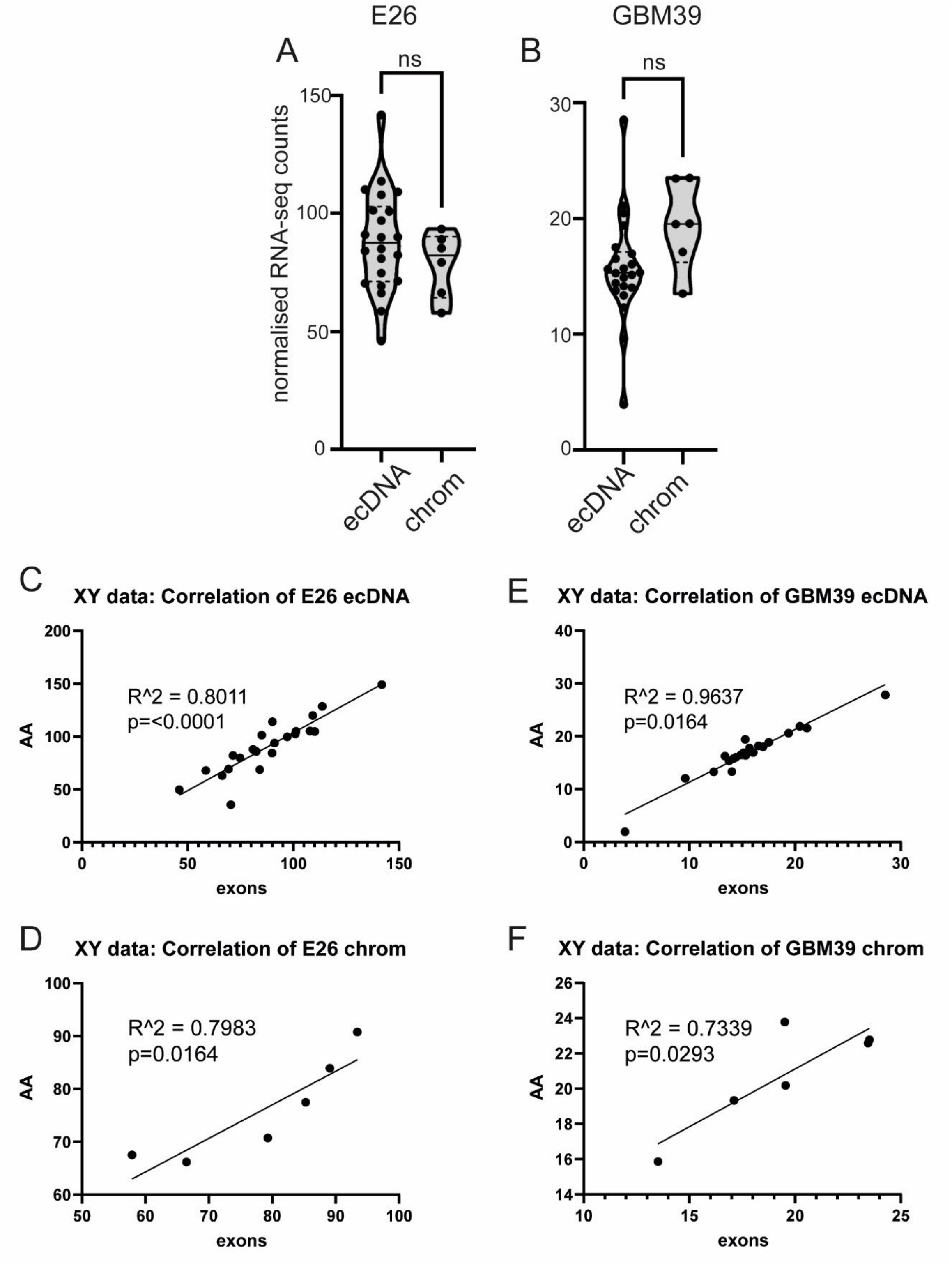


Figure 4.22 | Normalised RNA counts by direct exon RNA/WGS count

A) EGFR RNA-seq counts in E26 exons normalised by WGS read counts. Statistical significance examined by Mann-Whitney test. (B) As for (A) but for GBM39. C) Pearson correlation of normalised RNA counts in E26 cell line calculated using exon size-normalised RNAseq counts and WGS read counts normalised by AA region (AA – y axis) versus direct exon normalisation (exons – x axis). Shown for predominantly ecDNA exons (1, 8-28) D) as for C) but for chromosomal (chrom) exons (2-7). E) As for C) but for GBM39 cell line. F) As for D) but for GBM39 cell line. Adapted from (Purshouse et al., 2022b)

4.8.4 SNP analysis confirms comparable ec- and chromosomal transcription

A further approach is to calculate the allele frequency of polymorphic sites in the amplified sequence. Genomic events that give rise to ecDNA have been shown to occur in only a single parental copy of the chromosome, with the subsequent conclusion that ecDNA is derived from one parental homolog (Stephens *et al.*, 2011; Hung *et al.*, 2022). Therefore, analysis of polymorphic sites in the amplified sequence can capitalise on this without relying on mutations in *EGFR*, as was required in the *EGFRvIII*-based analysis. This is informative as co-amplified genes may not be under the same positive selection pressures as oncogenes, which drive cancer growth and progression via a range of up- and downstream pathways and therefore accumulate at higher rates than non-driver passenger mutations. We can perform this analysis in both *ecEGFR* cell lines (E26 and E28) as *EGFR* variant status is irrelevant and patient control (blood) WGS data is available.

This SNP analysis was performed by Dr Elias Friman, a Postdoctoral Fellow in the Bickmore lab.

Germline variants were called in patient control (blood) WGS samples in the amplicon region at 40-60% allele frequencies. As expected with the amplicon region being derived from one parental homolog, the amplicon frequencies of these SNPs were predominantly >80% in the corresponding glioblastoma WGS samples (Figure 4.23A). Exon-overlapping SNPs in the amplified region were used to calculate allele frequencies in WGS and RNAseq samples. In E26, *EGFR* and *LANCL2* were the only sufficiently expressed genes containing SNPs. In E28, there were polymorphisms in *VSTMA, EGFR* and *LANCL2.* The WGS allele frequencies for all these overlapping polymorphisms were >88%, again indicating that SNPs located on ecDNA originate from a single allele of the endogenous locus.

The RNAseq to WGS ratio was calculated for each SNP. An RNAseq:WGS allele frequency ratio >1 would indicate enhanced amplicon transcription. The ratio was approximately 1 for all SNPs, with the highest being 1.05 (SNP in *VSTMA*, E28 cells) and the lowest being 0.78 (SNP in *LANCL2*, E28 cells)

(Figure 4.23B). The latter is likely due to only a portion of the *LANCL2* gene being present on the amplicon, resulting in a truncated transcript.

Overall, both WGS and imaging analysis indicate that genes present on ecDNA are transcribed at a similar efficiency to those on chromosomes. Amplified gene expression is driven primarily by copy number amplification rather than augmented transcriptional mechanisms.

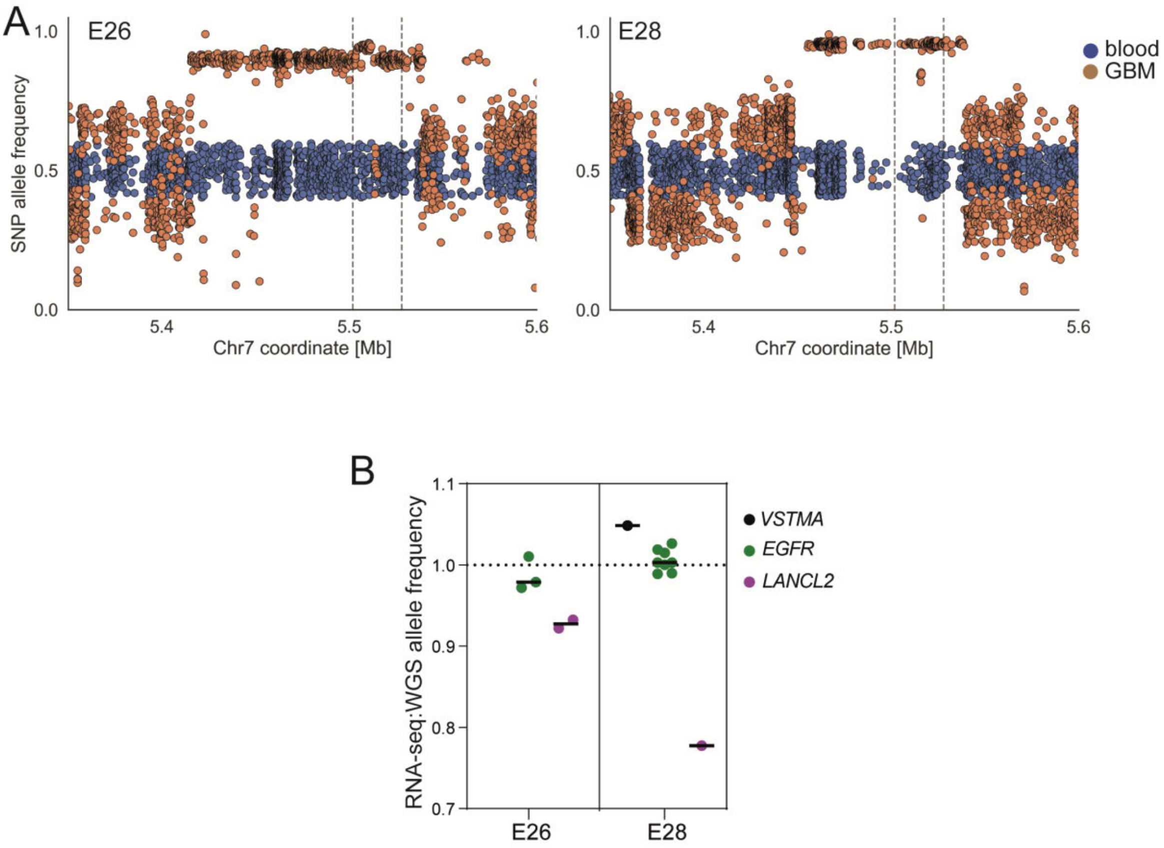


Figure 4.23 | SNP allele frequencies in E26 and E28 amplicons

A) SNP allele frequencies in E26 and E28 cell lines plotted in control (blood - blue) and glioblastoma (orange) WGS. Dotted lines = start/end of EGFR gene. B) RNA-seq/WGS allele frequency ratio for SNPs overlapping with expressed exons in the amplicon. Lines denote median values, dotted line denotes allele frequency of 1. Adapted from (Purshouse et al., 2022b). Analysis by Dr Elias Friman.

4.9 Similar enhancer/promoter proximity between ecDNA and chromosomal loci

Previous studies propose that ecDNA are associated with highly accessible chromatin (Wu *et al.*, 2019) yet I have shown that the transcriptional efficiency of their resident genes is comparable with their chromosomal counterparts. EcDNA from the E26 and E28 cell lines appear to carry the cognate enhancers for *EGFR* (Figure 4.24). To determine whether enhancer-promoter proximity might be different on the small circular ecDNA molecules compared to the endogenous chromosomal locus, I sought to quantify the spatial relationship between *EGFR* enhancers and promoters on ecDNA.

I defined the genomic regions to which DNA FISH probes would hybridise. I used Genecards to identify *EGFR* promoter regions, and data from Morton et al., 2019 to plot two functional *EGFR* super-enhancer regions reported in glioblastoma. The two super-enhancer regions are 128kb and 84kb from the *EGFR* promoter. Fosmids were selected that cover these regions (enhancer – M9; promoter – M3), with a 103,691bp gap (~100kb) separating the two (Figure 4.24).

DNA FISH on E26 metaphase spreads confirmed these enhancer (green)/promoter (red) regions were localised to ecDNA. Red and green foci colocalised, suggestive of ecDNA harbouring both the enhancer and promoter region. However, there were also foci that appeared clearly separate, and overall there were more promoter (red) foci than enhancer (green) foci (Figure 4.25A). It is notable that the WGS profile is suggestive of a further amplification over the ecDNA region to which this promoter FISH probe hybridises (Figure 4.24), which, taken together with the metaphase spread data, implies a subpopulation of ecDNA harbouring only the promoter segment.

I repeated this in E28 cells, where the WGS profile across the amplified region is more uniform (Figure 4.24). While a mixture of colocalised and separate foci were visible, the ratio of enhancer:promoter foci appeared more equal, suggesting this may be a better model for ecDNA enhancer:promoter distance calculation (Figure 4.25B).

3D DNA FISH was then performed on nuclei from E28 and NSC cells. In addition to the enhancer and promoter probes, a probe for centromere 7 (CEN7) was also used to allow identification of E28 nuclei without ecDNA - i.e. all EGFR signals are associated with chromosome 7 (Figure 4.26A). After spot calling (Imaris), the shortest distance from each enhancer (green) foci to a promoter (red) locus was calculated in Imaris (Figure 4.26B). When all distances from E28 and NSC cells were compared, there was no significant difference between the shortest enhancer/promoter distances (Median distance in µm – E28 = 0.37; NSC = 0.35) (Figure 4.26C).

To ensure this wasn't skewed by ecDNA where the enhancer and promoter were not on the same ecDNA (and therefore further apart), this analysis was repeated including only E28 distances that were shorter than the longest distance observed in the NSC nuclei. There remained no significant difference in the enhancer/promoter distances (Median distance in µm – E28 = 0.36; NSC = 0.35) (Figure 4.26D).

As a further control, a proportion of E28 nuclei (9/32) appeared to have no ecDNA, so all enhancer/promoter foci could be attributed to the chromosomal locus. The shortest enhancer/promoter distances in E28 cells with only chromosomal *EGFR* were compared to those harbouring *ecEGFR* (Median distance in µm – E28 ecDNA = 0.38; E28 chrom only = 0.30) (Figure 4.26E and F). An additional analysis included only ecDNA-containing E28 enhancer/promoter distances that were shorter than the longest distance observed in the chromosome-only *EGFR* E28 nuclei (Median distance in µm – E28 ecDNA = 0.33; E28 chrom only = 0.30) (Figure 4.26G). In both cases, there was no significant difference in enhancer/promoter distances, although shorter distances (<100nm) were observed in E28 nuclei harbouring ecDNA.

These data suggest that previously proposed *EGFR* super-enhancer and promoter regions share a similar spatial organisation whether they are on ecDNA or chromosomes. Even if these super-enhancer and promoter regions are not ubiquitous to these cell lines, this still suggests that chromatin organisation overall is similar between ecDNA and chromosomes.

Altogether, the data in this chapter suggests that the spatial organisation and transcriptional regulation of ecDNA is similar to that of chromosomes in glioblastoma stem cells.

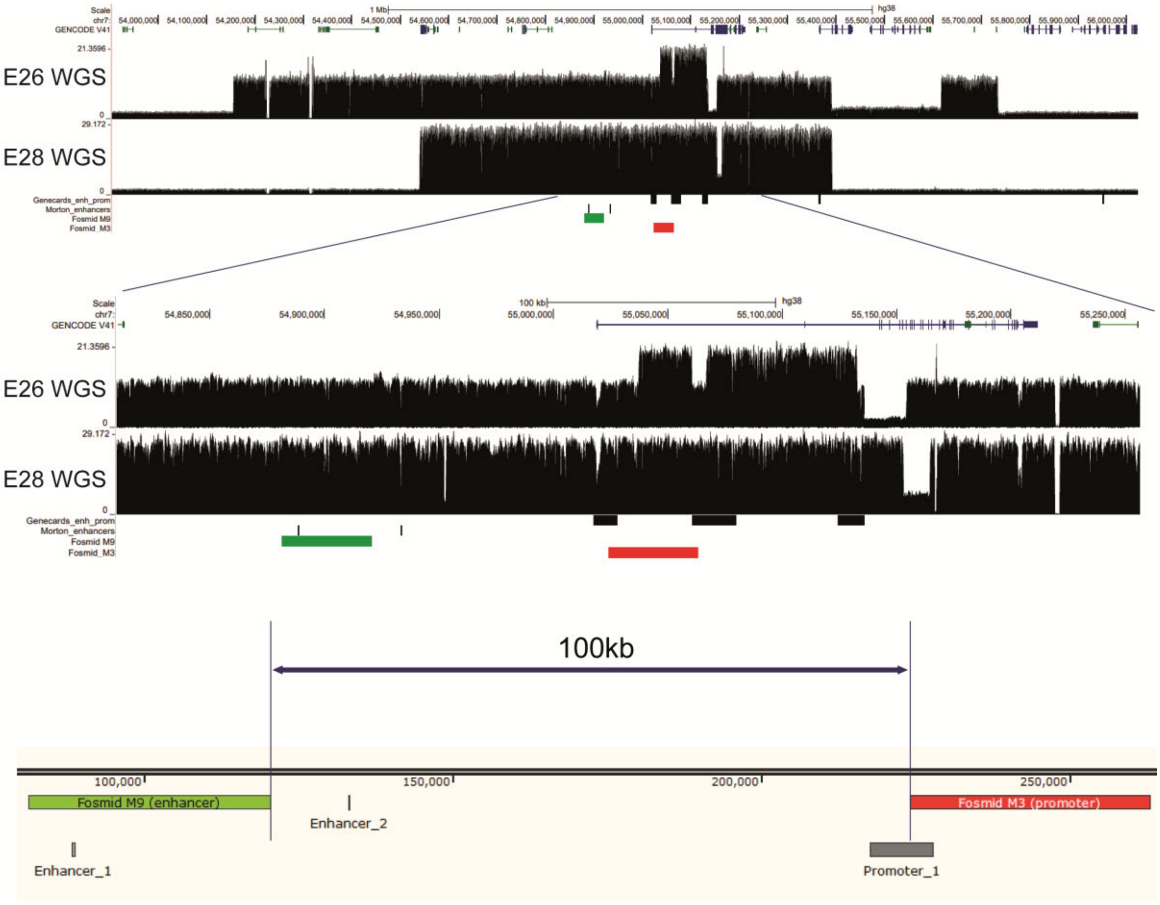


Figure 4.24 | Enhancer/promoter regions and corresponding DNA FISH probes

UCSC Genome Browser tracks showing E26 and E28 WGS, Genecards enhancer/promoter regions (Genecard_enh_prom), EGFR super-enhancers described by Morton et al., 2019 (Morton_enhancers), Enhancer DNA FISH probe (Fosmid M9 - Green) and Promoter DNA FISH probe (Fosmid M3 - Red). Inset shows region magnified. Snapgene visualisation to indicate Fosmid overlap between enhancer and promoter regions. The fosmids are ~100kb apart.

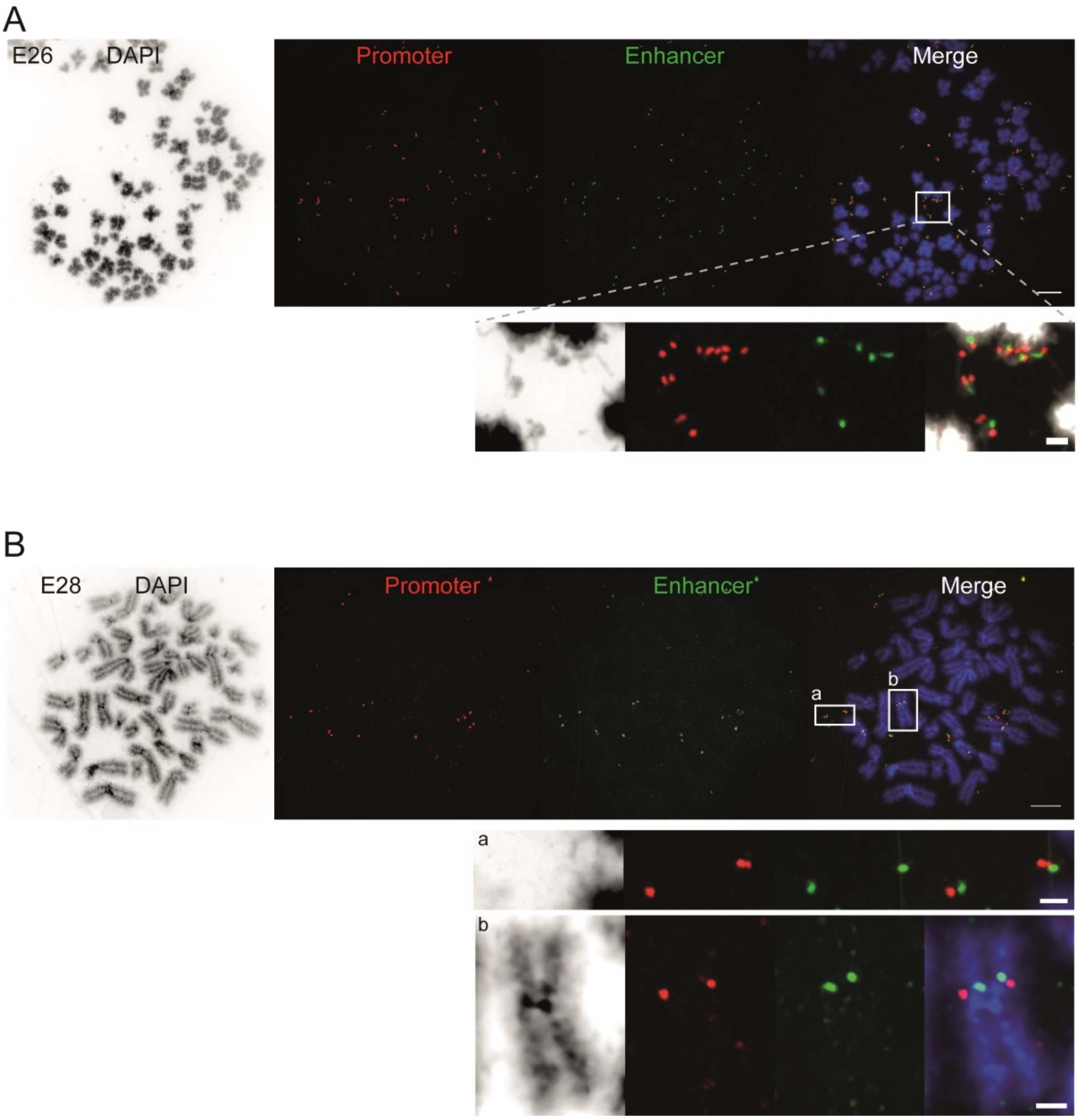


Figure 4.25 | E26 and E28 metaphase spreads with DNA FISH for enhancer and promoter regions

A) Representative metaphase spread of E26 cell line following hybridisation by DNA FISH for promoter (red) and enhancer (green) regions. DNA – blue (DAPI). Inset indicating multiple promoter foci relative to enhancer foci. B) As for A) but for E28 cell line, inset a) showing ecDNA with promoter/enhancer signals, inset b) as for a) but for chr7. Scale bars = 5µm (insets = 1µm). Biological replicates = 1.

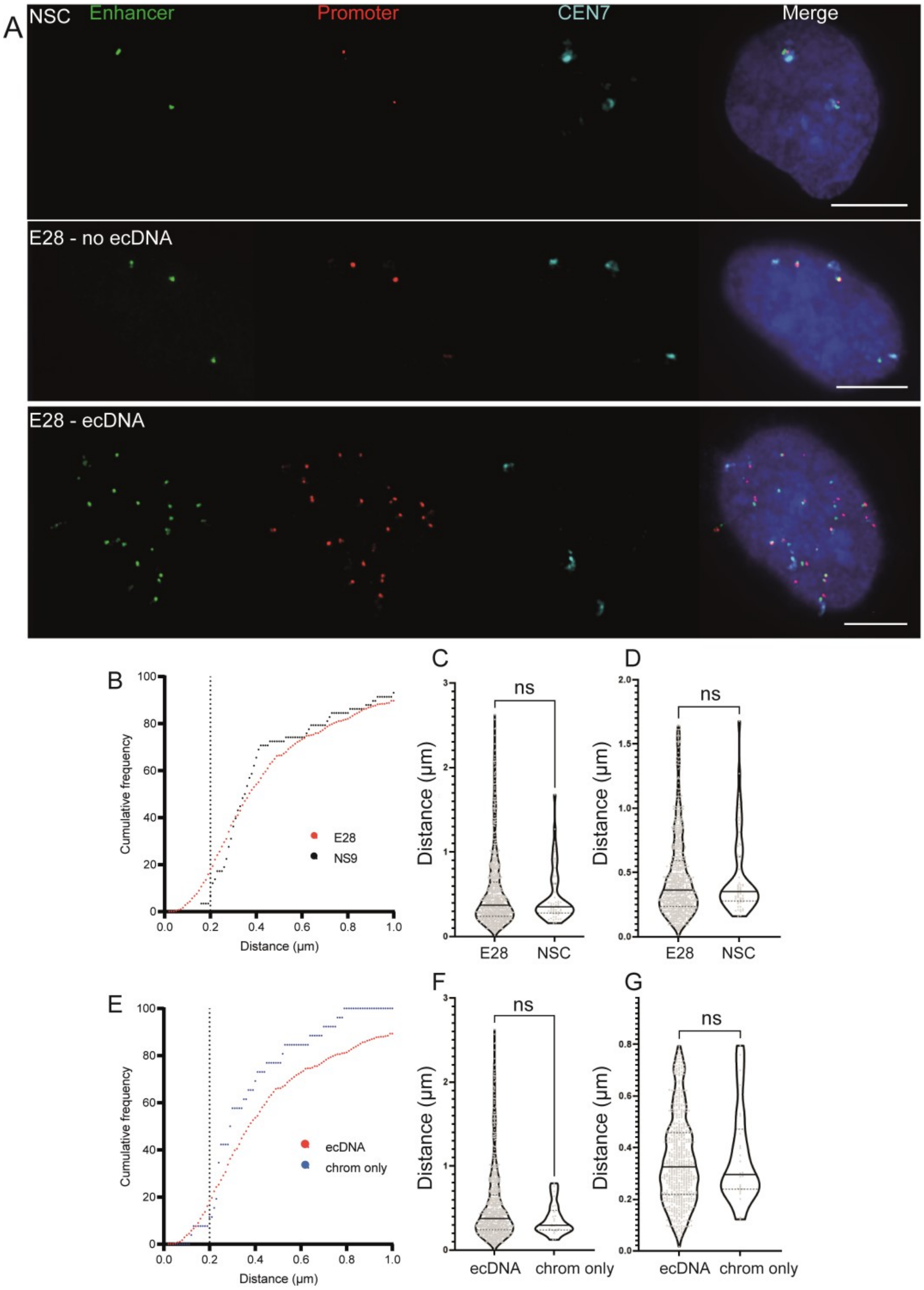


Figure 4.26 | Enhancer/promoter distances in E28 interphase nuclei

A) Representative images of DNA FISH of NSC and E28 nuclei (no ecDNA vs with ecDNA), Enhancer (enh - green), Promoter (prom - red), Chr7 (CEN7 – cyan), DNA = blue (DAPI). Scale bar = 5µm, MIP. B) Cumulative frequency distribution of shortest enh-prom distances across all E28 and NSC nuclei. Dotted line = 200nm. C) Shortest distance from each enh to prom locus in E28 (n=32) and NSC (n=29) nuclei. D) As for C) but for distances shorter than the longest NSC interprobe distance. E) As for B) but for E28 nuclei harbouring ecDNA (n=23) or no ecDNA (chrom only; n=9). F) As for C) but for E28 nuclei with ecDNA or chrom only. G) As for F) but for distances shorter than the longest interprobe distance in E28 cells without ecDNA. Violin plots show median and interquartile range. Statistical significance tested by Mann-Whitney test. Biological replicates = 1.

4.10 Discussion

EcDNA represent a key mechanism of oncogene amplification, with links to prognosis and treatment resistance (Nathanson *et al.*, 2014; Kim *et al.*, 2020). Analysis of chromatin accessibility using tools including ATACseq, MNase seq and ATAC-see have demonstrated that ecDNA harbour highly accessible chromatin (Wu *et al.*, 2019; Kim *et al.*, 2020). In addition, H3K4me1/H3K27ac ChIP-Seq data have proposed the presence of active, and absence of repressive, chromatin marks (Wu *et al.*, 2019). ChIA-PET and ChIA-Drop analysis further indicated interactions between ecDNA, chromosomes and RNA Pol II, proposing the role of ecDNA as a transcription-factor-like trans-activator (Zhu *et al.*, 2021). Taken together, these have led to the hypothesis of 'ecDNA hubs', where ecDNA and RNA Pol II form key interactions facilitated by this accessible chromatin conformation. Studies of established tumour cell lines have suggested that ecDNA cluster into hubs and colocalise with Pol II, thereby driving amplified oncogene transcription beyond that expected by copy number alone (Hung *et al.*, 2021; Yi *et al.*, 2021). This is an appealing hypothesis as it might provide enhanced transcriptional mechanisms to cancer cells, and these might be therapeutically targetable. However, our data provide no indication that this is occurring in the primary glioblastoma lines studied here.

I observed a more central nuclear distribution of *ecEGFR* than chromosomal *EGFR* in glioblastoma stem cells. This central distribution is known to correspond with an actively transcribing chromatin state (Croft *et al.*, 1999; Boyle *et al.*, 2001), and indeed previous data suggest ecDNA carry marks and characteristics of accessible chromatin (Wu *et al.*, 2019; Kim *et al.*, 2020; Zhu *et al.*, 2021). It highlights the freedom ecDNA enjoy from chromosomes and their associated chromosome territories, such as chr7 which has a more peripheral organisation (Boyle *et al.*, 2001; Mahy, Perry and Bickmore, 2002; Kalhor *et al.*, 2011).

My data support a regional organisation of ecDNA in glioblastoma stem cells, but where ecDNA do not cluster at distances expected to be associated with coordinated transcription ($\leq$200nm). This likely reflects the non-random organisation of the nucleus into different compartments and chromosome territories, around which ecDNA must be arranged. This analysis was designed to control for ecDNA number and nuclear size. In addition, I analysed single and dual oncogene-harbouring ecDNA glioblastoma cell lines as a means of controlling for two FISH foci labelled with the same fluorophore being too close together to be resolved as separate structures. I reanalysed the raw data with a smaller spot size and used super-resolution imaging which has an optic resolution of approx. 120nm as further controls. These data also suggest that previous concerns about 2D Ripley's K function and an association with over-counting (Veatch *et al.*, 2012) could not be seen in this 3D analysis, on the basis that we did not observe over-counting (which would manifest as clustering).

Comparison with previous data purporting to show ecDNA clustering is challenging owing to differences in imaging quality and analysis approach. Hung *et al.*, 2021 used a confocal microscope to image fixed nuclei with ecDNA-resident oncogenes labelled by DNA FISH, with Z stacks across approximately 8 µm (0.6 µm z step size), resulting in images where individual ecDNA cannot be defined (Figure 1.9). In contrast, I used the SoRa microscope to image nuclei across 3 µm in 100nm z steps. Previous analysis for clustering used an autocorrelation function which appears to have 2D data (either slices or as 2D images from combined 3D slices - this is not clear from the methods) and assigns a random distribution as a control, with output defined in pixels (Hung *et al.*, 2021). In another study where live-cell imaging was used, ecDNA (and ecDNA/Pol II) clustering is defined by visual (i.e. subjective) calling of fluorescent signal overlap or close proximity (Figure 1.10) (Yi *et al.*, 2021). The 3D analysis tool described here enables true characterisation of the spatial organisation of ecDNA within clearly defined distances, avoiding the limited data provided by 2D analysis, false positive clustering effects by 2D analysis of 3D images (e.g. MIP) or subjective calling of colocalisation. Existing cluster analysis strategies also do not account for chance colocalisation owing to increased copy number in a given space - the closest equivalent is the use of repetitive telomere targeting to reflect high copy number (Yi *et al.*, 2021). Ripley's K analysis allows for the ecDNA copy number and nuclear size to be controlled for in each individual nucleus.

The E20 cell line harboured a small number of colocalised *CDK4-PDGFRA* foci (~10%) which were present in similar proportions in metaphase spreads and 3D nuclear analysis. This suggests that this colocalised population represented a subclone of hybrid ecDNA too small to be identified in the bulk WGS performed to date, but can be confidently verified by DNA FISH of metaphase spreads. It also represents a positive control, giving further confidence in the ability of Ripley's K to identify colocalisation. Another positive control could have included adjacent DNA FISH probes on the same ecDNA. A dual-oncogene ecDNA cell line previously used as evidence of ecDNA colocalisation (Hung *et al.*, 2021) has subsequently been shown to harbour a subclone of oncogene-hybrid ecDNA via ecDNA enrichment, long read sequencing and reconstruction (Hung *et al.*, 2022). Confirmation of these hybrid *CDK4-PDGFRA* ecDNA is likely to require such an approach. Overall, analysis of two single-oncogene, and two dual-oncogene, ecDNA-harbouring glioblastoma cell lines suggest that close clustering is not a key feature of ecDNA.

A key difference between my data and previous reports of ecDNA clustering is the use of primary patient-derived cell cultures, whereas other studies have focused primarily on established cell lines (Hung *et al.*, 2021; Yi *et al.*, 2021). Such established cell lines may differ in terms of the size of the ecDNA, or the number of oncogene loci per ecDNA as outlined above. I attempted to reproduce my analysis in the widely-studied Colo320DM cell line, which has previously been reported to harbour clustered ecDNA (Hung *et al.*, 2021). My analysis of Colo320DM suggests this cell line is not well suited to oncogene-oriented cluster analysis. DAPI staining and DNA FISH for *MYC* on metaphase spreads showed that ecDNA in Colo320DM nuclei are large doublets approximately 1-2um in size and harbour multiple copies of *MYC* per ecDNA doublet. Consistent with this, Colo320DM ecDNA have previously been reconstructed at 4.328Mb in size, measuring approx 1.75µm in diameter via imaging and harbouring 3 copies of *MYC* following detailed multimodal sequencing-based reconstruction (Wu *et al.*, 2019; Hung *et al.*, 2021). As such, clustering of *MYC* hybridisation signals is inevitable owing to their structural colocalisation on the same DNA molecules. Previous studies defined clustering in terms of pixels rather than µm without defined cut-offs or definitions of clustering (Hung *et al.*, 2021) so any results would be difficult to compare. The GCGR cell lines presented previously are, however, more representative of the expected size of ecDNA at 1-3Mb

(Verhaak, Bafna and Mischel, 2019), suggesting that this clustering tool is directly relevant to primary cells. Cell lines with ecDNA in this size range only host one oncogene copy per ecDNA, as suggested by the data presented on these GCGR cell lines in Chapter 3. For example, the *EGFR*-harbouring ecDNA in the GBM39 cell line have been reconstructed following CRISPR-CATCH isolation, and although several ecDNA structures are proposed, each harbours only a single copy of *EGFR* (Hung *et al.*, 2022). In addition, the Colo320DM cell line represents a highly atypical cell line (Quinn *et al.*, 1979), and, given their duration in culture, may further behave differently to primary cells such as those presented here. Colo320DM cells have formed the basis for many studies of ecDNA biology; however, these data suggest their relevance to the true dynamics of ecDNA must be approached with caution.

It may be more relevant to compare these findings with other glioblastoma studies. HK359 is a glioblastoma cell line previously noted to have clustered ecDNA hubs (Hung *et al.*, 2021). HK359 has a 42kb insertion at the site of *EGFRvIII* (exon 2-7 deletion), again suggesting the presence of larger and atypical ecDNA (Koga *et al.*, 2018). A live cell analysis utilised a primary glioblastoma neurosphere culture grown in serum-based neurosphere media, HF3016, tagging *EGFR*-ecDNA breakpoints using the Casilio system to track ecDNA dynamics (Yi *et al.*, 2021). The Casilio system generates large aggregates of varying size, suggesting clumping of fluorescent elements resulting in larger-than-expected loci and increasing the chance of off-target fluorescent signal (Clow *et al.*, 2022). When observed in ecDNA, larger loci have been suggested to represent ecDNA hubs, with dual-colour ecDNA labelling of *ecEGFR* breakpoints used to subjectively call instances of colocalisation, although this does not appear consistent (Figure 1.10) (Yi *et al.*, 2021). Super-resolution microscopy of fixed cells and more accurate quantitative and statistical analysis provides more conclusive data, with a clear definition of clustering at distances relevant to gene function.

Live cell data has been proposed as a superior means of quantifying dynamic ecDNA interactions due to the transient nature of hubs. Using the Casilio system described above, hubs were observed in approximately 50% of cells over 48 hours, with MIP analysis (i.e. 2D analysis of 3D images) (Yi *et al.*, 2021). Unpublished live-cell data (presented by Mila Ilic at ecDNA EMBO Workshop, Switzerland, 2021) used an adapted version of the ANCHOR system (Meschichi *et al.*, 2021) to visualise ecDNA in interphase HeLa nuclei where the foci appeared more uniform and of a more expected size (comparable with DNA FISH data). Other studies have used fixed nuclei to identify ecDNA clusters, noting the limited definitions of the distances being studied (e.g. pixels rather than µm) (Hung *et al.*, 2021). Consequently, I would still expect to see hubs, if present and a major mechanism, using single time points with fixed cells.

Existing evidence from long-established cell lines may reflect widely differing characteristics, such as cell cycle length, growth patterns, mitotic rates and metabolic environment, in comparison with primary glioblastoma cells (Hung *et al.*, 2021). In order to directly compare these previously published cell lines with the Ripley's K approach here, one strategy would be to acquire these cell lines and individually perform the same analysis as I have performed on the GCGR cell lines. Overall, further studies of primary cell cultures would provide valuable data around ecDNA clustering and how these compare with established cell cultures.

Colocalisation of ecDNA with RNA Pol II has been proposed by ChIA-PET (Zhu *et al.*, 2021) and a live-cell ecDNA tagging strategy combined with immunofluorescence staining for RNA Pol II (Figure 1.10B) (Yi *et al.*, 2021). Previous imaging-based studies of ecDNA and RNA Pol II used subjective calling of colocalisation (Yi *et al.*, 2021) . This is challenging, given that RNA Pol II staining is pan-nuclear and diffuse (e.g. Figure 4.11). I analysed large Pol II hubs representing highly concentrated regions of the transcriptional machinery, on the basis that these could be clearly defined, and represented a biological structure of central relevance to the hub hypothesis. I did not observe a close relationship between ecDNA (DNA or nascent RNA) and large Pol II hubs. This difference in definition is important in explaining our results, as it is possible that smaller transcriptional hubs are associated with ecDNA, or indeed that hubs are not needed for ecDNA transcription.

My results suggest that the greater transcriptional output from oncogene-resident ecDNA is not due to increased ecDNA transcriptional activity, but is explained more simply by increased copy number. This is most definitively shown in the SNP analysis (Figure 4.23). As ecDNA are derived from one parental homolog (Stephens *et al.*, 2011; Hung *et al.*, 2022), the SNP analysis described here could be used to evaluate ecDNA-resident gene expression in any cell line where RNAseq and WGS (tumour and germline) data is available. Previous studies of large sequencing datasets reported that copy-normalised oncogene expression is higher on circular amplicons using RNA-seq plotted against WGS-derived CN (Kim *et al.*, 2020). RNA-seq and WGS from 36 clinical samples also concluded higher transcriptional efficiency, although only 3/11 ecDNA-resident genes (of which only one was a known oncogene), had significantly higher copy-number normalised gene expression (Wu *et al.*, 2019). Indeed, a recent study of primary neuroblastoma demonstrated a linear relationship between ecDNA copy number and gene expression (Stöber *et al.*, 2023). These studies therefore support my observation that copy number is the key driver of increased ecDNA gene transcription. In addition, my findings that ecDNA and chromosomal *EGFR* enhancers/promoter distances do not significantly differ further suggest that the gene regulatory environment is not markedly different. It is possible that the FISH-probed regions are not active enhancers/promoters in the E28 cell line; this could be established by ChIP-seq for associated proteins such as H3K27ac and H3K4me3. However, these data still suggest a similar chromatin organisation between the two structures, in alignment with the gene and polymorphism expression findings presented.

Recent data has suggested that while the overall relationship between ecDNA copy number and gene expression is linear, gene expression from individual ecDNA remains highly heterogeneous (Stöber *et al.*, 2023). This may be true of ecDNA in glioblastoma stem cells too. While such expression heterogeneity has been hypothesised to be secondary to mobile enhancers/hubs (Hung *et al.*, 2021; Zhu *et al.*, 2021; Stöber *et al.*, 2023), a first hypothesis might be that this simply reflects differences in regulatory elements on individual ecDNA. Targeted profiling by enrichment and long read sequencing of individual ecDNA may offer one means of analysing this in the cell lines described in this thesis as a future line of study. This may be challenging where multiple similar ecDNA are present, and still requires an underlying knowledge of likely ecDNA sequences for targeted CRISPR-CATCH guides with the assumption that these are shared between all ecDNA (Hung *et al.*, 2022).

Overall, the data presented in this chapter demonstrate a novel and robust strategy for evaluating ecDNA spatial organisation and transcriptional regulation. EcDNA copy number appears to be the primary driver of increased expression of any resident genes. Future studies could explore the heterogeneity of these ecDNA, and design tools to characterise this further.

Chapter 5 Effect of DNA damage and PARylation on ecDNA dynamics

5.1 Introduction

DNA damage is considered central to ecDNA formation, ecDNA-directed drug resistance and as a driver of chromosomal rearrangements (Rosswog *et al.*, 2021; Shoshani *et al.*, 2021). Further, studies of targeted and non-specific IR-induced DNA damage suggest a dynamic response of existing ecDNA to DNA strand breaks. An ecDNA-induced HeLa cell line showed a drive towards ecDNA integration following Cas9 and TALE-induced DNA damage, although the impact on stable, established ecDNA, and the impact on oncogene copy number following integration, is unclear (Shoshani *et al.*, 2021). Other studies of non-specific DNA damage via IR have shown a reduction in ecDNA number, with IR and PARPi driving ecDNA reintegration or micronuclei formation (Schoenlein *et al.*, 2003; Shoshani *et al.*, 2021).

All existing treatment options for glioblastoma induce DNA damage, with IR primarily leading to DNA strand breaks (SSB > DSBs) (Goodhead, 1994; Erasimus *et al.*, 2016). Previous data has also demonstrated excessive replication stress and high activation of the DDR in glioblastoma cells (Bartkova *et al.*, 2010; Carruthers *et al.*, 2018). SSBs can be repaired by poly-ADP-ribosylation (PARylation) leading to recruitment of DNA repair proteins (Saleh-Gohari *et al.*, 2005). PARP inhibitors enhance the radiosensitivity of glioblastoma cells, probably via their impact on DNA repair, leading to current clinical trials with the PARPi olaparib (Dungey, Löser and Chalmers, 2008; Hanna *et al.*, 2020). Given the role of DNA damage in current treatment modalities, it is important to understand how the primary mode of oncogene amplification, namely ecDNA, might be affected.

The presence of ecDNA-encoded oncogenes has already been shown to have a significant effect on glioblastoma cell characteristics. EGFRvIII-high GBM39 cells have a higher proliferation index (Ki67 staining), less apoptosis, higher glucose uptake and higher cell death in response to EGFR inhibitors than EGFRvIII-low FACS sorted cells (Nathanson *et al.*, 2014). While these results suggest possible downstream effects of ecDNA loss on glioblastoma characteristics, this is yet to be studied in cells where ecDNA copy number and characteristics are permanently altered by DNA damage.

This chapter describes preliminary work exploring the relationship between DNA damage, PARylation and ecDNA dynamics in glioblastoma stem cells. My hypothesis is that DNA damage results in ecDNA reorganisation driving them to form HSRs by chromosomal integration, which is exacerbated by PARP inhibition, leading to a reduction in ecDNA number. To explore this, I first characterised the baseline of DNA damage sites in GSC models and how these spatially relate to ecDNA. I explored the effect of IR on ecDNA dynamics, and how this is affected by the PARP inhibitor, olaparib. Finally, I used CRISPR-Cas9 with gRNAs targeting *EGFR* to evaluate the effect of targeted DNA damage on ecDNA dynamics, both on *ecEGFR* and non-*EGFR* oncogenes located on ecDNA.

5.2 DNA Damage in glioblastoma tumour cells

Phosphorylation of the histone variant H2AX to form γH2AX is a well-recognised marker of DNA DSBs, and has been shown to correlate well with sites of active DNA damage (Sedelnikova *et al.*, 2002). To establish the presence of DNA damage as a baseline, I assayed for γH2AX by IF in primary (E26, E28) and recurrent (E37) glioblastoma cells, using NSCs as a reference control (Figure 5.1A). In all cases, glioblastoma cells had a greater proportion of nuclei with large γH2AX foci. Approximately 40-50% of glioblastoma cells had >5 large foci per nucleus compared with 5-15% of NSCs (Figure 5.1B). This suggests that there is already a high baseline of breaks in GSCs, most likely due to replication stress.

The high amount of DNA damage, as detected by γH2AX staining, in glioblastoma cells harbouring ecDNA raised the question of whether there are specifically high levels of damage at ecDNA. I hypothesised that ecDNA and large γH2AX foci would be correlated both in terms of number and spatial colocalisation. I therefore performed immunoFISH in E26 and E37 nuclei, probing for γH2AX (IF) followed by DNA FISH for their respective ecDNA-residing oncogenes (*EGFR* and *c-MYC* respectively). There was neither a qualitative correlation between the number of ecDNA and number of large γH2AX foci, nor a close spatial relationship between the two (Figure 5.2). As this was apparent from direct visualisation and qualitative analysis, detailed quantitative analysis via the spot-defining and quantification tools described above was not performed. In addition, such analysis is extremely time intensive and technical, and qualitative data combined with the experience of data analysis from Chapter 3 provided enough evidence to exclude the hypothesis of ecDNA/γH2AX foci colocalisation. These data suggest that ongoing replication stress is a feature of glioblastoma cells, but active sites thereof are independent of existing ecDNA.

