Queen Square Institute of Neurology

Analysis of glioma cell heterogeneity by lineage tracing in human and murine model systems

James Alexander Innes

Queen Square Institute of Neurology

University College London

This dissertation is submitted for the degree of

Doctor of Philosophy

2021

Declaration

I, James Innes, hereby declare that this dissertation is the result of my own work except where specifically indicated in the acknowledgements and in the thesis text.

James Innes

2021

Acknowledgements

I would like to express a great deal of thanks to the charity “Brain Tumour Research” for funding my PhD project. BTR funds many quality research programmes as well as playing a vital role in furthering public awareness and support for families affected by brain tumours.

I would like to extend this thanks to my supervisors and colleagues for their kind help and support, without them this project would not have been possible. Sebastian Brandner’s ever-present guidance and immense knowledge has been invaluable. I have very much enjoyed our thought-provoking discussions and appreciate the trust I have been given to make important decisions. The additional supervision from Silvia Marino was vital for furthering the project and I would like to thank Silvia for sharing her knowledge and guidance.

Within the Brandner research lab, I would like to show appreciation to Tian Yu for her patient training, encouragement and experimental guidance. Joanne Lau helped immensely with logistical tasks, animal work and genotyping. Andrew Lowe also offered valuable insights and designed aspects of our clonal tracking analysis. Working alongside Raquel Fonseca made days go by easier and I would like to thank her for helping with technical issue and regularly covering lab tasks that I could not attend. I would also like to thank Dr Ying Zhang,, Ms Tasha Aley, Ms Tedani El-Hassan and Mrs Angela Richard-Londt for their technical and logistical support.

Within UCL, Dale moulding and Alan Greig at the ICH and Rockefeller core imaging facilities were vital for developing the imaging approaches reported. Dr Ayad Eddaoudi from the UCL, Institute of child health flow cytometry unit was another very encouraging and immensely helpful figure during the project.

Finally, I would like to extend a special thanks to my family and friends for their continued support, encouragement and interest.

Abstract

Intra-tumour heterogeneity and plasticity are key factors in treatment resistance and the recurrence of Glioblastoma (GBM), which is invariably fatal. Genetics, epigenetics, cell metabolism and plastic cancer stem cell (CSC) hierarchies interact with volatile micro-environmental forces to promote and shape cell identity. Improved understanding of these factors will inform more precise and effective GBM therapies. Here, we aim to develop a fluorescent tracking approach for patient derived GBM cells to investigate the relationship between clones, environment and CSC marker expression. Using a murine GBM model combined with Rosa26-confetti fluorescent labelling, we trialled suitable techniques for detection of labelled tumour clones and concluded fluorescent imaging and flow cytometry were the most effective. For patient-derived cells, we modified LeGO-vector fluorescent labelling with the aim of tracking a greater number of clones. We further optimised this technique for simultaneous flow cytometry detection of clones and their CSC marker expression. In the final chapter, we address the hypothesis that whole population CSC surface marker plasticity is a result of emergent clonal predominance. In two patient derived GBM lines, under steady-state environmental conditions, serial passaging and assessment of clonal marker expression detected distinct marker expression patterns between clones in the same culture dish. For both cell lines, transfer and culture of clonal mixtures to Matrigel® spheroids produced an expected plastic transition in population marker expression but also considerable predominance of certain clones. While the clonal surface marker dynamics of the two cell lines were markedly distinct, divergent surface marker plasticity between clones of the same cell line was a consistent observation. Taken together these results supported our hypothesis that population marker plasticity is in part a result of emergent clonal predominance. We propose our developed techniques are suitable for rapid and economic characterisation of patient specific gene disruption, therapeutic vulnerabilities and resistance mechanisms.

Impact statement

The development of cancer is a result of somatic mutations and the selection of clonal populations. Over the last few decades, the general principles of cancer evolution have been documented, and more recently, the molecular evolution of cancer cells, and the role of cancer stem cells has added to the understanding of cancer biology. Heterogeneity within individual tumours remains one of the major challenges, in particular in the context of the development of therapies targeting certain mutational combinations but also differentiation states of cancer cells. Understanding clonal evolution and fate mapping of cancer cells was the major objectives of this study. The scientific impact of this study is twofold: First, the systematic establishment of a model system aiming at the spatial mapping of clonal populations using a highly controlled genetically labelled in vivo system has shown the potential and limitations of a number of experimental strategies. Second, the generation of clonal populations from barcoded cancer stem cells provides a tool that can be replicated effectively, and affordably in the scientific community. This part of the study is an excellent example how computational analysis can significantly enhance the readout of data from genetic information, combined with marker phenotypes, on a single-cell level.

In conclusion, this study provides a systematic analysis to inform the scientific community of Pitfalls and potentials of different model system aiming at the characterisation of clonal populations in cancer. The most significant impact comes from establishing an experimental approach to rapidly establish and characterise large numbers of clonal populations derived from newly established GBM cell lines. This system offers a powerful and affordable method to identify specific therapeutic vulnerabilities in heterogenous tumour cell populations, and to characterise selective, experimentally introduced mutations.

Contents

List of figures

Abbreviations List

AKT	Protein kinase B
AML	Acute myeloid leukemia
APC	Astrocyte precurser cell
ARRIVE	Animal Research: Reporting of In Vivo Experiments
ATRX	Alpha-thalassemia/mental retardation, X-linked
C.elegans	Caenorhabditis elegans
CD133	Promonin-1
CD15	Stage-specific embryonic antigen/Sialyl lewisX/3-fucosyl-N-acetyl-lactosamine
CD44	HCAM/homing cell adhesion molecule
CDKN2A	Cyclin dependent kinase inhibitor 2A
CML	Chronic myeloid leukemia
CNS	Central nervous system
CNV	Copy number variation
Confetti-GBM	GLAST$^{CreERT2/CreERT2}$/PTEN$^{lox/lox}$/P53$^{lox/lox}$/ROSA$^{confetti/WT}$ mice
Cre	Cre recombinase
CreERT	Cre recombinase fused to a mutant estrogen ligand-binding domain
CRISPR	Clusters regularly interspace short palindromic repeats
CSC	Cancer stem cell
CSF	Cerebrospinal fluid
DDR	DNA-damage reponse
DNA	Deoxyribose nucleic acid
EBFP	Enhanced Blue fluorescent protein
ECFP	Enhanced cerulean fluorescent protein
EGF	Epidermal growth factor
EGFP	Enhance green fluorescent protein
EGFR	Epidermal growth factor receptor
ERK	Extracellular signal-regulated kinase
ESC	Embryonic stem cell
EYFP	Enhanced yellow fluorescent protein

FACS	Fluorescence activated cell sorting
FGF	Fibroblast growth factor
FRET	Forster resonance energy transfer
FSB	FACS staining buffer
FSC-A	Forward scatter area
FSC-H	Forward scatter height
FSC-W	Forward scatter width
GBM	GBM
GBO	GBM organoid
GEMM	Genetically engineered mouse model
GFAP	Glial fibrillary acidic protein
GIC	Glioma initiating cell
GLAST	Glutamate aspartate transporter
GLICO	Glioma cerebral organoid
GSC	Glioma stem-like cell
HPS	Hydrogel polymer solution
IDH	Isocitrate dehydrogenase
IRES	Internal ribosomal entry site
LeGO	Lentiviral gene ontology
Lgr5	Leucine rich repeat containing G protein-coupled receptor 5
Lsm	Laser scanning microscopy
MGMT	O-6-methylguanine-DNA methyltransferase
mO2F	mOrange2 fluorescent protein (from Brainbow 3.2 construct)
mK2F	mKate2 fluorescent protein (from Brianbow 3.2 construct)
MYC	MYC Proto-Oncogene
NeoCOR	Neoplastic cerebral organoids
NF1	Neurofibromatosis 1
NFIX	Nuclear factor 1X
NG2	Neuron-glial antigen 2
NOD/SCID	Non-obese diabetic/severe combined immunodeficient mice
NSC	Neural stem cell
NSLC	Non-small cell lung cancer

OPC	Oligodendrocyte precurser
P53	Transformation-related protein 53
PBS	Phosphate buffered saline
PCR	Polymerase chain reaction
PDGF	Platelet derived growth factor
PDGFR	Platelet derived growth factor receptor
PDGFRα	Platelet derived growth factor receptor α
PDGFβ	Platelet derived growth factor β
PFA	Paraformaldehyde
PFU	Plaque forming units
PI3K	Phosphoinositide 3-kinase
PIC	PDGFβ-Ires-Cre retrovirus
PIP2	Phosphatidylinositol 2, 5-bisphosphate
PIP3	Phosphatidylinositol 3, 5-bisphosphate
PTEN	Phosphatase and tensin homologue
RAS	RAS GTPase
RB1	Retinoblastoma transcriptional corepressor 1
RFP	Red fluorescent protein
RIMS	Refractive Index matching solution
RNA	Ribonucleic acid
Rosa26	Reverse orientation splice acceptor 26
RTK	Receptor tyrosine kinase
RTKI	Receptor tyrosine kinase inhibitor
scRNA-seq	Single cell RNA sequencing
SGZ	Sub-granular zone
SSC-A	Side scatter area
SVZ	Sub-ventricular zone
TCGA	The Cancer Genome Atlas
TERT	Telomere reverse transcriptase
TMZ	Temozolamide
VEGFR	Vascular endothelial derived growth factor
WHO	World health organisation