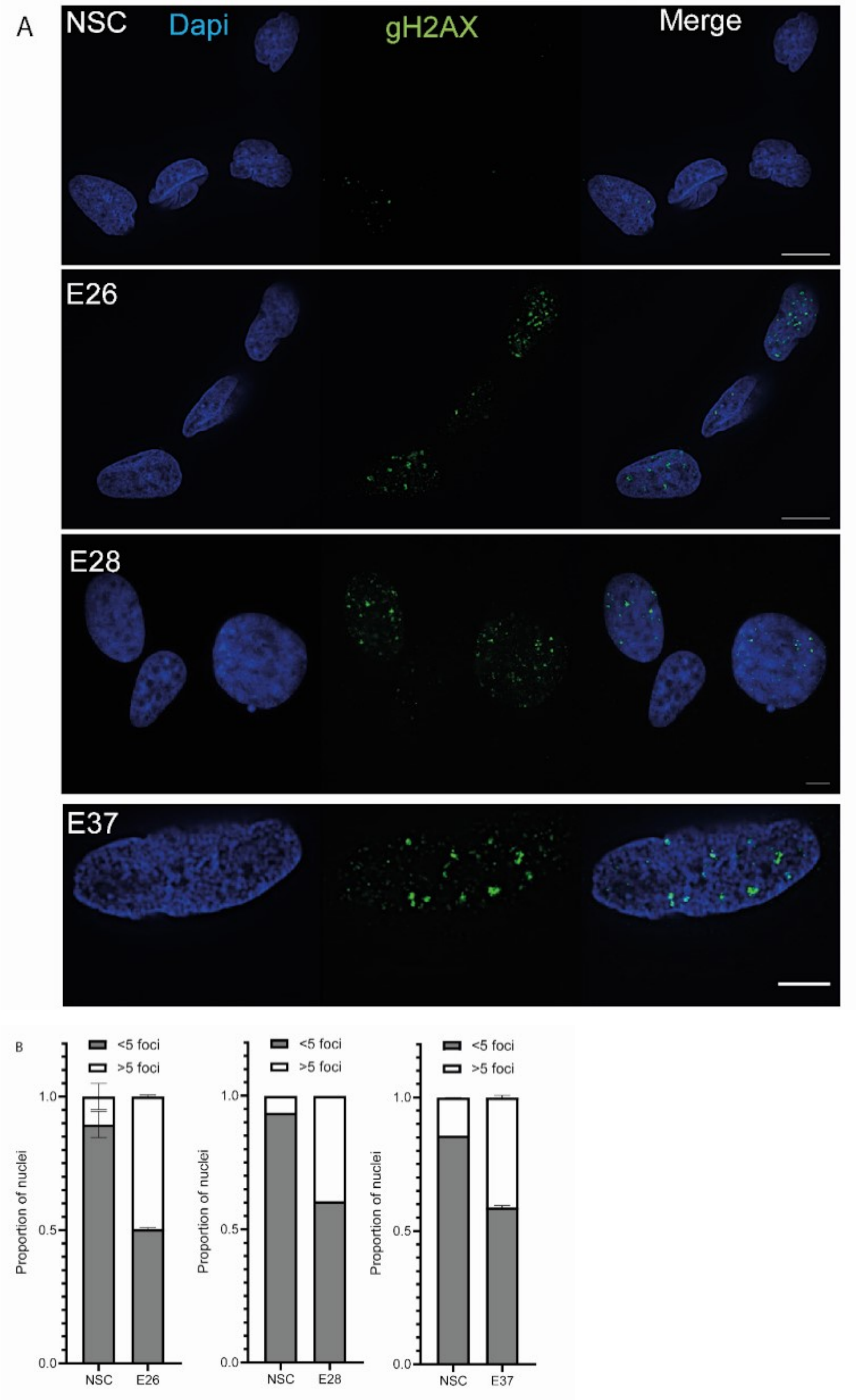
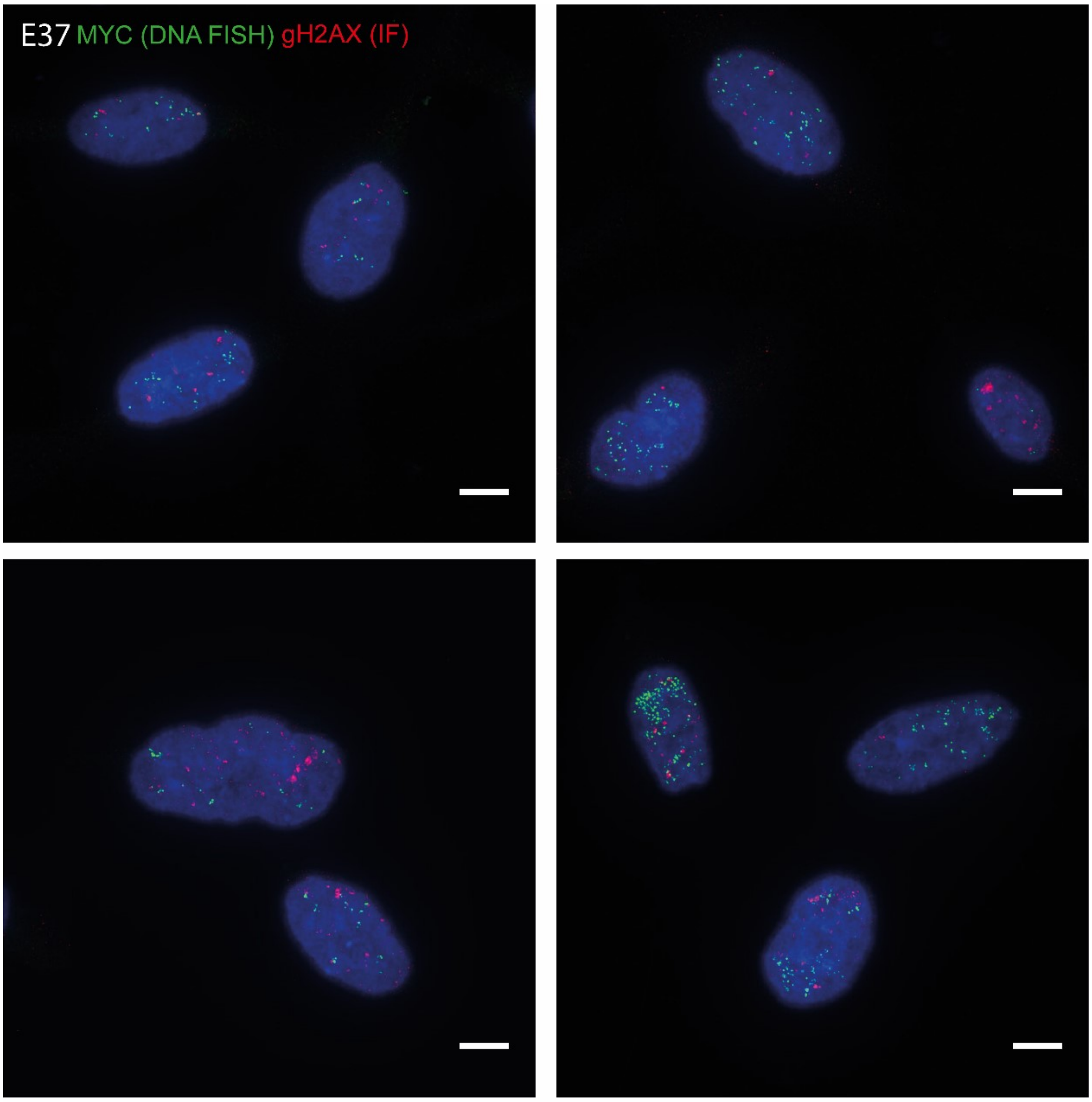
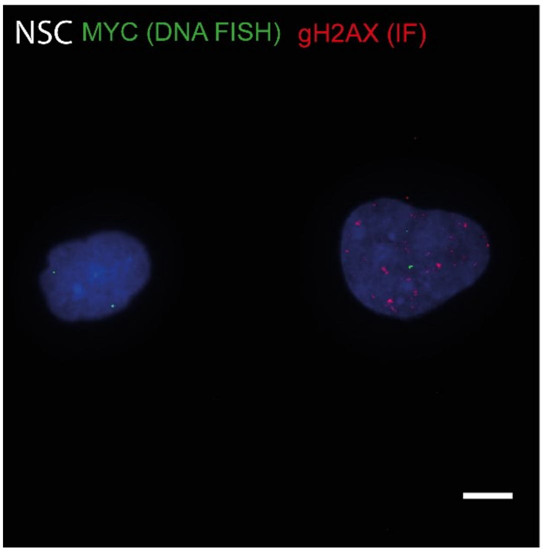


Figure 5.1 | γH2AX foci in glioblastoma stem cells

A) Representative images of nuclei from NSC, E26, E28 and E37 cells (DAPI, blue) and γH2AX immunofluorescence (green). Scale bar = 5 μm B) Proportion of nuclei with <5 and >5 γH2AX foci in E26 cells (n=2 biological replicates), E28 cells (n =1 biological replicate) and E37 cells (n=2 biological replicates)


E37 MYC (DNA FISH) gH2AX (IF)


NSC MYC (DNA FISH) gH2AX (IF)

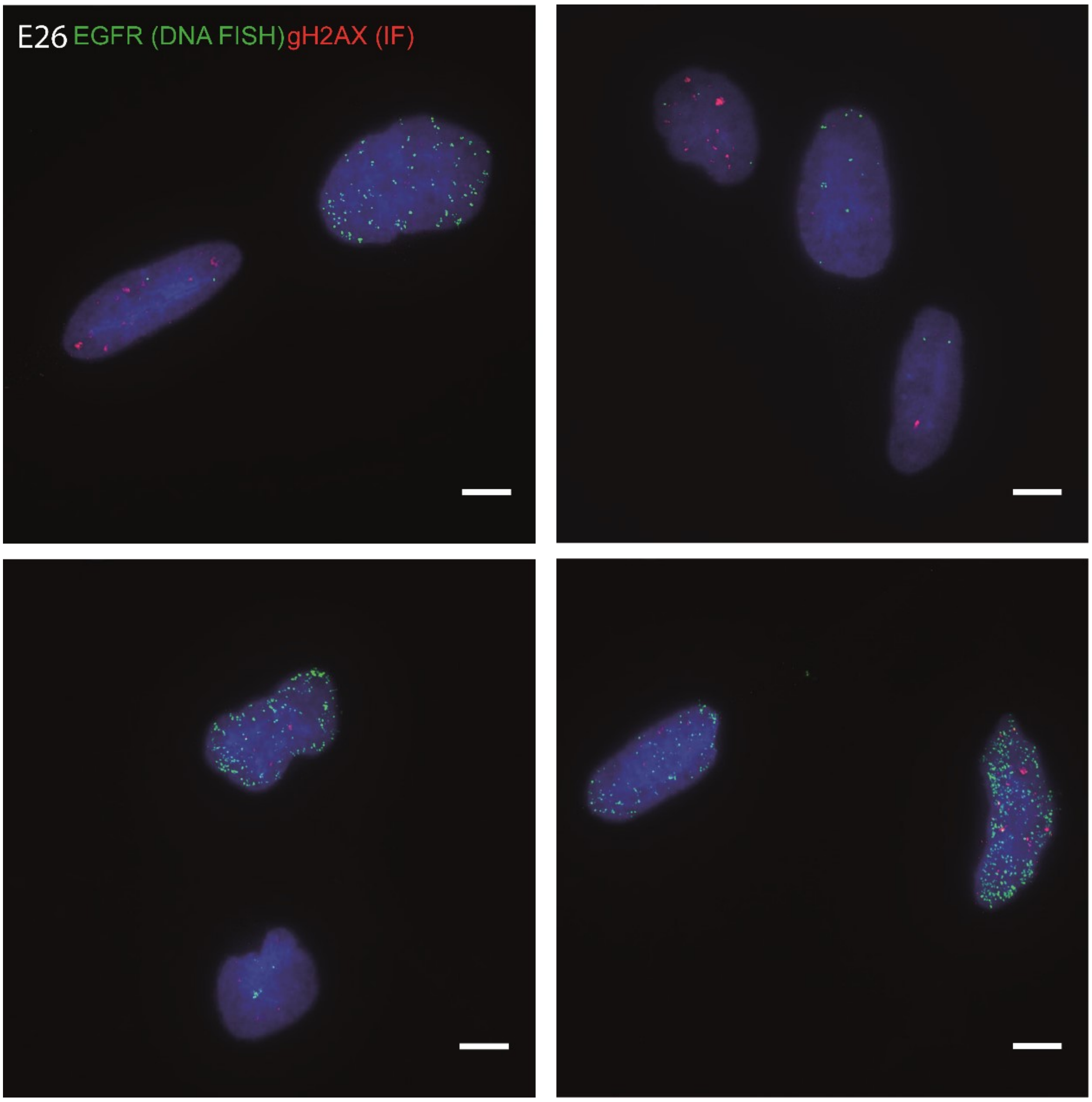


Figure 5.2 | γH2AX foci and ecDNA

Representative images of ImmunoFISH in E26, E37 and NSC cells probing for γH2AX (immunofluorescence, red) and the resident ecDNA oncogene (DNA FISH, green). Scale bar = 5 µm. Biological replicates = 1.

5.3 Non-specific DNA damage via Ionising Radiation (IR)

5.3.1 Effect of IR on ecDNA number and characteristics after 24 hours

I then wanted to examine the impact of non-specific random DNA damage using a clinically relevant strategy and first aimed to characterise IR responses in GSC cells. A dose of 4Gy has previously been shown to impair cell survival in glioblastoma stem cells and is a clinically relevant dose of IR (Carruthers et al., 2018). To confirm that the glioblastoma cell lines selected were radiosensitive and that a treatment effect could be seen with 4Gy IR, a modified colony survival assay was performed on E26 and E28 cells seeded at varying densities. Reduced cell survival was seen in both cell lines following a single dose of 4Gy IR, and this was more pronounced in E28 cells (Figure 5.3A and B). A range of alternative IR doses could also be performed to validate the optimal dose. This would be defined as the dose where a cell response is observed but sufficient cells survive for these to be analysed, which was successfully achieved with 4Gy.

First I wanted to assess if there was an effect of 4Gy IR on ecDNA copy number after short time period post-IR, as previous studies have evaluated ecDNA characteristics at 48h (Shoshani *et al.*, 2021) or ecDNA copy number after 2-3 weeks of multiple IR fractions (Schoenlein *et al.*, 2003). I probed for *EGFR* using DNA FISH 24 hours after either 4Gy or control conditions (Figure 5.4A). Using FITC nuclear signal intensity as a surrogate of the number of *EGFR* foci per nucleus, the number of *EGFR* foci reduced in both cell lines following 4Gy IR (Figure 5.4B). *EGFR* DNA FISH of metaphase spreads (Figure 5.4C) at the same dose and time point was used to establish if any structural changes in ecDNA could be detected. A reduction in the number of spreads with ecDNA was also observed in E28 (Figure 5.4D). No large ecDNA were observed in any E28 metaphase spreads, and although there was a small reduction in the % of metaphase spreads with large DMs in E26 cells, the numbers are too small to be confident (Figure 5.4E). This highlights the challenge of studying ecDNA evolution in cells such as the E28 cell line which have a relatively sparse number of ecDNA. Many nuclei and metaphase spreads will be needed to have the power to observe any changes in E28 cells. In addition, a time-course could add further important detail as to the timeframe of ecDNA copy number and structural changes.

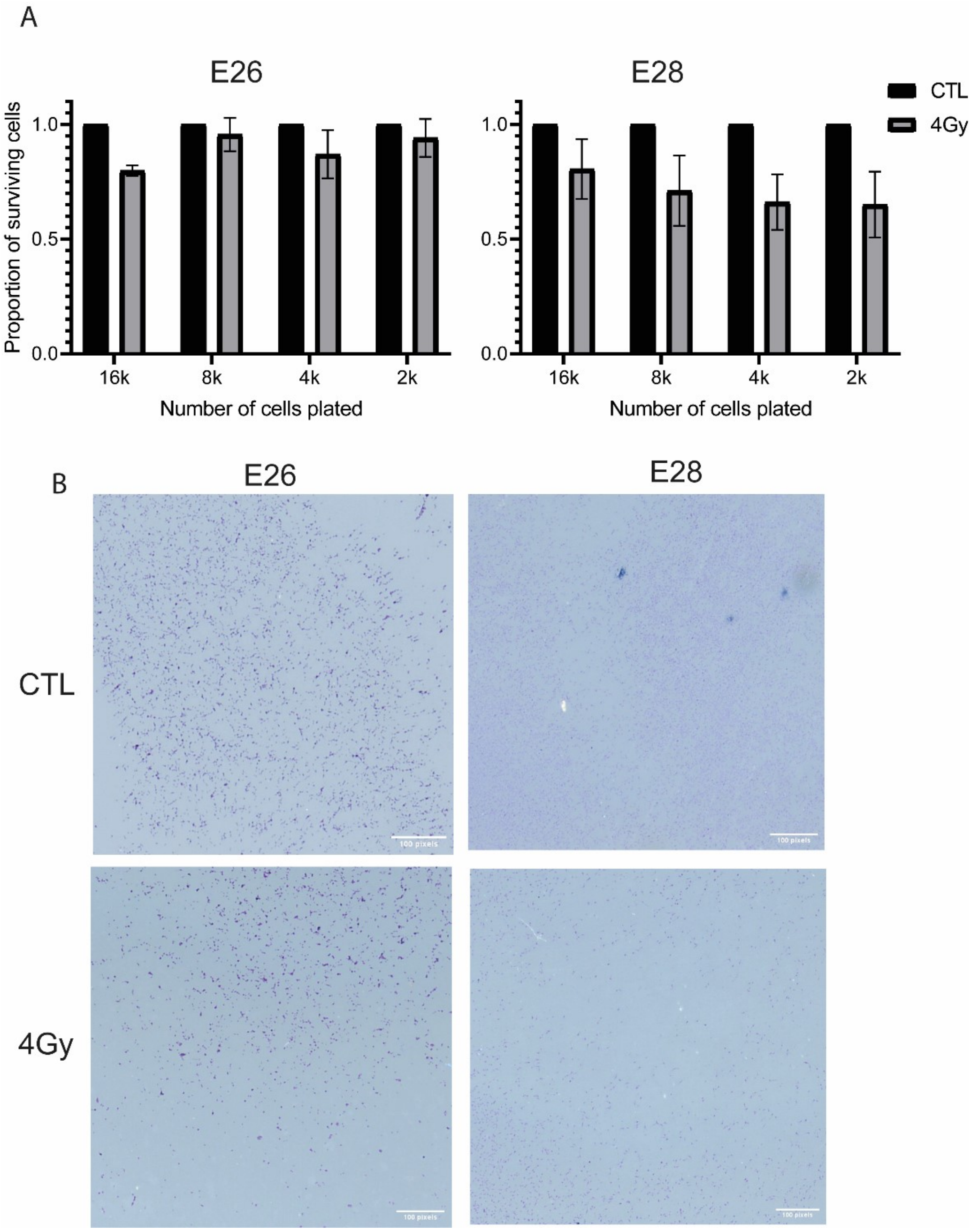


Figure 5.3 | Proportion of cell survival following Ionising Radiation (IR)

A) Proportion of surviving cells following the indicated conditions at different cell plating densities. Two technical replicates per condition, mean +- SD shown. Unpaired T-test p value – E26 = 0.034, E28 = 0.0003 B) Representative images of cells from 16k cell plating density in indicated conditions. Scale bar = 100 pixels. Biological replicates = 1.

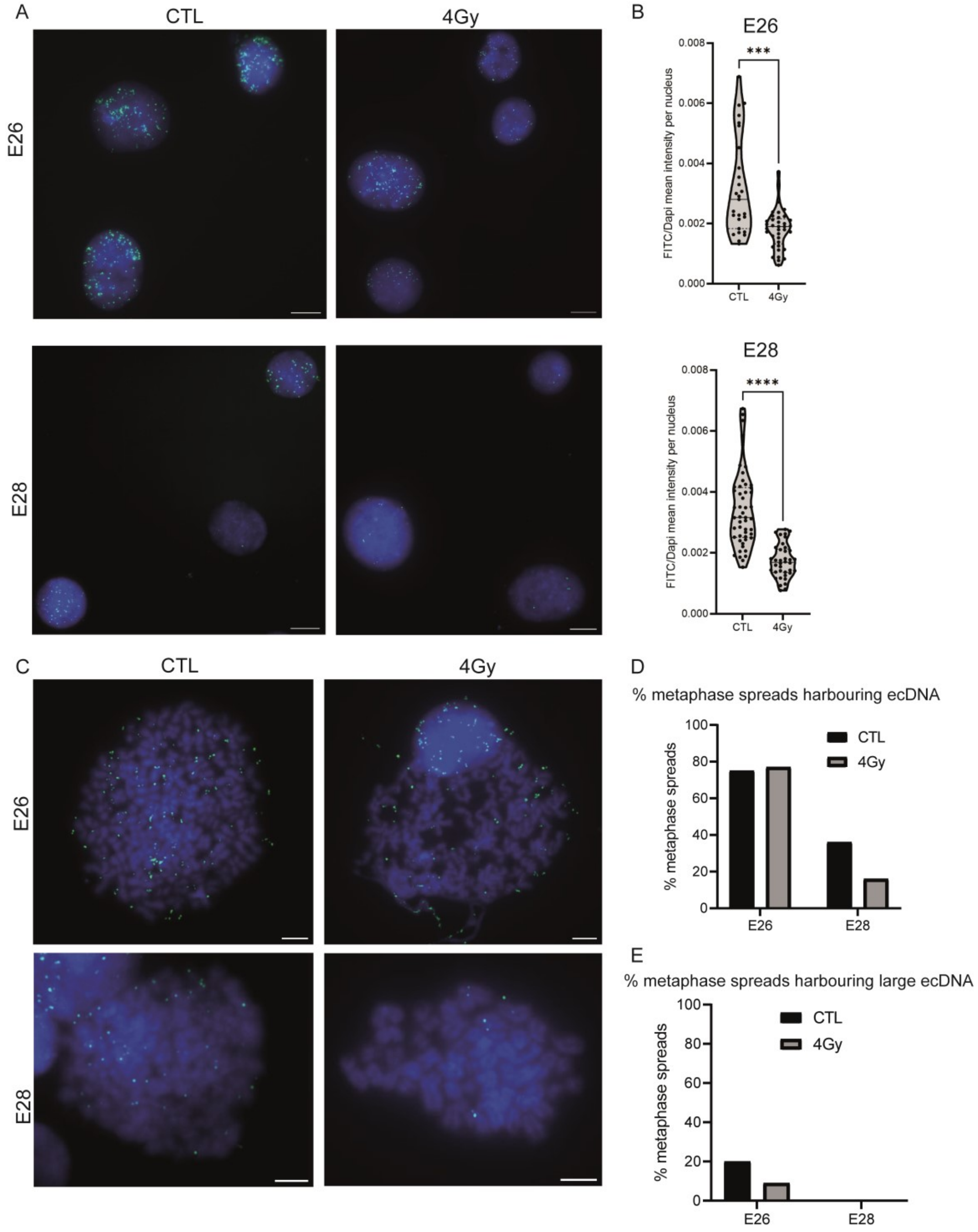


Figure 5.4 | Effect on ecDNA 24 hours after 4Gy IR

A) Representative images of nuclei from conditions indicated (MIP), scale bar = 10 µm. B) FITC (green – EGFR) signal intensity per nucleus, normalised by DAPI signal intensity and size of nucleus. E26 CTL (n=27) and 4Gy (n=37), Mann-Whitney p-value = 0.0001. E28 CTL (n=50) and 4Gy (n=38), Mann-Whitney p-value = <0.0001. C) Representative metaphase spreads from indications indicated, scale bar 10 µm. D) Percentage of metaphase spreads harbouring ecDNA. E26 CTL (16/20 metaphase spreads) and 4Gy (17/22), E28 CTL (7/19) and 4Gy (3/18). E) Percentage of metaphase spreads harbouring large ecDNA. E26 CTL (4/20) and 4Gy (2/22), E28 CTL (0/19) and 4Gy (0/18). Biological replicates = 1.

5.3.2 PARP inhibition and effect on EGFR foci copy number

To study the effects of PARP inhibition on ecDNA, I chose to use the PARPi olaparib - selected on the basis that it is clinically relevant and non-toxic dosing regimens are well established. Olaparib works by trapping covalent PARP-DNA complexes (Saleh-Gohari *et al.*, 2005). PARylation, detected by IF, was increased following induction of DNA damage with H_2O_2 (a positive control), and this was blocked in the presence of 1μm olaparib, suggesting successful inhibition of PARP at this dose (Figure 5.5A). To establish a baseline impact of olaparib on ecDNA copy number, *EGFR* DNA FISH foci were counted in E26 cells after 24 hours and 2 weeks exposure to 1μM olaparib. No significant change in the number of *EGFR* foci was observed, although a trend towards reduced number of foci was observed after 2 weeks of olaparib (median number of foci – CTL = 41 vs 4Gy = 25; Figure 5.5B). These data indicate that PARPi alone does not have a marked effect on ecDNA copy number.

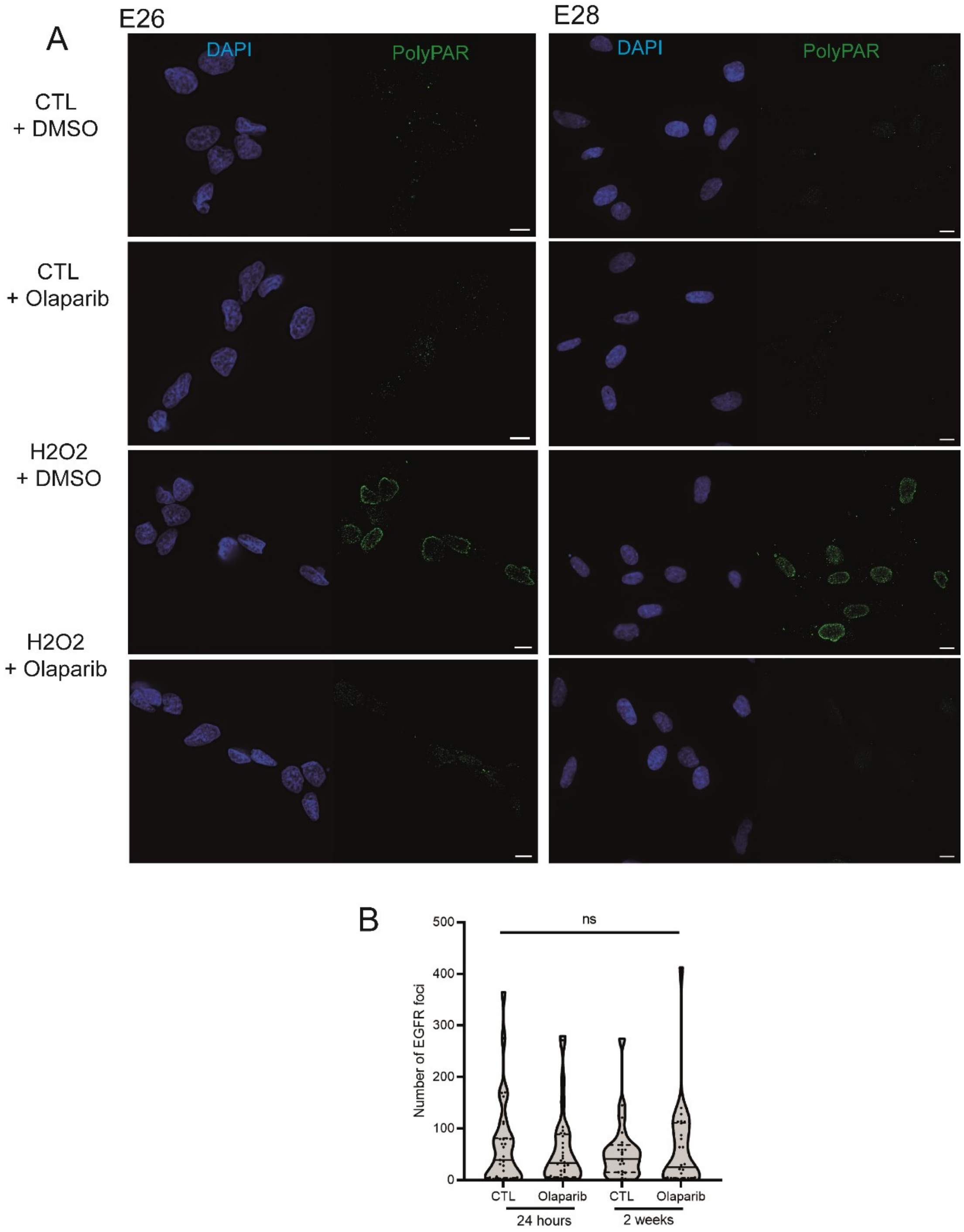


Figure 5.5 | PARP inhibition with Olaparib

A) Representative images of immunofluorescence for PolyPAR in indicated conditions in E26 and E28 cells, indicating PolyPAR expression following DNA damage with H2O2, and loss thereof in the presence of olaparib. Scale bar = 10 µm. B) Number of EGFR foci (by DNA FISH) following 24 hours and 2 weeks of CTL vs olaparib in E26 cells. Kruskall Wallis - 0.9618, ns. Biological replicates = 1.

5.3.3 Combined effect of IR and PARP inhibition on ecDNA characteristics

To evaluate the combined effect of IR and PARPi on ecDNA copy number on a larger cell population than could be assayed by DNA FISH alone, cells were first evaluated by flow cytometry with EGF-647.

E26 cells, which harbour many *EGFR* ecDNA, were treated as outlined in Figure 5.6. Flow cytometry demonstrated a reduction in EGF-647 signal following IR, with the highest shift to the left seen in IR + olaparib cells (n=2, replicate 1 shown in Figure 5.7A). This is reflected in the mean EGF-647 signal (2 biological replicates), which is significantly lower following IR, and trends lowest in the IR + olaparib cohort (Figure 5.7B). DNA FISH showed a reduction in number of *EGFR* foci with IR, but not when cells were also treated with olaparib (Replicate 1 shown in Figure 5.7C). Plotting the mean number of foci relative to control conditions across two replicates shows a trend of reduced *EGFR* foci number following IR that is lost when co-treated with olaparib (Figure 5.7D).

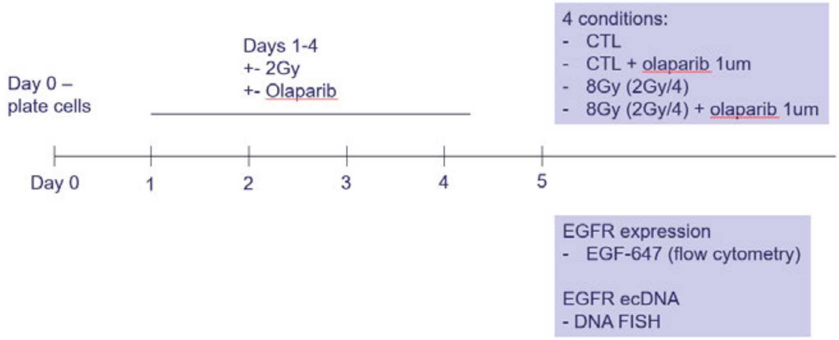


Figure 5.6 | Experiment outline for evaluating combined effect of IR and olaparib on ecDNA characteristics

This was repeated in a cell line with fewer EGFR ecDNA, E28. Although the EGF-647 left shift was not as apparent, once the EGF-647-negative control signal was accounted for, the same reduction in EGF-647 signal with IR and IR + olaparib could be observed (Figure 5.8A and B). Subsequent DNA FISH detected a significant difference in *EGFR* DNA foci number across the conditions (Kruskall-Wallis, p=0.0136, Figure 5.8C and D). The difference between the number of foci in CTL vs IR cohorts was not significant once multiple hypothesis testing was accounted for, although the difference between IR vs IR + olaparib was significant. Again, plotting the mean number of foci relative to control conditions repeats the trend seen in E26 cells of a drop in number of *EGFR* foci that is lost with co-treatment with olaparib (Figure 5.8E).

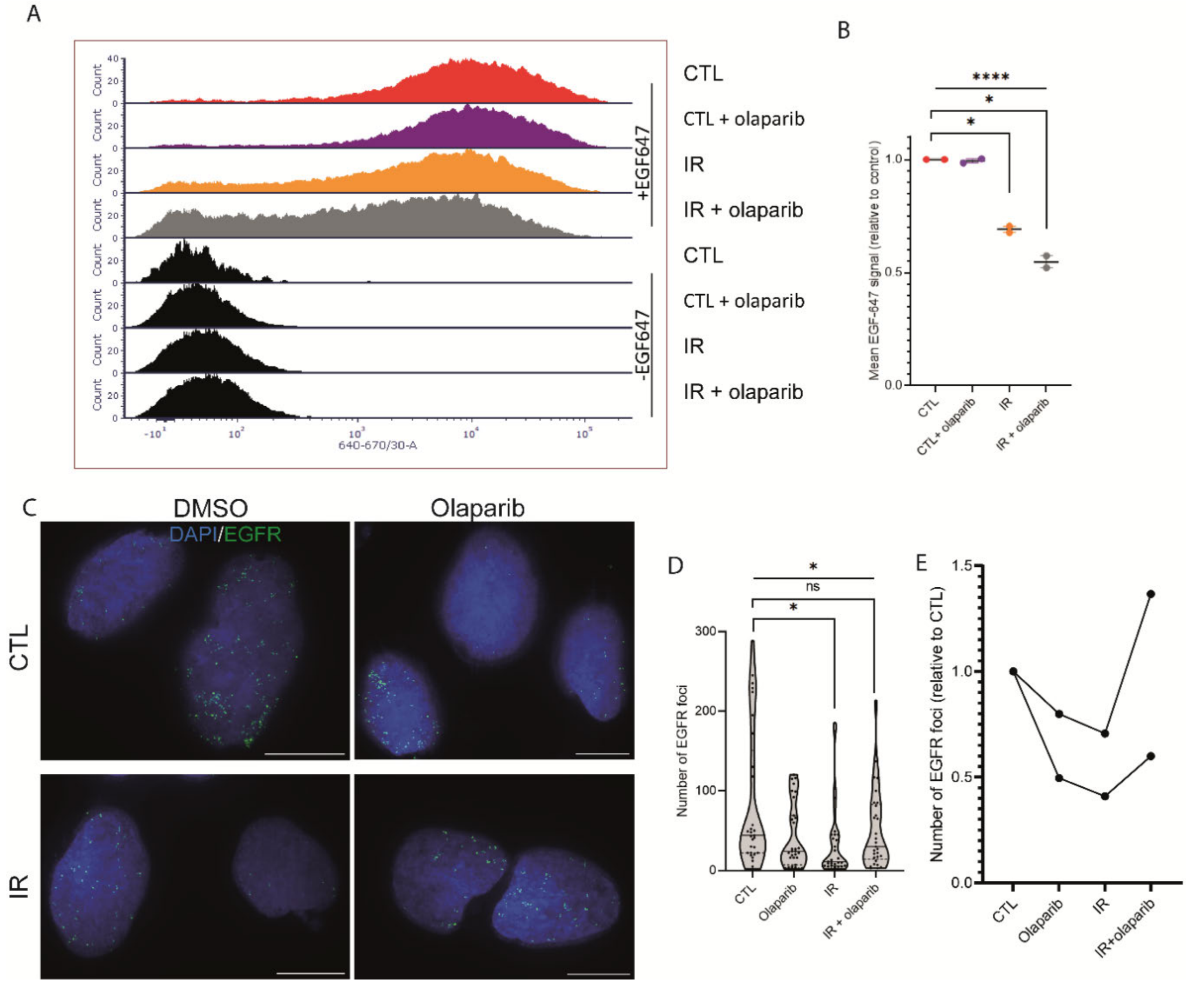


Figure 5.7 | EGFR expression by flow cytometry and DNA FISH in E26 cells treated with IR and olaparib

A) Histogram of flow cytometry of cells treated in indicated conditions (see Figure 3.14 for details), normalised to peak counts (y axis) and plotted again EGF-647 signal (x axis). Replicate 1 of two biological replicates shown. B) Mean EGF-647 signal, relative to EGF-647 negative control for each cell line, n=2 biological replicates. Mean +- SD C) Representative nuclei showing EGFR (green) DNA FISH, conditions shown, replicate 1 of two biological replicates shown. Scale bar = 10 µm D) Number of EGFR foci per condition from replicate 1. Flat line – Kruskall-Wallis p-value 0.0165. Hooked lines - Mann-Whitney p-value (normalised p value) - CTL vs IR = 0.0018 (0.0108), CTL vs IR + olaparib = 0.1546 (0.9276). E) Combined data from two biological replicates indicating mean number of EGFR foci per condition normalised to CTL for each replicate.

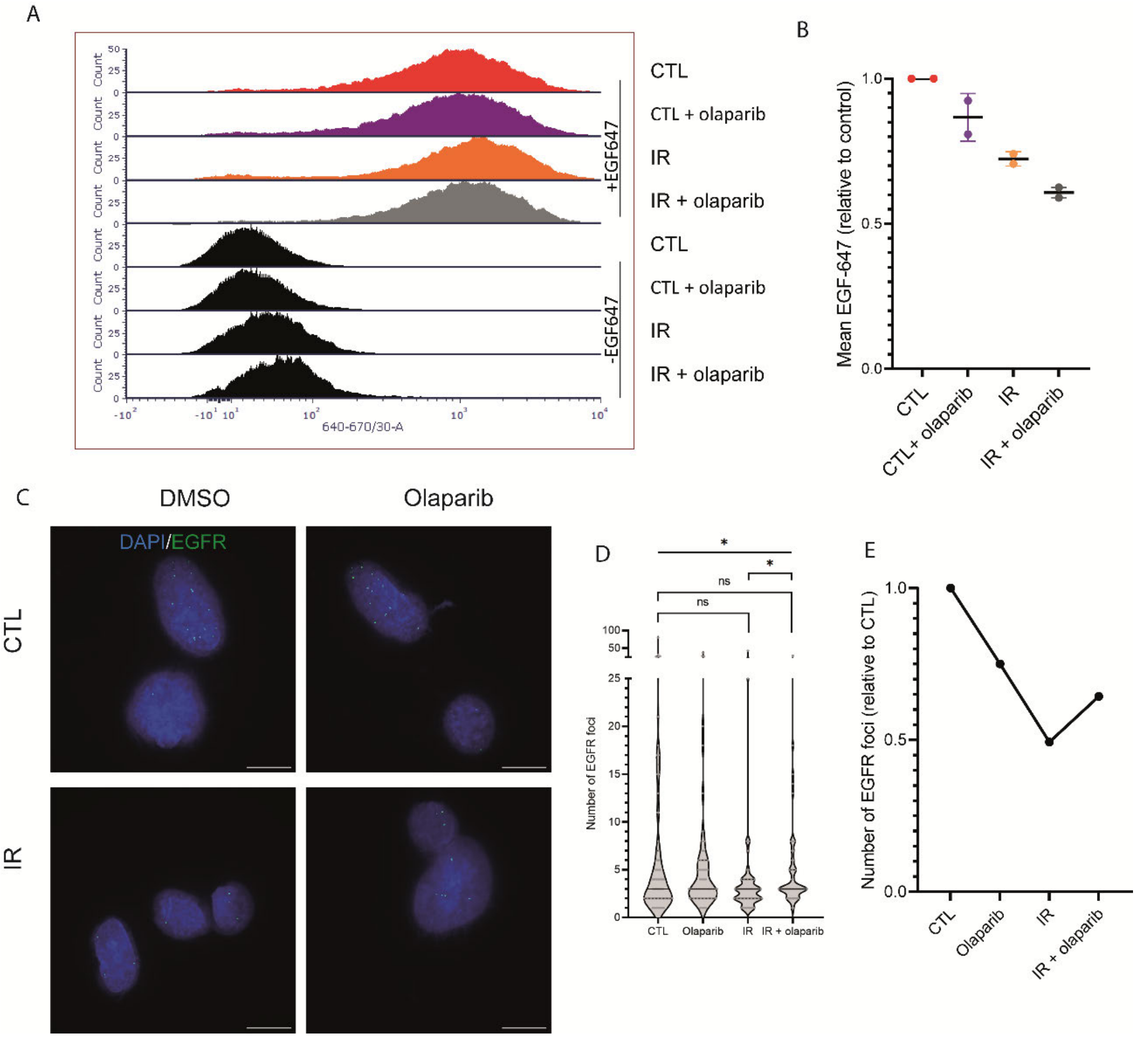


Figure 5.8 | EGFR expression by flow cytometry and DNA FISH in E28 cells treated with IR and olaparib

A) Histogram of flow cytometry of cells treated in indicated conditions (see Figure 3.14 for details), normalised to peak counts (y axis) and plotted again EGF-647 signal (x axis). Replicate 1 of two biological replicates shown. B) Mean EGF-647 signal, relative to EGF-647 negative control for each cell line, n=2 biological replicates. Mean +- SD C) Representative nuclei showing EGFR (green) DNA FISH, conditions shown. Scale bar = 10 µm D) Number of EGFR foci per condition by DNA FISH. Flat line – Kruskall-Wallis p-value 0.0136. Hooked lines - Mann-Whitney p-value (normalised p-value) - CTL vs IR =0.0266 (0.1596), CTL vs IR + olaparib = 0.7925 (1.0), IR vs IR + olaparib = 0.0032 (0.0192). E) Mean number of EGFR foci per condition normalised to CTL for each replicate.

5.4 Specific DNA damage via CRISPR-Cas9

5.4.1 Targeted CRISPR-Cas9 cleavage of EGFR in E26 cells harbouring ecEGFR

I next wanted to examine the impact of specific DNA damage targeting a precise genomic locus using CRISPR-Cas9.

The CRISPR-Cas9 system involves guide RNAs (gRNAs) targeting specific genomic loci that result in site-specific DSBs. gRNAs were designed targeting the 3’ and 5’ ends of *EGFR* (gEGFR) to assess the impact of *ecEGFR* cleavage in a cell line harbouring ec*EGFR* (Figure 5.9).

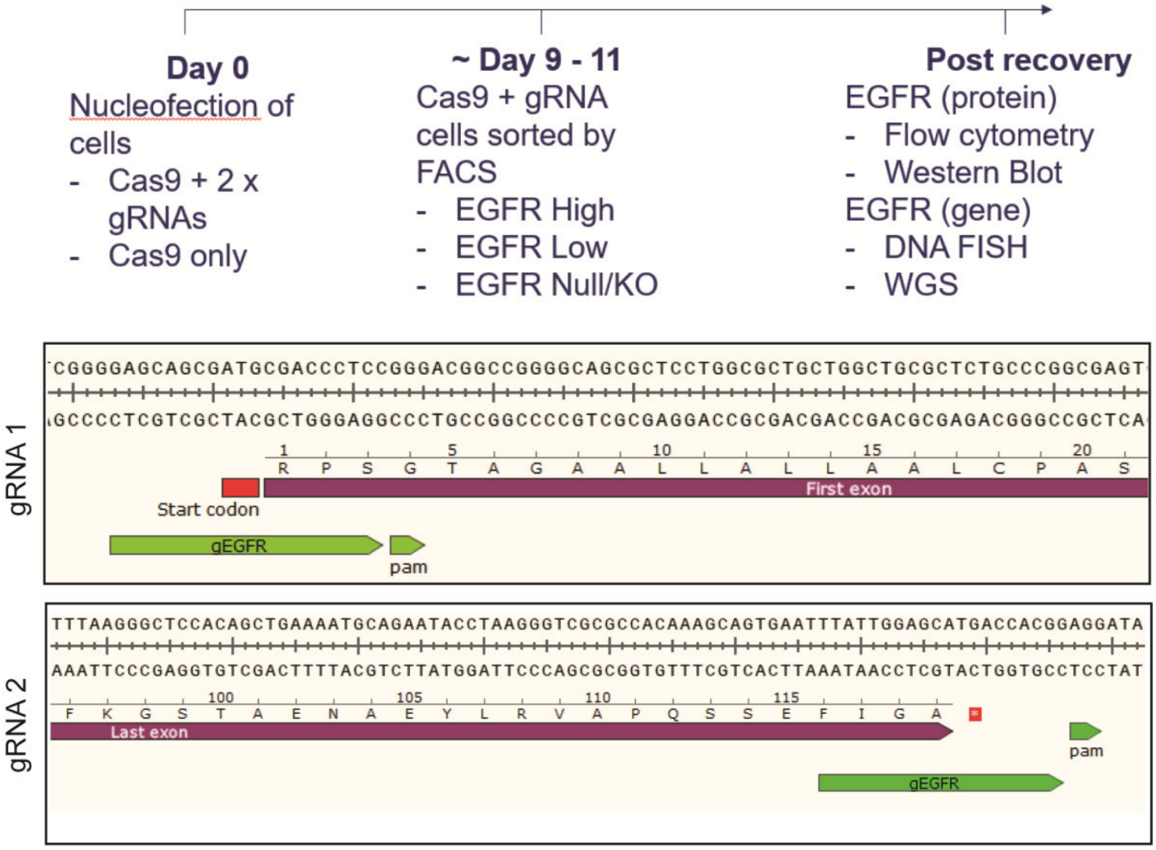


Figure 5.9 | Experiment timeline for CRISPR-Cas9-mediated *EGFR* targeting in E26 and E37 cells

Flow cytometry with the EGF-647 fluorescent ligand confirmed that Cas9_control cells had high levels of EGFR, and very few cells with negative signal (mean 0.31% cells gated as EGFR Negative across n=3 biological replicates, Figure 5.10). By contrast, Cas9_gRNA transfected cells had a wide distribution of EGFR expression, including a larger EGFR negative population suggestive of successful ablation of *EGFR* and loss of the protein on the cell surface (Figure 5.10) (mean 5.05% cells across n=3 biological replicates - note not significant (EGFR Negative population in Cas9_control vs Cas9_EGFR, Mann-Whitney *p* value = 0.100)).

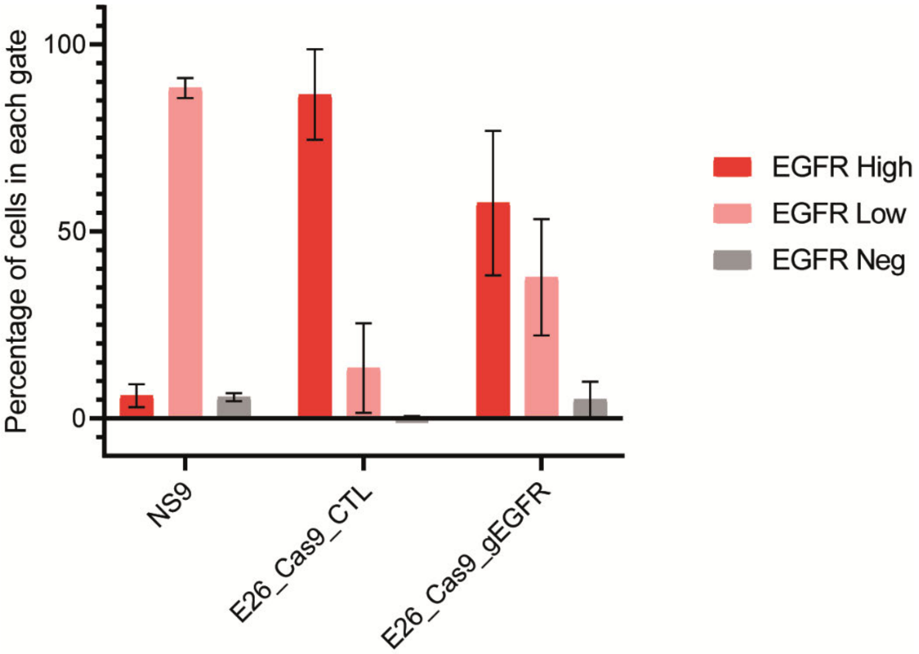


Figure 5.10 | Flow cytometry of E26_Cas9_CTL vs E26_Cas9_gEGFR

Percentage of cells in EGFR High, Low and Negative groups as measured by flow cytometry (see Figure 5.11A for gates) across n=3 biological replicates, mean+- SEM. Kruskall-Wallis p values – EGFR High = 0.05, EGFR Low = 0.011, EGFR Neg = 0.011.

Glioblastoma cells grow slowly, are challenging to clonally expand and have the risk of selection of unusual variants that are not representative of the high heterogeneity seen in glioblastoma. Given how well *EGFR* DNA and EGF-647 flow cytometry signals correlated (Chapter 4 Section 4.8.1), flow cytometry seemed an appropriate tool to generate pools of deleted cells that cover a diverse spectrum of the genetics from the parental populations. Cells were FACS sorted into EGFR High, Low and Negative groups, the gating of the latter two groups being determined using negative controls (NSC and EGF-647-negative) (Figure 5.11A). Following recovery, these pooled cell populations were validated by repeat flow cytometry with the EGF-647 ligand, and by immunoblot.

Flow cytometry confirmed an EGF-647 High cohort in EGFR-CTL cells, an intermediate EGF-647 cohort in the EGFR-Low cells, and a distinct EGF-647 negative cell population in EGFR null cells (Figure 5.11C). Immunoblotting detected a strong band at 145kDa, in keeping with EGFRvIII, and a fainter band at 175kDa, consistent with EGFR wt, in EGFR-CTL cells (Figure 5.11D). The EGFR wt band can be faintly seen in the EGFR-Low cells. No bands are visible in EGFR-null cells.

Having confirmed the loss of EGFR protein, DNA FISH on metaphase spreads was utilised to characterise *ecEGFR* in these sorted cell populations (Figure 5.11E). *EcEGFR* could still be detected on metaphase spreads from EGFR-CTL cells, but no *ecEGFR* were observed in EGFR-Low or EGFR-Null cells, although the endogenous chromosomal *EGFR* loci could be visualised (see arrows, Figure 5.11E).

These data suggest that CRISPR-Cas9 targeting an oncogene present on ecDNA can eliminate these ecDNA.

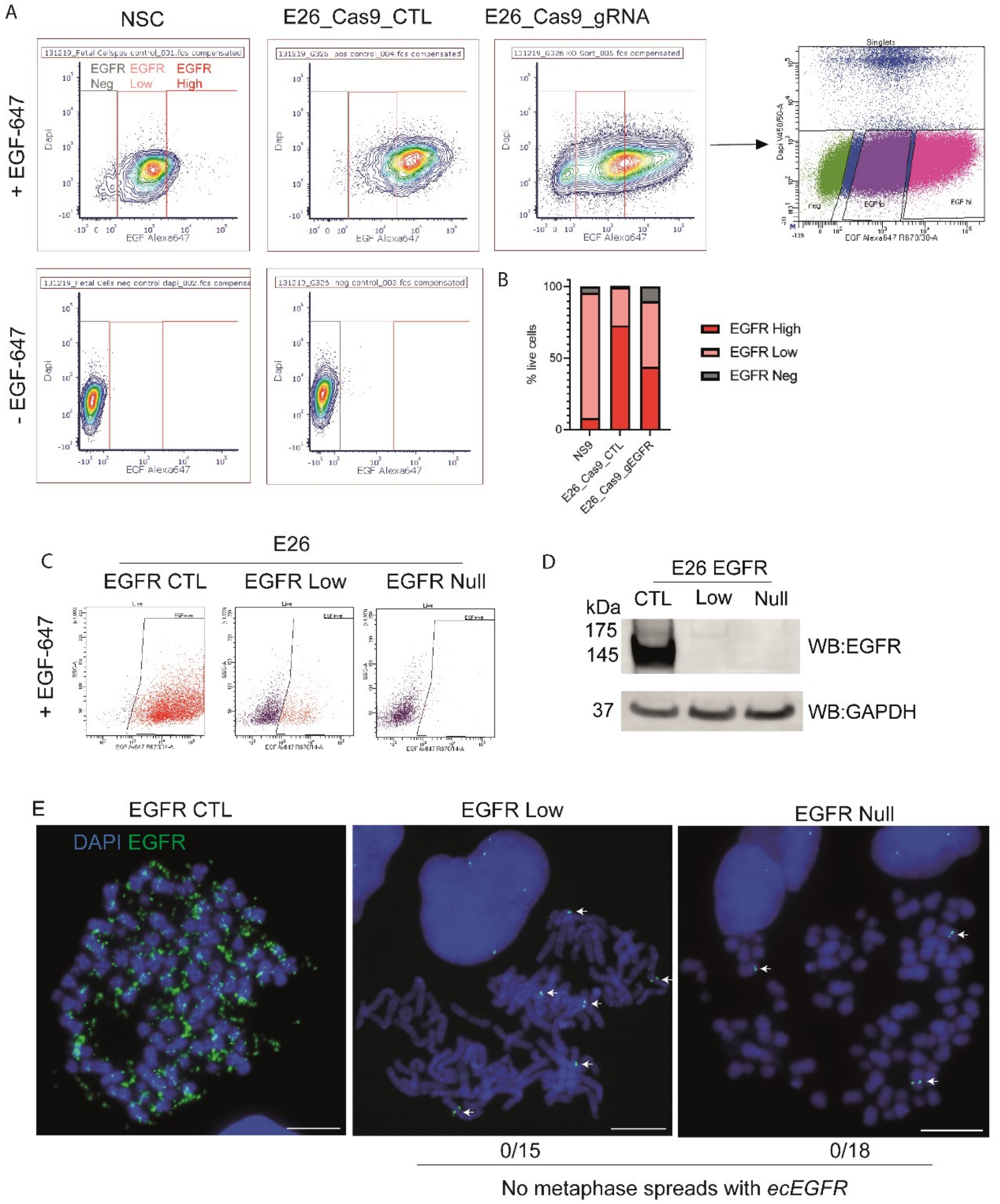


Figure 5.11 | E26 EGFR CTL, Low and Null cells

A) Flow cytometry for EGF-647 in indicated cell lines, arrow denotes FACS gating for cell sort into EGFR negative (green), low (purple) and CTL (pink) cohorts. B) Bar plot showing % of cells in each gate shown in EGF-647+ samples in A). C) Flow cytometry for EGF-647 confirming EGFR CTL, Low and Null cell pools recovered after FACS. D) Western blot for EGFR and GAPDH. E) Representative metaphase spreads from E26 cell lines shown. EGFR Low and Null populations = number of metaphases with ecEGFR/total number of metaphase spreads visualised. Arrows = chromosomal EGFR. Scale bar = 10 µm. Biological replicates = 1.

5.4.2 EcEGFR loss in E26 EGFR Low and Null cells confirmed by WGS

WGS was used to further characterise the genome of the sorted cell populations. Read count profiles of the E26 EGFR CTL, Low and Null cell lines reveal an overall increase in copy number of reads on chromosome 7 suggestive of some aneuploidy (Figure 5.12A). WGS data (Figure 5.12B) suggests that *EGFRvIII* amplicons with the same profile as the original cell line remain in the CTL cell line, but are lost in the EGFR Low or Null cell lines, in keeping with the loss of *EGFR* ecDNA seen by DNA FISH. This was verified by bioinformatics analysis using AA, in addition to confirming that no other new amplicons had emerged. Plotting the CNV regions identified by AA against their copy number confirmed that the only amplified regions were from chromosome 7 in the EGFR CTL cell line (Figure 5.12C and D), and these four regions mapped to the amplified region (see ‘CNV regions’, Figure 5.12A). AA profiled one amplicon region, classified as ‘complex cyclic’, which, allowing for differences in sequencing depth, appear similar to the amplicon of the original cell line (Figure 5.12E). No amplicons were identified by AA in the EGFR Low and Null cell lines.

These data suggest targeted cleavage of ecDNA amplicons via CRISPR-Cas9 can lead to loss of ecDNA and reduced expression of their resident oncogene. This is sustained in cell lines pooled by oncogene expression status, and ecDNA-negative cell lines can be produced by this method without introducing new amplicons.

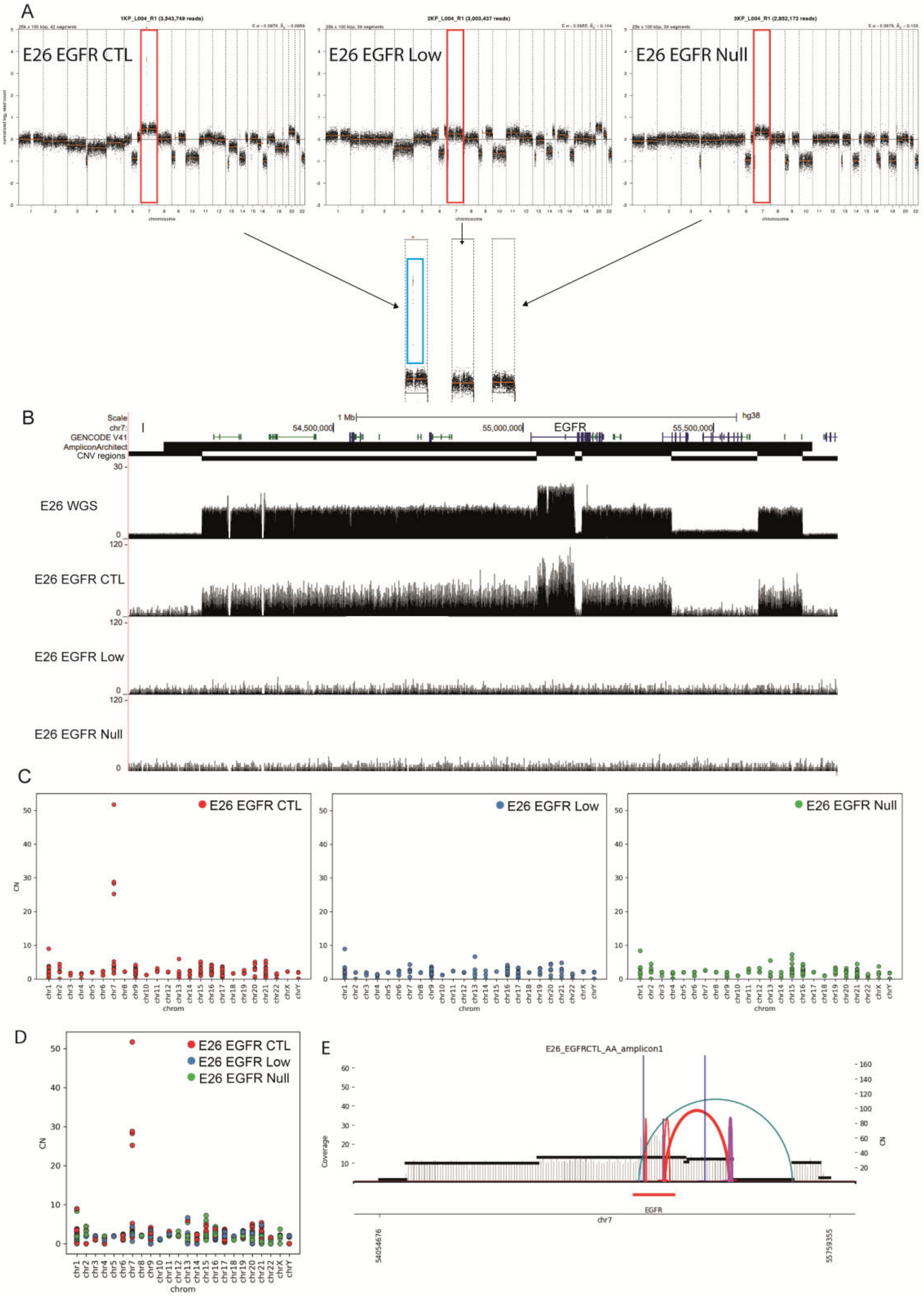


Figure 5.12 | WGS and AA analysis of E26 EGFR CTL, Low and Null cell lines

A) Copy number profile (15kbp normalised) of E26 EGFR CTL, Low and Null cell lines. Red box = chr7 region highlighted by arrows. Blue box = copy number amplifications seen in EGFR CTL. B) WGS of E26 (original cell line) and EGFR CTL, Low and Null cell lines. C) AA CNV regions by chromosome plotted against CN for EGFR CTL, Low and Null cell lines. D) As C) but plotted together. E) Amplicon detected in E26 EGFR CTL cell line via AA.

5.4.3 Targeted CRISPR-Cas9 cleavage of EGFR in E37 cells harbouring EGFR HSRs and ecMYC

I repeated this strategy in a recurrent glioblastoma cell line harbouring *EGFR* HSRs and *c-MYC* ecDNA. Flow cytometry confirmed higher EGFR expression in E37 cells in comparison to NSCs, in keeping with WGS data showing *EGFR* amplification (Figure 5.13A and B). As for E26 cells, E37 cells treated by CRISPR-Cas9 and the same *EGFR*-targeting gRNAs used above were initially sorted into EGFR CTL and Null cell pools (Figure 5.13A). A subsequent FACS sort of the original CRISPR-Cas9 pooled cells was performed to generate the EGFR Low population (See Appendix 8.8). Following cell recovery, these cells populations could be validated by flow cytometry; despite stepwise FACS sorting, EGFR CTL, Low and Null populations had a stepwise reduction in EGFR expression suggesting these were sustained and defined populations (Figure 5.13C). Immunoblotting confirmed loss of EGFR expression. Surprisingly, subsequent probing for c-MYC suggested loss of MYC expression in the EGFR Null population (Figure 5.13D).

As might be expected, E37 EGFR Null cells appeared to grow more slowly. To confirm this, cell growth was measured by live-cell time-lapse imaging and estimates of cell confluence (using Incucyte). EGFR null cells morphologically appeared to have longer extension processes than EGFR Low or CTL cells (Figure 5.14A) and, although they proliferated, they were sparser at day 5 and day 12.

I grew E37 and NSC cells in in media that can drive a proliferative (serum) and quiescent (BMP, -EGF/FGF) growth trajectory to establish the response of unedited cells. This showed that E37 cells grow more rapidly in serum, in keeping with a proliferative phenotype, and slower in conditions that drive quiescence (Figure 5.14B). E37 EGFR CTL and Low cells demonstrated very similar growth trajectories to the original E37 cell line. In contrast, E37 EGFR Null cells grew more slowly, although a proliferative response to serum was observed independent of EGFR signalling (Figure 5.14C).

To confirm that this did not represent a false confluence reading due to increased cell spreading and size, proliferation in the presence of serum was confirmed by the presence of Ki67 after 7 days in serum vs +EGF/FGF media (Figure 5.15). This suggests that, despite the loss of key oncogenes on ecDNA, cells can remain proliferative when in the astrocyte differentiation state and are not impacted by EGFR loss; but GSCs in the neural stem cell-like state require EGFR for maximum proliferative capacity.

Taken together, these data show that CRISPR-Cas9-targeting of a non-ecDNA-resident *EGFR* can generate a cell population where EGFR expression is lost, with concurrent loss of MYC expression despite *MYC* being untargeted by Cas9 and located separately on ecDNA. These cells can remain proliferative despite the loss of these oncogene drivers.

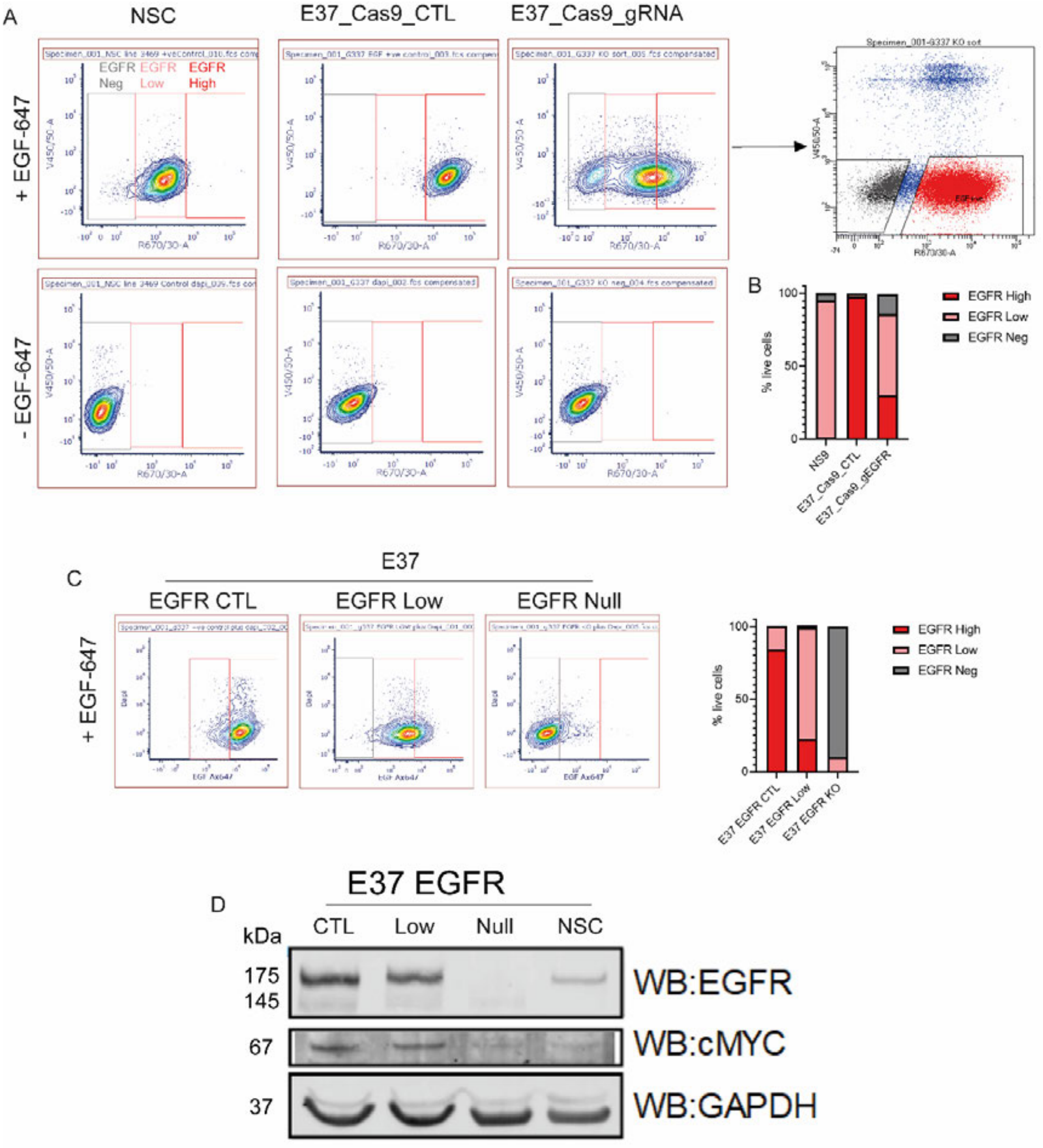


Figure 5.13 | E37 EGFR CTL, Low and Null cell lines (replicate one)

A) Flow cytometry for EGF-647 in indicated cell lines, arrow denotes FACS gating for cell sort into EGFR negative (black) and CTL (red) cohorts. See Appendix 8.8 for EGFR Low FACS. B) Bar plot showing % of cells in each gate shown in EGF-647+ samples in A). C) Flow cytometry for EGF-647 confirming EGFR CTL, Low and Null cell pools recovered after FACS, with bar plot indicating EGFR High, Low and Negative populations denoted in FACS gates. D) Western blot for EGFR, MYC and GAPDH. Biological replicates = 1.

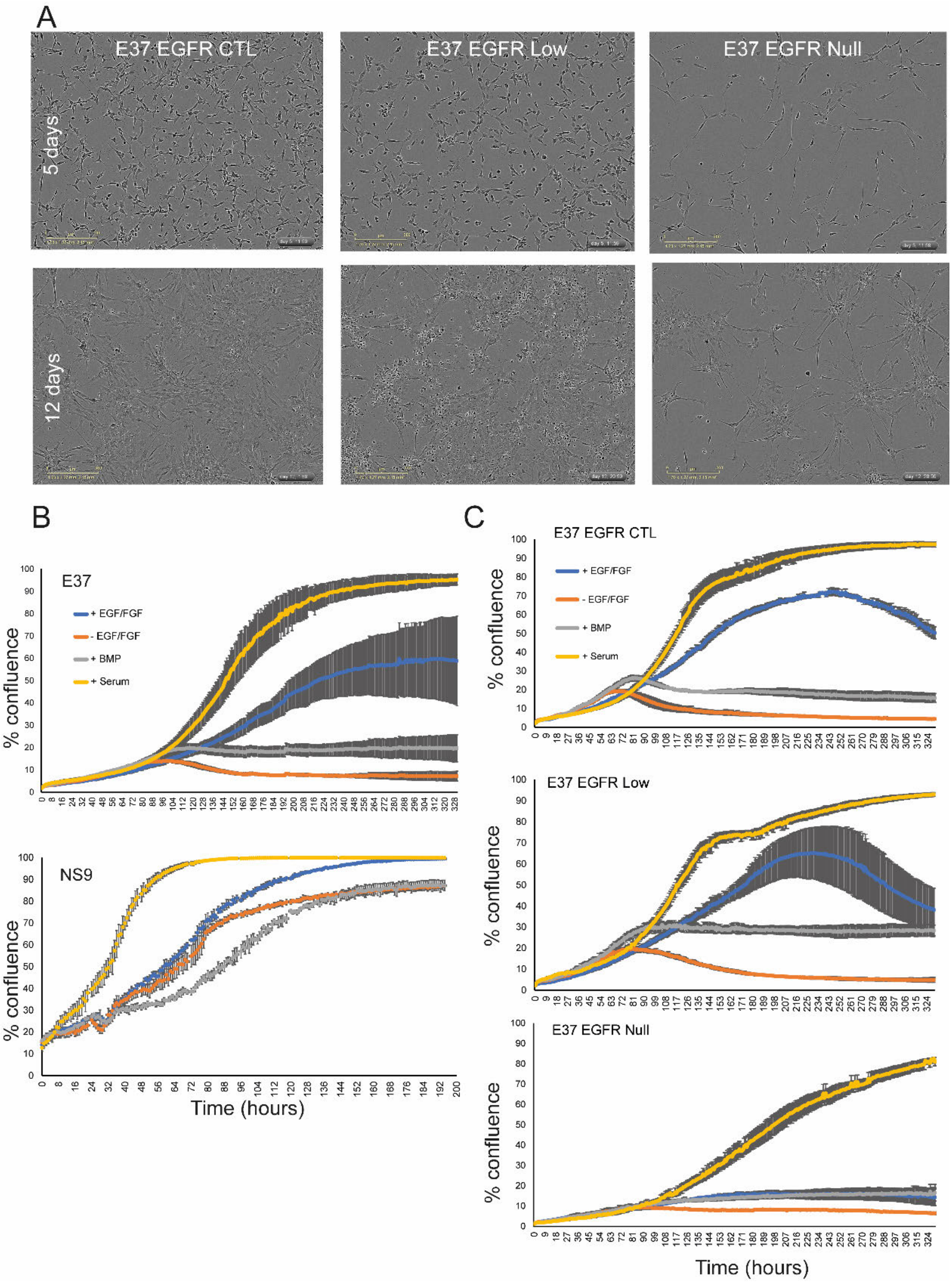


Figure 5.14 | Growth of E37 EGFR CTL, Low and Null cells

A) Images of cells 5 days and 12 days after plating, demonstrating slower growth of E37 EGFR Null cells. Scale bar = 300 µm. B) Growth curve of E37 and NS9 (NSC) cells C) Growth curves of cells grown in the indicated conditions. Biological replicates = 1.

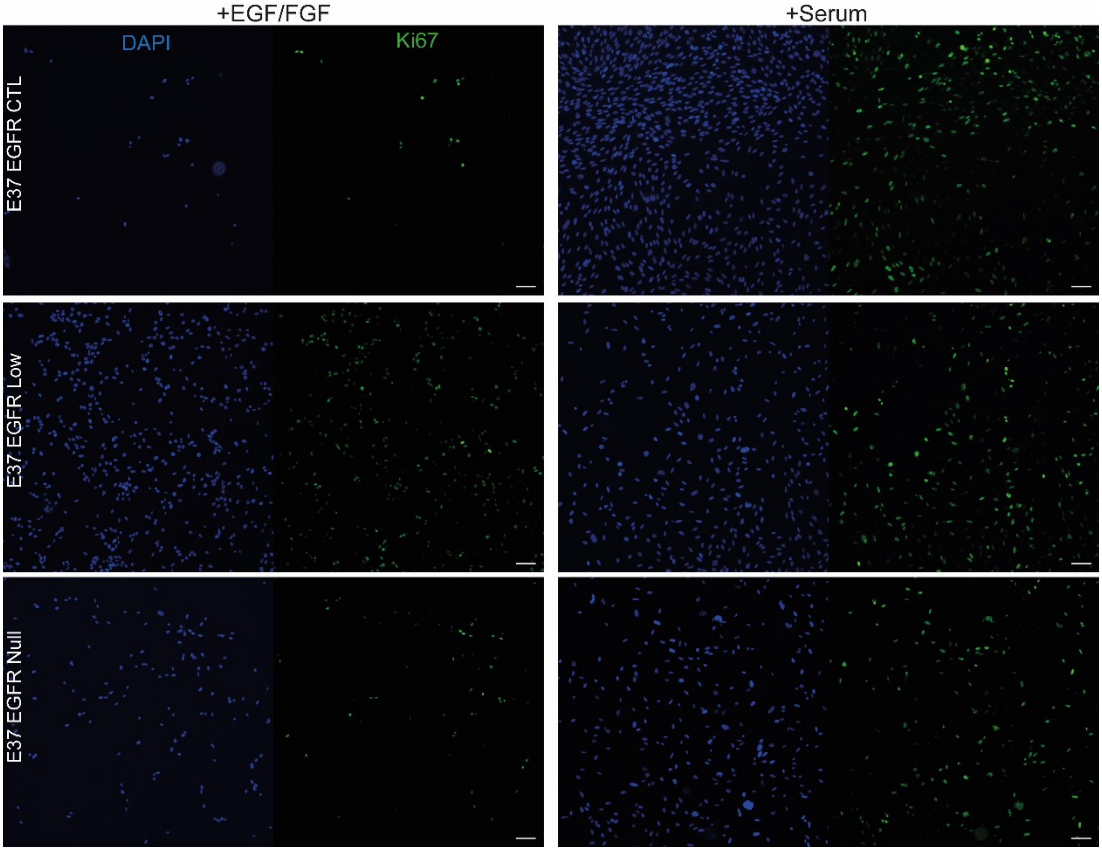


Figure 5.15 | Ki67 staining in E37 EGFR CTL, Low and Null cells

Immunocytochemistry for Ki67 (green) in cells grown in control media (+EGF/FGF) vs 5% serum. DNA = DAPI (blue). Scale bar =100 µm. Biological replicates = 1.

5.4.4 DNA FISH of EGFR Null cells indicates major ecDNA rearrangements

Having characterised EGFR and MYC protein expression, I then sought to characterise these loci using DNA FISH. I expected to see a loss of *EGFR* and *MYC* foci on metaphase spreads in the E37 EGFR Null population. DNA FISH for *EGFR* and *MYC* suggested not only a reduction in the number of clear *EGFR* HSRs but also a reduction of *ecMYC* in E37 treated with Cas9-gEGFR. An increase in *MYC* HSRs and foci of *MYC* signal on chromosomes was also detected in the EGFR-null cell population (Figure 5.16). This suggests that CRISPR-Cas9 targeting EGFR likely ablated *EGFR* expression in the EGFR Null cells via the introduction of indels, but more unexpectedly, resulted in major ecDNA rearrangements in a non-target ecDNA locus.

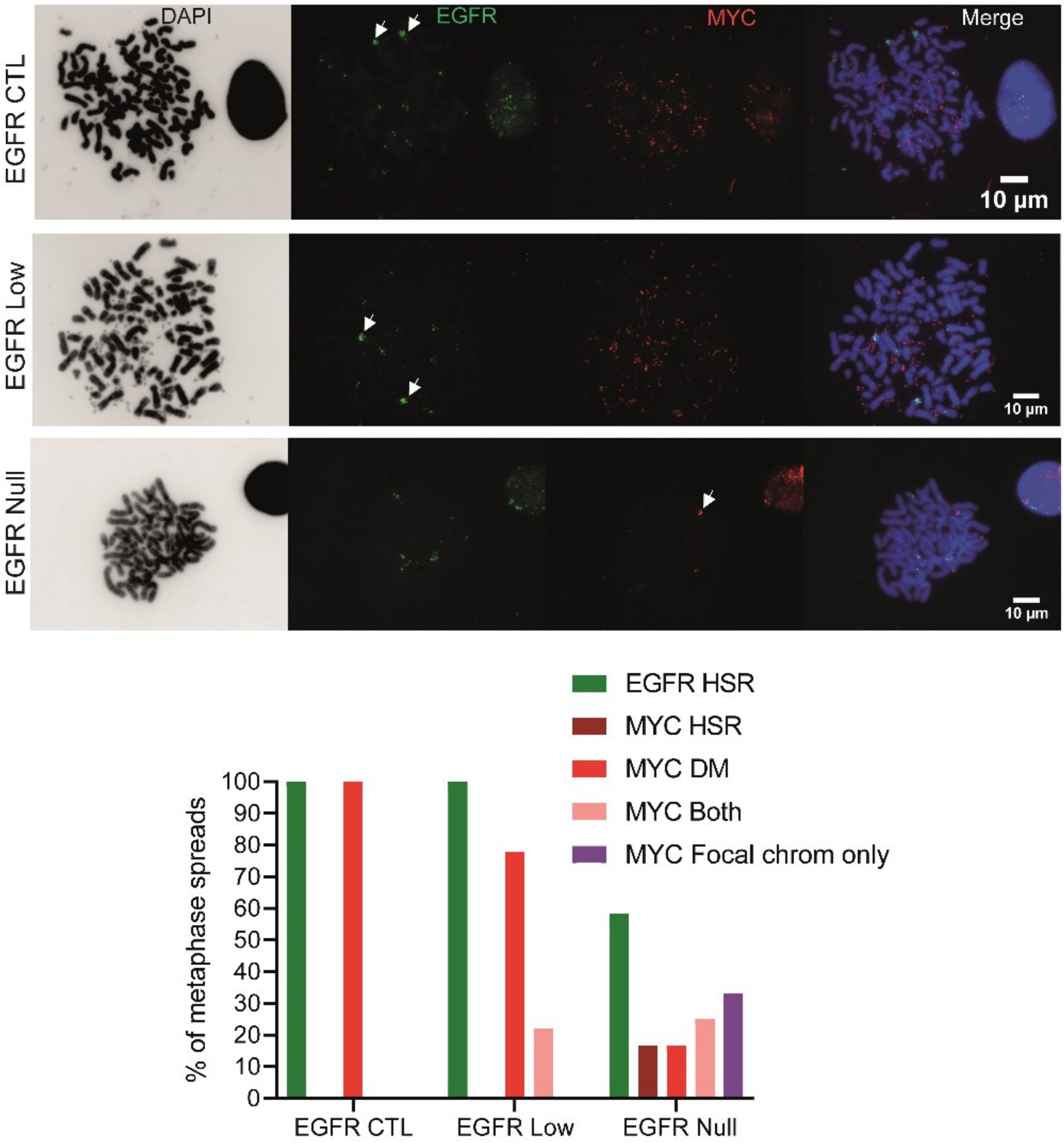


Figure 5.16 | E37 EGFR CTL, Low and Null metaphase spreads (replicate one)

Representative images of FISH signals on metaphase spreads from E37 sorted populations, with white arrows indicating HSRs. Scale bar = 10 µm. Bar chart showing percentage of metaphase spreads with indicated features. Number of metaphases counted per cell line: EGFR CTL = 9, EGR Low = 9, EGFR KO = 12. Biological replicates = 1.

5.4.5 Repeat targeting of EGFR via CRISPR-Cas9

To validate this result, a biological replicate was performed (Figure 5.17A and B). For this replicate, the EGFR Null population was re-sorted by FACS as a broad population recovered following initial FACS sorting (Figure 5.17C), suggesting less efficient CRISPR-Cas9 cleavage than in the first experiment. On subsequent validation by flow cytometry, the three EGFR CTL/Low/Null populations were clearly sustained and separate (Figure 5.17D). Immunoblot showed progressive loss of EGFR across the three populations (Figure 5.17E).

DNA FISH for *EGFR* and *MYC* showed a reduction in the number of *EGFR* HSRs in the EGFR Null population, consistent with the first biological repeat. I observed changes in *MYC* characteristics, including presence of HSRs and larger ecDNA seen in 37.5% (6/16) of metaphase spreads (Figure 5.18).

Overall, this is suggestive of major *ecMYC* rearrangements following *EGFR*-targeted CRISPR/Cas9 cleavage. This is an unexpected result and suggests ecDNA are vulnerable to a broader DDR even when another genomic locus is targeted.

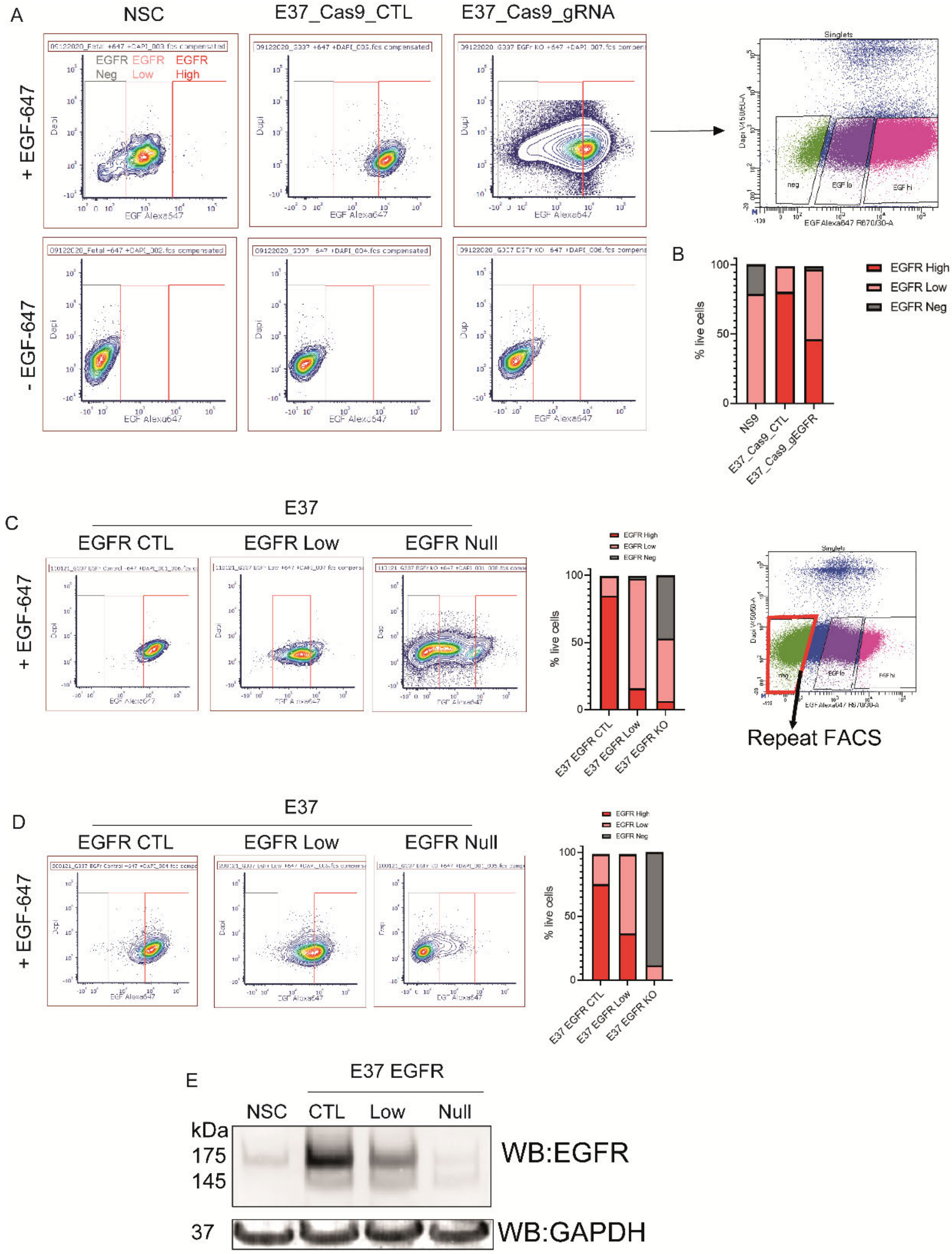


Figure 5.17 | E37 EGFR CTL, Low and Null cell lines (replicate two)

A) Flow cytometry for EGF-647 in indicated cell lines, arrow denotes FACS gating for cell sort into EGFR negative (green), low (purple) and CTL (pink) cohorts. B) Bar plot showing % of cells in each gate shown in EGF-647+ samples in A). C) Flow cytometry for EGF-647 confirming EGFR CTL, Low and Null cell pools recovered after FACS, with bar plot indicating EGFR High, Low and Negative populations. Inset of FACS gating of red gate (EGFR negative) population that was re-sorted. D) Following recovery, flow cytometry for EGF-647 confirming EGFR CTL, Low and Null cell pools. E) Western blot for EGFR and GAPDH. Biological replicates = 1.

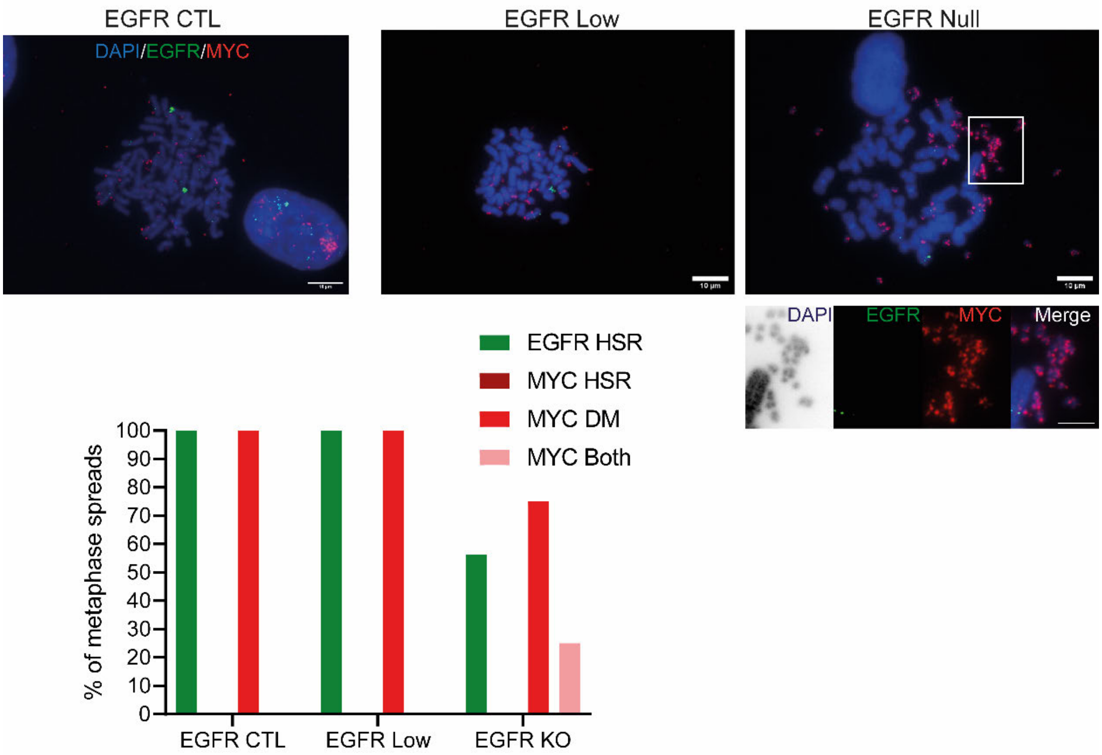


Figure 5.18 | E37 EGFR CTL, Low and Null metaphase spreads (replicate two)

Representative images of metaphase spreads, with white arrows indicating HSRs. Inset highlighting large ecDNA. Scale bar = 10 µm, inset 5 µm. Bar chart showing percentage of metaphase spreads with indicated features. Number of metaphases counted – EGFR CTL = 4, EGFR Low = 3, EGR KO = 16. Biological replicates = 1.

5.4.6 WGS and AmpliconArchitect analysis of E37 EGFR Null cells indicate changes in MYC ecDNA copy number

Rearrangements of *c-MYC* were further evaluated by WGS. Read count profile of both biological repeats suggested higher read counts in regions on chromosome 7 and 8 only (Figure 5.19A and Figure 5.20A).

WGS read number profile is shown in comparison to E37 WGS of the original cell line, noting the lower depth of sequencing of the CRISPR-Cas9 samples (Figure 5.19B). The amplicon regions containing *EGFR* (Figure 5.19B; left) and *MYC* (Figure 5.19B; right) are both lowest in copy number in the EGFR Null cell line. Analysis with AA suggests that the CNV regions most highly amplified in all three cell lines are the region including *c-MYC* (Figure 5.19C). The two populations identified in the original E37 cell line (chr7-chr8 hybrid (containing *EGFR*) and chr8 circular (containing *c-MYC*)) were all present in the EGFR CTL, Low and Null cell lines (Figure 5.19D). Additional amplicons were also identified

(Total number of amplicons called: EGFR CTL – 5, EGFR Low – 4, EGFR Null – 4), although these may be reflective of shallow sequencing rather than true amplicons. When the copy number of the three amplicons were plotted, while there is a subtle reduction in the copy number of the hybrid amplicon containing *EGFR*, the notable change is in chr8-*MYC*-containing amplicon across the three cell populations (Figure 5.19E).

To validate this finding, I repeated the analysis in biological replicates. WGS profile in the EGFR Null cell line was similar to the other cell lines across the *EGFR* amplicon region, suggesting less efficient knock-out of EGFR, and a less marked reduction in the *MYC* amplicon is also apparent (Figure 5.20B). This may explain why I observed less efficient EGFR KO following CRISPR-Cas9-gEGFR targeting when initially analysed by flow cytometry with EGF-647 (Figure 5.17). Again, the *MYC*-containing CNV region was the most amplified across all three cell lines (Figure 5.20C). The *EGFR*-hybrid and *MYC*-chr8 amplicons were identified in all three cell lines along with other amplified regions (see caveats above) (Figure 5.20D). The copy number of the *MYC*-chr8 amplicon is highest in EGFR Low cells and lowest in EGFR KO cells (Figure 5.20E). Note in these biological replicates, AA only predicted one additional amplicon in the EGFR Low cell line (Total number of amplicons called: EGFR CTL – 3, EGFR Low – 4, EGFR Null – 3).

Overall, WGS and AA analysis suggest that CRISPR/Cas9-induced DNA damage targeting a locus (*EGFR*) leads to rearrangement of a non-targeted, (*c-MYC*) oncogene harbouring ecDNA population, despite less efficient *EGFR* targeting in the biological repeat. This was surprising and suggests that the maintenance of ecDNA is compromised by induced DSBs in a cell. Activation of the DDR in one ecDNA population can lead to secondary effects that affect other ecDNA that are not directly cleaved.

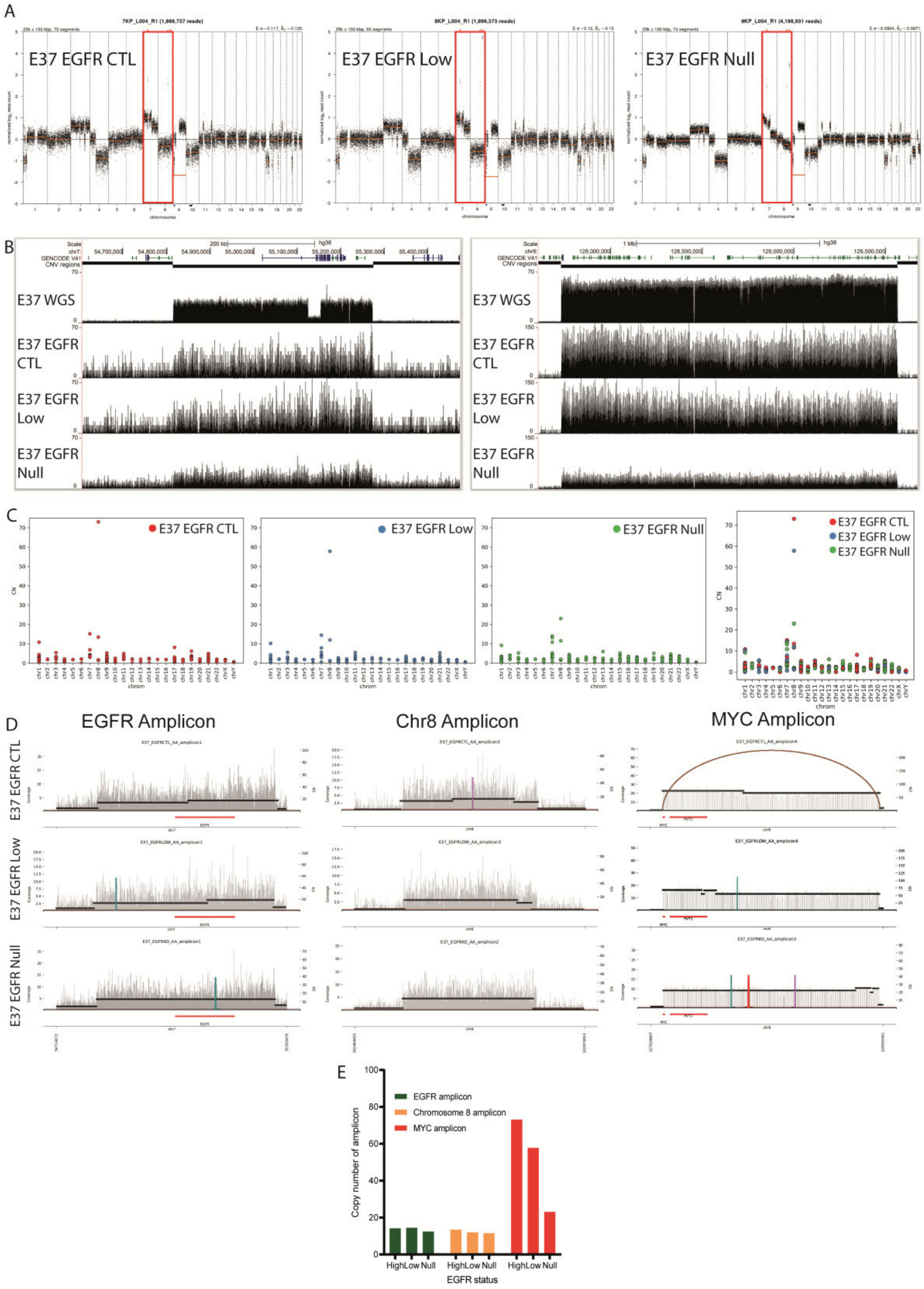


Figure 5.19 | WGS and AA analysis of E37 EGFR CTL, Low and Null cell lines (replicate one)

A) Copy number profile (15kbp normalised) of E37 EGFR CTL, Low and Null cell lines. Red box = chr7 and 8 regions. B) WGS of E37 (original cell line) and EGFR CTL, Low and Null cell lines. C) AA CNV regions by chromosome plotted against CN for EGFR CTL, Low and Null cell lines, plotted separately and together. D) Amplicon in E37 EGFR CTL, Low and Null cell lines via AA. E) Copy number of regions shown in D)

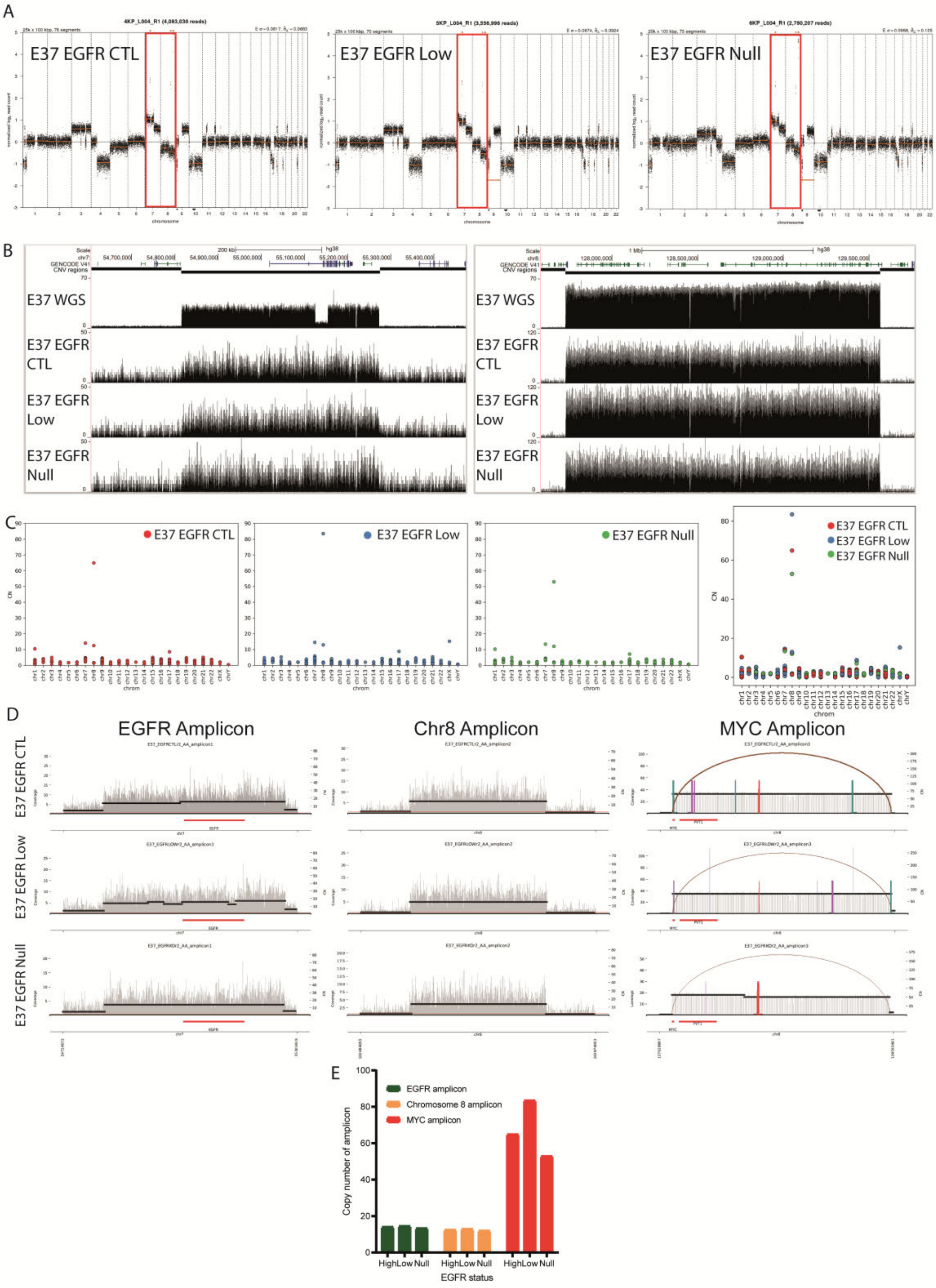


Figure 5.20 | WGS and AA analysis of E37 EGFR CTL, Low and Null cell lines (replicate two)

A) Copy number profile (15kbp normalised) of E37 EGFR CTL, Low and Null cell lines. Red box = chr7 and 8 regions. B) WGS of E37 (original cell line) and EGFR CTL, Low and Null cell lines. C) AA CNV regions by chromosome plotted against CN for EGFR CTL, Low and Null cell lines, plotted separately and together. D) Amplicon in E37 EGFR CTL, Low and Null cell lines via AA. E) Copy number of regions shown in D)

5.5 Discussion

5.5.1 DNA damage in glioblastoma stem cells

DNA strand breaks play a central role in ecDNA formation and subsequent dynamics. It is thought that ecDNA originate primarily through chromothripsis leading to rearrangements and circular recombination, with more recent work also highlighting the role played by DNA damage in ecDNA reintegration (Stephens *et al.*, 2011; Rosswog *et al.*, 2021; Shoshani *et al.*, 2021). The only standard non-surgical treatments of glioblastoma, namely radiotherapy and TMZ chemotherapy, induce DNA damage (Erasimus *et al.*, 2016). Activation of the DDR and altered DNA repair mechanisms are both implicated in tumour recurrence and poor prognosis in glioblastoma (Lozinski *et al.*, 2021). The data in this chapter give some preliminary indications of how ecDNA in glioblastoma stem cells respond to different types of DNA damage – non-specific DNA damage in the form of IR, and targeted DNA damage via CRISPR-Cas9. Identification of some unique vulnerability of ecDNA, and an ability to clear these through modulating DNA repair pathways, might be of great interest therapeutically.

Previous analysis of glioblastoma stem cells have shown constitutive activation of the DDR machinery using γH2AX as a marker, a characteristic which has been linked with replication stress, radioresistance and tumour recurrence (Bao *et al.*, 2006; Bartkova *et al.*, 2010; Carruthers *et al.*, 2018). My data confirmed that our GSC cultures harbour ongoing sites of DNA damage as visualised by staining for γH2AX, consistent with these previous studies. Cells with ecDNA have been suggested to harbour increased markers of replication stress, although it is not reported whether this is in a single or pooled cell analysis (conference abstract, (Chowdhry *et al.*, 2022). I did not observe any obvious spatial relationship between ecDNA and γH2AX foci (sites of active DNA damage), suggesting that while these features are present, there is no enrichment of DNA damage on ecDNA. This is supported by data suggesting ecDNA are formed as a result of a single catastrophic event rather than ongoing DNA damage and cumulative events (Stephens *et al.*, 2011).

5.5.2 Non-specific DNA damage via IR

Using IR to induce non-specific DNA damage is not only clinically relevant to glioblastoma, but also allows dsDNA strand breaks to be generated in a time and dose-controlled manner. I sought to characterise more immediate changes in ecDNA dynamics within days of DNA damage induction. My data suggest a reduction in ecDNA number only 24 hours after irradiation. Further, combining IR with the PARP inhibitor olaparib results in a reduction in EGFR expression but an apparent protection in ecDNA copy number.

Previous studies have also recognised a loss of ecDNA after IR (Schoenlein *et al.*, 2003), although my data suggest this happens early following IR, and in primary glioblastoma cells. Another analysis at a similar early time point noted integration of ecDNA to chromosomes (as focal and ectopic HSRs) and generation of larger ecDNA, which was further enhanced in the presence of PARPi (Shoshani *et al.*, 2021). However, integration alone would not explain the reduction in copy number. Are ecDNA specifically damaged, leading to their linearisation and destruction, or are cells with ecDNA more genomically unstable and therefore fail to survive by another mechanism? IR would not be expected to have an effect on cell survival 24h after IR (and indeed there was no subjective cell loss observed) but this could be further evaluated using an alternative assay for cell survival (these are outlined in Section 2.6.1.1). Cells could be separated by FACS using the EGF-647 fluorescent ligand into ecDNA high and low populations, and compare cell survival to establish if ecDNA copy number is correlated with radiosensitivity.

In addition, the synergistic reintegration of ecDNA following IR and PARPi would not explain the contrasting observation between EGFR copy number (DNA FISH) and EGFR expression (flow cytometry), given that the data presented in Chapter 4 suggest expression of oncogenes is not significantly different at ecDNA vs chromosomal loci. PARP1 is known to regulate transcription in addition to its DNA damage response (Lee *et al.*, 2012; Rose *et al.*, 2020). My results may represent the combined effects of IR on ecDNA reintegration, and PARP-induced transcriptional repression of *EGFR*. Another possibility is that PARP-trapping on ecDNA prevents the copy number loss of ecDNA following DNA damage, while simultaneously resulting in transcriptional repression.

Subsequent experiments will be needed to validate these observations and increase sample numbers. In order to characterise EGFR expression following IR and PARP inhibition, FlowFISH, a tool that allows RNA expression to be quantified by flow cytometry, could be used (Fulco *et al.*, 2019). This would confirm the loss of *ecEGFR* following IR and whether PARPi results in its transcriptional repression. I could also use RNA FISH to visualise RNA transcripts, although the sample number would be smaller.

A strategy for higher throughput genomic analysis is needed to better characterise ecDNA dynamics. WGS analysis can, as shown in relation to targeted ecDNA damage, give some indication of ecDNA characteristics and copy number. I would propose to utilise shallow WGS in the first instance to validate the observations from DNA FISH, specifically ecDNA copy number. Having validated these findings, I would then optimise tools to characterise these further. Metaphase spreads offer a low throughput, subjective means of characterising DNA arrangements that current WGS analysis tools such as AmpliconArchitect cannot (e.g. HSR vs ecDNA). Long-read sequencing using Oxford Nanopore Technology (ONT) also offers a means of more deeply characterising the likely highly diverse population of ecDNA present in these glioblastoma cell lines. Isolating, enriching and characterising ecDNA by long read sequencing (Hung *et al.*, 2022) could be one strategy for characterising individual ecDNA reads within a cell population.

5.5.3 Targeted DNA damage via CRISPR-Cas9

Non-specific DNA damage via IR generated loss of ecDNA copy number. I used a targeted DNA damage strategy to evaluate whether targeting this DNA damage specifically to ecDNA resulted in comparable observations, and indeed whether this might represent an ecDNA-specific therapeutic approach. I used CRISPR-Cas9 to target the *EGFR* oncogene present on ecDNA in E26 cells. One might have expected this to simply create indels, thereby affecting *EGFR* expression and function. Unexpectedly, this generated a large population of cells where ecDNA harbouring *EGFR* were eliminated, while retaining chromosomal *EGFR* foci on metaphase spreads in-keeping with the known chromosome 7p11.2 localisation of endogenous *EGFR* (3-6 copies of chromosome 7 per cell). This might be due to linearisation and subsequent degradation of the ecDNA by nucleases, or a broader DDR. This is an entirely novel result not previously reported. (Shoshani *et al.*, 2021) designed TALENs and CRISPR-Cas9 tools to target DHFR ecDNA generated in HeLa cells via MTX selection. They used a sgRNA approximately 3.5kb centromeric to the *DHFR* locus and reported approx. 6-22% DM reintegration depending on the exact strategy used, and no comment is made suggesting ecDNA deletion or loss without integration. This may reflect the type of cells studied, with ecDNA induced by drug selection (and requiring ongoing drug treatment for their maintenance) as opposed to the glioblastoma cells here where ecDNA are already established and stable. A study of mitochondrial eccDNA (mtDNA) used CRISPR/Cas9-guided cleavage with subsequent exonuclease degradation to remove most (>85%) eccDNA from a range of mouse and human cells (Feng *et al.*, 2022), although mtDNA is notably smaller than ecDNA at only 14-20kb. My data provides the first evidence that ecDNA might be removed via CRISPR-Cas9 targeting, eliminating a major source of oncogene amplification.

A possible caveat to my experiments is that I have inadvertently selected for a small proportion of pre-existing ecDNA-null E26 cells rather than creating new ecDNA null populations. This seems unlikely as in all flow replicates (n=3), <1% of the Cas9_CTL existed in the EGFR-negative population, and the EGFR-low population is also sparse (Figure 5.10). In addition, previous data using FACS to sort GBM39 cells harbouring ecEGFRvIII into EGFRvIII-high and -low populations saw the re-establishment of a mixed population within 2 weeks of sorting by EGFRvIII status (Nathanson *et al.*, 2014), in contrast with the sustained change in EGFR status shown here. The experiment could be repeated with dead Cas9 (dCas9), a non-catalytically (and therefore non-cleaving) version of Cas9, or dCas9-KRAB (Li *et al.*, 2021) which would silence EGFR expression without cleaving *EGFR.* If the results showed that there was no *ecEGFR* loss without active Cas9-driven cleavage, this would support the hypothesis that *ecEGFR* are being targeted, cleaved and lost.

Overall, the effect of the DDR in reducing ecDNA copy number despite the DDR being responsible for ecDNA genesis via chromothripsis raises the question of whether ecDNA are a critical entity in established glioblastoma cells or merely a side-effect of upstream events. Therefore further experiments should explore the effect of ecDNA loss on tumorigenesis and treatment susceptibility. Tumorigenesis could be evaluated by characteristics such as proliferation (e.g. Ki67, EdU incorporation) and the ability to generate mouse tumour xenografts. This would robustly demonstrate whether loss of ecDNA affects the ability of tumours to develop and thrive. Treatment susceptibility to IR and other anti-cancer therapeutics (e.g. TMZ) could be assessed via cell survival assay and cell death. These are important tools for assessing the true importance of ecDNA in cancer maintenance, and therefore whether they represent a meaningful potential therapeutic target.

I also found that targeted CRISPR-Cas9-induced DNA damage of a non-ecDNA oncogene generated an unexpected copy number and structural change on the *c-MYC*-containing ecDNA population in E37 cells. This is another hugely exciting and unexpected finding. This may represent an off-target effect of CRISPR-Cas9 - it has been suggested that ≥50% RNA-guided endonuclease mutations occur at non-target sites with CRISPR-Cas9 in a manner that cannot solely be resolved by reducing gRNA or CRISPR-Cas9 concentrations (Fu *et al.*, 2013; Zhang *et al.*, 2015). Another possibility is that ecDNA loss is due to the induction of a large number of DSBs and the DDR, or is representative of the inherent flexibility of ecDNA-harbouring cells. It would be interesting to observe whether random DNA damage via IR generated a similar effect on the MYC locus.

My findings suggest that there may be off-target effects of inducing focused DNA damage and targeting ecDNA. Again, it is possible that I have selected a small pre-existing subpopulation of cells. To validate these observations, the experiment could be repeated with nuclease dCas9, or dCas9-KRAB (Li *et al.*, 2021). I could then FACS-sort cells using the same gating strategy and use the analysis tools presented here to characterise the ecDNA population. This would confirm that the observed changes in *c-MYC* ecDNA were a result of off-target DNA damage effects rather than due to cell selection. In addition, I would propose to repeat this experiment in another cell line with another non-*EGFR* ecDNA-harbouring glioblastoma cell line, such as E20, to confirm this is not a cell line-specific effect. As above, further studies could expand on the phenotypic changes observed in E37 cells to characterise features such as tumorigenesis and treatment susceptibility. A more detailed study of DNA damage and repair proteins could then elucidate the mechanism behind this exciting observation.

Both experiments highlight some limitations of existing WGS analysis tools in evaluating ecDNA amplicons, but the impact is more significant in the analysis of E37. AA is not designed to differentiate HSRs and ecDNA, and indeed was unable to do so in E37, which FISH shows harbours *EGFR* HSRs, despite AA predicting a circular ecDNA amplicon. In addition, shallower sequencing data, as is presented here in the CRISPR-Cas9 samples, results in broken circles that cannot be identified by AA. Overall, this limits the interpretation of AA data in characterising ecDNA structural changes, although it remains useful in identifying amplified regions and quantifying the copy number of these regions. Other tools, such as long read sequencing and enrichment of ecDNA prior to sequencing (Hung *et al.*, 2022), offer methods to overcome these limitations.

In conclusion, these data indicate ecDNA are vulnerable to DNA strand breaks which have a major impact on ecDNA copy number, morphology and function. Crucially, this appears to occur via direct targeting of ecDNA and an as-yet undefined broader DDR affecting ecDNA. Understanding ecDNA maintenance, and their corresponding role in tumour maintenance and evolution, is crucial for their potential therapeutic exploitation.

Chapter 6 Conclusions

Glioblastoma is a cancer with limited treatment options and poor prognosis, characterised by intratumoural genomic heterogeneity and molecular subtypes defined by core oncogene amplifications (Cancer Genome Atlas Research Network, 2008; Verhaak *et al.*, 2010; Brennan *et al.*, 2013; Stupp *et al.*, 2014; Wang *et al.*, 2017). Similarly, ecDNA are associated with treatment resistance and poor prognosis, intratumoural copy number heterogeneity and are the location of all key oncogene amplifications (Turner *et al.*, 2017; Kim *et al.*, 2020). Given that ecDNA are particularly common in glioblastoma compared with other cancers (Turner *et al.*, 2017; Kim *et al.*, 2020), it creates an attractive hypothesis that many of the characteristic features of glioblastoma could be attributed to ecDNA. At a minimum, it implies that glioblastoma represents an important model in which to study the fundamental characteristics of ecDNA and whether they deploy some regulatory mechanisms that might be targeted therapeutically.

6.1 Patient-derived glioblastoma cells as a model for the study of ecDNA

During my PhD, I had access to primary and recurrent glioblastoma cell lines harbouring characteristic glioblastoma oncogene amplifications via the GCGR. These included three of the classical subtype (E26, E28, E37) and two of the proneural subtype (E20, E25), in addition to a primary NSC forebrain-derived cell line cultured as part of the same GCGR collection. Having selected these glioblastoma cell lines on the basis of copy number profiles of key oncogenes (*EGFR, PDGFRA, CDK4*), I showed that ecDNA copy number is dynamic, and varies by passage number and cell line, in agreement with previous studies (Turner *et al.*, 2017; Lange *et al.*, 2022). I characterised the ecDNA in these cell lines using DNA FISH and WGS analysis, showing that all amplified oncogenes were located on ecDNA or HSRs. This supports previous findings that all amplified oncogenes are found on ecDNA (Turner *et al.*, 2017).

Bulk WGS analysed using AA suggested that some ecDNA represent quite simple structures, such as the EGFR amplicon in E28 cells or the *PDGFRA* amplicon in E20 cells. In contrast, other amplicons appeared far more complex, such as E25 amplicons, and the *CDK4* amplicon in E20 cells. Whilst not the focus of this thesis, this raises some interesting questions about the origins of ecDNA, with recent studies citing chromothripsis as being both very common in glioblastoma and the predominant source of ecDNA (Cortés-Ciriano *et al.*, 2020; Rosswog *et al.*, 2021; Shoshani *et al.*, 2021). The mechanism of chromothripsis might explain the more complex, multi-fragmented nature of some amplicons, whereas simpler ecDNA structures might be more easily explained by the post-replicative episome model. It may be that multiple mechanisms can result in ecDNA formation, or that simple ecDNA can be further fragmented by micronuclei-driven chromothripsis (C.-Z. Zhang *et al.*, 2015; Shoshani *et al.*, 2021). Studies using primary cell lines, including gliomas, mainly support the episome model of ecDNA formation (Vogt *et al.*, 2004; Storlazzi *et al.*, 2006; Hung *et al.*, 2022), whereas those using established cell lines or drug-induced ecDNA cell models suggest chromothripsis as the primary initiating mechanism (Stephens *et al.*, 2011; Shoshani *et al.*, 2021), suggesting the model system may be important. An exception to this is a study which used a combination of primary neuroblastoma samples, established cell lines and WGS data from over 2500 cancer samples, which collectively supported the chromothripsis model (Rosswog *et al.*, 2021). Uncertainty as to the origins of ecDNA remain. Overall, my data suggests that more than one mechanism may be involved, and that this represents an important area of future study.

Comparing genomic analysis tools (bulk WGS and visualisation of genomic loci by DNA FISH) also highlighted some of the relative merits and limitations of each approach. Bulk WGS analysis using AA resulted in some errors, such as calling an HSR as ecDNA (E37 - *EGFR*), calling a circular amplicon as linear (E20 - *PDGFRA*), and being unable to identify rare subclones (E20 - *CDK-PDGFRA* amplicon). These limitations of AA have been recognised previously (Deshpande *et al.*, 2019; Kim *et al.*, 2020; Hung *et al.*, 2022), although they are not widely discussed in ecDNA literature. In all cases, these were validated by DNA FISH on metaphase spreads, a tool that is low throughput and requires good quality metaphase spreads, DNA FISH probes and high resolution imaging, all of which can be challenging. Both approaches will significantly underestimate the diversity of ecDNA species present in a single nucleus, let alone the many cells comprising a whole tumour. Methods such as isolation and enrichment of ecDNA prior to sequencing (Hung *et al.*, 2022), long read sequencing (Wu *et al.*, 2019) or multi-modality tools such as spatial transcriptomics (Walentynowicz *et al.*, 2023) could represent key technologies in the future study of ecDNA, particularly at a single cell level. Selection of an appropriate *in vitro* cellular model both for glioblastoma and ecDNA is important to ensure my findings are biologically and clinically relevant. Overall, I demonstrated that glioblastoma cell lines from the GCGR harboured ecDNA and represented a tractable and disease-relevant model for the study of ecDNA dynamics.

6.2 Spatial organisation and transcriptional regulation of ecDNA

Next, I addressed the core question of transcriptional efficiency of ecDNA by evaluating the spatial organisation and transcriptional regulation of ecDNA in glioblastoma cells. I used quantitative image analysis of super-resolution imaging to show that ecDNA are not clustered, either with each other or with transcriptional hubs, at distances associated with coordinated transcription (Figure 6.1). My data supported a more central organisation, suggestive of the spatial freedom of ecDNA relative to chromosomes, and a regional localisation more reflective of the non-random organisation of the nucleus into chromosome territories and A/B compartments. This contrast with other published data may be due to differences between the primary cell cultures used here and previously studied established cell lines, including the Colo320DM cell line which harbours multiple copies of MYC on each ecDNA (Wu *et al.*, 2019; Hung *et al.*, 2021). Additionally, the analysis presented here represents 3D analysis of well-resolved DNA foci, rather than the potential artifacts caused by poor quality imaging or current live cell imaging approaches (Hung *et al.*, 2021; Yi *et al.*, 2021). This new analysis methodology represents an important tool now available to others in the field.

I also showed that transcriptional hubs are not frequent and do not colocalise with ecDNA, including nascent ecDNA-resident oncogene transcripts. Hubs may not be important in GSCs, and indeed this has been suggested in other cancers. The number of hubs (described as ‘RNA Pol II Condensates (PCs)’) in a study of HCT116 (colorectal cancer) cells demonstrated a similar range of PCs per nucleus as observed here (Imada *et al.*, 2021). In Ewing sarcoma, it has been shown that an unexpected effect of a characteristic chromosomal translocation could abrogate TF activity required for phase separation, with the subsequent hypothesis that failure of hub formation could represent a broader pathogenic phenomenon (Boulay *et al.*, 2017; Palacio and Taatjes, 2022). Overall, cancer cells likely harbour markedly different characteristics, including transcriptional regulation, as compared with the models used in many studies of phase separation and transcriptional hubs, regardless of the presence of ecDNA. Given that RNA Pol II-ecDNA colocalisation has thus far only been proposed by ChIA-PET and subjective calling of imaging (Yi *et al.*, 2021; Zhu *et al.*, 2021), further studies should establish whether RNA Pol II binding with ecDNA binding differs from that of chromosomal loci.

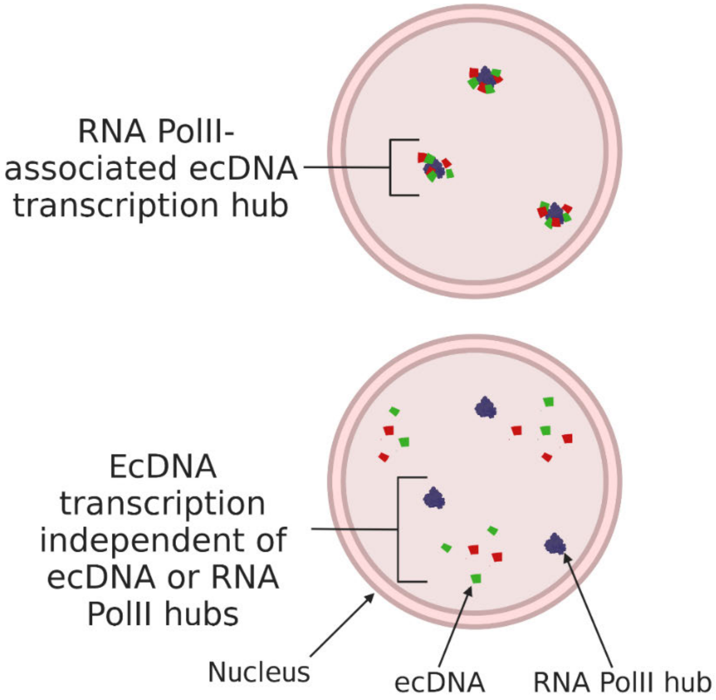


Figure 6.1 | Hypotheses for ecDNA transcription

Top – Existing hypothesis for ecDNA transcription driven by ecDNA-ecDNA and ecDNA-RNA Pol II hubs. Bottom – Hypothesis proposed by this thesis – that ecDNA transcription is independent of ecDNA-ecDNA and ecDNA-RNA Pol II hubs.

I subsequently demonstrated via super-resolution imaging and sequencing analysis that the greater transcriptional output from oncogene-resident ecDNA is primarily driven by copy number rather than increased transcription at individual ecDNA loci. This was most definitively demonstrated by SNP analysis, which included genes not under the same selective pressure as oncogenes. Although other studies have proposed that transcription from ecDNA exceeds that observed by copy number alone, closer examination of copy number-normalised data suggests overall agreement with my findings (Wu *et al.*, 2019; Kim *et al.*, 2020; Stöber *et al.*, 2023).

Overall, my data suggest that the transcriptional regulation of genes on ecDNA in GSCs is very similar to that of chromosomal loci. Future research might explore the varying intranuclear heterogeneity of ecDNA transcription, with the recent observation from single cell sequencing that copy-number normalised ecDNA gene expression is highly variable (Stöber *et al.*, 2023). This has been supported by single-cell extrachromosomal circular DNA and transcriptome sequencing (scEC&T-seq) which also showed intercellular differences in ecDNA-resident oncogene expression (Chamorro González *et al.*, 2023). Developing such tools to combine long-read sequencing with chromosome conformation capture techniques might reveal if the individual combination of genes and CREs on single ecDNA is important for ecDNA-resident gene transcription. For example, Hi-C, which has already been used to show ecDNA-chromosomal interactions (Zhu *et al.*, 2021), can be combined with enriched ecDNA to characterise CREs (Hung *et al.*, 2022) indicating the feasibility of such an approach.

6.3 The effect of DNA damage on ecDNA

GSCs have a constitutively activated DDR, and I also observed this in the glioblastoma cell lines studied in this thesis. That these DDR sites are not spatially related to ecDNA suggests this constitutive DDR activity is unrelated to existing ecDNA dynamics. As such, I proceeded to evaluate the effect of exogenous DNA strand breaks, first in the form of random IR-induced DNA damage, and second through targeted DNA damage using CRISPR-Cas9. Overall, this revealed a clear effect of DNA strand breaks on ecDNA copy number and structure (Figure 6.2).

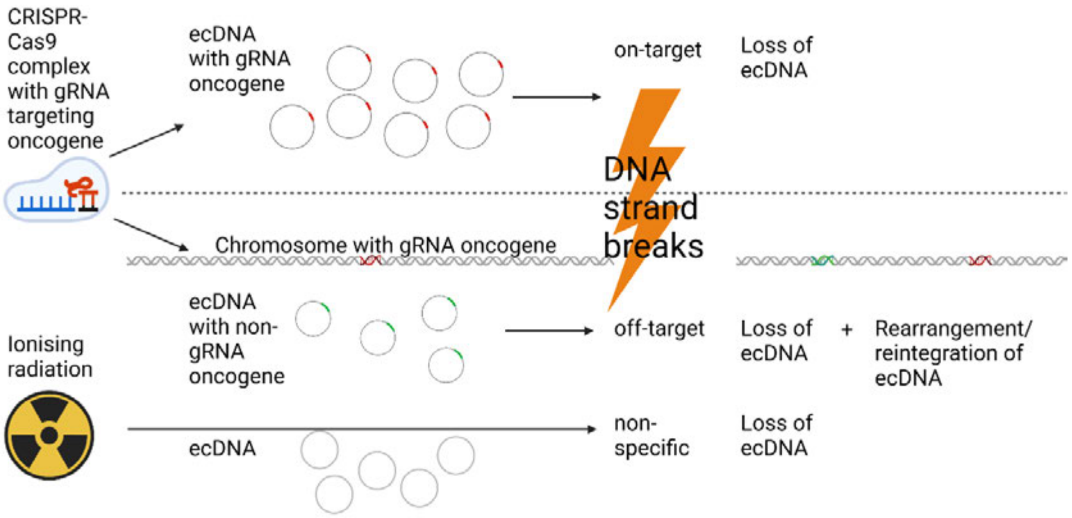


Figure 6.2 | Summary of responses to DNA Damage

DNA damage that is targeted (CRISPR-Cas9) and non-targeted (IR) both result in DNA strand breaks and subsequent loss of ecDNA. In addition, where CRISPR-Cas9 is targeting a non-ecDNA oncogene, rearrangement and reintegration of ecDNA was observed.

IR-induced DNA damage resulted in reduced ecDNA copy number, even after the relatively short period of 24h. In addition, PARP inhibition in addition to IR appeared to inhibit EGFR expression, while protecting ecDNA copy number from IR-associated loss. Previous studies suggested ecDNA may reintegrate or form micronuclei as mechanisms of reduced ecDNA copy number in response to IR (Sanchez, Barrett and Schoenlein, 1998; Schoenlein *et al.*, 2003; Shoshani *et al.*, 2021). These data are the first time this has been studied in primary patient-derived cells with established ecDNA. Characterising this more thoroughly through repeat DNA FISH analysis of metaphase spreads and interphase nuclei for micronuclei is therefore a vital next step. More thorough characterisation of the effect of IR and PARylation through higher throughput modalities such as FlowFISH, WGS and long read sequencing will also provide greater validity to the preliminary data presented here. I chose to focus on PARylation owing to the current glioblastoma clinical trial landscape and data suggesting that treatment with a PARPi, but not a DNA-PKc (i.e. NHEJ) inhibitor, resulted in ecDNA integration following DNA damage (Shoshani *et al.*, 2021). However, given the latter was performed in Colo320DM cells and a drug-induced ecDNA model, there is value in revisiting the role of different DSB repair mechanisms in the primary patient-derived cell models presented in this thesis. This could be extended to compare random SSBs vs DSBs using TOPI and TOPII inhibitors respectively.

I used CRISPR-Cas9 to target and eliminate *ecEGFR*, with chromosomal loci observed in cells where EGFR expression was either low (comparable to NSC EGFR expression) or null. Whether these represent reintegrated loci, as proposed in a study of drug-induced ecDNA (Shoshani *et al.*, 2021), or existing chromosomal loci is unclear. The mechanism of this ecDNA loss is also unknown, and may be relevant to both random (e.g. IR induced) and targeted DNA strand breaks. The two primary hypotheses could be that ecDNA are linearised and degraded following DNA strand breaks, or that the huge DDR and/or genomic instability generated in ecDNA-harbouring cells results in cells without ecDNA harbouring a selection advantage. Elucidating the mechanism may require clonal tracking via optical or genetic barcoding (Serrano *et al.*, 2022).

It remains unclear whether ecDNA represent an important therapeutic target entity, or are merely a downstream epiphenomena of upstream cellular events. While many glioblastoma have ecDNA, ~40-50% do not (Turner *et al.*, 2017; Kim *et al.*, 2020), implying ecDNA are not a ubiquitous and therefore essential feature for all glioblastoma to form and flourish. Previous tools have suggested that ecDNA-'high' cells are more tumorigenic, aggressive and proliferative than ecDNA-'low' cells, although crucially these were GBM39 cells sorted into ecDNA 'high' and 'low' populations rather than sustained cell lines created by CRISPR-Cas9 (Nathanson *et al.*, 2014). In addition, proliferation may be less important than quiescence in identifying treatment resistant glioblastoma cells that result in glioblastoma relapse (Sachdeva *et al.*, 2019). The cell lines developed here provide an ideal tool with which to study the effect of ecDNA loss on tumorigenesis and treatment susceptibility in a non-clonal primary cell model.

Finally, I showed that, surprisingly, CRISPR-Cas9 targeting a non-ecDNA-resident oncogene resulted in major copy number and structural rearrangements of ecDNA, with relatively minor genomic changes at the target locus. This suggests that ecDNA may be vulnerable to an excessive DDR, even if not directed specifically at ecDNA themselves. If this is reproducible in another cell line, exploring the underlying mechanism will explain how ecDNA evolve in response to damage, both in terms of current glioblastoma treatments and future treatments targeting ecDNA.

Overall, ecDNA represent a major mechanism of oncogene amplification in many cancers, including glioblastoma. Despite being described for the first time almost 60 years ago, novel scientific tools are enabling their study in greater detail. This thesis has utilised super-resolution imaging, quantitative image analysis and bioinformatics strategies to address important questions about ecDNA organisation and regulation in glioblastoma stem cells, as well as explore their response to DNA damage. Further study will identify if ecDNA represent an important targetable vulnerability in glioblastoma, and potentially other cancers.

Chapter 7 References

Ahmed, S.U. *et al.* (2015) ‘Selective Inhibition of Parallel DNA Damage Response Pathways Optimizes Radiosensitization of Glioblastoma Stem-like Cells’, *Cancer research*, 75(20), pp. 4416–4428.

Ahnesorg, P., Smith, P. and Jackson, S.P. (2006) ‘XLF interacts with the XRCC4-DNA ligase IV complex to promote DNA nonhomologous end-joining’, *Cell*, 124(2), pp. 301–313.

Alcantara Llaguno, S. *et al.* (2019) ‘Cell-of-origin susceptibility to glioblastoma formation declines with neural lineage restriction’, *Nature neuroscience*, 22(4), pp. 545–555.

Aldape, K. *et al.* (2019) ‘Challenges to curing primary brain tumours’, *Nature reviews. Clinical oncology*, 16(8), pp. 509–520.

Al-Hajj, M. *et al.* (2003) ‘Prospective identification of tumorigenic breast cancer cells’, *Proceedings of the National Academy of Sciences of the United States of America*, 100(7), pp. 3983–3988.

Alitalo, K. *et al.* (1983) ‘Homogeneously staining chromosomal regions contain amplified copies of an abundantly expressed cellular oncogene (c-myc) in malignant neuroendocrine cells from a human colon carcinoma’, *Proceedings of the National Academy of Sciences of the United States of America*, 80(6), pp. 1707–1711.

Alt, F.W. *et al.* (1978) ‘Selective multiplication of dihydrofolate reductase genes in methotrexate-resistant variants of cultured murine cells’, *The Journal of biological chemistry*, 253(5), pp. 1357–1370.

An, Z. *et al.* (2018) ‘Epidermal growth factor receptor and EGFRvIII in glioblastoma: signaling pathways and targeted therapies’, *Oncogene*, 37(12), pp. 1561–1575.

Balaban-Malenbaum, G. and Gilbert, F. (1980) ‘The proposed origin of double minutes from Homogeneously Staining Region (HSR)-marker chromosomes in human neuroblastoma hybrid cell lines’, *Cancer genetics and cytogenetics*, 2(4), pp. 339–348.

Bao, S. *et al.* (2006) ‘Glioma stem cells promote radioresistance by preferential activation of the DNA damage response’, *Nature*, 444(7120), pp. 756–760.

Barker, N., Bartfeld, S. and Clevers, H. (2010) ‘Tissue-resident adult stem cell populations of rapidly self-renewing organs’, *Cell stem cell*, 7(6), pp. 656–670.

Barker, P.E. *et al.* (1980) ‘Double minutes replicate once during S phase of the cell cycle’, *Experimental cell research*, 130(2), pp. 353–360.

Bartkova, J. *et al.* (2006) ‘Oncogene-induced senescence is part of the tumorigenesis barrier imposed by DNA damage checkpoints’, *Nature*, 444(7119), pp. 633–637.

Bartkova, J. *et al.* (2010) ‘Replication stress and oxidative damage contribute to aberrant constitutive activation of DNA damage signalling in human gliomas’, *Oncogene*, 29(36), pp. 5095–5102.

Beier, D. *et al.* (2007) ‘CD133(+) and CD133(-) glioblastoma-derived cancer stem cells show differential growth characteristics and molecular profiles’, *Cancer research*, 67(9), pp. 4010–4015.

Benabdallah, N.S. *et al.* (2019) ‘Decreased Enhancer-Promoter Proximity Accompanying Enhancer Activation’, *Molecular cell* [Preprint]. Available at: https://doi.org/10.1016/j.molcel.2019.07.038.

Benjamin, R.C. and Gill, D.M. (1980) ‘ADP-ribosylation in mammalian cell ghosts. Dependence of poly(ADP-ribose) synthesis on strand breakage in DNA’, *The Journal of biological chemistry*, 255(21), pp. 10493–10501.

Benner, S.E., Wahl, G.M. and Von Hoff, D.D. (1991) ‘Double minute chromosomes and homogeneously staining regions in tumors taken directly from patients versus in human tumor cell lines’, *Anti-cancer drugs*, 2(1), pp. 11–25.

Berg, S. *et al.* (2019) ‘ilastik: interactive machine learning for (bio)image analysis’, *Nature methods*, 16(12), pp. 1226–1232.

Biedler, J.L. and Spengler, B.A. (1976) ‘A novel chromosome abnormality in human neuroblastoma and antifolate-resistant Chinese hamster cell lives in culture’, *Journal of the National Cancer Institute*, 57(3), pp. 683–695.

Bigner, S.H. *et al.* (1987) ‘Relationship between gene amplification and chromosomal deviations in malignant human gliomas’, *Cancer genetics and cytogenetics*, 29(1), pp. 165–170.

Bigner, S.H. *et al.* (1988) ‘Specific chromosomal abnormalities in malignant human gliomas’, *Cancer research*, 48(2), pp. 405–411.

Bishop, S.C. and Abel, E.A. (1985) ‘DNA repair elicited by UVB during PUVA therapy for psoriasis’, *Archives for dermatological research. Archiv fur dermatologische Forschung*, 278(1), pp. 25–30.

Blajeski, A.L. *et al.* (2002) ‘G(1) and G(2) cell-cycle arrest following microtubule depolymerization in human breast cancer cells’, *The Journal of clinical investigation*, 110(1), pp. 91–99.

Boland, C.R. and Goel, A. (2010) ‘Microsatellite instability in colorectal cancer’, *Gastroenterology*, 138(6), pp. 2073–2087.e3.

Boulay, G. *et al.* (2017) ‘Cancer-Specific Retargeting of BAF Complexes by a Prion-like Domain’, *Cell*, 171(1), pp. 163–178.e19.

Boveri, T. (1909) ‘Die Blastomerenkerne von Ascaris megalocephala und dieTheorie der Chromosomenindividualitit. Arch. Zellforschung. 3: 181-268’, *Boveri1813Arch. Zellforschung* [Preprint].

Boyle, S. *et al.* (2001) ‘The spatial organization of human chromosomes within the nuclei of normal and emerin-mutant cells’, *Human molecular genetics*, 10(3), pp. 211–219.

Boyle, S. *et al.* (2020) ‘A central role for canonical PRC1 in shaping the 3D nuclear landscape’, *Genes & development*, 34(13-14), pp. 931–949.

Bradley, M.O. and Kohn, K.W. (1979) ‘X-ray induced DNA double strand break production and repair in mammalian cells as measured by neutral filter elution’, *Nucleic acids research*, 7(3), pp. 793–804.

Brennan, C.W. *et al.* (2013) ‘The somatic genomic landscape of glioblastoma’, *Cell*, 155(2), pp. 462–477.

Bressan, R.B. *et al.* (2017) ‘Efficient CRISPR/Cas9-assisted gene targeting enables rapid and precise genetic manipulation of mammalian neural stem cells’, *Development* , 144(4), pp. 635–648.

Bryant, H.E. *et al.* (2005) ‘Specific killing of BRCA2-deficient tumours with inhibitors of poly(ADP-ribose) polymerase’, *Nature*, 434(7035), pp. 913–917.

Bryant, H.E. *et al.* (2009) ‘PARP is activated at stalled forks to mediate Mre11-dependent replication restart and recombination’, *The EMBO journal*, 28(17), pp. 2601–2615.

Buis, J. *et al.* (2008) ‘Mre11 nuclease activity has essential roles in DNA repair and genomic stability distinct from ATM activation’, *Cell*, 135(1), pp. 85–96.

Cahuzac, M. *et al.* (2022) ‘Development of Olaparib-Resistance Prostate Cancer Cell Lines to Identify Mechanisms Associated with Acquired Resistance’, *Cancers*, 14(16). Available at: https://doi.org/10.3390/cancers14163877.

Cai, M. *et al.* (2019) ‘Inhibiting homologous recombination decreases extrachromosomal amplification but has no effect on intrachromosomal amplification in methotrexate-resistant colon cancer cells’, *International journal of cancer. Journal international du cancer*, 144(5), pp. 1037–1048.

Caldecott, K.W. (2008) ‘Single-strand break repair and genetic disease’, *Nature reviews. Genetics*, 9(8), pp. 619–631.

Cancer Genome Atlas Research Network (2008) ‘Comprehensive genomic characterization defines human glioblastoma genes and core pathways’, *Nature*, 455(7216), pp. 1061–1068.

Cancer Genome Atlas Research Network (2011) ‘Integrated genomic analyses of ovarian carcinoma’, *Nature*, 474(7353), pp. 609–615.

Cannan, W.J. *et al.* (2014) ‘Nucleosomes suppress the formation of double-strand DNA breaks during attempted base excision repair of clustered oxidative damages’, *The Journal of biological chemistry*, 289(29), pp. 19881–19893.

Carroll, S.M. *et al.* (1988) ‘Double minute chromosomes can be produced from precursors derived from a chromosomal deletion’, *Molecular and cellular biology*, 8(4), pp. 1525–1533.

Carruthers, R.D. *et al.* (2018) ‘Replication Stress Drives Constitutive Activation of the DNA Damage Response and Radioresistance in Glioblastoma Stem-like Cells’, *Cancer research*, 78(17), pp. 5060–5071.

Carvalho, C. *et al.* (2001) ‘Chromosomal G-dark bands determine the spatial organization of centromeric heterochromatin in the nucleus’, *Molecular biology of the cell*, 12(11), pp. 3563–3572.

Chamorro González, R. *et al.* (2023) ‘Parallel sequencing of extrachromosomal circular DNAs and transcriptomes in single cancer cells’, *Nature genetics*, 55(5), pp. 880–890. Available at: https://doi.org/10.1038/s41588-023-01386-y.

Chen, J. *et al.* (2012) ‘A restricted cell population propagates glioblastoma growth after chemotherapy’, *Nature*, 488(7412), pp. 522–526.

Chiocca, E.A. *et al.* (2022) ‘Combined immunotherapy with controlled interleukin-12 gene therapy and immune checkpoint blockade in recurrent glioblastoma: An open-label, multi-institutional phase I trial’, *Neuro-oncology*, 24(6), pp. 951–963.

Chong, S. *et al.* (2018) ‘Imaging dynamic and selective low-complexity domain interactions that control gene transcription’, *Science*, 361(6400). Available at: https://doi.org/10.1126/science.aar2555.

Chowdhry, S. *et al.* (2022) ‘Program Planner’, in *Replication stress and the inability to repair damaged DNA, the potential ‘Achilles’ heel’ of ecDNA+ tumor cells*. *Proceedings of the 113th Annual Meeting of the American Association for Cancer Research*, AACR. Available at: https://www.abstractsonline.com/pp8/#!/10517/presentation/21207 (Accessed: 20 January 2023).

Cho, W.-K. *et al.* (2018) ‘Mediator and RNA polymerase II clusters associate in transcription-dependent condensates’, *Science*, 361(6400), pp. 412–415.

Chuntova, P. *et al.* (2021) ‘Unique challenges for glioblastoma immunotherapy-discussions across neuro-oncology and non-neuro-oncology experts in cancer immunology. Meeting Report from the 2019 SNO Immuno-Oncology Think Tank’, *Neuro-oncology*, 23(3), pp. 356–375.

Ciccia, A. and Elledge, S.J. (2010) ‘The DNA damage response: making it safe to play with knives’, *Molecular cell*, 40(2), pp. 179–204.

Clow, P.A. *et al.* (2022) ‘CRISPR-mediated multiplexed live cell imaging of nonrepetitive genomic loci with one guide RNA per locus’, *Nature communications*, 13(1), p. 1871.

Conti, L. *et al.* (2005) ‘Niche-independent symmetrical self-renewal of a mammalian tissue stem cell’, *PLoS biology*, 3(9), p. e283.

Coquelle, A. *et al.* (2002) ‘Induction of multiple double-strand breaks within an hsr by meganucleaseI-SceI expression or fragile site activation leads to formation of double minutes and other chromosomal rearrangements’, *Oncogene*, 21(50), pp. 7671–7679.

Cortés-Ciriano, I. *et al.* (2020) ‘Comprehensive analysis of chromothripsis in 2,658 human cancers using whole-genome sequencing’, *Nature genetics*, 52(3), pp. 331–341.

Cowell, J.K. and Rupniak, H.T. (1983) ‘Chromosome analysis of human neuroblastoma cell line TR14 showing double minutes and an aberration involving chromosome 1’, *Cancer genetics and cytogenetics*, 9(3), pp. 273–280.

Cox, D., Yuncken, C. and Spriggs, A.I. (1965) ‘MINUTE CHROMATIN BODIES IN MALIGNANT TUMOURS OF CHILDHOOD’, *The Lancet*, 1(7402), pp. 55–58.

Cremer, T. and Cremer, M. (2010) ‘Chromosome territories’, *Cold Spring Harbor perspectives in biology*, 2(3), p. a003889.

Croft, J.A. *et al.* (1999) ‘Differences in the localization and morphology of chromosomes in the human nucleus’, *The Journal of cell biology*, 145(6), pp. 1119–1131.

CRUK (2021) *Brain, other CNS and intracranial tumours incidence statistics*, *Cancer Research UK*. Available at: https://www.cancerresearchuk.org/health-professional/cancer-statistics/statistics-by-cancer-type/brain-other-cns-and-intracranial-tumours/incidence (Accessed: 14 March 2023).

D’Amours, D. *et al.* (1999) ‘Poly(ADP-ribosyl)ation reactions in the regulation of nuclear functions’, *Biochemical Journal*, 342 (Pt 2)(Pt 2), pp. 249–268.

Danecek, P. *et al.* (2021) ‘Twelve years of SAMtools and BCFtools’, *GigaScience*, 10(2). Available at: https://doi.org/10.1093/gigascience/giab008.

deCarvalho, A.C. *et al.* (2018) ‘Discordant inheritance of chromosomal and extrachromosomal DNA elements contributes to dynamic disease evolution in glioblastoma’, *Nature genetics*, 50(5), pp. 708–717.

Della-Maria, J. *et al.* (2011) ‘Human Mre11/Human Rad50/Nbs1 and DNA Ligase IIIα/XRCC1 Protein Complexes Act Together in an Alternative Nonhomologous End Joining Pathway*’, *The Journal of biological chemistry*, 286(39), pp. 33845–33853.

Desai, A., Webb, B. and Gerson, S.L. (2014) ‘CD133+ cells contribute to radioresistance via altered regulation of DNA repair genes in human lung cancer cells’, *Radiotherapy* *and oncology: journal of the European Society for Therapeutic Radiology and Oncology*, 110(3), pp. 538–545.

Deshpande, V. *et al.* (2019) ‘Exploring the landscape of focal amplifications in cancer using AmpliconArchitect’, *Nature communications*, 10(1), p. 392.

Dewari, P.S. *et al.* (2018) ‘An efficient and scalable pipeline for epitope tagging in mammalian stem cells using Cas9 ribonucleoprotein’, *eLife*, 7. Available at: https://doi.org/10.7554/eLife.35069.

Dixon, J.R. *et al.* (2012) ‘Topological domains in mammalian genomes identified by analysis of chromatin interactions’, *Nature*, 485(7398), pp. 376–380.

Dixon, J.R., Gorkin, D.U. and Ren, B. (2016) ‘Chromatin Domains: The Unit of Chromosome Organization’, *Molecular cell*, 62(5), pp. 668–680.

Dobin, A. *et al.* (2013) ‘STAR: ultrafast universal RNA-seq aligner’, *Bioinformatics* , 29(1), pp. 15–21.

Downs, J.A. and Jackson, S.P. (2004) ‘A means to a DNA end: the many roles of Ku’, *Nature reviews. Molecular cell biology*, 5(5), pp. 367–378.

Dungey, F.A., Löser, D.A. and Chalmers, A.J. (2008) ‘Replication-dependent radiosensitization of human glioma cells by inhibition of poly(ADP-Ribose) polymerase: mechanisms and therapeutic potential’, *International journal of radiation oncology, biology, physics*, 72(4), pp. 1188–1197.

Durkacz, B.W. *et al.* (1980) ‘(ADP-ribose)n participates in DNA excision repair’, *Nature*, 283(5747), pp. 593–596.

Dutta, A. *et al.* (2017) ‘Microhomology-mediated end joining is activated in irradiated human cells due to phosphorylation-dependent formation of the XRCC1 repair complex’, *Nucleic acids research*, 45(5), pp. 2585–2599.

Erasimus, H. *et al.* (2016) ‘DNA repair mechanisms and their clinical impact in glioblastoma’, *Mutation Research-Reviews in Mutation Research*, 769, pp. 19–35.

Esnault, C. *et al.* (2008) ‘Mediator-dependent recruitment of TFIIH modules in preinitiation complex’, *Molecular cell*, 31(3), pp. 337–346.

Eyler, C.E. and Rich, J.N. (2008) ‘Survival of the fittest: cancer stem cells in therapeutic resistance and angiogenesis’, *Journal of clinical oncology: official journal of the American Society of Clinical Oncology*, 26(17), pp. 2839–2845.

Fan, Y. *et al.* (2011) ‘Frequency of double minute chromosomes and combined cytogenetic abnormalities and their characteristics’, *Journal of applied genetics*, 52(1), pp. 53–59.

Farmer, H. *et al.* (2005) ‘Targeting the DNA repair defect in BRCA mutant cells as a therapeutic strategy’, *Nature*, 434(7035), pp. 917–921.

Feng, W. *et al.* (2022) ‘Targeted removal of mitochondrial DNA from mouse and human extrachromosomal circular DNA with CRISPR-Cas9’, *Computational and structural biotechnology journal*, 20, pp. 3059–3067.

Ferro, A.M. and Oppenheimer, N.J. (1978) ‘Structure of a poly (adenosine diphosphoribose) monomer: 2’-(5"-hosphoribosyl)-5'-adenosine monophosphate’, *Proceedings of the National Academy of Sciences of the United States of America*, 75(2), pp. 809–813.

Fong, P.C. *et al.* (2009) ‘Inhibition of poly(ADP-ribose) polymerase in tumors from BRCA mutation carriers’, *The New England journal of medicine*, 361(2), pp. 123–134.

Fudenberg, G. *et al.* (2016) ‘Formation of Chromosomal Domains by Loop Extrusion’, *Cell reports*, 15(9), pp. 2038–2049.

Fulco, C.P. *et al.* (2019) ‘Activity-by-contact model of enhancer-promoter regulation from thousands of CRISPR perturbations’, *Nature genetics*, 51(12), pp. 1664–1669.

Fulton, B. *et al.* (2018) ‘PARADIGM-2: Two parallel phase I studies of olaparib and radiotherapy or olaparib and radiotherapy plus temozolomide in patients with newly diagnosed glioblastoma, with treatment stratified by MGMT status’, *Clinical and translational radiation oncology*, 8, pp. 12–16.

Furnari, F.B. *et al.* (2015) ‘Heterogeneity of epidermal growth factor receptor signalling networks in glioblastoma’, *Nature reviews. Cancer*, 15(5), pp. 302–310.

Fu, Y. *et al.* (2013) ‘High-frequency off-target mutagenesis induced by CRISPR-Cas nucleases in human cells’, *Nature biotechnology*, 31(9), pp. 822–826.

Galli, R. *et al.* (2004) ‘Isolation and characterization of tumorigenic, stem-like neural precursors from human glioblastoma’, *Cancer research*, 64(19), pp. 7011–7021.

Gangemi, R.M.R. *et al.* (2009) ‘SOX2 silencing in glioblastoma tumor-initiating cells causes stop of proliferation and loss of tumorigenicity’, *Stem cells*, 27(1), pp. 40–48.

Gangoso, E. *et al.* (2021) ‘Glioblastomas acquire myeloid-affiliated transcriptional programs via epigenetic immunoediting to elicit immune evasion’, *Cell*, 184(9), pp. 2454–2470.e26.

Garsed, D.W. *et al.* (2014) ‘The architecture and evolution of cancer neochromosomes’, *Cancer cell*, 26(5), pp. 653–667.

Gilard, V. *et al.* (2021) ‘Diagnosis and Management of Glioblastoma: A Comprehensive Perspective’, *Journal of personalized medicine*, 11(4). Available at: https://doi.org/10.3390/jpm11040258.

Gilbert, M.R. *et al.* (2014) ‘A randomized trial of bevacizumab for newly diagnosed glioblastoma’, *The New England journal of medicine*, 370(8), pp. 699–708.

Gisselsson, D. *et al.* (2000) ‘Chromosomal breakage-fusion-bridge events cause genetic intratumor heterogeneity’, *Proceedings of the National Academy of Sciences of the United States of America*, 97(10), pp. 5357–5362.

Goodhead, D.T. (1994) ‘Initial events in the cellular effects of ionizing radiations: clustered damage in DNA’, *International journal of radiation biology*, 65(1), pp. 7–17.

Guérin, T.M. and Marcand, S. (2022) ‘Breakage in breakage-fusion-bridge cycle: an 80-year-old mystery’, *Trends in genetics: TIG*, 38(7), pp. 641–645.

Haaf, T. *et al.* (1995) ‘Nuclear foci of mammalian Rad51 recombination protein in somatic cells after DNA damage and its localization in synaptonemal complexes’, *Proceedings of the National Academy of Sciences of the United States of America*, 92(6), pp. 2298–2302.

Haber, D.A. *et al.* (1981) ‘Properties of an altered dihydrofolate reductase encoded by amplified genes in cultured mouse fibroblasts’, *The Journal of biological chemistry*, 256(18), pp. 9501–9510.

Haber, D.A. and Schimke, R.T. (1981) ‘Unstable amplification of an altered dihydrofolate reductase gene associated with double-minute chromosomes’, *Cell*, 26(3 Pt 1), pp. 355–362.

Halappanavar, S.S. and Shah, G.M. (2004) ‘Defective control of mitotic and post-mitotic checkpoints in poly(ADP-ribose) polymerase-1(-/-) fibroblasts after mitotic spindle disruption’, *Cell cycle* , 3(3), pp. 335–342.

Halazonetis, T.D., Gorgoulis, V.G. and Bartek, J. (2008) ‘An oncogene-induced DNA damage model for cancer development’, *Science*, 319(5868), pp. 1352–1355.

Hamkalo, B.A. *et al.* (1985) ‘Ultrastructural features of minute chromosomes in a methotrexate-resistant mouse 3T3 cell line’, *Proceedings of the National Academy of Sciences of the United States of America*, 82(4), pp. 1126–1130.

Hanahan, D. and Weinberg, R.A. (2011) ‘Hallmarks of cancer: the next generation’, *Cell*, 144(5), pp. 646–674.

Hanna, C. *et al.* (2020) ‘Pharmacokinetics, safety, and tolerability of olaparib and temozolomide for recurrent glioblastoma: results of the phase I OPARATIC trial’, *Neuro-oncology*, 22(12), pp. 1840–1850.

Hansen, J.C., Maeshima, K. and Hendzel, M.J. (2021) ‘The solid and liquid states of chromatin’, *Epigenetics & chromatin*, 14(1), pp. 1–33.

Hatch, E.M. *et al.* (2013) ‘Catastrophic nuclear envelope collapse in cancer cell micronuclei’, *Cell*, 154(1), pp. 47–60.

Hayashi, K. *et al.* (1983) ‘Size and shape of poly(ADP-ribose): examination by gel filtration, gel electrophoresis and electron microscopy’, *Biochemical and biophysical research communications*, 112(1), pp. 102–107.

Hegi, M.E. *et al.* (2005) ‘MGMT gene silencing and benefit from temozolomide in glioblastoma’, *The New England journal of medicine*, 352(10), pp. 997–1003.

Hegi, M.E. *et al.* (2008) ‘Correlation of O6-methylguanine methyltransferase (MGMT) promoter methylation with clinical outcomes in glioblastoma and clinical strategies to modulate MGMT activity’, *Journal of clinical oncology: official journal of the American Society of Clinical Oncology*, 26(25), pp. 4189–4199.

Heinz, S. *et al.* (2010) ‘Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities’, *Molecular cell*, 38(4), pp. 576–589.

Helmsauer, K. *et al.* (2020) ‘Enhancer hijacking determines extrachromosomal circular MYCN amplicon architecture in neuroblastoma’, *Nature communications*, 11(1), p. 5823.

van der Hout, A.H. *et al.* (1989) ‘Localization of amplified c-myc and n-myc in small cell lung cancer cell lines’, *Cancer genetics and cytogenetics*, 38(1), pp. 1–8.

Hsieh, T.-H.S. *et al.* (2015) ‘Mapping Nucleosome Resolution Chromosome Folding in Yeast by Micro-C’, *Cell*, 162(1), pp. 108–119.

Hung, K.L. *et al.* (2021) ‘ecDNA hubs drive cooperative intermolecular oncogene expression’, *Nature*, pp. 1–6.

Hung, K.L. *et al.* (2022) ‘Targeted profiling of human extrachromosomal DNA by CRISPR-CATCH’, *Nature genetics* [Preprint]. Available at: https://doi.org/10.1038/s41588-022-01190-0.

Ignatova, T.N. *et al.* (2002) ‘Human cortical glial tumors contain neural stem-like cells expressing astroglial and neuronal markers in vitro’, *Glia*, 39(3), pp. 193–206.

Imada, T. *et al.* (2021) ‘RNA polymerase II condensate formation and association with Cajal and histone locus bodies in living human cells’, *Genes to cells: devoted to molecular & cellular mechanisms*, 26(5), pp. 298–312.

Inda, M.-M. *et al.* (2010) ‘Tumor heterogeneity is an active process maintained by a mutant EGFR-induced cytokine circuit in glioblastoma’, *Genes & development*, 24(16), pp. 1731–1745.

Itoh, N. and Shimizu, N. (1998) ‘DNA replication-dependent intranuclear relocation of double minute chromatin’, *Journal of cell science*, 111 (Pt 22), pp. 3275–3285.

Jackson, S.P. and Bartek, J. (2009) ‘The DNA-damage response in human biology and disease’, *Nature*, 461(7267), pp. 1071–1078.

Jain, R.K. *et al.* (2007) ‘Angiogenesis in brain tumours’, *Nature reviews. Neuroscience*, 8(8), pp. 610–622.

Kaighn, M.E. *et al.* (1979) ‘Establishment and characterization of a human prostatic carcinoma cell line (PC-3)’, *Investigative urology*, 17(1), pp. 16–23.

Kalhor, R. *et al.* (2011) ‘Genome architectures revealed by tethered chromosome conformation capture and population-based modeling’, *Nature biotechnology*, 30(1), pp. 90–98.

Kane, L. *et al.* (2022) ‘Cohesin is required for long-range enhancer action at the Shh locus’, *Nature structural & molecular biology*, 29(9), pp. 891–897.

Kathe, S.D., Shen, G.-P. and Wallace, S.S. (2004) ‘Single-stranded breaks in DNA but not oxidative DNA base damages block transcriptional elongation by RNA polymerase II in HeLa cell nuclear extracts’, *The Journal of biological chemistry*, 279(18), pp. 18511–18520.

Kaufman, R.J., Brown, P.C. and Schimke, R.T. (1979) ‘Amplified dihydrofolate reductase genes in unstably methotrexate-resistant cells are associated with double minute chromosomes’, *Proceedings of the National Academy of Sciences of the United States of America*, 76(11), pp. 5669–5673.

Kent, W.J. *et al.* (2002) ‘The human genome browser at UCSC’, *Genome research*, 12(6), pp. 996–1006.

Kim, H. *et al.* (2020) ‘Extrachromosomal DNA is associated with oncogene amplification and poor outcome across multiple cancers’, *Nature genetics* [Preprint]. Available at: https://doi.org/10.1038/s41588-020-0678-2.

Kim, M.Y. *et al.* (2004) ‘NAD+-dependent modulation of chromatin structure and transcription by nucleosome binding properties of PARP-1’, *Cell*, 119(6), pp. 803–814.

Kim, S. *et al.* (2018) ‘Strelka2: fast and accurate calling of germline and somatic variants’, *Nature methods*, 15(8), pp. 591–594.

Kim, Y.J. *et al.* (1994) ‘A multiprotein mediator of transcriptional activation and its interaction with the C-terminal repeat domain of RNA polymerase II’, *Cell*, 77(4), pp. 599–608.

Koga, T. *et al.* (2018) ‘Mapping of genomic EGFRvIII deletions in glioblastoma: insight into rearrangement mechanisms and biomarker development’, *Neuro-oncology*, 20(10), pp. 1310–1320.

Koga, T. *et al.* (2020) ‘Longitudinal assessment of tumor development using cancer avatars derived from genetically engineered pluripotent stem cells’, *Nature communications*, 11(1), p. 550.

Kohl, N.E. *et al.* (1983) ‘Transposition and amplification of oncogene-related sequences in human neuroblastomas’, *Cell*, 35(2 Pt 1), pp. 359–367.

Koster, D.A. *et al.* (2007) ‘Antitumour drugs impede DNA uncoiling by topoisomerase I’, *Nature*, 448(7150), pp. 213–217.

Kraus, W.L. and Lis, J.T. (2003) ‘PARP goes transcription’, *Cell*, 113(6), pp. 677–683.

Kreisl, T.N. *et al.* (2009) ‘Phase II trial of single-agent bevacizumab followed by bevacizumab plus irinotecan at tumor progression in recurrent glioblastoma’, *Journal of clinical oncology: official journal of the American Society of Clinical Oncology*, 27(5), pp. 740–745.

Kreso, A. and Dick, J.E. (2014) ‘Evolution of the cancer stem cell model’, *Cell stem cell*, 14(3), pp. 275–291.

Kuehner, J.N., Pearson, E.L. and Moore, C. (2011) ‘Unravelling the means to an end: RNA polymerase II transcription termination’, *Nature reviews. Molecular cell biology*, 12(5), pp. 283–294.

Kuzminov, A. (2001) ‘Single-strand interruptions in replicating chromosomes cause double-strand breaks’, *Proceedings of the National Academy of Sciences of the United States of America*, 98(15), pp. 8241–8246.

L’Abbate, A. *et al.* (2014) ‘Genomic organization and evolution of double minutes/homogeneously staining regions with MYC amplification in human cancer’, *Nucleic acids research*, 42(14), pp. 9131–9145.

Laks, D.R. *et al.* (2016) ‘Large-scale assessment of the gliomasphere model system’, *Neuro-oncology*, 18(10), pp. 1367–1378.

Lam, I. and Keeney, S. (2014) ‘Mechanism and regulation of meiotic recombination initiation’, *Cold Spring Harbor perspectives in biology*, 7(1), p. a016634.

Lange, J.T. *et al.* (2022) ‘The evolutionary dynamics of extrachromosomal DNA in human cancers’, *Nature genetics* [Preprint]. Available at: https://doi.org/10.1038/s41588-022-01177-x.

Langelier, M.-F. *et al.* (2018) ‘NAD+ analog reveals PARP-1 substrate-blocking mechanism and allosteric communication from catalytic center to DNA-binding domains’, *Nature communications*, 9(1), p. 844.

Lang, F.F. *et al.* (2018) ‘Phase I Study of DNX-2401 (Delta-24-RGD) Oncolytic Adenovirus: Replication and Immunotherapeutic Effects in Recurrent Malignant Glioma’, *Journal of clinical oncology: official journal of the American Society of Clinical Oncology*, 36(14), pp. 1419–1427.

Lathia, J.D. *et al.* (2015) ‘Cancer stem cells in glioblastoma’, *Genes & development*, 29(12), pp. 1203–1217.

Lee, A. *et al.* (2020) ‘Anti-epidermal growth factor receptor therapy for glioblastoma in adults’, *Cochrane database of systematic reviews* , 5(5), p. CD013238.

Lee, J. *et al.* (2006) ‘Tumor stem cells derived from glioblastomas cultured in bFGF and EGF more closely mirror the phenotype and genotype of primary tumors than do serum-cultured cell lines’, *Cancer cell*, 9(5), pp. 391–403.

Lee, M.-H. *et al.* (2012) ‘Poly(ADP-ribosyl)ation of p53 induces gene-specific transcriptional repression of MTA1’, *Oncogene*, 31(49), pp. 5099–5107.

Li, A. *et al.* (2021) ‘Using the dCas9-KRAB system to repress gene expression in hiPSC-derived NGN2 neurons’, *STAR protocols*, 2(2), p. 100580.

Lieberman-Aiden, E. *et al.* (2009) ‘Comprehensive mapping of long-range interactions reveals folding principles of the human genome’, *Science*, 326(5950), pp. 289–293.

Lieber, M.R. (2010a) ‘NHEJ and its backup pathways in chromosomal translocations’, *Nature structural & molecular biology*, pp. 393–395.

Lieber, M.R. (2010b) ‘The mechanism of double-strand DNA break repair by the nonhomologous DNA end-joining pathway’, *Annual review of biochemistry*, 79, pp. 181–211.

Li, H. *et al.* (2009) ‘The Sequence Alignment/Map format and SAMtools’, *Bioinformatics* , 25(16), pp. 2078–2079.

Li, H. and Durbin, R. (2009) ‘Fast and accurate short read alignment with Burrows-Wheeler transform’, *Bioinformatics* , 25(14), pp. 1754–1760.

Lin, D. *et al.* (2021) 'Trends in Intracranial Glioma Incidence and Mortality in the United States, 1975-2018', *Frontiers in oncology*, 11, p. 748061.

Lindahl, T. and Barnes, D.E. (2000) 'Repair of endogenous DNA damage', *Cold Spring Harbor symposia on quantitative biology*, 65, pp. 127–133.

Ljungman, M. and Lane, D.P. (2004) 'Transcription - guarding the genome by sensing DNA damage', *Nature reviews. Cancer*, 4(9), pp. 727–737.

Lord, C.J. and Ashworth, A. (2012) 'The DNA damage response and cancer therapy', *Nature*, 481(7381), pp. 287–294.

Lord, C.J. and Ashworth, A. (2017) 'PARP inhibitors: Synthetic lethality in the clinic', *Science*, 355(6330), pp. 1152–1158.

Louis, D.N. *et al.* (2016) 'The 2016 World Health Organization Classification of Tumors of the Central Nervous System: a summary', *Acta neuropathologica*, 131(6), pp. 803–820.

Louis, D.N. *et al.* (2021) 'The 2021 WHO Classification of Tumors of the Central Nervous System: a summary', *Neuro-oncology*, 23(8), pp. 1231–1251.

Lovén, J. *et al.* (2013) 'Selective inhibition of tumor oncogenes by disruption of super-enhancers', *Cell*, 153(2), pp. 320–334.

Lozinski, M. *et al.* (2021) 'DNA damage repair in glioblastoma: current perspectives on its role in tumour progression, treatment resistance and PIKKing potential therapeutic targets', *Cellular oncology* , 44(5), pp. 961–981.

Lubs, H.A., Jr and Salmon, J.H. (1965) 'THE CHROMOSOMAL COMPLEMENT OF HUMAN SOLID TUMORS. II. KARYOTYPES OF GLIAL TUMORS', *Journal of neurosurgery*, 22, pp. 160–168.

Ludwig, A. *et al.* (1988) 'Immunoquantitation and size determination of intrinsic poly(ADP-ribose) polymerase from acid precipitates. An analysis of the in vivo status in mammalian species and in lower eukaryotes', *The Journal of biological chemistry*, 263(15), pp. 6993–6999.

Luebeck, J. *et al.* (2020) 'AmpliconReconstructor integrates NGS and optical mapping to resolve the complex structures of focal amplifications', *Nature communications*, 11(1), p. 4374.

Luebeck, J. *et al.* (2022) ‘Extrachromosomal DNA in the cancerous transformation of Barrett’s esophagus’, *bioRxiv*. Available at: https://doi.org/10.1101/2022.07.25.501144.

Lu, G. *et al.* (2016) ‘Ligase I and ligase III mediate the DNA double-strand break ligation in alternative end-joining’, *Proceedings of the National Academy of Sciences of the United States of America*, 113(5), pp. 1256–1260.

Lundberg, G. *et al.* (2008) ‘Binomial mitotic segregation of MYCN-carrying double minutes in neuroblastoma illustrates the role of randomness in oncogene amplification’, *PloS one*, 3(8), p. e3099.

Ly, P. and Cleveland, D.W. (2017) ‘Rebuilding Chromosomes After Catastrophe: Emerging Mechanisms of Chromothripsis’, *Trends in cell biology*, 27(12), pp. 917–930.

Maciejowski, J. *et al.* (2015) ‘Chromothripsis and Kataegis Induced by Telomere Crisis’, *Cell*, 163(7), pp. 1641–1654.

MacLeod, R.A.F., Kaufmann, M. and Drexler, H.G. (2011) ‘Cytogenetic Analysis of Cancer Cell Lines’, in I.A. Cree (ed.) *Cancer Cell Culture: Methods and Protocols*. Totowa, NJ: Humana Press, pp. 57–78.

Mahy, N.L., Perry, P.E. and Bickmore, W.A. (2002) ‘Gene density and transcription influence the localization of chromatin outside of chromosome territories detectable by FISH’, *The Journal of cell biology*, 159(5), pp. 753–763.

Malhotra, A. *et al.* (2013) ‘Breakpoint profiling of 64 cancer genomes reveals numerous complex rearrangements spawned by homology-independent mechanisms’, *Genome research*, 23(5), pp. 762–776.

Marquart, J., Chen, E.Y. and Prasad, V. (2018) ‘Estimation of the Percentage of US Patients With Cancer Who Benefit From Genome-Driven Oncology’, *JAMA oncology*, 4(8), pp. 1093–1098.

Ma, W. *et al.* (2012) ‘Differential effects of poly(ADP-ribose) polymerase inhibition on DNA break repair in human cells are revealed with Epstein-Barr virus’, *Proceedings of the National Academy of Sciences of the United States of America*, 109(17), pp. 6590–6595.

McClintock, B. (1938) ‘The Production of Homozygous Deficient Tissues with Mutant Characteristics by Means of the Aberrant Mitotic Behavior of Ring-Shaped Chromosomes’, *Genetics*, 23(4), pp. 315–376.

McClintock, B. (1941) ‘The Stability of Broken Ends of Chromosomes in Zea Mays’, *Genetics*, 26(2), pp. 234–282.

McGranahan, N. and Swanton, C. (2017) ‘Clonal Heterogeneity and Tumor Evolution: Past, Present, and the Future’, *Cell*, 168(4), pp. 613–628.

McSwiggen, D.T. *et al.* (2019) ‘Evaluating phase separation in live cells: diagnosis, caveats, and functional consequences’, *Genes & development* [Preprint]. Available at: https://doi.org/10.1101/gad.331520.119.

Mehta, A. and Haber, J.E. (2014) ‘Sources of DNA double-strand breaks and models of recombinational DNA repair’, *Cold Spring Harbor perspectives in biology*, 6(9), p. a016428.

Meng, X. *et al.* (2015) ‘Novel role for non-homologous end joining in the formation of double minutes in methotrexate-resistant colon cancer cells’, *Journal of medical genetics*, 52(2), pp. 135–144.

Meschichi, A. *et al.* (2021) ‘ANCHOR: A Technical Approach to Monitor Single-Copy Locus Localization in Planta’, *Frontiers in plant science*, 12, p. 677849.

Meyer, M. *et al.* (2015) ‘Single cell-derived clonal analysis of human glioblastoma links functional and genomic heterogeneity’, *Proceedings of the National Academy of Sciences of the United States of America*, 112(3), pp. 851–856.

Mir, M. *et al.* (2018) ‘Dynamic multifactor hubs interact transiently with sites of active transcription in Drosophila embryos’, *eLife*, 7. Available at: https://doi.org/10.7554/eLife.40497.

Mir, M. *et al.* (2019) ‘Chromatin topology, condensates and gene regulation: shifting paradigms or just a phase?’, *Development* , 146(19). Available at: https://doi.org/10.1242/dev.182766.

Miwa, M. *et al.* (1974) ‘Purification and properties of glycohydrolase from calf thymus splitting ribose-ribose linkages of poly(adenosine diphosphate ribose)’, *The Journal of biological chemistry*, 249(11), pp. 3475–3482.

Møller, H.D. *et al.* (2016) ‘Genome-wide Purification of Extrachromosomal Circular DNA from Eukaryotic Cells’, *Journal of visualized experiments: JoVE*, (110), p. e54239 |.

Morton, A.R. *et al.* (2019) ‘Functional Enhancers Shape Extrachromosomal Oncogene Amplifications’, *Cell*, 179(6), pp. 1330–1341.e13.

Moynahan, M.E. and Jasin, M. (2010) ‘Mitotic homologous recombination maintains genomic stability and suppresses tumorigenesis’, *Nature reviews. Molecular cell biology*, 11(3), pp. 196–207.

Munnur, D. and Ahel, I. (2017) ‘Reversible mono-ADP-ribosylation of DNA breaks’, *The FEBS journal*, 284(23), pp. 4002–4016.

Nathanson, D.A. *et al.* (2014) ‘Targeted therapy resistance mediated by dynamic regulation of extrachromosomal mutant EGFR DNA’, *Science*, 343(6166), pp. 72–76.

Neftel, C. *et al.* (2019) ‘An Integrative Model of Cellular States, Plasticity, and Genetics for Glioblastoma’, *Cell*, 178(4), pp. 835–849.e21.

Nones, K. *et al.* (2014) ‘Genomic catastrophes frequently arise in esophageal adenocarcinoma and drive tumorigenesis’, *Nature communications*, 5, p. 5224.

Ogawa, K. *et al.* (2013) ‘Radiotherapy targeting cancer stem cells: current views and future perspectives’, *Anticancer research*, 33(3), pp. 747–754.

Omuro, A. *et al.* (2023) ‘Radiotherapy combined with nivolumab or temozolomide for newly diagnosed glioblastoma with unmethylated MGMT promoter: An international randomized phase III trial’, *Neuro-oncology*, 25(1), pp. 123–134.

Orkin, S.H. and Hochedlinger, K. (2011) ‘Chromatin connections to pluripotency and cellular reprogramming’, *Cell*, 145(6), pp. 835–850.

Ortega, E. *et al.* (2018) ‘Transcription factor dimerization activates the p300 acetyltransferase’, *Nature*, 562(7728), pp. 538–544.

Ostrom, Q.T. *et al.* (2019) ‘CBTRUS Statistical Report: Primary Brain and Other Central Nervous System Tumors Diagnosed in the United States in 2012-2016’, *Neuro-oncology*, 21(Suppl 5), pp. v1–v100.

Palacio, M. and Taatjes, D.J. (2022) ‘Merging Established Mechanisms with New Insights: Condensates, Hubs, and the Regulation of RNA Polymerase II Transcription’, *Journal of molecular biology*, 434(1), p. 167216.

Palazzo, L. *et al.* (2018) ‘Serine is the major residue for ADP-ribosylation upon DNA damage’, *eLife*, 7. Available at: https://doi.org/10.7554/eLife.34334.

Palazzo, L., Suskiewicz, M.J. and Ahel, I. (2021) ‘Serine ADP-ribosylation in DNA-damage response regulation’, *Current opinion in genetics & development*, 71, pp. 106–113.

Pandey, N. and Black, B.E. (2021) ‘Rapid Detection and Signaling of DNA Damage by PARP-1’, *Trends in biochemical sciences*, 46(9), pp. 744–757.

Park, J.G. *et al.* (1990) ‘Characteristics of cell lines established from human gastric carcinoma’, *Cancer research*, 50(9), pp. 2773–2780.

Patel, A.P. *et al.* (2014) ‘Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma’, *Science*, 344(6190), pp. 1396–1401.

Petrenko, N. *et al.* (2016) ‘Mediator Undergoes a Compositional Change during Transcriptional Activation’, *Molecular cell*, 64(3), pp. 443–454.

Philips, A. *et al.* (2018) ‘Brain Tumours: Rise in Glioblastoma Multiforme Incidence in England 1995-2015 Suggests an Adverse Environmental or Lifestyle Factor’, *Journal of environmental and public health*, 2018, p. 7910754.

Phillips, H.S. *et al.* (2006) ‘Molecular subclasses of high-grade glioma predict prognosis, delineate a pattern of disease progression, and resemble stages in neurogenesis’, *Cancer cell*, 9(3), pp. 157–173.

Piper, R.J. *et al.* (2018) ‘Neuroimaging classification of progression patterns in glioblastoma: a systematic review’, *Journal of neuro-oncology*, 139(1), pp. 77–88.

Poirier, G.G. *et al.* (1982) ‘Poly(ADP-ribosyl)ation of polynucleosomes causes relaxation of chromatin structure’, *Proceedings of the National Academy of Sciences of the United States of America*, 79(11), pp. 3423–3427.

Pollard, S.M. *et al.* (2009) ‘Glioma stem cell lines expanded in adherent culture have tumor-specific phenotypes and are suitable for chemical and genetic screens’, *Cell stem cell*, 4(6), pp. 568–580.

Pommier, Y., O'Connor, M.J. and de Bono, J. (2016) 'Laying a trap to kill cancer cells: PARP inhibitors and their mechanisms of action', *Science translational medicine*, 8(362), p. 362ps17.

Ponticelli, C., Escoli, R. and Moroni, G. (2018) 'Does cyclophosphamide still play a role in glomerular diseases?', *Autoimmunity reviews*, 17(10), pp. 1022–1027.

Prasad, C.B. *et al.* (2017) 'Olaparib modulates DNA repair efficiency, sensitizes cervical cancer cells to cisplatin and exhibits anti-metastatic property', *Scientific reports*, 7(1), p. 12876.

Puchalski, R.B. *et al.* (2018) 'An anatomic transcriptional atlas of human glioblastoma', *Science*, 360(6389), pp. 660–663.

Purshouse, K. *et al.* (2022a) 'Oncogene expression from extrachromosomal DNA is driven by copy number amplification and does not require spatial clustering', *bioRxiv*. Available at: https://doi.org/10.1101/2022.01.29.478046.

Purshouse, K. *et al.* (2022b) 'Oncogene expression from extrachromosomal DNA is driven by copy number amplification and does not require spatial clustering in glioblastoma stem cells', *eLife*, 11. Available at: https://doi.org/10.7554/eLife.80207.

Quinn, L.A. *et al.* (1979) 'Cell lines from human colon carcinoma with unusual cell products, double minutes, and homogeneously staining regions', *Cancer research*, 39(12), pp. 4914–4924.

Rabl, C. (1885) 'Uber zellthilung', *Morphol Jahrb*, 10, pp. 214–330.

Rack, J.G.M., Perina, D. and Ahel, I. (2016) 'Macrodomains: Structure, Function, Evolution, and Catalytic Activities', *Annual review of biochemistry*, 85, pp. 431–454.

Rai, A.K. *et al.* (2018) 'Kinase-controlled phase transition of membraneless organelles in mitosis', *Nature*, 559(7713), pp. 211–216.

Ramírez, F. *et al.* (2016) 'deepTools2: a next generation web server for deep-sequencing data analysis', *Nucleic acids research*, 44(W1), pp. W160–5.

Rao, S.S.P. *et al.* (2014) 'A 3D map of the human genome at kilobase resolution reveals principles of chromatin looping', *Cell*, 159(7), pp. 1665–1680.

Rausch, T. *et al.* (2012) ‘Genome sequencing of pediatric medulloblastoma links catastrophic DNA rearrangements with TP53 mutations’, *Cell*, 148(1-2), pp. 59–71.

Ray Chaudhuri, A. *et al.* (2012) ‘Topoisomerase I poisoning results in PARP-mediated replication fork reversal’, *Nature structural & molecular biology*, 19(4), pp. 417–423.

Reardon, D.A. *et al.* (2020) ‘Effect of Nivolumab vs Bevacizumab in Patients With Recurrent Glioblastoma: The CheckMate 143 Phase 3 Randomized Clinical Trial’, *JAMA oncology*, 6(7), pp. 1003–1010.

Reardon, D.A., Wen, P.Y. and Mellinghoff, I.K. (2014) ‘Targeted molecular therapies against epidermal growth factor receptor: past experiences and challenges’, *Neuro-oncology*, 16 Suppl 8(Suppl 8), pp. viii7–13.

Richards, L.M. *et al.* (2021) ‘Gradient of Developmental and Injury Response transcriptional states defines functional vulnerabilities underpinning glioblastoma heterogeneity’, *Nature Cancer*, 2(2), pp. 157–173.

Rich, J.N. *et al.* (2001) ‘A genetically tractable model of human glioma formation’, *Cancer research*, 61(9), pp. 3556–3560.

Robertson, F.L. *et al.* (2019) ‘Experimental models and tools to tackle glioblastoma’, *Disease models & mechanisms*, 12(9). Available at: https://doi.org/10.1242/dmm.040386.

Röijer, E. *et al.* (2002) ‘Translocation, deletion/amplification, and expression of HMGIC and MDM2 in a carcinoma ex pleomorphic adenoma’, *The American journal of pathology*, 160(2), pp. 433–440.

Rose, M. *et al.* (2020) ‘PARP Inhibitors: Clinical Relevance, Mechanisms of Action and Tumor Resistance’, *Frontiers in cell and developmental biology*, 8, p. 564601.

Rosswog, C. *et al.* (2021) ‘Chromothripsis followed by circular recombination drives oncogene amplification in human cancer’, *Nature genetics*, 53(12), pp. 1673–1685.

Rothkamm, K. *et al.* (2003) ‘Pathways of DNA double-strand break repair during the mammalian cell cycle’, *Molecular and cellular biology*, 23(16), pp. 5706–5715.

Rouse, J. and Jackson, S.P. (2002) ‘Interfaces between the detection, signaling, and repair of DNA damage’, *Science*, 297(5581), pp. 547–551.

Sabari, B.R. *et al.* (2018) ‘Coactivator condensation at super-enhancers links phase separation and gene control’, *Science*, 361(6400). Available at: https://doi.org/10.1126/science.aar3958.

Sachdeva, R. *et al.* (2019) ‘BMP signaling mediates glioma stem cell quiescence and confers treatment resistance in glioblastoma’, *Scientific reports*, 9(1), p. 14569.

Saleh-Gohari, N. *et al.* (2005) ‘Spontaneous homologous recombination is induced by collapsed replication forks that are caused by endogenous DNA single-strand breaks’, *Molecular and cellular biology*, 25(16), pp. 7158–7169.

Sanborn, A.L. *et al.* (2015) ‘Chromatin extrusion explains key features of loop and domain formation in wild-type and engineered genomes’, *Proceedings of the National Academy of Sciences of the United States of America*, 112(47), pp. E6456–65.

Sanchez, A.M., Barrett, J.T. and Schoenlein, P.V. (1998) ‘Fractionated ionizing radiation accelerates loss of amplified MDR1 genes harbored by extrachromosomal DNA in tumor cells’, *Cancer research*, 58(17), pp. 3845–3854.

Sanchez, J.N., Wang, T. and Cohen, M.S. (2018) ‘BRAF and MEK Inhibitors: Use and Resistance in BRAF-Mutated Cancers’, *Drugs*, 78(5), pp. 549–566.

Sarkaria, J.N. *et al.* (2006) ‘Use of an orthotopic xenograft model for assessing the effect of epidermal growth factor receptor amplification on glioblastoma radiation response’, *Clinical cancer research: an official journal of the American Association for Cancer Research*, 12(7 Pt 1), pp. 2264–2271.

Sarkaria, J.N. *et al.* (2007) ‘Identification of molecular characteristics correlated with glioblastoma sensitivity to EGFR kinase inhibition through use of an intracranial xenograft test panel’, *Molecular cancer therapeutics*, 6(3), pp. 1167–1174.

Satoh, M.S. and Lindahl, T. (1992) ‘Role of poly(ADP-ribose) formation in DNA repair’, *Nature*, 356(6367), pp. 356–358.

Schimke, R.T. (1984) ‘Gene amplification in cultured animal cells’, *Cell*, 37(3), pp. 705–713.

Schoenlein, P.V. *et al.* (2003) ‘Radiation therapy depletes extrachromosomally amplified drug resistance genes and oncogenes from tumor cells via micronuclear capture of episomes and double minute chromosomes', *International journal of radiation oncology, biology, physics*, 55(4), pp. 1051–1065.

Schultz, N. *et al.* (2003) 'Poly(ADP-ribose) polymerase (PARP-1) has a controlling role in homologous recombination', *Nucleic acids research*, 31(17), pp. 4959–4964.

Schützenhofer, K., Rack, J.G.M. and Ahel, I. (2021) 'The Making and Breaking of Serine-ADP-Ribosylation in the DNA Damage Response', *Frontiers in cell and developmental biology*, 9, p. 745922.

Serrano, A. *et al.* (2022) 'Mastering the use of cellular barcoding to explore cancer heterogeneity', *Nature reviews. Cancer*, 22(11), pp. 609–624.

Shapiro, D.N. *et al.* (1993) 'Detection of N-myc gene amplification by fluorescence in situ hybridization. Diagnostic utility for neuroblastoma', *The American journal of pathology*, 142(5), pp. 1339–1346.

Shimizu, N. *et al.* (1998) 'Selective entrapment of extrachromosomally amplified DNA by nuclear budding and micronucleation during S phase', *The Journal of cell biology*, 140(6), pp. 1307–1320.

Shoshani, O. *et al.* (2021) 'Chromothripsis drives the evolution of gene amplification in cancer', *Nature*, 591(7848), pp. 137–141.

Simonin, F. *et al.* (1993) 'Identification of potential active-site residues in the human poly(ADP-ribose) polymerase', *The Journal of biological chemistry*, 268(12), pp. 8529–8535.

Singh, S.K. *et al.* (2003) 'Identification of a cancer stem cell in human brain tumors', *Cancer research*, 63(18), pp. 5821–5828.

Singh, S.K. *et al.* (2004) 'Identification of human brain tumour initiating cells', *Nature*, 432(7015), pp. 396–401.

Slade, D. *et al.* (2011) 'The structure and catalytic mechanism of a poly(ADP-ribose) glycohydrolase', *Nature*, 477(7366), pp. 616–620.

Snuderl, M. *et al.* (2011) 'Mosaic amplification of multiple receptor tyrosine kinase genes in glioblastoma', *Cancer cell*, 20(6), pp. 810–817.

Song, K. *et al.* (2021) ‘Plasticity of extrachromosomal and intrachromosomal BRAF amplifications in overcoming targeted therapy dosage challenges’, *Cancer discovery* [Preprint]. Available at: https://doi.org/10.1158/2159-8290.CD-20-0936.

Sonoda, Y. *et al.* (2001) ‘Formation of intracranial tumors by genetically modified human astrocytes defines four pathways critical in the development of human anaplastic astrocytoma’, *Cancer research*, 61(13), pp. 4956–4960.

Sottoriva, A. *et al.* (2013) ‘Intratumor heterogeneity in human glioblastoma reflects cancer evolutionary dynamics’, *Proceedings of the National Academy of Sciences of the United States of America*, 110(10), pp. 4009–4014.

Soutourina, J. *et al.* (2011) ‘Direct interaction of RNA polymerase II and mediator required for transcription in vivo’, *Science*, 331(6023), pp. 1451–1454.

Soutourina, J. (2018) ‘Transcription regulation by the Mediator complex’, *Nature reviews. Molecular cell biology*, 19(4), pp. 262–274.

Spell, R.M. and Holm, C. (1994) ‘Nature and distribution of chromosomal intertwinings in Saccharomyces cerevisiae’, *Molecular and cellular biology*, 14(2), pp. 1465–1476.

Stark, G.R. *et al.* (1989) ‘Recent progress in understanding mechanisms of mammalian DNA amplification’, *Cell*, 57(6), pp. 901–908.

Stephens, P.J. *et al.* (2011) ‘Massive genomic rearrangement acquired in a single catastrophic event during cancer development’, *Cell*, 144(1), pp. 27–40.

Stevens, T.J. *et al.* (2017) ‘3D structures of individual mammalian genomes studied by single-cell Hi-C’, *Nature*, 544(7648), pp. 59–64.

Stöber, M.C. *et al.* (2023) ‘Intercellular extrachromosomal DNA copy number heterogeneity drives cancer cell state diversity’, *bioRxiv*. Available at: https://doi.org/10.1101/2023.01.21.525014.

Storlazzi, C.T. *et al.* (2006) ‘MYC-containing double minutes in hematologic malignancies: evidence in favor of the episome model and exclusion of MYC as the target gene’, *Human molecular genetics*, 15(6), pp. 933–942.

Storlazzi, C.T. *et al.* (2010) ‘Gene amplification as double minutes or homogeneously staining regions in solid tumors: origin and structure’, *Genome research*, 20(9), pp. 1198–1206.

Strom, A.R. and Brangwynne, C.P. (2019) ‘The liquid nucleome - phase transitions in the nucleus at a glance’, *Journal of cell science*, 132(22). Available at: https://doi.org/10.1242/jcs.235093.

Stupp, R. *et al.* (2005) ‘Radiotherapy plus concomitant and adjuvant temozolomide for glioblastoma’, *The New England journal of medicine*, 352(10), pp. 987–996.

Stupp, R. *et al.* (2014) ‘High-grade glioma: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up’, *Annals of oncology: official journal of the European Society for Medical Oncology / ESMO*, 25 Suppl 3, pp. iii93–101.

Stupp, R. *et al.* (2017) ‘Effect of Tumor-Treating Fields Plus Maintenance Temozolomide vs Maintenance Temozolomide Alone on Survival in Patients With Glioblastoma: A Randomized Clinical Trial’, *JAMA: the journal of the American Medical Association*, 318(23), pp. 2306–2316.

Sung, P. and Klein, H. (2006) ‘Mechanism of homologous recombination: mediators and helicases take on regulatory functions’, *Nature reviews. Molecular cell biology*, 7(10), pp. 739–750.

Suskiewicz, M.J. *et al.* (2020) ‘HPF1 completes the PARP active site for DNA damage-induced ADP-ribosylation’, *Nature*, 579(7800), pp. 598–602.

Suvà, M.L. *et al.* (2014) ‘Reconstructing and reprogramming the tumor-propagating potential of glioblastoma stem-like cells’, *Cell*, 157(3), pp. 580–594.

Szerlip, N.J. *et al.* (2012) ‘Intratumoral heterogeneity of receptor tyrosine kinases EGFR and PDGFRA amplification in glioblastoma defines subpopulations with distinct growth factor response’, *Proceedings of the National Academy of Sciences of the United States of America*, 109(8), pp. 3041–3046.

Takahashi, K. and Yamanaka, S. (2006) ‘Induction of pluripotent stem cells from mouse embryonic and adult fibroblast cultures by defined factors’, *Cell*, 126(4), pp. 663–676.

Talevich, E. *et al.* (2016) ‘CNVkit: Genome-Wide Copy Number Detection and Visualization from Targeted DNA Sequencing’, *PLoS computational biology*, 12(4), p. e1004873.

Tan, B.T. *et al.* (2006) ‘The cancer stem cell hypothesis: a work in progress’, *Laboratory investigation; a journal of technical methods and pathology*, 86(12), pp. 1203–1207.

Turner, K.M. *et al.* (2017) ‘Extrachromosomal oncogene amplification drives tumor evolution and genetic heterogeneity’, *Nature*, 543(7643), pp. 122–125.

Van Roy, N. *et al.* (2006) ‘Translocation-excision-deletion-amplification mechanism leading to nonsyntenic coamplification of MYC and ATBF1’, *Genes, chromosomes & cancer*, 45(2), pp. 107–117.

Veatch, S.L. *et al.* (2012) ‘Correlation functions quantify super-resolution images and estimate apparent clustering due to over-counting’, *PloS one*, 7(2), p. e31457.

Verhaak, R.G.W. *et al.* (2010) ‘Integrated genomic analysis identifies clinically relevant subtypes of glioblastoma characterized by abnormalities in PDGFRA, IDH1, EGFR, and NF1’, *Cancer cell*, 17(1), pp. 98–110.

Verhaak, R.G.W., Bafna, V. and Mischel, P.S. (2019) ‘Extrachromosomal oncogene amplification in tumour pathogenesis and evolution’, *Nature reviews. Cancer*, 19(5), pp. 283–288.

Vernimmen, D. and Bickmore, W.A. (2015) ‘The Hierarchy of Transcriptional Activation: From Enhancer to Promoter’, *Trends in genetics: TIG*, 31(12), pp. 696–708.

Vogt, N. *et al.* (2004) ‘Molecular structure of double-minute chromosomes bearing amplified copies of the epidermal growth factor receptor gene in gliomas’, *Proceedings of the National Academy of Sciences of the United States of America*, 101(31), pp. 11368–11373.

Von Hoff, D.D. *et al.* (1992) ‘Elimination of extrachromosomally amplified MYC genes from human tumor cells reduces their tumorigenicity’, *Proceedings of the National Academy of Sciences of the United States of America*, 89(17), pp. 8165–8169.

Walentynowicz, K.A. *et al.* (2023) ‘Single-cell heterogeneity of EGFR and CDK4 co-amplification is linked to immune infiltration in glioblastoma’, *Cell reports*, 42(3), p. 112235.

Wang, J.C. (2002) ‘Cellular roles of DNA topoisomerases: a molecular perspective’, *Nature reviews. Molecular cell biology*, 3(6), pp. 430–440.

Wang, L.-B. *et al.* (2021) ‘Proteogenomic and metabolomic characterization of human glioblastoma’, *Cancer cell*, 39(4), pp. 509–528.e20.

Wang, Q. *et al.* (2017) ‘Tumor Evolution of Glioma-Intrinsic Gene Expression Subtypes Associates with Immunological Changes in the Microenvironment’, *Cancer cell*, 32(1), pp. 42–56.e6.

Wang, T. *et al.* (2021) ‘Extrachromosomal circular DNA: a new potential role in cancer progression’, *Journal of translational medicine*, 19(1), p. 257.

Wang, Y. *et al.* (2021) ‘Nanopore sequencing technology, bioinformatics and applications’, *Nature biotechnology*, 39(11), pp. 1348–1365.

Waskom, M. (2021) ‘seaborn: statistical data visualization’, *Journal of open source software*, 6(60), p. 3021.

Weber, G.F. (2015) ‘DNA Damaging Drugs’, in G.F. Weber (ed.) *Molecular Therapies of Cancer*. Cham: Springer International Publishing, pp. 9–112.

Whyte, W.A. *et al.* (2013) ‘Master transcription factors and mediator establish super-enhancers at key cell identity genes’, *Cell*, 153(2), pp. 307–319.

Williamson, I. *et al.* (2016) ‘Shh and ZRS enhancer colocalisation is specific to the zone of polarising activity’, *Development* , 143(16), pp. 2994–3001.

Williamson, I. *et al.* (2019) ‘Developmentally regulated Shh expression is robust to TAD perturbations’, *Development* , 146(19). Available at: https://doi.org/10.1242/dev.179523.

de Wit, E. and de Laat, W. (2012) ‘A decade of 3C technologies: insights into nuclear organization’, *Genes & development*, 26(1), pp. 11–24.

Wu, S. *et al.* (2019) ‘Circular ecDNA promotes accessible chromatin and high oncogene expression’, *Nature*, 575(7784), pp. 699–703.

Yaneva, M., Kowalewski, T. and Lieber, M.R. (1997) ‘Interaction of DNA-dependent protein kinase with DNA and with Ku: biochemical and atomic-force microscopy studies’, *The EMBO journal*, 16(16), pp. 5098–5112.

Yang, G. *et al.* (2018) ‘Super-resolution imaging identifies PARP1 and the Ku complex acting as DNA double-strand break sensors’, *Nucleic acids research*, 46(7), pp. 3446–3457.

Yan, J.-L. *et al.* (2019) ‘Multimodal MRI characteristics of the glioblastoma infiltration beyond contrast enhancement’, *Therapeutic advances in neurological disorders*, 12, p. 1756286419844664.

Yi, E. *et al.* (2021) ‘Live-cell imaging shows uneven segregation of extrachromosomal DNA elements and transcriptionally active extrachromosomal DNA hubs in cancer’, *Cancer discovery* [Preprint]. Available at: https://doi.org/10.1158/2159-8290.CD-21-1376.

Ying, S. *et al.* (2016) ‘DNA-PKcs and PARP1 Bind to Unresected Stalled DNA Replication Forks Where They Recruit XRCC1 to Mediate Repair’, *Cancer research*, 76(5), pp. 1078–1088.

Yu, S.-W. *et al.* (2002) ‘Mediation of poly(ADP-ribose) polymerase-1-dependent cell death by apoptosis-inducing factor’, *Science*, 297(5579), pp. 259–263.

Zack, T.I. *et al.* (2013) ‘Pan-cancer patterns of somatic copy number alteration’, *Nature genetics*, 45(10), pp. 1134–1140.

Zhang, C.-Z. *et al.* (2015) ‘Chromothripsis from DNA damage in micronuclei’, *Nature*, 522(7555), pp. 179–184.

Zhang, X.-H. *et al.* (2015) ‘Off-target Effects in CRISPR/Cas9-mediated Genome Engineering’, *Molecular therapy. Nucleic acids*, 4(11), p. e264.

Zhu, Y. *et al.* (2021) ‘Oncogenic extrachromosomal DNA functions as mobile enhancers to globally amplify chromosomal transcription’, *Cancer cell*, 39(5), pp. 694–707.e7.

Chapter 8 Appendices

8.1 Fosmid clones

Locus (chr:coordinates)	Gene	Start (bp)	End (bp)	Fosmid ID	Clone name
7: 55019017-55211628	EGFR	55024189	55063180	G248P88704G2	WI2-2910M03
7: 55019017-55211628	EGFR	54881239	54920527	G248P87040G5	WI2-3534M09
12: 57747727-57752310	CDK4	57746054	57783419	G248P80931E4	WI2-0793J08
4: 54,229,293-54298245	PDGFRA	54230802	54268615	G248P86466H11	WI2-2022O22
8: 127,736,231-127,742,951	MYC	127720911	127761795	G248P802009C12	WI2-3746F23

8.2 crRNA sequences

Gene	CrRNA sequence
EGFR (5’ Homology arm)	AGCGATGCGACCCTCCGGGA
EGFR (3’ Homology arm)	TTATTGGAGCATGACCACGG
POLR2G	ACCAAGGGTAGGAGGCCACC

8.3 mCherry dsDNA sequence for POLR2G locus

dsDNA donor DNA block sequence for target locus POLR2G
TCCCTGATGGACGATTACTTGGGTGAGTGCCTGATCATAGGTGCTGGGGTTATTGCCTGG AGAAGGGATGTGTGGGGGTGGGGAGTAATATAGGATTCAATGCCCAAATCAGAGAGACA GAAGAAACTTTCATGCTGTCTGCTTGAAAGATCCAGGACATTTGCCTTGGGATGAGGAGTA CATGGTTGTGGCTACCCTAAATTCCGGTTCTAACTGATATGCTTTTTCTGGTTTCGCAGGG CTTGTAAGCGGCGACGGCGGCAGCGGCGGCGGCAGCATGGTGAGCAAGGGCGAGGAG GATAACATGGCCATCATCAAGGAGTTCATGCGCTTCAAGGTGCACATGGAGGGCTCCGTG AACGGCCACGAGTTCGAGATCGAGGGCGAGGGCGAGGGCCGCCCCTACGAGGGCACCC AGACCGCCAAGCTGAAGGTGACCAAGGGTGGCCCCCTGCCCTTCGCCTGGGACATCCTG TCCCCTCAGTTCATGTACGGCTCCAAGGCCTACGTGAAGCACCCCGCCGACATCCCCGAC TACTTGAAGCTGTCCTTCCCCGAGGGCTTCAAGTGGGAGCGCGTGATGAACTTCGAGGAC GGCGGCGTGGTGACCGTGACCCAGGACTCCTCCCTGCAGGACGGCGAGTTCATCTACAA GGTGAAGCTGCGCGGCACCAACTTCCCCTCCGACGGCCCCGTAATGCAGAAGAAGACCA TGGGCTGGGAGGCCTCCTCCGAGCGGATGTACCCCGAGGACGGCGCCCTGAAGGGCGA GATCAAGCAGAGGCTGAAGCTGAAGGACGGCGGCCACTACGACGCTGAGGTCAAGACCA CCTACAAGGCCAAGAAGCCCGTGCAGCTGCCCGGCGCCTACAACGTCAACATCAAGTTG GACATCACCTCCCACAACGAGGACTACACCATCGTGGAACAGTACGAACGCGCCGAGGG CCGCCACTCCACCGGCGGCATGGACGAGCTGTACAAGTGAGGCTGGTGCCCTCCTACCC TTGGTCCTACTCTAGGAAGTGTGATTGTCACACTTATCATGTTGTCCAGAGGTCCAGTCTG GCTGCTGTTGTGGAGGCAAGGAAGGCAACTCATCCCAGAAGGCATCTGGTGCTTCTTGTA GCTTAACTACTGCCTCCTCATTTTTCAGTATGTGTTCTAAGTATAAAAAGTCCTTGG

8.4 Custom Stellaris® RNA FISH probe sequences

cggccgaagaacgaaacgtc	ctcagttgggaattcgactc	ctgctgaaactggtaggagt
aacgaggaaactaaccgccg	tactgaatgggaacgtctcc	aaactaagtcagggccatgt
cttggataggcaagacgagg	tctctttaagacagctcgga	tccagacacttaagggcaat
tgtcaagtgggtttatggtc	cattaatatccccactttgg	caaaagtcatcctttgtcct
taaaccttccaaggcatgtt	gaagcctgaagggtcaaagt	ccccatgtagaattaaaggc
agccaatgacatcaacagca	agcaagagcagaaaccctta	ttccatttgtggctaaacgt
acaattccagatgtcctcaa	ccagtttcttgtattctgtg	gagccaggaaactgtgagag
gagaagaaacctttccacca	aatggactttcggcaatcgg	tcaatttcgatgctgacggt
ctgacacctgactaacagga	tagagttggtttgtcactct	agagagaagctggtatccag
ggagcattctgttttgagtg	tgcccataatcatcaattgc	tgcaactaatcagctgctga
taatttgcccctgaaaatcc	acaggcgaaatgtcctcaac	tccgagtatttgcagcaata
aagttagcatgtcccagaac	aagcaacatgtgggtgaacc	tgtggtcttaagcagggaac
tacctctaagagcagcaagt	atggcctacaggaacaagtt	aacatgtaatccactcctca
tctcctgaaaccagaactcg	caatctagagatgtcgtcca	gatttctttcgttacgagct
catttgcacacacctgattt	ggattcttgcccaaattcaa	aatgggggtctgagagcaaa
aacagagggctgaggaactc	tctaggggctcagtaaacac	ctgtttctggatctcgaagg

8.5 Antibodies

Description	**Source or reference**	**Identifiers**	**Additional information**
Phospho-Histone H3 (ser28) eFluor 660 (Rat monoclonal)	eBioscience	h9908	IF (1 in 50)
Ki67 (Rabbit mono-clonal)	Thermo Fisher Scientific	rm9106	IF (1 in 100)
mCherry (Rabbit poly-clonal)	abcam	ab167453	IF (1 in 500)
mCherry (16D7) (Rat mono-clonal	Thermo Fisher Scientific	# M11217	IF (1 in 500)
Rpb1 NTD (D8L4Y) (Rabbit mono-clonal)	Cell Signalling Technology	#14958	IF (1 in 1000)
PolyPAR (Rabbit)	Sigma-Aldrich	MABE1031	IF (1 in 250)
Phospho-histone H2A.X (Ser139) clone JBW301 (Mouse mono-clonal)	Sigma-Aldrich	05-636	IF (1 in 2000)
Anti-Digoxigenin (Sheep poly-clonal)	Roche	11333089001	DNA FISH (1 in 10)
Secondary Antibody – Alexa Fluor 647 (Donkey anti-Sheep IgG poly-clonal)	Thermo Fisher Scientific	A-21448	DNA FISH (1 in 10)
Secondary Antibody – Alexa Fluor 568 (Donkey anti-Rabbit IgG poly-clonal)	Thermo Fisher Scientific	A-10042	IF (1 in 1000)
Secondary Antibody – Alexa Fluor 488 (Donkey anti-Rabbit IgG poly-clonal)	Thermo Fisher Scientific	A-21206	IF (1 in 1000)
Secondary Antibody – Alexa Fluor 488 (Donkey anti-Rat IgG poly-clonal)	Thermo Fisher Scientific	A-21208	IF (1 in 1000)
Secondary Antibody - Alexa Fluor 488 (Goat anti-Rabbit IgG (H+L) poly-clonal)	Invitrogen	A11034	IF (1 in 1000)
Secondary Antibody - Alexa Fluor 568 (Donkey anti-Mouse IgG poly-clonal)	Thermo Fisher Scientific	A10037	IF (1 in 1000)
EGFR (D38B1) (Rabbit mono-clonal)	Cell Signalling	#4267	Western blotting (1 in 1000)
c-Myc (9E10) (Mouse mono-clonal)	Santa Cruz Biotechnology, Inc.	Sc-40	Western blotting (1 in 500)

GAPDH 6C5 (Mouse mono-clonal)	Thermo Fisher Scientific	# AM4300	Western blotting (1 in 10,000)
Secondary Antibody - Goat anti-Rabbit IgG StarBright™ Blue 700 poly-clonal	BioRad	12004161	Western blotting (1 in 10,000)
Secondary Antibody - Goat anti-Mouse IgG StarBright™ Blue 520 poly-clonal	BioRad	12005866	Western blotting (1 in 10,000)
Alexa Fluor™ 647 EGF complex	Thermo Fisher Scientific	E35351	Flow cytometry

8.6 Reagents and solutions

Description	**Source or reference**	**Identifiers**	**Additional information**
DMEM/HAMS-F12	Sigma Aldrich	Cat#: D8437	Cell culture, media
DMEM	Thermo Fisher Scientific	Cat#: 11995073	Cell culture, media (Colo320DM only)
Pen/Strep	GIBCO	Cat#:15140-122	Cell culture, media supplement
BSA Solution	GIBCO	Cat#:15260-037	Cell culture, media supplement
Goat serum	Sigma-Aldrich	#G6767	IF, blocking buffer
B27 Supplement (50X)	LifeTech/GIBCO	Cat#: 17504-044	Cell culture, media supplement
N2 Supplement (100X)	LifeTech/GIBCO	Cat#: 17502-048	Cell culture, media supplement
Laminin	Cultrex	Cat#: 3446-005-01	Cell culture, media supplement and pre-lamination of culture vessels
EGF	Peprotech	Cat: 315-09	Cell culture, media supplement
FGF-2	Peprotech	100-18B	Cell culture, media supplement
Accutase	Sigma Aldrich	Cat#: A6964	Cell culture, cell dissociation agent
0.25% (v/v) Trypsin-EDTA solution	GIBCO	Cat#: 25200072	Cell culture, cell dissociation agent (Colo320DM only)
DMSO	Sigma Aldrich	Cat#: 276855	Cell culture, freeze media and drug diluent
BMP	Peprotech	# 120-05	Cell culture, media supplement where indicated
Triton X-100	Merck Life Sciences	Cat#: X-100	Cell permeabilisation agent following cell fixation
Olaparib	Selleck Chemicals	S1060-SEL	PARP inhibitor
Hydrogen peroxide	Merck	H1009	DNA damage agent

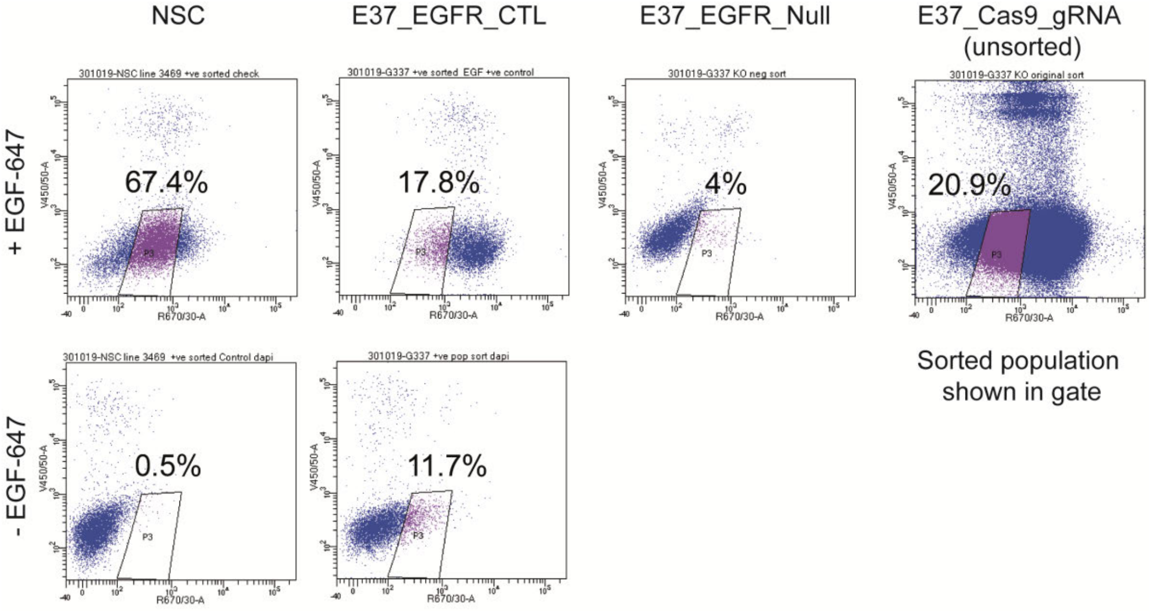
Paraformaldehyde Powder 95%	Sigma	Cat#: 158127	Cell fixation agent
Tween 20	Cambridge Bioscience	Cat#: TW0020	DNA FISH (hybridization mix)
PBS Tablets	Sigma Aldrich	Cat#: P4417	Diluent and washing agent
Ethanol	VWR	Cat#: 20821-330	DNA FISH
Methanol	Fisher Chemical	M/4000/17	Used 3:1 with acetic acid for metaphase spreads
Acetic acid	Honeywell Research Chemicals	33209-1L	See above
Green496-dUTP	ENZO life sciences	ENZ-42831L	Direct labelling of Fosmid DNA FISH probes via nick translation
ChromaTide AlexaFluor 594-5-dUTP	Thermo Fisher Scientific	C11400	Direct labelling of Fosmid DNA FISH probes via nick translation
Qubit dsDNA broad range assay	Thermo Fisher Scientific	Q32850	Quantification of DNA
DNA Polymerase 1	Invitrogen	18010-017	Nick Translation
DNase I recombinant, RNase-free	Roche	04716728001	Nick Translation
RNase A 20mg/ml	Invitrogen	12091021	DNA FISH
Human Cot-1 DNA	Thermo Fisher Scientific	15279011	DNA FISH
Salmon Sperm DNA	Invitrogen	15632011	DNA FISH
Dextran sulphate	Sigma	D8906	DNA FISH hybridisation mix
Sephadex Quick Spin Columns	Roche	11273973001	Nick Translation purification
Colcemid	Thermo Fisher Scientific	15210040	1µg/ml for 30 mins (E37)
Paclitaxel	Cambridge Bioscience	CAY10461	10-100nM
Nocodazole	Sigma-Aldrich	SML1665	50-100ng/ml
XCP 7 Orange Chromosome Paint	MetaSystems Probes	D-0307-100-OR	DNA FISH

Stellaris RNA-FISH probes ((Custom Assay with Quasar® 570 Dye)	LGC Biosearch Technologies	SMF-1063-5	RNA FISH
Stellaris® RNA FISH Hybridization Buffer	LGC Biosearch Technologies	SMF-HB1-10	RNA FISH
Alt-R® CRISPR-Cas9 crRNA	IDT-Technologies	Alt-R® CRISPR-Cas9 crRNA	CRISPR/Cas9 reagent
Alt-R® CRISPR-Cas9 tracrRNA	IDT-Technologies	1072532	CRISPR/Cas9 reagent
SG Cell Line 4D-NucleofectorTM X Kit S	Lonza Bioscience	V4XC-3032	CRISPR/Cas9 nucleofection reagents
Chromosome 7 Control Probe	Pisces Scientific	CHR07-10-DIG	Probe and hybridization mix
DAPI (4',6-Diamidino-2-Phenylindole, Dihydrochloride)	Thermo Fisher Scientific	D1306	Nuclear staining; 50ng/ml and 5ng/ml (as indicated in methods)
Complete Protease Inhibitors	Roche	11697498001	Protein lysate for western blotting
Pierce BCA Protein Kit	Thermo Fisher Scientific	23227	Protein quantification for western blotting
LDS Sample Buffer 4X	Thermo Fisher Scientific	84788	Protein lysate preparation for gel electrophoresis
DTT	Thermo Fisher Scientific	P2325	Protein lysate preparation for gel electrophoresis
MES	Thermo Fisher Scientific	NP0002	Gel electrophoresis solute
Precision Plus Protein Dual Color Standards	BioRad	#1610374	Protein standard for gel electrophoresis
10X Tris/Glycine Transfer Buffer	BioRad	#1610734	Transfer buffer for wet electroblotting
Vectashield mounting medium	Vector laboratories	H-1000-10	Mounting slides

8.7 Software and Bioinformatics resources

Name	**Reference**	**Source**
GraphPad Prism 9.0	GraphPad Software, Inc	https://www.graphpad.com/
FCS Express	FCS Express 7	https://denovosoftware.com/
BD FACSDiva	BD Biosciences	https://www.bdbiosciences.com/
Fiji/ImageJ	Open Source	https://imagej.net/Fiji
BioRender	BioRender	https://biorender.com/
Python v3.9	Open Source	https://www.python.org
Algorithm - RipleyK package	Python Package Index	https://pypi.org/project/ripleyk/
Imaris x64 v9.4.0	Imaris Microscopy Image Analysis Software	https://imaris.oxinst.com/
Ilastik v1.3.3	(Berg *et al.*, 2019)	https://www.ilastik.org/
UCSC Genome Browser	(Kent *et al.*, 2002)	https://genome.cshlp.org/content/12/6/996
STAR 2.7.1a	(Dobin *et al.*, 2013)	https://github.com/alexdobin/STAR
Picard	Broad Institute	https://broadinstitute.github.io/picard/
AmpliconSuite	(Luebeck *et al.*, 2022)	https://github.com/jluebeck/AmpliconSuite-pipeline
AmpliconArchitect	(Deshpande *et al.*, 2019)	https://github.com/virajbdeshpande/AmpliconArchitect
AmpliconClassifier	(Kim *et al.*, 2020)	https://github.com/jluebeck/AmpliconClassifier
deepTools v3.4	(Ramírez *et al.*, 2016)	https://deeptools.readthedocs.io/en/develop/
HOMER2 4.10	(Heinz *et al.*, 2010)	http://homer.ucsd.edu/homer/
SAMtools v1.10	(Li *et al.*, 2009)	http://www.htslib.org
BEDTools v2.3	(Quinlan and Hall, 2010)	http://code.google.com/p/bedtools
bcftools	(Danecek et al. 2021)	https://doi.org/10.1093/gigascience/giab008
strelka v2.9.10	(Kim et al. 2018)	https://doi.org/10.1038/s41592-018-0051-x
BWA MEM	(Li and Durbin, 2009)	https://github.com/lh3/bwa
CNVkit v0.9.9	(Talevich *et al.*, 2016)	https://github.com/etal/cnvkit
Seaborn v12.0	(Waskom, 2021)	https://seaborn.pydata.org/

8.8 FACS of E37 EGFR-Low population



FACS for E37 cells already sorted into EGFR-CTL and EGFR-Null populations. Using the NSC cell line as the primary reference point for a normal EGFR expression profile, and EGFR-Low population was sorted and recovered from the pooled E37_Cas9_gRNA cell pool as shown.

8.9 Publication

RESEARCH ARTICLE

Oncogene expression from extrachromosomal DNA is driven by copy number amplification and does not require spatial clustering in glioblastoma stem cells

Karin Purshouse[1,2], Elias T Friman[1], Shelagh Boyle[1], Pooran Singh Dewari[2], Vivien Grant[2], Alhafidz Hamdan[2], Gillian M Morrison[2], Paul M Brennan[2,3], Sjoerd V Beentjes[1,4], Steven M Pollard[2]*, Wendy A Bickmore[1]*†

[1]MRC Human Genetics Unit, Institute of Genetics and Cancer, The University of Edinburgh, Edinburgh, United Kingdom; [2]Centre for Regenerative Medicine and Cancer Research UK Edinburgh Centre, Institute for Regeneration and Repair, The University of Edinburgh, Edinburgh, United Kingdom; [3]Centre for Clinical Brain Sciences, University of Edinburgh, Edinburgh, United Kingdom; [4]School of Mathematics, University of Edinburgh, Edinburgh, United Kingdom

***For correspondence:** steven.pollard@ed.ac.uk (SMP); wendy.bickmore@ed.ac.uk (WAB)

†Lead contact

Competing interest: The authors declare that no competing interests exist.

Funding: See page 20

Preprinted: 29 January 2022
Received: 16 May 2022
Accepted: 21 October 2022
Published: 07 December 2022

Reviewing Editor: Jessica K Tyler, Weill Cornell Medicine, United States

Abstract Extrachromosomal DNA (ecDNA) are frequently observed in human cancers and are responsible for high levels of oncogene expression. In glioblastoma (GBM), ecDNA copy number correlates with poor prognosis. It is hypothesized that their copy number, size, and chromatin accessibility facilitate clustering of ecDNA and colocalization with transcriptional hubs, and that this underpins their elevated transcriptional activity. Here, we use super-resolution imaging and quantitative image analysis to evaluate GBM stem cells harbouring distinct ecDNA species (*EGFR, CDK4, PDGFRA*). We find no evidence that ecDNA routinely cluster with one another or closely interact with transcriptional hubs. Cells with *EGFR*-containing ecDNA have increased *EGFR* transcriptional output, but transcription per gene copy is similar in ecDNA compared to the endogenous chromosomal locus. These data suggest that it is the increased copy number of oncogene-harbouring ecDNA that primarily drives high levels of oncogene transcription, rather than specific interactions of ecDNA with each other or with high concentrations of the transcriptional machinery.

Editor's evaluation

This study convincingly shows that, in contrast to recent reports, the transcriptional output of oncogenes carried on extrachromosomal DNA (ecDNA) in glioblastoma cell lines is driven by the copy number of the ecDNA, rather than their spatial localization into transcriptional hubs. This study is relevant to researchers interested in nuclear function, particularly transcriptional organization within malignant cells.

Introduction

Glioblastoma (GBM) is characterized by intra-tumoural heterogeneity and stem cell-like properties that underpin treatment resistance and poor prognosis (***Bulstrode et al., 2017***; ***Suvà et al., 2014***). GBM is divided into distinct transcriptional subtypes that span a continuum of stem cell/developmental and injury response/immune evasion cell states (**Richards et al., 2021**; **Verhaak et al., 2010**; **Wang et al., 2021**). Genetically, activation or amplification of *EGFR* (chr7) is altered in almost two-thirds of GBM (**Brennan et al., 2013**). Other commonly amplified genes include *PDGFRA* (chr4), *CDK4*, *MDM2* (chr12), *MET,* and *CDK6* (chr7) with multicopy extrachromosomal DNA (ecDNA) considered a major mechanism for oncogene amplification (**Brennan et al., 2013**; **Kim et al., 2020**; **Snuderl et al., 2011**; **Szerlip et al., 2012**).

Although a long-recognized feature of cancer (**Cox et al., 1965**), ecDNA are particularly common in GBM, with 90% of patient-derived GBM tumour models harbouring ecDNA (**Turner et al., 2017**). However, there is much broader interest in mechanisms of ecDNA function across many solid tumours, as ecDNA enable rapid oncogene amplification in response to selective pressures, and have been shown to correlate with poor prognosis and treatment resistance (**Kim et al., 2020**; **Nathanson et al., 2014**; **Vicario et al., 2015**). EcDNA are centromere-free DNA circles of around 1–3 Mb in size that frequently exist as doublets (double minutes), but also as single elements (**Hamkalo et al., 1985**; **Verhaak et al., 2019**; **Vogt et al., 2004**). EcDNA can be composed of multiple genetic fragments generated as a result of chromothripsis (**Gibaud et al., 2010**; **Shoshani et al., 2021**; **Rosswog et al., 2021**). Although ecDNA were previously identified in 1.4% of cancers, more recent studies have shown their prevalence to be significantly higher (**Fan et al., 2011**; **Kim et al., 2020**; **Turner et al., 2017**). EcDNA can lead to oncogene copy number being amplified to >100 in any given cell, with significant copy number heterogeneity between cells (**Lange et al., 2022**; **Turner et al., 2017**). Freed from the constraints imposed by being embedded within a chromosome, ecDNA have spatial freedom and can adapt to targeted therapeutics (**Lange et al., 2022**; **Nathanson et al., 2014**). For example, the *EGFR* variant *EGFRvIII* (exon 2–7 deletion) is found on ecDNA, and is associated with an aggressive disease course and resistance mechanisms against EGFR inhibitors (**Brennan et al., 2013**; **Inda et al., 2010**; **Nathanson et al., 2014**; **Turner et al., 2017**).

As well as their resident oncogenes, ecDNA also harbour regulatory elements (enhancers) required to drive oncogene expression (**Morton et al., 2019**; **Zhu et al., 2021**). Consistent with this, ecDNA have been found to have regions of largely accessible chromatin (assayed by ATAC-seq), indicative of nucleosome displacement by bound transcription factors, and to be decorated with histone modifications associated with active chromatin (**Wu et al., 2019**). Transcription factors densely co-bound at enhancers have been suggested to nucleate condensates or ‘hubs’ (**Cho et al., 2018**; **Rai et al., 2018**; **Strom and Brangwynne, 2019**), enriched with key transcriptional components such as mediator and RNA polymerase II (PolII) to drive high levels of gene expression (**Cho et al., 2018**; **Chong et al., 2018**; **Sabari et al., 2018**). Given the colocation of enhancers and driver oncogenes on ecDNA, it has therefore been suggested that ecDNA cluster together in the nucleus, driving the recruitment of a high concentration of RNA PolII and creating ecDNA-driven nuclear hubs that in turn enhance the transcriptional output from ecDNA (**Adelman and Martin, 2021**; **Hung et al., 2021**; **Yi et al., 2021**; **Zhu et al., 2021**).

Here, using super-resolution imaging of primary GBM cell lines, we find that ecDNA are widely dispersed throughout the nucleus and we find neither evidence of ecDNA clustering together nor any significant spatial overlap between ecDNA and large PolII hubs. As expected, we show that expression from genes on ecDNA, both at mRNA and protein level, correlates with ecDNA copy number in the tumour cell lines. However, transcription of genes present on each individual ecDNA molecule appears to occur at a similar efficiency (transcripts per copy number) to that of the equivalent endogenous chromosomally located gene. These data suggest that it is primarily the increased copy number of ecDNA in GBM stem cells, and not a specific property of nuclear colocalization, that drives the increased transcriptional capacity of their resident oncogenes.

Results

EcDNA are more frequently located centrally in the nucleus in GBM stem cells

We characterized two GBM-derived glioma stem cell (GSC) primary cell lines containing multiple *EGFR*-harbouring ecDNA (ecEGFR) populations (GCGR-E26 and GCGR-E28, referred to here as E26 and E28). Whole genome sequencing (WGS) analysis using Amplicon Architect (**Deshpande et al., 2019**) indicated that E26 ecDNA harbour an *EGFRvIII* (exon 2–7 deletion), and E28 have a subpopulation of ecDNA with *EGFR* exon 7–14 deleted (***Figure 1A***). The presence of *EGFR* on ecDNA was confirmed by DNA FISH on metaphase spreads (***Figure 1B and C***). E26 harboured more ecDNA per cell than E28 (***Figure 1D***), with approximately 10% of metaphases also indicating the presence of a chromosomal homogeneously staining region (HSR) (***Figure 1B***; arrow). Endogenous *EGFR* is located on human chromosome 7, and metaphase spreads of the two tumour lines showed 3–6 copies of chromosome 7 in E26 and frequently 3 copies in E28 (***Figure 1E***).

Human chromosomes have non-random nuclear organization, with active regions preferentially located towards the central regions of the nucleus (***Boyle et al., 2001***; ***Croft et al., 1999***). We sought to determine the nuclear localization of ecDNA in GBM cell lines as compared with the endogenous chromosomal *EGFR*. DNA FISH for chromosome 7 and *EGFR* in nuclei from human fetal neural stem cells (NSCs) confirmed the trend for human chromosome 7 to be generally found towards the periphery of the nucleus (***Boyle et al., 2001 Figure 1F and G***, ***Figure 1—figure supplement 1***, ***Figure 1—source data 1***). Signal intensity analysis for equally sized bins eroded from the edge to the centre of each nucleus indicated that chromosome 7 and *EGFR* signal intensity were preferentially located towards the nuclear periphery in each cell line (***Figure 1—figure supplement 1***, ***Figure 1—source data 1***). Even once chromosome 7 signal was accounted for, *EGFR* DNA FISH signal was still highest at the periphery of NSC nuclei and lowest in the central regions ($p<0.0001$) (***Figure 1G***), likely reflecting the centromere proximal localization of endogenous *EGFR* on chromosome 7 (***Carvalho et al., 2001***). This radial organization was still significant ($p=0.012$), but much less marked, in E28 cells which have on average a modest number of EGFR ecDNA compared to endogenous copies (***Figure 1D***). In E26 cells, which have a very high copy number of ecDNA, this preference for a more peripheral localization is lost ($p=0.06$). These data suggest that, freed of the constraints on nuclear localization imposed by human chromosome 7, *EGFR* genes located on ecDNA can access more central regions of the nucleus.

EGFR-containing ecDNA in GBM stem cells do not cluster in the nucleus

It has been suggested that ecDNA cluster into ‘ecDNA hubs’ within nuclei of cancer cells, including for *EGFRvIII*-containing ecDNA in other GBM cell lines (HK359 and GBM39) (***Hung et al., 2021***; ***Yi et al., 2021***). We sought to quantify this using our E26 and E28 GBM cells with a single oncogene-harbouring ecDNA population (*EGFR* variant amplicons). Previous studies exploring genomic loci proximity and contact domains (***Williamson et al., 2016***; ***Williamson et al., 2019***; ***Hansen et al., 2021***), and the proximity of super-enhancers to BRD4/MED1 puncta (***Sabari et al., 2018***), would suggest that ecDNAs clustering together at a transcriptional hub should be located within ~200 nm or less of each other. We used 3D image-based analysis of the *EGFR* DNA FISH signals (***Figure 2A***) to determine if there is clustering of ecDNA. The relative frequency of all shortest *EGFR-EGFR* distances per nucleus did not suggest frequent ecDNA-ecDNA interactions at ≤200 nm in either cell line (***Figure 2B***, ***Figure 2—figure supplement 1A***). The mean shortest interprobe distances per nucleus were also not suggestive of close interactions, with no values <500 nm (***Figure 2—figure supplement 1B, C***; ***Figure 2—source data 1***). The single shortest interprobe distance per nucleus was also larger (0.24 µm, E26; 0.25 µm, E28) than would be expected if there were clustering of ecDNA in the close proximity required for coordinated transcription in hubs (***Figure 2—figure supplement 1D, E***; ***Figure 2—source data 1***).

The analysis above quantified distances between FISH hybridization signals but does not determine whether there is a non-random distribution of foci in the nuclei at distances in keeping with transcription hubs. We therefore used 3D Ripley’s K function to determine the observed spatial pattern of the foci in each nucleus and compared this with a random null distribution of 10,000 simulations of the same number of foci in the same volume. We powered this to identify any significant clustering at each radius in 0.1 µm increments between 0.1 and 1 µm (examples of E26 and E28 nuclei and their corresponding Ripley’s K function in ***Figure 2C***). The E26 cell line had some nuclei with significant non-random distribution of ecDNA, but only at ≥400 nm radial distances, and E28 only had occasional nuclei with significant non-random distribution of ecDNA at ≥700 nm (***Figure 2D***). We repeated this analysis, reducing the focus spot size from 300 to 150 nm diameter to ensure no small FISH foci were omitted that might skew our analysis. No significant clustering was observed at <300 nm (***Figure 2—figure supplement 1F***).

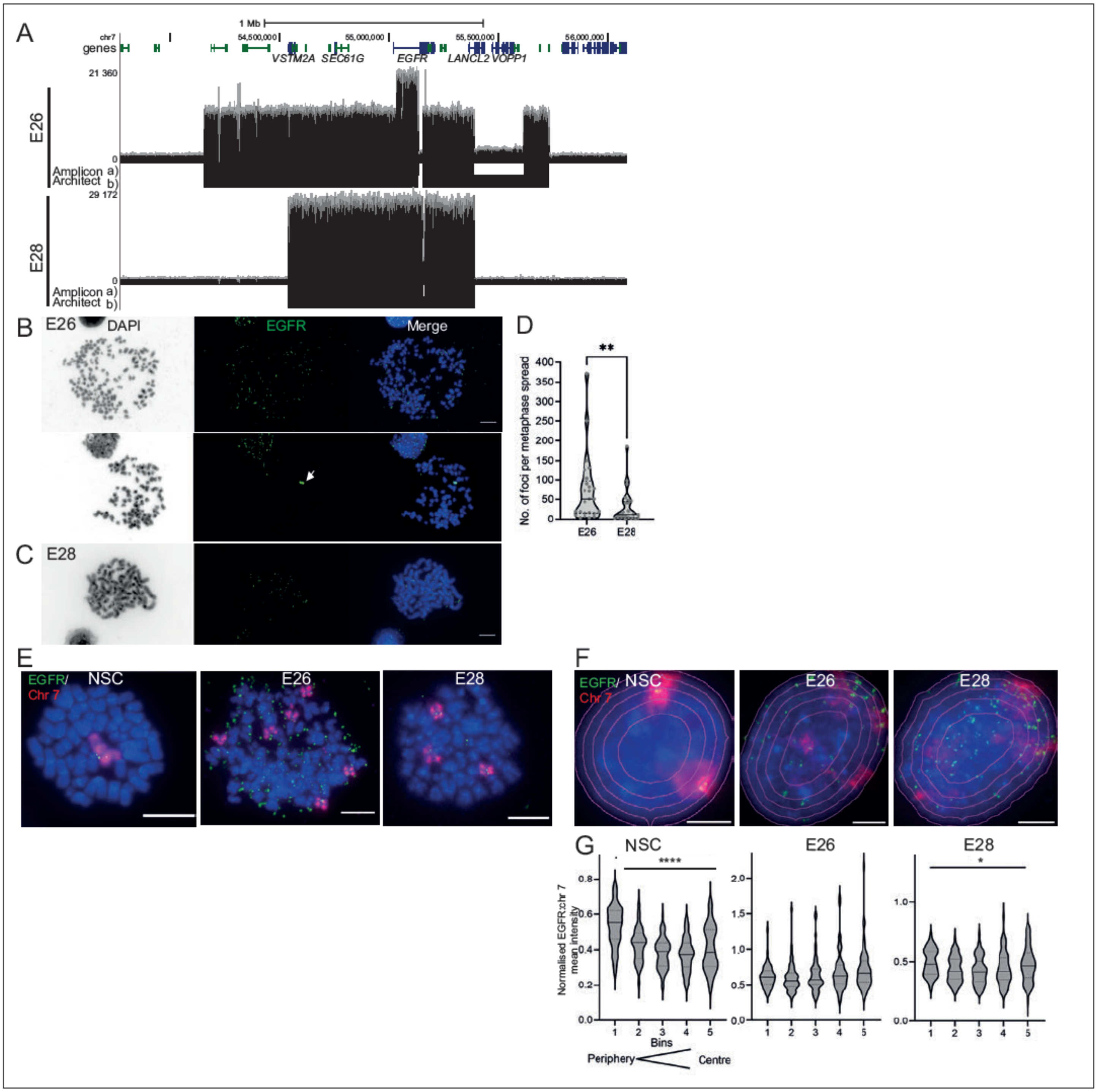


Figure 1. The nuclear localization of extrachromosomal DNA (ecDNA) in glioblastoma (GBM) cell lines. (**A**) Whole genome sequencing (WGS) and AmpliconArchitect analysis for ecDNA regions for E26 and E28 cell lines showing an *EGFR* exon 2–7 deletion in all ecDNA in E26 cells (seen in WGS and AmpliconArchitect regions a and b), and a subpopulation of ecDNA in E28 with a deletion across *EGFR* exons 7–14 (seen in WGS and Amplicon Architect region a – no deletion in E28 AmpliconArchitect region b). Genome coordinates (bp) are from the hg38 assembly of the human genome. (**B**) DNA FISH on metaphase spread of the E26 cell line showing *EGFR* (green) present on ecDNA, and on a homogeneously staining region (HSR) (arrowed) detected in ~10% of metaphases. Scale bar = 10 µm. (**C**) As for (**B**) but for the E28 cell line. (**D**) Violin plot of the number of *EGFR* DNA FISH signals per metaphase spread of E26 and E28 cells. Median and quartiles are shown. ** p=0.008 (Mann-Whitney test). Median values are 51 (E26)and 12 (E28), n=25 (E26) and 24 (E28) spreads. (**E and F**) Representative DNA FISH images of metaphase spread (**E**) and 2D nuclei (**F**) for neural stem cell (NSC), E26, and E28 cells showing signals for chromosome 7 (red) and EGFR (green). Blue = DNA (DAPI). Scale bar = 10 µm. The five erosions bins from the periphery to the centre of the nucleus are shown in F. (**G**) *EGFR* FISH signal intensity normalized to that for chromosome 7 (EGFR:Chr7 Mean Intensity)

Figure 1 continued on next page

Figure 1 continued

across five bins of equal area eroded from the peripheral (Bin 1) to the centre (Bin 5) of the nucleus for NSC, E26, and E28 cell lines. Median and quartiles shown. **** p<0.0001, * p<0.05. Kruskall-Wallis test. EGFR and chr7 signal normalized to DAPI shown in ***Figure 1—figure supplement 1***. n=66 (NSC), 59 (E26), 64 (E28) nuclei. Statistical data relevant for this figure are in ***Figure 1—source data 1***.

The online version of this article includes the following source data and figure supplement(s) for figure 1:

Source data 1. Statistical data for ***Figure 1*** and ***Figure 1—figure supplement 1***.

Figure supplement 1. Additional *EGFR* and chromosome 7 signal intensity data.

Different ecDNA populations do not cluster in the nucleus of GBM stem cells

To ensure that multiple ecDNAs are not so tightly clustered that they cannot be resolved by FISH, we analysed another primary GBM cell line (E25) which has two different oncogenes carried on separate ecDNA populations: *CDK4* and *PDGFRA* (***Figure 3—figure supplement 1A, B***). There was no obvious clustering of the two ecDNA populations in the nuclei of E25 cells (***Figure 3A***). The relative frequency of *CDK4-CDK4*, *PDGFRA-PDGFRA,* and *CDK4-PDGFRA* distances of ≤200 nm was low (***Figure 3B***). Indeed, the mean shortest interprobe distances per nucleus were overwhelmingly >1 μm, suggesting ecDNA were generally not in close proximity (***Figure 3—figure supplement 1C***). The shortest interprobe distances for *CDK4-CDK4* and *CDK4-PDGFRA* were shorter than for *PDGFRA-PDGFRA* foci, as expected given the higher copy number of *CDK4* ecDNA (***Figure 3—figure supplement 1B***); however, there was no significant difference in the shortest distance between *CDK4-CDK4* and *CDK4-PDGFRA* foci (***Figure 3—figure supplement 1D***). No two *CDK4* or two *PDGFRA* foci were <200 nm apart, and only four *CDK4-PDGFRA* distances were <200 nm (4/1011 [0.39%] *CDK4* foci, 4/518 [0.77%] *PDGFRA* foci) (***Figure 3—figure supplement 1D***). These data suggest that clustering is not a significant feature of two separate populations of ecDNA.

We used 3D Ripley's K function to evaluate point patterns in the E25 dual ecDNA oncogene cell line (***Figure 3C***). Some nuclei had a significant non-random distribution of *PDGFRA* ecDNA at ≥400 nm, and most nuclei had non-random distribution of *CDK4* ecDNA at >400 nm (***Figure 3D***). When both foci were combined, there was no significant clustering at <300 nm in any nucleus, and the number of nuclei with a significant non-random distribution at a given radius rose with increasing radial distance (***Figure 3D***). As previously, a repeat analysis with a smaller (150 nm diameter) spot size identified no instances of significant clustering at <300 nm (***Figure 3—figure supplement 1E***).

To further validate this, we repeated 3D Ripley's function analysis in a second GBM cell line (E20) harbouring *CDK4* and *PDGFRA* ecDNAs. Whilst in the majority of metaphase spreads these two oncogenes were on clearly separate ecDNAs, in approximately 10% of metaphase spreads we noted colocalization of *CDK4* and *PDGFRA* hybridization signals indicating a subset of ecDNA harbouring both oncogenes (***Figure 3—figure supplement 2A, B***). This colocalization could be observed in a similar proportion of interphase nuclei (***Figure 3E and F***). However, as observed in E25 cells the relative frequency of *CDK4-CDK4*, *PDGFRA-PDGFRA,* and *CDK4-PDGFRA* distances of ≤200 nm was low in the nucleus of E20 cells (***Figure 3H***). Ripley's K function analysis of hybridization signals in most E20 nuclei (22/24) showed no evidence for significant clustering of *CDK4* or *PDGFRA* at <300 nm (***Figure 3I***). We noted 2/24 (8.3%) of interphase nuclei (e.g. ***Figure 3F***, see inset) where Ripley's K function indicated clustering of *CDK4* and *PDGFRA* foci at 100–200 nm and we suggest that these represent cells, as seen at metaphase, where the two oncogenes are located on the same ecDNA molecule. Doublets of *CDK4* foci (200 nm) were detected in 4/24 (16.7%) nuclei (***Figure 3G***, see inset).

Our analysis of two independent GBM cell lines harbouring different ecDNA populations (*CDK4* and *PDGFRA*) provides no evidence for systematic clustering of ecDNA molecules in the nucleus at distances <200 nm.

ecDNA do not colocalize with large RNA PolII hubs in GBM stem cells

DNA FISH detects all ecDNA, so it might be that only transcriptionally active elements cluster. Therefore, we used RNA FISH to detect nascent *EGFR* transcripts in the nuclei of GBM cells. As expected, nascent RNA FISH foci were more frequent in the *EGFR* ecDNA-harbouring cell lines than in NSCs and were more frequent in the E26 GBM cell line than in E28 (***Figure 4—figure supplement 1A and B***). As for DNA FISH, we found no evidence of clustering of sites of *EGFR* nascent transcription at <400 nm in

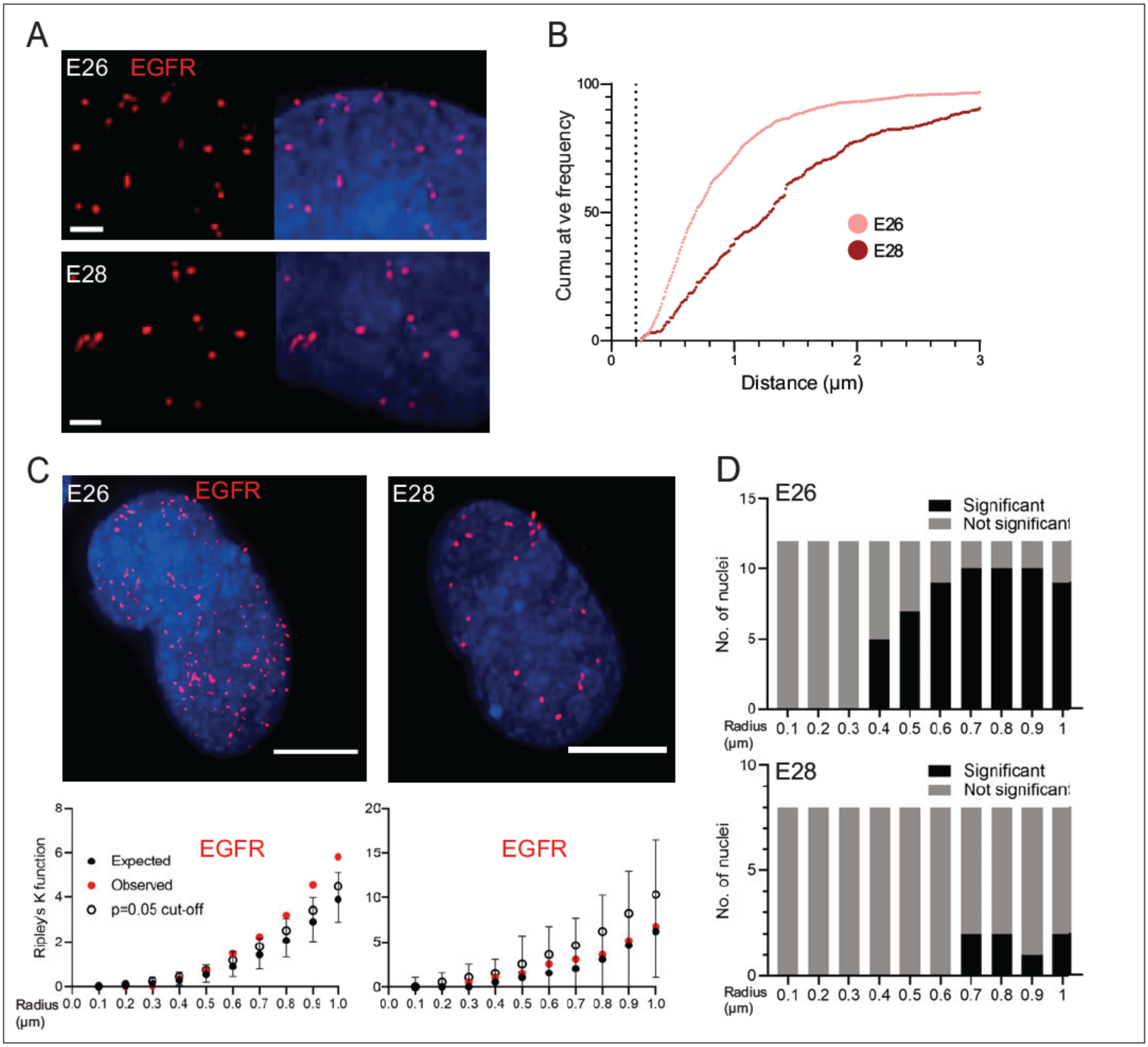


Figure 2. EGFR-containing extrachromosomal DNA (ecDNA) do not cluster in the nucleus. (**A**) Representative images shown as maximum intensity projection of DNA FISH for *EGFR* (red) in the nuclei of E26 (top) and E28 (bottom) glioblastoma (GBM) cell lines, scale bar = 1 µm. (**B**) Cumulative frequency distribution of shortest *EGFR-EGFR* distances between all foci in each nucleus across all E26 (n=37) and E28 (n=36) nuclei. Dotted line = 200nm. (**C**) (Top) Representative maximum intensity projection images of *EGFR* DNA FISH (red) in nuclei of E26 and E28 cells (blue=DNA). Scale bar = 5 µm. (Bottom) Associated 3D Ripley's K function for these nuclei showing observed K function (red), max/min/median (black) of 10,000 null samples with p=0.05 significance cut-off shown (empty black circle). (**D**) Ripley's K function for *EGFR* DNA FISH signals showing number of E26 (n=12) and E28 (n=8) nuclei with significant and non-significant clustering at each given radius. p-values were calculated using Neyman-Pearson lemma with optimistic estimate p-value where required (see Materials and methods), and Benjamini-Hochberg procedure (BHP, FDR = 0.05).

The online version of this article includes the following source data and figure supplement(s) for figure 2:

Source data 1. Statistical data for ***Figure 2—figure supplement 1***.

Figure supplement 1. Additional analysis of *EGFR-EGFR* distances in E26 and E28 cell lines.

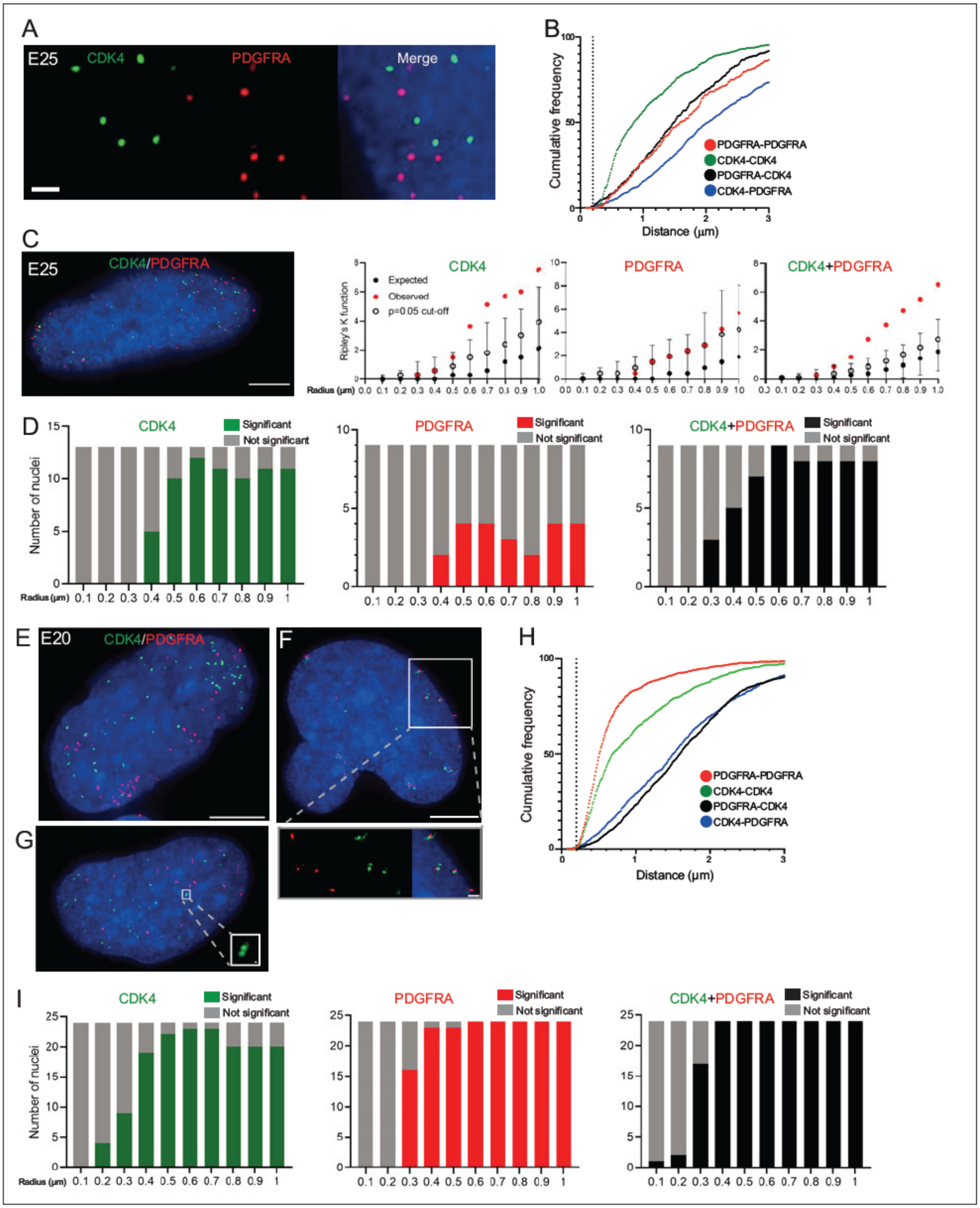


Figure 3. Two separate extrachromosomal DNA (ecDNA) populations do not cluster in the nucleus. (**A**) Representative maximum intensity projection images of DNA FISH for *CDK4* (green) and *PDGFRA* (red) in an E25 nucleus. Blue=DNA (DAPI) . Scale bar = 1 µm. (**B**) Cumulative frequency distribution of shortest interprobe distances (*CDK-CDK, PDGFRA-PDGFRA, CDK4-PDGFRA*, and *PDGFRA-CDK4*) between all foci in each nucleus across all E25 nuclei (n=26). (**C**) (Left) Representative maximum intensity projection image shown of E25 nucleus hybridized with probes for *CDK4* (green) and

Figure 3 continued on next page

Figure 3 continued

PDGFRA (red). Blue=DNA (DAPI). Scale bar = 5 µm. (Right) Ripley's K function for this nucleus showing observed K function (red), max/min/median (black) of 10,000 null samples with p=0.05 significance cut-off shown (empty black circle) for CDK4, *PDGFRA*, and *CDK4* and *PDGFRA* spots combined. (**D**) Ripley's K function for E25 nuclei showing number of nuclei with significant and non-significant clustering at each given radius for *CDK4* spots (n=13 nuclei), *PDGFRA* spots (n=9 nuclei), and *CDK4* and *PDGFRA* spots combined (n=9 nuclei). p-values were calculated using Neyman-Pearson lemma with optimistic estimate p-value where required (see Materials and methods), and Benjamini-Hochberg procedure (BHP, FDR = 0.05). Metaphase analysis of E25 cells and Ripley's K analysis with smaller foci are in ***Figure 3—figure supplement 1***. (**E**) Representative maximum intensity projection image of E20 interphase nuclei hybridized with probes for *CDK4* (green) and *PDGFRA* (red). Scale bar = 5 µm. (**F**) As in (**E**) but for a nucleus where the close association of CDK4 and PDGFA signal in doublets is indicative of ecDNAs harbouring both oncogenes. Scale bar = 1 µm in main panel (**G**) as in (**E**) but showing an E20 nucleus with doublets of CDK4 foci. Metaphase analysis of E20 cells with *CDK4* and *PDGFRA* probes in ***Figure 3—figure supplement 2***. (**H**) As in (**B**) but for E20 nuclei (n=24) (noting all nuclei shown here harbored >20 foci of each oncogene). (**I**) As in (**D**) but for E20 nuclei.

The online version of this article includes the following source data and figure supplement(s) for figure 3:

Source data 1. Statistical data for ***Figure 3—figure supplement 1***.

Figure supplement 1. Additional analysis of the distribution of *CDK4* and *PDGFRA* ecDNAs in the E25 cell line.

Figure supplement 2. DNA FISH on metaphase spreads of the E20 cell line showing hybridization signal for *PDGFRA* (red) and *CDK4* (green).

E26 cells (***Figure 4A and B***). These data suggest that ecDNA actively transcribing a driver oncogene do not colocalize in the nucleus of GBM cells more than expected by chance.

We next assessed whether ecDNA foci, albeit not clustered with each other, colocalize with high focal concentrations of the transcriptional machinery to create ecDNA/large PolII transcription hubs. First, we examined the presence of such hubs by immunofluorescence for RPB1 (POLR2A), the largest subunit of RNA PolII. The large RPB1 foci we detected were sparse with only a few clearly visible per nucleus (***Figure 4—figure supplement 1C***).

We used 3D analysis of immunoFISH in NSCs and compared this to E26 and E28 GBM cells to establish whether ecDNA and large RPB1 foci colocalized. There was no obvious overlap between foci of RPB1 and *EGFR* (***Figure 4C***) and no correlation between the number of large RPB1 foci and the number of *EGFR* foci (***Figure 4D***). Indeed, the mean shortest distance between *EGFR* foci and large RPB1 foci per nucleus was routinely >1 µm in all cell lines, despite the greater number of *EGFR* foci in the GBM cell lines (***Figure 4E***). The single shortest distance per nucleus between an *EGFR* locus and a large RPB1 locus was not significantly different across NSC and tumour lines (***Figure 4F***). There were no instances where the distance between *EGFR* and large RPB1 foci was <200 nm. To test if this was also the case for the nascent *EGFR* RNA transcript, we repeated this analysis using nascent RNA FISH, with the same result (***Figure 4—figure supplement 1D–F***). As the distance distributions to large RPB1 foci were similar for DNA and RNA FISH, this suggests that proximity to large PolII hubs does not alter the probability that ecDNA are transcribed.

To ensure this result was not specific to this PolII antibody, we repeated this analysis using E28 cells in which mCherry was fused by knock-in to endogenous POLR2G, a key subunit of RNA PolII (***Cramer et al., 2000***). The mean distance between *EGFR* foci and large POLR2G foci and the shortest minimum distance in any given nucleus (***Figure 4—figure supplement 1G–I***) further support that there is no close spatial relationship apparent between ecDNA and large PolII hubs.

Levels of EGFR transcription from ecDNA reflect copy number, not enhanced transcriptional efficiency

Having shown a lack of colocalization of ecDNA, either with each other or with large PolII foci, we proceeded to characterize the levels of *EGFR* expression from ecDNA. Flow cytometry using a fluorophore-conjugated EGFR ligand (EGF-647) revealed consistently higher levels of EGFR in the GBM cells than NSC, with highest signal in E26 (***Figure 5—figure supplement 1A, B***), consistent with their higher ecDNA copy number compared with E28 (***Figure 1C***). To confirm this link between ecDNA number and levels of EGFR, E26 and E28 cells were sorted by fluorescence activated cell sorting (FACS) into EGFR-high and EGFR-low populations. In both tumour cell lines, DNA FISH demonstrated that EGFR-high cells had a significantly higher number of *EGFR* DNA foci than EGFR-low (***Figure 5—figure supplement 1C–E***).

Previous studies have reported that ecDNA have greater transcript production per oncogene than chromosomal loci (***Wu et al., 2019***). We therefore sought to characterize the transcriptional efficiency (per copy number) of chromosomal and ecDNA-located *EGFR* genes in our GBM cell lines, by

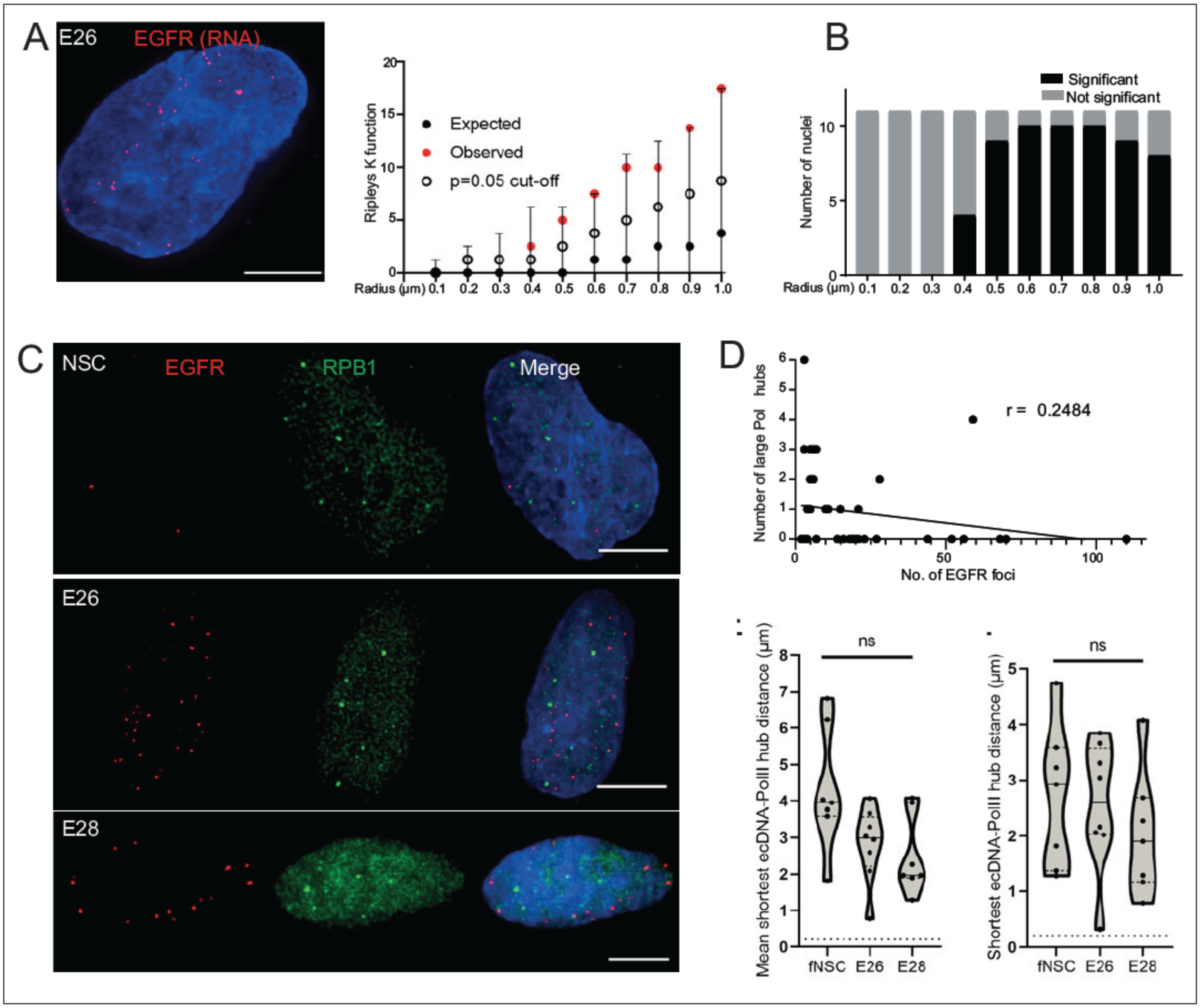


Figure 4. Extrachromosomal DNA (ecDNA) do not colocalize with large foci of the transcriptional machinery. (**A**) Representative maximum intensity projection image of nascent *EGFR* RNA FISH (red) in E26 cell nucleus,(blue=DNA). Scale bar = 5 µm. Associated Ripley's K function for this nucleus showing observed K function (red), max/min/median (black) of 10,000 null samples with p=0.05 significance cut-off shown (empty black circle). (**B**) Ripley's K function for E26 nuclei (n=11) after *EGFR* nascent RNA FISH showing number of nuclei with significant and non-significant clustering at each given radius. All p-values for Ripley's K function calculated using Neyman-Pearson lemma with optimistic estimate p-value where required, and Benjamini-Hochberg procedure (BHP, FDR = 0.05). (**C**) Representative maximum intensity projection images of immunoFISH in neural stem cell (NSC), E26 and E28 cell lines: Immunofluorescence for RPB1 (green) and *EGFR* DNA FISH (red). Scale bar = 5 µm. (**D**) Spearman's correlation between number of *EGFR* foci and number of RPB1 foci, p = 0.13, E26 and E28 cell line data combined. (**E**) Violin plot of distribution of mean shortest interprobe distance per nucleus between *EGFR* foci and PolII foci in NSC (n=7), E26 (n=8) and E28 (n=7) cell lines. (**F**) As for (**E**) but for shortest single distance in each nucleus. ns, not significant. Kruskall-Wallis test. Statistical data relevant for this figure are in ***Figure 4—source data 1***.

The online version of this article includes the following source data and figure supplement(s) for figure 4:

Source data 1. Statistical data for ***Figure 4*** and ***Figure 4—figure supplement 1***.

Figure supplement 1. Analysis of sites of EGFR nascent transcription relative to RNA polymerase II in GBM cell lines.

assaying the RNA:DNA *EGFR* FISH foci ratio. We performed nascent *EGFR* RNA FISH using a probe targeting the first intron of *EGFR* and *EGFR* DNA FISH to test this hypothesis (***Figure 5A***).

When comparing the RNA:DNA ratio of all nuclei, only E26 had a higher ratio than NSCs (***Figure 5B***). To explore whether *EGFR* transcription in these cell lines could be due to ec*EGFR*-driven increased transcriptional efficiency, we used chromosome 7 copy number (evaluated by CEN7 probe) to account for chromosomal *EGFR* copy number. We correlated the RNA:DNA FISH ratio with the proportion of *ecEGFR* (number of *EGFR* foci minus number of CEN7 foci, divided by the total number of EGFR foci). We observed no correlation in either cell line (***Figure 5C***), suggesting that *EGFR* transcription from ecDNA and chromosomes occurs at similar levels when normalized to chromosome 7 copy number. There is no increased transcriptional efficiency from ecDNA compared to chromosomal DNA based on these analyses.

To test this using an independent method, we took advantage of WGS and RNA-seq data (***Figure 5D***) and called SNPs present in the amplicon region at 40% to 60% allele frequencies in patient control blood WGS (control) samples. Most of the allele frequencies of these SNPs were >80% in GBM samples in the main part of the amplicons, in line with the amplification being derived from one parental allele (***Figure 5—figure supplement 1F***). We then selected those SNPs located in expressed exons of the amplicon, including several in *EGFR*. The WGS allele frequencies of these were all >88%, that is, predominantly from amplicons. If genes on the ecDNA are more highly transcribed than chromosomal counterparts, we expect the ratio of RNA-seq to WGS reads of the amplicon-derived SNP to be above 1. Consistent with genes on ecDNA and on chromosomes being transcribed with similar efficiencies, these values were very close to 1, the highest being 1.05 (***Figure 5D, E***). The lower values for *LANCL2*, 3′ of *EGFR*, are likely because only part of this gene is present on the amplicon such that the transcript is truncated. As an additional approach, we utilized the large exon 2–7 deletion present on E26 *EGFR* ecDNA to compare the copy number-normalized RNA expression of exons present only on the endogenous chromosomal *EGFR* locus (exons 2–7) with those predominantly on ecDNA (exons 1, 8–28) (***Figure 5E, D***). Copy number normalized *EGFR* RNA counts were not significantly different between exons 2–7 and those located predominantly on ecDNA (***Figure 5F***). EcDNA with *EGFR* in another established GBM cell line, GBM39, also contain a deletion spanning exons 2–7. We therefore repeated this analysis using previously published WGS and RNA-seq data from this cell line (***Wu et al., 2019***). The normalized RNA read count of primarily *ecEGFR* exons was not significantly different than that of chromosomal *EGFR* exons (***Figure 5G***). Altogether, RNA:DNA FISH and sequencing analyses suggest that *EGFR* on each ecDNA is transcribed at a similar level to that of the corresponding endogenous chromosomal *EGFR* locus. Increased output of oncogenes in GBM stem cells with ecDNA appears to be primarily driven by increased copy number, rather than inherent features of their chromatin state, transcriptional control, or spatial localization.

Discussion

Understanding the importance of ecDNA in the etiology of cancer, and whether this poses an interesting target for therapeutic interventions, depends on deeper analysis of ecDNA activity (***Nathanson et al., 2014***; ***Kim et al., 2020***). Clustering of ecDNA into ‘ecDNA hubs’ based on imaging and chromosome conformation capture data has been reported in a range of established cancer cell lines, and has been suggested to underlie the ability of ecDNA to drive very high levels of transcription (***Hung et al., 2021***; ***Yi et al., 2021***; ***Zhu et al., 2021***). However, in multiple primary human GBM cells studied here, we observe no significant colocalization at distances (~200 nm) thought to be functionally important in driving transcription. We reach this conclusion for both cells with single ecDNA species, as well as with heterogeneous ecDNA harbouring different oncogenes. EcDNA were not colocalized with, or notably close to, large PolII foci. Moreover, taking advantage of the unique transcripts from ecDNA, and the presence of SNPs in these transcripts, to compare ecDNA-derived and chromosomal transcripts, we demonstrate that increased copy number primarily drives increased transcription of ecDNA-located genes rather than increased transcriptional efficiency of ecDNA in GBM stem cells.

Our data support a regional, rather than clustered, spatial organization of ecDNA in GBM stem cells. We observe that oncogenes on ecDNA are distributed more towards the centre of the nucleus than the corresponding endogenous gene loci. This is consistent with an actively transcribing state (***Boyle et al., 2001***; ***Croft et al., 1999***) and independence from the constraints of chromosome territories (***Kalhor et al., 2011***; ***Mahy et al., 2002***).

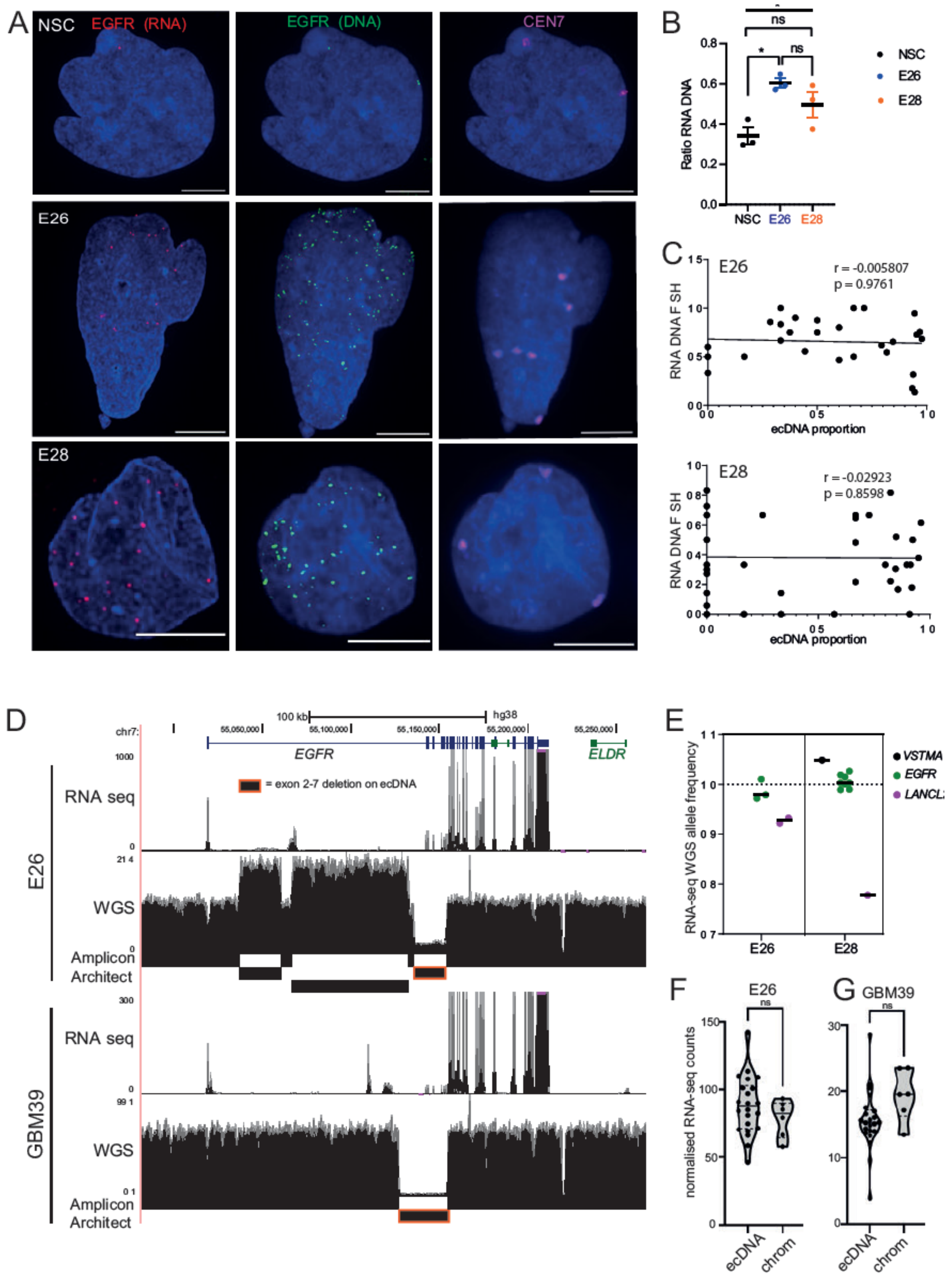


Figure 5. Levels of transcription from extrachromosomal DNA (ecDNA) reflect copy number but not enhanced transcriptional efficiency. (**A**) Representative maximum intensity projection (MIP) images of nascent *EGFR* RNA, *EGFR,* and centromere 7 (CEN7) DNA FISH in neural stem cell (NSC), E26 and E28 cell lines (scale bar = 5 μm). (**B**) Ratio of RNA:DNA foci per nucleus in NSC, E26 and E28 cell lines. * $p<0.05$, n.s. not significant. Flat line – one-way ANOVA, hooked lines – unpaired t-test. Mean and standard error of the mean (SEM) plotted, with 3 biological replicates for NSC (total

Figure 5 continued on next page

Figure 5 continued

n=67), E26 (98) and E28 (95) nuclei. (**C**) Representative Spearman r correlation (ρ) and p-values shown for E26 (n=29) and E28 (n=39) cells. RNA:DNA ratio = number of RNA foci/number of DNA foci. EcDNA proportion = (number of *EGFR* DNA foci – number of CEN7 foci)/number of *EGFR* DNA foci. Three biological replicates performed, data from replicate 1 shown here. (**D**) UCSC genome browser tracks showing E26 and GBM39 RNA-seq and WGS aligned sequences in the region of chromosome 7 where *EGFR* is located, *EGFR* exons (GENCODE) and the exon deletion predicted by AmpliconArchitect. Note that RNA-seq counts in some ecDNA regions go above the maximum value. Genome coordinates (Mb) are from the hg38 assembly of the human genome. (**E**) RNA-seq/WGS allele frequency ratio for SNPs overlapping with expressed exons in the amplicon. Lines denote median values. (**F**) *EGFR* RNA-seq counts normalized by WGS read count per *EGFR* exon in E26, with exons defined as extrachromosomal (exons 1,8-28) or chromosomal (exons 2-7). Statistical significance examined by Mann-Whitney test. ns, not significant. (**G**) As for (**F**) but for GBM39. Statistical data relevant for this figure are in ***Figure 5—source data 1***.

The online version of this article includes the following source data and figure supplement(s) for figure 5:

Source data 1. Statistical data for ***Figure 5*** and ***Figure 5—figure supplement 1***.

Figure supplement 1. EGFR levels, ecDNA number, and ecDNA SNP allele frequency in E26 and E28 cell lines.

We sought to maximize our opportunity of observing ecDNA clustering at close distances by performing 3D spot analysis, using Ripley's K to call instances of significant clustering at given distances using ecDNA x,y,z coordinates, and utilizing cells with two distinct ecDNA species to ensure we were not under-scoring colocalization. 3D analysis ensures a false positive clustering effect is avoided that might be seen when 3D images are combined via tools such as maximum intensity projection (MIP). Other tools to assess clustering have noted the possibility of the 2D Ripley's K function resulting in over-counting, leading to the development of alternative auto-correlation tools, but this was not observed in this 3D Ripley's K analysis (***Veatch et al., 2012***). It is possible that multiple clustered DNA/RNA foci appear as a single DNA/RNA FISH signal that we cannot resolve. We controlled for this by repeating cluster analysis with smaller spot sizes, analyzing cell lines with two ecDNA populations and using super-resolution imaging (optical resolution ~120 nm). We did observe ecDNA clustering at close distances (≤200 nm) in a small proportion of E20 dual-ecDNA cells, but in the case of *CDK4-PDGFRA* colocalization this was at a similar proportion to that observed in metaphase spreads, indicative of ecDNA molecules harbouring both *CDK* and *PDGFRA*. The incidence of *CDK4* doublets (which appeared in keeping with double minutes) was also low. Overall, this suggests that close clustering is not a major contributor to increased ecDNA transcriptional output in GBM stem cells.

Our findings may reflect fundamentally different functional characteristics of the ecDNA in patient-derived primary GBM cell cultures used in our experiments versus previously published studies (***Hung et al., 2021***; ***Yi et al., 2021***). These might include the size of the ecDNA, or the number of oncogene loci per ecDNA (which was singular in our cell lines, with the exception of ~10% E20 *CDK4/PDGFRA* colocalized ecDNA). For example, the COLO320-DM cell line, used in a recent study of ecDNA hubs, harbours 3 copies of *MYC* on each of its ecDNA, and results in large (4.328 Mb, approx. 1.75 µm diameter) ecDNA (***Hung et al., 2021***; ***Wu et al., 2019***). The HK359 GBM cell line, previously noted to have clustered ecDNA hubs, has a 42 kb insertion at the site of *EGFR*vIII (exon 2–7 deletion), again suggesting a large ecDNA quite different in character to those described here (***Hung et al., 2021***; ***Koga et al., 2018***). More quantitative analysis across a larger set of primary cancer cells will be needed to determine if long-term established cell lines have unusual ecDNA features and are unrepresentative of primary GBM cells.

Recent work proposing that ecDNA act as mobile super-enhancers for chromosomal targets has raised the possibility that ecDNA can actively recruit RNA PolII to drive 'ecDNA-associated phase separation' (***Zhu et al., 2021***). A live-cell ecDNA-labelling strategy reported colocalization of ecDNA and RNA PolII (***Yi et al., 2021***). We did not detect evidence of a close relationship between ecDNA, or their nascent transcript, with large PolII foci, but cannot exclude that there are smaller, sub-diffraction limit sized transcriptional hubs associated with our ecDNA.

We observe that while the copy number of *EGFR* ecDNAs positively correlates with greater transcriptional output, this is likely due to copy number increases, rather than increased transcriptional activity on individual ecDNA. It has been proposed that ecDNA increase transcription of their resident oncogenes partly due to their increased DNA copy number, but also due to their more accessible chromatin structure, and that gene transcription from circular amplicons is greater than that of linear amplicons once copy number normalized (***Kim et al., 2020***; ***Wu et al., 2019***). An analysis of RNA-seq and WGS data from a cohort of 36 independent clinical samples found that only 3 out of

11 ecDNA-encoded genes produced significantly more transcripts when normalized to gene copy number, only one of which is a key oncogene (***Wu et al., 2019***). In agreement with this, our analysis of both oncogene and amplicon-resident polymorphisms suggests that copy number is the dominant driver of ecDNA gene transcription.

Overall, our data suggest that in primary GBM stem cells, ecDNA can succeed at driving oncogene expression without requiring close colocalization with each other, or with transcriptional hubs. It is the increased copy number that is primarily responsible for higher levels, rather than ecDNA-intrinsic features or nuclear sub-localization.

Materials and methods

Key resources table

Reagent type (species) or resource	Designation	Source or reference	Identifiers	Additional information
Antibody	mCherry (Rabbit poly-clonal)	abcam	ab167453	IF (1 in 500)
Antibody	Rpb1 NTD (D8L4Y) (Rabbit mono-clonal)	Cell Signaling Technology	#14958	IF (1 in 1000)
Antibody	Anti-Digoxigenin (Sheep poly-clonal)	Roche	Ref 11333089001	DNA FISH (1 in 10)
Antibody	Secondary Antibody – Alexa Fluor 647 (Donkey anti-Sheep IgG poly-clonal)	Thermo Fisher Scientific	A-21448	DNA FISH (1 in 10)
Antibody	Secondary Antibody – Alexa Fluor 568 (Donkey anti-Rabbit IgG poly-clonal)	Thermo Fisher Scientific	A-10042	IF (1 in 1000)
Antibody	Secondary Antibody – Alexa Fluor 488 (Donkey anti-Rabbit IgG poly-clonal)	Thermo Fisher Scientific	A-21206	IF (1 in 1000)
Antibody	Secondary Antibody – Alexa Fluor 488 (Donkey anti-Rat IgG poly-clonal)	Thermo Fisher Scientific	A-21208	IF (1 in 1000)
Genetic reagent (human)	Fosmid FISH probe (Human)	BACPAC resource	https://bacpacresources.org/library.php?id=275	See Materials and methods - ***Supplementary file 1***
Cell line (*Homo sapiens*)	E20, E25, E26, E28, NSC – GCGR Human Glioma Stem Cells	This paper, Glioma Cellular Genetics Resource, CRUK, UK	http://gcgr.org.uk; pending publication	
Other	DMEM/HAMS-F12	Sigma-Aldrich	Cat#: D8437	Cell culture, media
Chemical compound, drug	Pen/Strep	GIBCO	Cat#: 15140–122	Cell culture, media supplement
Other	BSA Solution	GIBCO	Cat#: 15260–037	Cell culture, media supplement
Other	B27 Supplement (×50)	LifeTech/GIBCO	Cat#: 17504–044	Cell culture, media supplement
Other	N2 Supplement (×100)	LifeTech/GIBCO	Cat#: 17502–048	Cell culture, media supplement
Other	Laminin	Cultrex	Cat#: 3446-005-01	Cell culture, media supplement, and pre-lamination of culture vessels
Peptide, recombinant protein	EGF	Peprotech	Cat: 315–09	Cell culture, media supplement
Peptide, recombinant protein	FGF-2	Peprotech	100-18B	Cell culture, media supplement
Other	Accutase	Sigma-Aldrich	Cat#: A6964	Cell culture, cell dissociation agent
Other	DMSO	Sigma-Aldrich	Cat#: 276855	Cell culture, freeze media, and drug diluent

Continued on next page

Continued

Reagent type (species) or resource	Designation	Source or reference	Identifiers	Additional information
Other	Triton X-100	Merck Life Sciences	Cat#: X-100	Cell permeabiliz-ation agent following cell fixation
Other	Paraformaldehyde Powder 95%	Sigma	Cat#: 158127	Cell fixation agent
Other	Tween 20	Cambridge Bioscience	Cat#: TW0020	DNA FISH (hybridization mix)
Other	PBS Tablets	Sigma-Aldrich	Cat#: P4417	Diluent and washing agent
Other	Ethanol	VWR	Cat#: 20821–330	DNA FISH
Other	Methanol	Fisher Chemical	M/4000/17	Used 3:1 with acetic acid for metaphase spreads
Other	Acetic acid	Honeywell Research Chemicals	33209-1L	See above
Peptide, recombinant protein	Alexa Fluor 647 EGF complex	Thermo Fisher Scientific	E35351	Flow cytometry
Other	Green496-dUTP	ENZO Life Sciences	ENZ-42831L	Direct labelling of Fosmid DNA FISH probes via nick translation
Other	ChromaTide Alexa Fluor 594–5-dUTP	Thermo Fisher Scientific	C11400	Direct labelling of Fosmid DNA FISH probes via nick translation
Peptide, recombinant protein	DNA Polymerase 1	Invitrogen	18010–017	
Peptide, recombinant protein	DNase I recombinant, RNase-free	Roche	04716728001	
Genetic reagent (human)	Human Cot-1 DNA	Thermo Fisher Scientific	15279011	
Genetic reagent (salmon)	Salmon Sperm DNA	Invitrogen	15632011	
Chemical compound, drug	Paclitaxel	Cambridge Bioscience	CAY10461	10–100 nM
Chemical compound, drug	Nocodazole	Sigma-Aldrich	SML1665	50–100 ng/ml
Other	XCP 7 Orange Chromosome Paint	MetaSystems Probes	D-0307-100-OR	DNA FISH (see ***Figure 1*** and Materials and methods referring to this)
Commercial assay or kit	Stellaris RNA-FISH probes (Custom Assay with Quasar 570 Dye)	LGC Biosearch Technologies	SMF-1063–5	RNA FISH
Commercial assay or kit	Stellaris RNA FISH Hybridization Buffer	LGC Biosearch Technologies	SMF-HB1-10	RNA FISH
Genetic reagent (human)	Alt-R CRISPR-Cas9 crRNA	IDT-Technologies	Alt-R CRISPR-Cas9 crRNA	
Genetic reagent (human)	Alt-R CRISPR-Cas9 tracrRNA	IDT-Technologies	1072532	
Commercial assay or kit	SG Cell Line 4D-NucleofectorTM X Kit S	Lonza Bioscience	V4XC-3032	
Genetic reagent (human)	Chromosome 7 Control Probe	Pisces Scientific	CHR07-10-DIG	Probe and hybridization mix
Other	DAPI (4',6-Diamidino-2-Phenylindole, Dihydrochloride)	Thermo Fisher Scientific	D1306	Nuclear staining; 50 ng/ml and 5 ng/ml (as indicated in Materials and methods)

Continued on next page

Continued

Reagent type (species) or resource	Designation	Source or reference	Identifiers	Additional information
Sequence-based reagent	mCherry_PolR2G crRNA and dsDNA (donor)	Twist Bioscience		See Materials and methods and ***Supplementary file 1***
Other	WGS and RNAseq	This paper Glioma Cellular Genetics Resource, CRUK, UK	GEO: GSE215420 See also: https://gcgr.org.uk	See Materials and methods
Other	Erosion Territories analysis	This paper		Code available at: https://github.com/IGC-Advanced-Imaging-Resource/Purshouse2022_paper
Other	Cluster analysis	This paper		Code available at: https://github.com/SjoerdVBeentjes/ripleyk
Other	RNA-seq/WGS analysis	This paper		Code available at: https://github.com/kpurshouse/ecDNAcluster
Software, algorithm	GraphPad Prism 9.0	GraphPad Software, Inc	https://www.graphpad.com/	
Software, algorithm	FCS Express	FCS Express 7	https://denovosoftware.com/	
Software, algorithm	Fiji/ImageJ	Open Source	https://imagej.net/Fiji	
Software, algorithm	BioRender	BioRender	https://biorender.com/	
Software, algorithm	Python v3.9	Open Source	https://www.python.org	
Software, algorithm	Algorithm - RipleyK package	Python Package Index	https://pypi.org/project/ripleyk/	
Software, algorithm	Imaris x64 v9.4.0	Imaris Microscopy Image Analysis Software	https://imaris.oxinst.com/	
Software, algorithm	UCSC Genome Browser	***Kent et al., 2002***	https://genome.cshlp.org/content/12/6/996	
Software, algorithm	STAR 2.7.1a	***Dobin et al., 2013***	https://github.com/alexdobin/STAR; ***Dobin et al., 2013***	
Software, algorithm	Picard	Broad Institute	https://broadinstitute.github.io/picard/ RRID:SCR_006525, Version 2.23.2	
Software, algorithm	AmpliconArchitect	***Deshpande et al., 2019***	https://github.com/virajbdeshpande/AmpliconArchitect; ***Deshpande et al., 2019*** (with Python v2.7)	
Software, algorithm	AmpliconClassifier	***Kim et al., 2020***	https://github.com/jluebeck/AmpliconClassifier (with Python v2.7)	
Software, algorithm	deepTools v3.4	***Ramírez et al., 2016***	https://deeptools.readthedocs.io/en/develop/	
Software, algorithm	HOMER2 4.10	***Heinz et al., 2010***	http://homer.ucsd.edu/homer/	
Software, algorithm	SAMtools v1.10	***Li et al., 2009***	http://www.htslib.org	
Software, algorithm	BEDTools v2.3	***Quinlan and Hall, 2010***	http://code.google.com/p/bedtools	
Software, algorithm	bcftools	***Danecek et al., 2021***	https://doi.org/10.1093/gigascience/giab008	
Software, algorithm	strelka v2.9.10	***Kim et al., 2018***	https://doi.org/10.1038/s41592-018-0051-x	

Lead contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the lead contacts, Wendy Bickmore (wendy.bickmore@ed.ac.uk) and Steven Pollard (steven.

pollard@ed.ac.uk).

Materials availability

This study generated a new CRISPR engineered knock-in reporter cell line – E28 mCherry_POLR2G.

Experimental model and subject details

GSC and NSC lines from the Glioma Cellular Genetics Resource (GCGR) (https://gcgr.org.uk) were cultured in serum-free basal DMEM/F12 medium (Sigma) supplemented with N2 and B27 (Life Technologies), 2 µg/ml laminin (Cultrex), and 10 ng/ml growth factors EGF and FGF-2 (Peprotech) (***Pollard et al., 2009***). Cells were split with Accutase solution (Sigma), and centrifuged approximately weekly as previously reported. All GBM cell lines were derived from treatment-naive patients, and the NSC cell line GCGR-NS9FB_B was derived from 9 week of gestation forebrain. GSC cell lines were selected on the basis of predominantly (E26) or entirely (E28, E25, and E20) harbouring oncogenes on ecDNAs (rather than HSRs) via metaphase spread analysis (see Materials and method below). Human GBM tissue was obtained with informed consent and ethical approval (East of Scotland Research Ethics service, REC reference 15/ES/0094). Human embryonic brain tissue was obtained with informed consent and ethical approval (South East Scotland Research Ethics Committee, REC reference 08/S1101/1). Cell lines were regularly tested for mycoplasma.

Method details

Metaphase spreads and interphase nuclei

Cell lines were optimized to generate metaphase spreads. Briefly, cells at near confluence in a T75 flask were incubated between 4 and 16 hr in the presence of 10–100 nm paclitaxel (Cambridge BioScience) with or without 50–100 ng/ml nocodazole (Sigma-Aldrich). Along with the media, cells dissociated with accutase were centrifuged, washed in PBS, and resuspended in 10 ml potassium chloride (KCl) 0.56%, with sodium citrate dihydrate (0.9%) if required, for 20 min. After further centrifugation, cells were resuspended in methanol:acetic acid 3:1 and dropped onto humidified slides.

For all other fixed cell experiments described below, cells were seeded overnight onto glass cover-slips or poly-L-lysine coated glass slides (Sigma-Aldrich). Cells were fixed with 4% paraformaldehyde (PFA – 10 min) and permeabilized with 0.5% Triton X-100 (15 min) with thorough PBS washes in-between. Where cells were dried (see FISH methods), this only occurred following PFA fixation in order to preserve 3D structures and minimize cell and nuclear flattening.

DNA FISH

A detailed method for DNA FISH has been described elsewhere (***Jubb and Boyle, 2020***). Briefly, DNA stocks of fosmid clones targeting EGFR (WI2-2910M03), CDK4 (WI2-0793J08), and PDGFRA (WI2-2022O22) (***Supplementary file 1***) were prepared via an alkaline lysis miniprep protocol (***Jubb and Boyle, 2020***). Each fosmid DNA probe was labelled via Nick Translation directly to a fluorescent dUTP (Green496-dUTP, ENZO Life Sciences; ChromaTide Alexa Fluor 594-5-dUTP, Thermo Fisher Scientific) and incubated with unlabelled dATP, dCTP, and dGTP, ice-cold DNase and DNA PolI for 90 min at 16°C. The reaction was quenched with EDTA and 20% SDS, TE buffer added, and the reaction mix run through a Quick Spin Sephadex G50 column.

Cells on slides or cover-slips were prepared by incubating for 1 hr in ×2 trisodium citrate and sodium chloride (SSC)/RNaseA 100 µg/ml at 37°C, then dehydrated in 70%, 90%, and 100% ethanol. Slides were warmed at 70°C prior to immersion in a denaturing solution (×2 SSC/70% formamide, pH 7.5) heated to 70°C (methanol:acetic acid-fixed cells) or 80°C (PFA-fixed cells), the duration of which was optimized to each cell line. After denaturing, slides were immersed in ice-cold 70% ethanol, then 90% and 100% ethanol at room temperature before air drying.

FISH probes were prepared by combining 100 ng of each directly labelled fosmid probe (per slide), 6 µg Human Cot-1 DNA (per probe), 5 µg sonicated salmon sperm (per slide), and 100% ethanol. Once completely dried, the resulting pellet was suspended in hybridization mix (50% deionized formamide [DF], ×2 SSC, 10% dextran sulfate, 1% Tween 20) for 1 hr at room temperature, denatured for 5 min at >70°C and annealed at 37°C for 15 min. Where relevant, FISH probes were instead hybridized in Chromosome 7 paint (XCP 7 Orange, Metasystems). The probes were incubated overnight at 37°C. The following day, the slides were washed in ×2 SSC (45°C), 0.1% SSC (60°C) and finally in ×4

SSC/0.1% Tween 20 with 50 ng/ml 4′,6-diamidino-2-phenylindole (DAPI). Slides were mounted with Vectashield.

RNA FISH

RNA FISH probes (Custom Assay with Quasar 570 Dye) targeting the first intron (pool of 48 22-mer probes) of *EGFR* were designed and ordered via the Stellaris probe designer (Biosearch Technologies, Inc, Petaluma, CA) (https://www.biosearchtech.com/support/tools/design-software/stellaris-probe-designer, version 4.2). Cells were seeded, fixed, and permeabilized as above. Slides were immersed in ×2 SSC, 10% DF in DEPC-treated water for 2–5 min before applying the hybridization mix (Stellaris RNA FISH hyb buffer, 10% DF, 125 nm RNA FISH probe) for incubation at 37°C. After overnight incubation, slides were incubated in ×2 SSC, 10% formamide in DEPC-treated water for 30 min, and then stained with DAPI (5 ng/ml). Slides were washed with PBS before mounting with Vectashield.

Combined RNA:DNA FISH

Nascent EGFR RNA FISH was performed as above, and nuclei imaged as described below. The x,y,z coordinates for each image were recorded via NIS software at the time of imaging. After removing the cover-slips and washing the slides in PBS, EGFR DNA FISH was performed whereby the probe preparation was as above. Centromere 7 (CEN7 – CHR07-Dig Control) FISH probe (Pisces Scientific) was prepared, denatured for 5 min at 80°C and snap-frozen on crushed ice. Slides were transferred from PBS wash to denaturing solution at 80°C for 15–30 min, washed in ×2 SSC, and incubated overnight with the probe(s) at 37°C. The subsequent stringency washes were as described above. Slides were then incubated in blocking buffer (×4 SSC/5% Marvel) for 5 min, followed by anti-digoxigenin antibody (Roche; 1 in 10; 1 hr at humidified 37°C) and anti-sheep Alexa Fluor 647 secondary antibody (Thermo Fisher Scientific; 1 in 10; 1 hr at humidified 37°C) with ×4 SSC/0.1% Tween 20 washes in between. After the final washes, slides were stained with DAPI and mounted as described above. The stored x,y,z coordinates were used to relocate and image each nucleus. Owing to the irregularity of the tumour nuclei, it was possible to be confident in re-imaging the correct nucleus – nuclei were excluded where this was not the case, or where nuclei were lost between RNA and DNA FISH. Spot counting was subsequently performed as described below with RNA and DNA foci being defined and counted separately to avoid influencing the outcome. For CEN7, nuclei were excluded if the number of foci could not be clearly identified.

Immunofluorescence and immuno-FISH

Slides were blocked in 1%BSA/PBS/Triton X-100 0.1% for 30 min at 37°C before overnight incubation with the primary antibody at 4°C (Rpb1 NTD (D8L4Y) #14958, Cell Signaling Technology, 1 in 1000; mCherry [ab167453], abcam, 1 in 500). The following day, slides were washed in PBS before incubation with an appropriate secondary antibody (1 in 1000 Alexa Fluor) for 1 hr at 37°C. After further PBS washes and DAPI staining, slides were mounted with Vectashield.

For immuno-FISH (DNA), the IF signal was fixed via incubation with 4% PFA for 30 min. Following thorough PBS washes, the DNA FISH protocol was then followed as above.

For immuno-FISH (RNA), the antibodies were added at the same concentration as described above to the hybridization mix (primary antibody) and ×2 SSC/10% DF washes (secondary antibody).

Flow cytometry and FACS

Cells were prepared by adding EGF-free media for 30 min before lifting and suspending cells in 0.1% BSA/PBS. Cells were incubated in 100 ng/ml EGF-647 (E35351, Thermo Fisher Scientific) in 0.1%BSA/PBS, with cells incubated in 0.1% BSA/PBS as a negative control, for 25 min. Cells were washed three times in 0.1%BSA/PBS before being analysed on the BD FACSAria III FUSION. Where indicated, cells were sorted by EGF-647 gated into high and low groups, and a sort check was performed to verify these were true populations prior to expanding these cells onto 22×22 mm^2 cover-slips. Fifteen days after the cells were sorted, the slides were fixed, permeabilized, and DNA FISH performed as above.

mCherry_POLR2G knock-in cell line

crRNA and donor DNA was designed using the previously reported TAG-IN tool (**Dewari et al., 2018**), with the corresponding fluorescent reporter gene sequences for mCherry implemented into the existing tool (**Supplementary file 1**). Output sequences from the TAG-IN tool were manufactured by Twist Bioscience. Gene-specific crRNA (100 pmoles – IDT Technologies, Coralville, IA, USA) and universal tracrRNA (100 pmoles, IDT Technologies, Coralville, IA, USA) were assembled to a cr:tracrRNA complex by annealing at the following settings on a PCR block: 95°C for 5 min, step down cooling from 95°C to 85°C at 0.5°C/s, step down cooling from 85°C to 20°C at 0.1°C/s, store at 4°C. Recombinant Cas9 protein (10 µg, purified in house – see **Dewari et al., 2018**) was added to form the ribonucleoprotein (RNP) complex at room temperature for 10 min, then stored on ice; 300 ng of donor dsDNA were denatured in 30% DMSO by incubating at 95°C for 5 min followed by immediate immersion in ice. The donor dsDNA and RNPs were electroporated into E28 cells using the 4D Amaxa X Unit (programme DN-100). After 2 weeks of serial expansion of cells in 2D culture, assessment of knock-in efficiency was assessed by suspending 5–7 × 10^5 cells in 0.2% BSA/PBS and analysed on BD LSRFortessa Cell Analyzer, with cells electroporated with tracrRNA:Cas9 only as a negative control. Cells were then further sorted into a pure KI population, and mCherry KI was verified by immunofluorescence for mCherry and Rpb1.

Imaging

Slides were imaged on epifluorescence microscopes (Zeiss AxioImager 2 and Zeiss AxioImager.A1) and the SoRa spinning disk confocal microscope (Nikon CSU-W1 SoRa). For 3D image analysis, images were taken with the SoRa microscope and a 3 µm section across each nucleus was imaged in 0.1 µm steps. Images were denoised and deconvolved using NIS deconvolution software (blind preset or Lucy-Richardson) (Nikon). 3D images are shown in the figures as MIP prepared using ImageJ.

Quantification and statistical analysis

Image analysis of nuclear localization

Images were analysed using Imarisv9.7 and Fiji. The scripts used to perform nuclear territory analysis have been described elsewhere (**Boyle et al., 2001**; **Croft et al., 1999**; see also Data availability). Briefly, single-slice images were taken with a ×20 lens using the Zeiss AxioImager 2, imaging at least 50 nuclei per cell line. The images were segmented first to individual nuclei, and subsequently the area of the DAPI signal was segmented to define the nuclear area. This area was segmented into concentric shells of equal area from the periphery to the centre of each nucleus. The signal intensity of each FISH probe or chromosome paint signal was calculated, with normalization for the DAPI signal in each shell.

Image analysis of ecDNA and large PolII foci

For 3D analysis, deconvolved images were analysed using Imaris (v9.7) and all analysis was performed on the full 3D image. RNA and DNA FISH foci, and where relevant, large PolII foci, were defined, counted and distances between them calculated, using the Spots function within Imaris. Imaris spot size diameter was selected by single plane measurement of representative foci and this defined diameter was applied to all nuclei of a given experiment for 3D analysis. For DNA FISH analysis, E26, E28, and E25 spot size was 300 nm diameter, and where indicated in the text, reanalysed with 150 nm spot diameter. For E20 and all RNA FISH experiments, a spot size diameter of 200 nm was used. For RPB1 and POLR2G foci (IF), large foci were defined as those ≥500 nm diameter (**Cho et al., 2018**; **Sabari et al., 2018**).

For 3D cluster analysis of FISH spots, Ripley's K function was performed using the x,y,z coordinates for each FISH spot using the Imaris Spots function to determine observed and null distribution values.

$$K(r) = \frac{1}{\lambda}\sum_{i\neq j}\frac{\mathbb{1}\left\{d(i,j)\leq r\right\}}{n}$$

Ripley's K function compares the number of points at a distance smaller than a given radius r, relative to the average number of points in the volume. This average is the density lambda, in this case the number of foci, n, divided by the volume. In the above equation, is the indicator function which equals 1 if the distance between points i and j is no larger than r, and 0 otherwise. A high value of Ripley's K function represents clustering at the given radius r, whereas a low value represents dispersion. Consequently, a high Ripley's K function at a given radius is indicative of clustering at this radius. By comparing the observed value of Ripley's K function at a given radius with that computed on the same number of foci and with the same volume but drawn from a uniform null distribution, the presence of significant clustering in the given cluster at the given radius can be detected.

The code written to perform this analysis was formed using a script written in Python (v3.9) and has been made available on GitHub (see Data availability). Ripley's K function was determined across a radius of 0.1–1 µm in 0.1 µm increments. After calculating the observed Ripley's K function value, a null distribution of no clustering, estimated on uniformly distributed samples with the same number of spots, was generated using the coordinates for each given nucleus to calculate 10,000 Ripley's K function values at each radial increment. We tested a sample of nuclei with 50,000 values and confirmed that 10,000 values would provide sufficient accuracy. Having sampled that nucleus shape and size did not affect the significance of a result at each increment in the given range of radii, a bounding radius of 5 was used for all samples. Only nuclei with greater than 20 *EGFR* foci were included to ensure both that the majority of foci were ec*EGFR,* to allow adequate granularity and minimize the risk of a false negative result due to lack of foci. The p-value for each observed K function was established against the expected values using the Neyman-Pearson lemma. Where the observed and expected K function at $p=0.05$ were the same, a randomized binomial test was performed to determine if $p<0.05$ for the observed value, weighting the probability of success as the ratio of the number of values $p<0.05$ and the total number of equal values. Having determined this, the most optimistic estimate of p-value was made which would favour identification of a significant result, that is, a bias in favour of significant clustering. A Benjamini-Hochberg procedure was performed to control for the false discovery rate (FDR = 0.05).

All other statistical analysis was performed with GraphPad Prism v9.0. The statistical details for each experiment can be found in the relevant figure legends and in the Source Data. For figures, p-values are represented as follows: *<0.05, **<0.01, ***<0.001, ****<0.0001. Where appropriate, Bonferroni correction for multiple hypothesis testing was performed, and, where relevant, corrected p-values are those plotted in the figures and are given in the Source Data in brackets next to the uncorrected p value.

RNA and WGS sequencing sample preparation, analysis, and processing

The preparation of these cell lines for RNA-seq has been described in detail elsewhere (***Gangoso et al., 2021***). WGS was undertaken by BGI Tech Solutions with PE100 and normal library construction. WGS, RNA-seq, and AmpliconArchitect data for GBM39 was taken from data made available via publication and in the NCBI Sequence Read Archive (BioProject: PRJNA506071) (***Wu et al., 2019***).

Sequences were aligned to hg38 with STAR 2.7.1a with settings '--outFilterMultimapNmax 1' used for WGS and RNA-seq data and settings '`--alignMatesGapMax 2000 --alignIntronMax 1 --alignEndsType` EndToEnd' used only for WGS data (***Dobin et al., 2013***). Duplicate reads were removed using Picard (Broad Institute). AmpliconArchitect (***Deshpande et al., 2019***) and AmpliconClassifier (***Kim et al., 2020***) were used to predict the ecDNA regions and classify circular amplicons for E26 and E28, and to classify EGFR exons as being located primarily on ecDNA or only on chromosomal DNA in E26 and E28. Exon coordinates were extracted from Ensembl (isoform:EGFR-201, Ensembl Transcript ID: ENST00000275493.7). Alignments were converted to bigWig files using deepTools bamCoverage with setting '--normalizeUsingRPKM' (***Ramírez et al., 2016***) and visualized using the UCSC genome browser (***Kent et al., 2002***). HOMER2 (***Heinz et al., 2010***) makeTagDirectory and annotatePeaks.pl (settings '-len 0 -size given') were used for read counting of WGS and RNA in EGFR exons. Analysis of RNA-seq counts per copy number was performed using scripts written in Python (v3.9). We normalized the RNA-seq read counts to the WGS read count in each EGFR exon, and analysed in GraphPad Prism v9.0. SNP calling was done using strelka v2.9.10 (***Kim et al., 2018***) using the configureStrelkaGermlineWorkflow.py command on all samples (WGS blood, WGS tumour, and RNA-seq tumour) for each cell line (E26 and E28) separately. SNPs that passed all filters were extracted using bcftools (***Danecek et al., 2021***) and selected for those that had an allele frequency in the WGS blood between 40% and 60%. The ratio of allele frequencies between the RNA-seq and WGS tumour samples were determined for those SNPs overlapping expressed exons with at least 20 reads in the RNA-seq samples . See Data availability.

Source data

Source data regarding the statistical tests applied, the exact sample number, p-values of tests (and adjustments for multiple hypothesis testing), and details of replicates are included where indicated in the article. N=number of nuclei.

Acknowledgements

SVB would like to thank Dr Tim Cannings for helpful suggestions on statistical analysis.

We acknowledge the Advanced Imaging Resource at the Institute of Genetics and Cancer and the Edinburgh Super-Resolution Imaging Consortium (ESRIC), and the Flow Cytometry team at the Centre for Regenerative Medicine, University of Edinburgh, for their technical support. This work has made use of the resources provided by the Edinburgh Compute and Data Facility (ECDF) (http://www.ecdf.ed.ac.uk/). KP was supported by a Wellcome PhD Training Fellowship (220399/Z/20/Z). ETF was supported by the Swiss National Science Foundation (P2ELP3_191695). GMM and the Glioma Cellular Genetics Resource (https://www.gcgr.org.uk/) were supported by the Cancer Research UK (CRUK) Centre Accelerator Award (A21922). AH was supported by a CRUK PhD Fellowship (C157/A29279). SMP is a Cancer Research UK Senior Research Fellow (A17368). Work in the group of WAB is supported by MRC University Unit grant MC_UU_00007/2.

Funding sources were not involved in study design, data collection, data interpretation, or the decision to submit the work for publication.

Additional information

Funding

Funder	Grant reference number	Author
Wellcome Trust	220399/Z/20/Z	Karin Purshouse
Swiss National Science Foundation	P2ELP3_191695	Elias T Friman
Cancer Research UK	DRCNPG-Nov21\100002	Steven M Pollard
Cancer Research UK	C157/A29279	Alhafidz Hamdan
Cancer Research UK	A17368	Karin Purshouse
Medical Research Foundation	MC_UU_00007/2	Wendy A Bickmore

The funders had no role in study design, data collection and interpretation, or the decision to submit the work for publication. For the purpose of Open Access, the authors have applied a CC BY public copyright license to any Author Accepted Manuscript version arising from this submission.

Author contributions

Karin Purshouse, Investigation, Methodology, Writing – original draft, Writing – review and editing; Elias T Friman, Sjoerd V Beentjes, Formal analysis, Methodology, Writing – review and editing; Shelagh Boyle, Pooran Singh Dewari, Vivien Grant, Gillian M Morrison, Investigation; Alhafidz Hamdan, Formal analysis, Investigation; Paul M Brennan, Resources; Steven M Pollard, Wendy A Bickmore, Conceptualization, Supervision, Funding acquisition, Writing – original draft, Project administration, Writing – review and editing

Author ORCIDs

Karin Purshouse http://orcid.org/0000-0003-0942-6342

Alhafidz Hamdan http://orcid.org/0000-0002-0794-5504
Sjoerd V Beentjes http://orcid.org/0000-0002-7998-4262
Steven M Pollard http://orcid.org/0000-0001-6428-0492
Wendy A Bickmore http://orcid.org/0000-0001-6660-7735

Decision letter and Author response

Decision letter https://doi.org/10.7554/eLife.80207.sa1
Author response https://doi.org/10.7554/eLife.80207.sa2

Additional files

Supplementary files

- Supplementary file 1. Genomic information for FISH probes and CRISPR knockin. (A)Fosmid probes for DNA FISH related to STAR methods. Genome coordinates (Mb) are from the hg38 assembly of the human genome. (B) CrRNA sequence and dsDNA sequence for mCherry_PolR2G CRISPR knock-in.
- MDAR checklist

Data availability

WGS and RNAseq data have been deposited on NCBI GEO under study accession number GSE215420 and is publicly available as of the date of publication As indicated in the Key Resources, all original code has been deposited as: https://github.com/IGC-Advanced-Imaging-Resource/Purshouse2022_paper (copy archived at swh:1:rev:5b1a3920afa8e85132c94bcc6dfce94575f939ce) https://github.com/SjoerdVBeentjes/ripleyk (copy archived at swh:1:rev:1303af539403303786b6460fabef355e345ea6c9) https://github.com/kpurshouse/ecDNAcluster (copy archived at swh:1:rev:9162a39f3c8e19e973eaedc50ad4e1d3dc570e90).

The following dataset was generated:

Author(s)	Year	Dataset title	Dataset URL	Database and Identifier
Purshouse et al	2022	WGS and RNA-seq data E26,E28	https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE215420	NCBI Gene Expression Omnibus, GSE215420

References

Adelman K, Martin BJE. 2021. EcDNA Party bus: bringing the enhancer to you. *Molecular Cell* **81**:1866–1867. DOI: https://doi.org/10.1016/j.molcel.2021.04.017, PMID: 33961776

Boyle S, Gilchrist S, Bridger JM, Mahy NL, Ellis JA, Bickmore WA. 2001. The spatial organization of human chromosomes within the nuclei of normal and emerin-mutant cells. *Human Molecular Genetics* **10**:211–219. DOI: https://doi.org/10.1093/hmg/10.3.211, PMID: 11159939

Brennan CW, Verhaak RGW, McKenna A, Campos B, Noushmehr H, Salama SR, Zheng S, Chakravarty D, Sanborn JZ, Berman SH, Beroukhim R, Bernard B, Wu CJ, Genovese G, Shmulevich I, Barnholtz-Sloan J, Zou L, Vegesna R, Shukla SA, Ciriello G, et al. 2013. The somatic genomic landscape of glioblastoma. *Cell* **157**:753. DOI: https://doi.org/10.1016/j.cell.2014.04.004

Bulstrode H, Johnstone E, Marques-Torrejon MA, Ferguson KM, Bressan RB, Blin C, Grant V, Gogolok S, Gangoso E, Gagrica S, Ender C, Fotaki V, Sproul D, Bertone P, Pollard SM. 2017. Elevated FOXG1 and Sox2 in glioblastoma enforces neural stem cell identity through transcriptional control of cell cycle and epigenetic regulators. *Genes & Development* **31**:757–773. DOI: https://doi.org/10.1101/gad.293027.116, PMID: 28465359

Carvalho C, Pereira HM, Ferreira J, Pina C, Mendonça D, Rosa AC, Carmo-Fonseca M. 2001. Chromosomal G-dark bands determine the spatial organization of centromeric heterochromatin in the nucleus. *Molecular Biology of the Cell* **12**:3563–3572. DOI: https://doi.org/10.1091/mbc.12.11.3563, PMID: 11694589

Cho WK, Spille JH, Hecht M, Lee C, Li C, Grube V, Cisse II. 2018. Mediator and RNA polymerase II clusters associate in transcription-dependent condensates. Science **361**:412–415. DOI: https://doi.org/10.1126/science.aar4199, PMID: 29930094

Chong S, Dugast-Darzacq C, Liu Z, Dong P, Dailey GM, Cattoglio C, Heckert A, Banala S, Lavis L, Darzacq X, Tjian R. 2018. Imaging dynamic and selective low-complexity domain interactions that control gene transcription. Science **361**:eaar2555. DOI: https://doi.org/10.1126/science.aar2555, PMID: 29930090

Cox D, Yuncken C, Spriggs AI. 1965. Minute chromatin bodies in malignant tumours of childhood. *Lancet* **1**:55–58. DOI: https://doi.org/10.1016/s0140-6736(65)90131-5, PMID: 14304929

Cramer P, Bushnell DA, Fu J, Gnatt AL, Maier-Davis B, Thompson NE, Burgess RR, Edwards AM, David PR, Kornberg RD. 2000. Architecture of RNA polymerase II and implications for the transcription mechanism. Science **288**:640–649. DOI: https://doi.org/10.1126/science.288.5466.640, PMID: 10784442

Croft JA, Bridger JM, Boyle S, Perry P, Teague P, Bickmore WA. 1999. Differences in the localization and morphology of chromosomes in the human nucleus. *The Journal of Cell Biology* **145**:1119–1131. DOI: https://doi.org/10.1083/jcb.145.6.1119, PMID: 10366586

Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, Whitwham A, Keane T, McCarthy SA, Davies RM, Li H. 2021. Twelve years of samtools and bcftools. *GigaScience* **10**:giab008. DOI: https://doi.org/10.1093/gigascience/giab008, PMID: 33590861

Deshpande V, Luebeck J, Nguyen NPD, Bakhtiari M, Turner KM, Schwab R, Carter H, Mischel PS, Bafna V. 2019. Exploring the landscape of focal amplifications in cancer using ampliconarchitect. *Nature Communications* **10**:392. DOI: https://doi.org/10.1038/s41467-018-08200-y, PMID: 30674876

Dewari PS, Southgate B, Mccarten K, Monogarov G, O'Duibhir E, Quinn N, Tyrer A, Leitner M-C, Plumb C, Kalantzaki M, Blin C, Finch R, Bressan RB, Morrison G, Jacobi AM, Behlke MA, von Kriegsheim A, Tomlinson S, Krijgsveld J, Pollard SM. 2018. An efficient and scalable pipeline for epitope tagging in mammalian stem cells using Cas9 ribonucleoprotein. *eLife* **7**:e35069. DOI: https://doi.org/10.7554/eLife.35069, PMID: 29638216

Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. 2013. STAR: ultrafast universal RNA-seq aligner. *Bioinformatics* **29**:15–21. DOI: https://doi.org/10.1093/bioinformatics/bts635, PMID: 23104886

Fan Y, Mao R, Lv H, Xu J, Yan L, Liu Y, Shi M, Ji G, Yu Y, Bai J, Jin Y, Fu S. 2011. Frequency of double minute chromosomes and combined cytogenetic abnormalities and their characteristics. *Journal of Applied Genetics* **52**:53–59. DOI: https://doi.org/10.1007/s13353-010-0007-z, PMID: 21107781

Gangoso E, Southgate B, Bradley L, Rus S, Galvez-Cancino F, McGivern N, Güç E, Kapourani CA, Byron A, Ferguson KM, Alfazema N, Morrison G, Grant V, Blin C, Sou I, Marques-Torrejon MA, Conde L, Parrinello S, Herrero J, Beck S, et al. 2021. Glioblastomas acquire myeloid-affiliated transcriptional programs via epigenetic immunoediting to elicit immune evasion. *Cell* **184**:2454–2470.. DOI: https://doi.org/10.1016/j.cell.2021.03.023, PMID: 33857425

Gibaud A, Vogt N, Hadj-Hamou NS, Meyniel JP, Hupé P, Debatisse M, Malfoy B. 2010. Extrachromosomal amplification mechanisms in a glioma with amplified sequences from multiple chromosome loci. *Human Molecular Genetics* **19**:1276–1285. DOI: https://doi.org/10.1093/hmg/ddq004, PMID: 20056677

Hamkalo BA, Farnham PJ, Johnston R, Schimke RT. 1985. Ultrastructural features of minute chromosomes in a methotrexate-resistant mouse 3T3 cell line. *PNAS* **82**:1126–1130. DOI: https://doi.org/10.1073/pnas.82.4.1126, PMID: 3856243

Hansen JC, Maeshima K, Hendzel MJ. 2021. The solid and liquid states of chromatin. *Epigenetics & Chromatin* **14**:50. DOI: https://doi.org/10.1186/s13072-021-00424-5, PMID: 34717733

Heinz S, Benner C, Spann N, Bertolino E, Lin YC, Laslo P, Cheng JX, Murre C, Singh H, Glass CK. 2010. Simple combinations of lineage-determining transcription factors prime cis-regulatory elements required for macrophage and B cell identities. *Molecular Cell* **38**:576–589. DOI: https://doi.org/10.1016/j.molcel.2010.05.004, PMID: 20513432

Hung KL, Yost KE, Xie L, Shi Q, Helmsauer K, Luebeck J, Schöpflin R, Lange JT, Chamorro González R, Weiser NE, Chen C, Valieva ME, Wong IT-L, Wu S, Dehkordi SR, Duffy CV, Kraft K, Tang J, Belk JA, Rose JC, et al. 2021. EcDNA hubs drive cooperative intermolecular oncogene expression. *Nature* **600**:731–736. DOI: https://doi.org/10.1038/s41586-021-04116-8, PMID: 34819668

Inda MM, Bonavia R, Mukasa A, Narita Y, Sah DWY, Vandenberg S, Brennan C, Johns TG, Bachoo R, Hadwiger P, Tan P, Depinho RA, Cavenee W, Furnari F. 2010. Tumor heterogeneity is an active process maintained by a mutant EGFR-induced cytokine circuit in glioblastoma. *Genes & Development* **24**:1731–1745. DOI: https://doi.org/10.1101/gad.1890510, PMID: 20713517

Jubb A, Boyle S. 2020. Visualizing genome reorganization using 3D DNA FISH. Nielsen BS, Jones J (Eds). *In Situ Hybridization Protocols*. New York, NY: Springer. p. 85–95. DOI: https://doi.org/10.1007/978-1-0716-0623-0_5

Kalhor R, Tjong H, Jayathilaka N, Alber F, Chen L. 2011. Genome architectures revealed by tethered chromosome conformation capture and population-based modeling. *Nature Biotechnology* **30**:90–98. DOI: https://doi.org/10.1038/nbt.2057, PMID: 22198700

Kent WJ, Sugnet CW, Furey TS, Roskin KM, Pringle TH, Zahler AM, Haussler D. 2002. The human genome browser at UCSC. *Genome Research* **12**:996–1006. DOI: https://doi.org/10.1101/gr.229102, PMID: 12045153

Kim S, Scheffler K, Halpern AL, Bekritsky MA, Noh E, Källberg M, Chen X, Kim Y, Beyter D, Krusche P, Saunders CT. 2018. Strelka2: fast and accurate calling of germline and somatic variants. *Nature Methods* **15**:591–594. DOI: https://doi.org/10.1038/s41592-018-0051-x, PMID: 30013048

Kim H, Nguyen NP, Turner K, Wu S, Gujar AD, Luebeck J, Liu J, Deshpande V, Rajkumar U, Namburi S, Amin SB, Yi E, Menghi F, Schulte JH, Henssen AG, Chang HY, Beck CR, Mischel PS, Bafna V, Verhaak RGW. 2020. Extrachromosomal DNA is associated with oncogene amplification and poor outcome across multiple cancers. *Nature Genetics* **52**:891–897. DOI: https://doi.org/10.1038/s41588-020-0678-2, PMID: 32807987

Koga T, Li B, Figueroa JM, Ren B, Chen CC, Carter BS, Furnari FB. 2018. Mapping of genomic EGFRvIII deletions in glioblastoma: insight into rearrangement mechanisms and biomarker development. *Neuro-Oncology* **20**:1310–1320. DOI: https://doi.org/10.1093/neuonc/noy058, PMID: 29660021

Lange JT, Rose JC, Chen CY, Pichugin Y, Xie L, Tang J, Hung KL, Yost KE, Shi Q, Erb ML, Rajkumar U, Wu S, Taschner-Mandl S, Bernkopf M, Swanton C, Liu Z, Huang W, Chang HY, Bafna V, Henssen AG, et al. 2022. The evolutionary dynamics of extrachromosomal DNA in human cancers. *Nature Genetics* **54**:1527–1533. DOI: https://doi.org/10.1038/s41588-022-01177-x, PMID: 36123406

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, 1000 Genome Project Data Processing Subgroup. 2009. The sequence alignment/map format and samtools. *Bioinformatics* **25**:2078–2079. DOI: https://doi.org/10.1093/bioinformatics/btp352, PMID: 19505943

Mahy NL, Perry PE, Bickmore WA. 2002. Gene density and transcription influence the localization of chromatin outside of chromosome territories detectable by fish. *The Journal of Cell Biology* **159**:753–763. DOI: https://doi.org/10.1083/jcb.200207115, PMID: 12473685

Morton AR, Dogan-Artun N, Faber ZJ, MacLeod G, Bartels CF, Piazza MS, Allan KC, Mack SC, Wang X, Gimple RC, Wu Q, Rubin BP, Shetty S, Angers S, Dirks PB, Sallari RC, Lupien M, Rich JN, Scacheri PC. 2019. Functional enhancers shape extrachromosomal oncogene amplifications. *Cell* **179**:1330–1341.. DOI: https://doi.org/10.1016/j.cell.2019.10.039, PMID: 31761532

Nathanson DA, Gini B, Mottahedeh J, Visnyei K, Koga T, Gomez G, Eskin A, Hwang K, Wang J, Masui K, Paucar A, Yang H, Ohashi M, Zhu S, Wykosky J, Reed R, Nelson SF, Cloughesy TF, James CD, Rao PN, et al. 2014. Targeted therapy resistance mediated by dynamic regulation of extrachromosomal mutant EGFR DNA. Science **343**:72–76. DOI: https://doi.org/10.1126/science.1241328, PMID: 24310612

Pollard SM, Yoshikawa K, Clarke ID, Danovi D, Stricker S, Russell R, Bayani J, Head R, Lee M, Bernstein M, Squire JA, Smith A, Dirks P. 2009. Glioma stem cell lines expanded in adherent culture have tumor-specific phenotypes and are suitable for chemical and genetic screens. *Stem Cell* **4**:568–580. DOI: https://doi.org/10.1016/j.stem.2009.03.014, PMID: 19497285

Quinlan AR, Hall IM. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics* **26**:841–842. DOI: https://doi.org/10.1093/bioinformatics/btq033, PMID: 20110278

Rai AK, Chen JX, Selbach M, Pelkmans L. 2018. Kinase-Controlled phase transition of membraneless organelles in mitosis. *Nature* **559**:211–216. DOI: https://doi.org/10.1038/s41586-018-0279-8, PMID: 29973724

Ramírez F, Ryan DP, Grüning B, Bhardwaj V, Kilpert F, Richter AS, Heyne S, Dündar F, Manke T. 2016. DeepTools2: a next generation web server for deep-sequencing data analysis. *Nucleic Acids Research* **44**:W160–W165. DOI: https://doi.org/10.1093/nar/gkw257, PMID: 27079975

Richards LM, Whitley OKN, MacLeod G, Cavalli FMG, Coutinho FJ, Jaramillo JE, Svergun N, Riverin M, Croucher DC, Kushida M, Yu K, Guilhamon P, Rastegar N, Ahmadi M, Bhatti JK, Bozek DA, Li N, Lee L, Che C, Luis E, et al. 2021. Gradient of developmental and injury response transcriptional states defines functional vulnerabilities underpinning glioblastoma heterogeneity. *Nature Cancer* **2**:157–173. DOI: https://doi.org/10.1038/s43018-020-00154-9, PMID: 35122077

Rosswog C, Bartenhagen C, Welte A, Kahlert Y, Hemstedt N, Lorenz W, Cartolano M, Ackermann S, Perner S, Vogel W, Altmüller J, Nürnberg P, Hertwig F, Göhring G, Lilienweiss E, Stütz AM, Korbel JO, Thomas RK, Peifer M, Fischer M. 2021. Chromothripsis followed by circular recombination drives oncogene amplification in human cancer. *Nature Genetics* **53**:1673–1685. DOI: https://doi.org/10.1038/s41588-021-00951-7, PMID: 34782764

Sabari BR, Dall'Agnese A, Boija A, Klein IA, Coffey EL, Shrinivas K, Abraham BJ, Hannett NM, Zamudio AV, Manteiga JC, Li CH, Guo YE, Day DS, Schuijers J, Vasile E, Malik S, Hnisz D, Lee TI, Cisse II, Roeder RG, et al. 2018. Coactivator condensation at super-enhancers links phase separation and gene control. Science **361**:eaar3958. DOI: https://doi.org/10.1126/science.aar3958, PMID: 29930091

Shoshani O, Brunner SF, Yaeger R, Ly P, Nechemia-Arbely Y, Kim DH, Fang R, Castillon GA, Yu M, Li JSZ, Sun Y, Ellisman MH, Ren B, Campbell PJ, Cleveland DW. 2021. Chromothripsis drives the evolution of gene amplification in cancer. *Nature* **591**:137–141. DOI: https://doi.org/10.1038/s41586-020-03064-z, PMID: 33361815

Snuderl M, Fazlollahi L, Le LP, Nitta M, Zhelyazkova BH, Davidson CJ, Akhavanfard S, Cahill DP, Aldape KD, Betensky RA, Louis DN, Iafrate AJ. 2011. Mosaic amplification of multiple receptor tyrosine kinase genes in glioblastoma. *Cancer Cell* **20**:810–817. DOI: https://doi.org/10.1016/j.ccr.2011.11.005, PMID: 22137795

Strom AR, Brangwynne CP. 2019. The liquid nucleome-phase transitions in the nucleus at a glance. *Journal of Cell Science* **132**:jcs235093. DOI: https://doi.org/10.1242/jcs.235093, PMID: 31754043

Suvà ML, Rheinbay E, Gillespie SM, Patel AP, Wakimoto H, Rabkin SD, Riggi N, Chi AS, Cahill DP, Nahed BV, Curry WT, Martuza RL, Rivera MN, Rossetti N, Kasif S, Beik S, Kadri S, Tirosh I, Wortman I, Shalek AK, et al. 2014. Reconstructing and reprogramming the tumor-propagating potential of glioblastoma stem-like cells. *Cell* **157**:580–594. DOI: https://doi.org/10.1016/j.cell.2014.02.030, PMID: 24726434

Szerlip NJ, Pedraza A, Chakravarty D, Azim M, McGuire J, Fang Y, Ozawa T, Holland EC, Huse JT, Jhanwar S, Leversha MA, Mikkelsen T, Brennan CW. 2012. Intratumoral heterogeneity of receptor tyrosine kinases EGFR and PDGFRA amplification in glioblastoma defines subpopulations with distinct growth factor response. *PNAS* **109**:3041–3046. DOI: https://doi.org/10.1073/pnas.1114033109, PMID: 22323597

Turner KM, Deshpande V, Beyter D, Koga T, Rusert J, Lee C, Li B, Arden K, Ren B, Nathanson DA, Kornblum HI, Taylor MD, Kaushal S, Cavenee WK, Wechsler-Reya R, Furnari FB, Vandenberg SR, Rao PN, Wahl GM, Bafna V, et al. 2017. Extrachromosomal oncogene amplification drives tumour evolution and genetic heterogeneity. *Nature* **543**:122–125. DOI: https://doi.org/10.1038/nature21356, PMID: 28178237

Veatch SL, Machta BB, Shelby SA, Chiang EN, Holowka DA, Baird BA. 2012. Correlation functions quantify super-resolution images and estimate apparent clustering due to over-counting. *PLOS ONE* **7**:e31457. DOI: https://doi.org/10.1371/journal.pone.0031457, PMID: 22384026

Verhaak RGW, Hoadley KA, Purdom E, Wang V, Qi Y, Wilkerson MD, Miller CR, Ding L, Golub T, Mesirov JP, Alexe G, Lawrence M, O'Kelly M, Tamayo P, Weir BA, Gabriel S, Winckler W, Gupta S, Jakkula L, Feiler HS, et al. 2010. Integrated genomic analysis identifies clinically relevant subtypes of glioblastoma characterized by abnormalities in PDGFRA, IDH1, EGFR, and NF1. *Cancer Cell* **17**:98–110. DOI: https://doi.org/10.1016/j.ccr.2009.12.020, PMID: 20129251

Verhaak RGW, Bafna V, Mischel PS. 2019. Extrachromosomal oncogene amplification in tumour pathogenesis and evolution. *Nature Reviews Cancer* **19**:283–288. DOI: https://doi.org/10.1038/s41568-019-0128-6, PMID: 30872802

Vicario R, Peg V, Morancho B, Zacarias-Fluck M, Zhang J, Martínez-Barriocanal Á, Navarro Jiménez A, Aura C, Burgues O, Lluch A, Cortés J, Nuciforo P, Rubio IT, Marangoni E, Deeds J, Boehm M, Schlegel R, Tabernero J, Mosher R, Arribas J. 2015. Patterns of HER2 gene amplification and response to anti-HER2 therapies. *PLOS ONE* **10**:e0129876. DOI: https://doi.org/10.1371/journal.pone.0129876, PMID: 26075403

Vogt N, Lefèvre SH, Apiou F, Dutrillaux AM, Cör A, Leuraud P, Poupon MF, Dutrillaux B, Debatisse M, Malfoy B. 2004. Molecular structure of double-minute chromosomes bearing amplified copies of the epidermal growth factor receptor gene in gliomas. *PNAS* **101**:11368–11373. DOI: https://doi.org/10.1073/pnas.0402979101, PMID: 15269346

Wang L-B, Karpova A, Gritsenko MA, Kyle JE, Cao S, Li Y, Rykunov D, Colaprico A, Rothstein JH, Hong R, Stathias V, Cornwell M, Petralia F, Wu Y, Reva B, Krug K, Pugliese P, Kawaler E, Olsen LK, Liang W-W, et al. 2021. Proteogenomic and metabolomic characterization of human glioblastoma. *Cancer Cell* **39**:509–528. DOI: https://doi.org/10.1016/j.ccell.2021.01.006, PMID: 33577785

Williamson I, Lettice LA, Hill RE, Bickmore WA. 2016. Shh and ZRS enhancer colocalisation is specific to the zone of polarising activity. *Development* **143**:2994–3001. DOI: https://doi.org/10.1242/dev.139188, PMID: 27402708

Williamson I, Kane L, Devenney PS, Flyamer IM, Anderson E, Kilanowski F, Hill RE, Bickmore WA, Lettice LA. 2019. Developmentally regulated shh expression is robust to TAD perturbations. *Development* **146**:dev179523. DOI: https://doi.org/10.1242/dev.179523, PMID: 31511252

Wu S, Turner KM, Nguyen N, Raviram R, Erb M, Santini J, Luebeck J, Rajkumar U, Diao Y, Li B, Zhang W, Jameson N, Corces MR, Granja JM, Chen X, Coruh C, Abnousi A, Houston J, Ye Z, Hu R, et al. 2019. Circular ecdna promotes accessible chromatin and high oncogene expression. *Nature* **575**:699–703. DOI: https://doi.org/10.1038/s41586-019-1763-5, PMID: 31748743

Yi E, Gujar AD, Guthrie M, Kim H, Zhao D, Johnson KC, Amin SB, Costa ML, Yu Q, Das S, Jillette N, Clow PA, Cheng AW, Verhaak RGW. 2021. Live-cell imaging shows uneven segregation of extrachromosomal DNA elements and transcriptionally active extrachromosomal DNA hubs in cancer. *Cancer Discovery* **12**:468–483. DOI: https://doi.org/10.1158/2159-8290.CD-21-1376, PMID: 34819316

Zhu Y, Gujar AD, Wong CH, Tjong H, Ngan CY, Gong L, Chen YA, Kim H, Liu J, Li M, Mil-Homens A, Maurya R, Kuhlberg C, Sun F, Yi E, AC, Ruan Y, Verhaak RGW, Wei CL. 2021. Oncogenic extrachromosomal DNA functions as mobile enhancers to globally amplify chromosomal transcription. *Cancer Cell* **39**:694–707.. DOI: https://doi.org/10.1016/j.ccell.2021.03.006, PMID: 33836152