Chapter 1: Introduction

1.1 Tumour heterogeneity and plasticity as a barrier to treatment

Refinement of surgical approaches and more precise application of radio/chemotherapeutic treatments have continued to improve survival times for many cancer patients. Identification of characteristic mutations and downstream aberrations to molecular pathways vital for cancer cell survival and proliferation have further improved survival outcomes (Urruticoechea et al., 2010). An illustrative example are receptor tyrosine kinases (RTK) such as epidermal growth factor receptor (EGFR), Platelet derived growth factor receptor (PDGFR), vascular endothelial derived growth factor receptor (VEGFR) and their downstream effector molecules such as RAS, ERK and mTOR that are hijacked by many cancer types to support cell survival, proliferation, differentiation and apoptosis resistance (Montor et al., 2018). In the case of chronic myeloid leukaemia (CML), the protein product of the well characterised gene fusion Philadelphia chromosome, leads to a constitutively active tyrosine kinase signalling molecule termed bcr-abl that controls activation and suppression of various cell cycle genes. The RTK inhibitor (RTKI) imatinib is a specific blocker of this fusion protein's kinase domain, quenching its kinase activity and preventing cell proliferation. As a result, the 10-year survival rate of CML patients treated with imatinib is 83.3% (Hochhaus et al., 2017). Instead, in the case of other cancers, especially solid tumour such as non-small cell lung cancers (NSCLC), RTKI's have had only modest effects on patient survival (Planchard et al., 2018, Sequist et al., 2013). Mutations leading to constitutively active EGFR are common in NSCLC but tumours harbour multiple defence mechanisms for achieving resistance to RTKI treatment (Morgillo et al., 2016). Primary resistance happens when a tumour bulk contains EGFR mutations that are not sensitive to an applied RTKI, reflecting the variety of patient specific EGFR alterations with the same tumour promoting effect (Eck and Yun, 2010). Prolonged treatment (> 6 months) with these drugs can lead to acquired resistance through mechanisms including secondary

mutations to the EGFR gene (Tan et al., 2016), a phenotypic transformation of tumour cells (Sequist et al., 2011) or activation of different pathways to overcome the inhibitory effects of RTKIs on survival and proliferation (Engelman et al., 2007). The processes of resistance are relatively well characterised in NSCLC and these mechanisms are likely consistent with other cancers, where RTKIs are only suitable for tumours with certain molecular characteristics and still report modest efficacy. RTKIs have been used as an example here but indeed tumour characteristics for evading resection, radiotherapy, alkylating chemotherapies and a variety of immunotherapies have been described (Vasan et al., 2019). Tumours recurring after treatment can be molecularly profiled to investigate changes associated with resistance, but addressing the dynamic processes of tumour biology allowing adaption and resistance acquisition can be more easily assessed in advanced experimental models. For example, when a mutation arises in tumour cells that imparts resistance to an ongoing RTKI therapy, how has this occurred? Is this the result of a spontaneous mutation acquired during the treatment course, or, had this mutation already occurred during tumour evolution and been selected for as a result of treatment? A resistance conferring alteration occurring in only a small minority of tumour cells during tumour development would not be detectable through bulk sample genomic approaches used to characterise tumour genetic and molecular profiles. Furthermore, in the circumstance of phenotypic adaptation to overcome treatment, is this plasticity a property of many cells in the tumour or a reprieve of a small set of cells with appropriate genetic and epigenetic conditions? Along with the failures of many promising phase III drug trials (ref), these questions underpin the drive to better understand concepts of tumour heterogeneity and plasticity. With the ultimate hope that an increased basic understanding can inform more precise, multi-target treatment regimes.

In this thesis, we explore fluorescent cell tracking to better understand clonal evolution and plasticity in Glioblastoma (GBM), the most aggressive form of malignant glioma. The following sections will outline the diagnostic and molecular parameters of gliomas with a focus on GBM. Subsequently describing how tumour origin, genomic instability and clonal evolution create cellular diversity. Leading on

to a discussion of how clonality and cancer stem-like cells (CSC) interact with the tumour micro-environment to further diversify cells. Cell tracking, where labels can identify clonal populations of tumours cells, is a powerful tool for investigating cell intrinsic processes related to heterogeneity. We discuss these methods focusing on the transgenic and viral based fluorescent approaches applied in this work.

1.2 GBM

GBM is a highly malignant and invasive form of diffuse glioma. Gliomas are the most common form of primary brain tumour in adults and have been traditionally classified on their histological resemblance to parenchymal brain cells. Gliomas are graded on a malignancy scale; grade I tumours are non-malignant and can often be resolved through surgical resection. While grades II – IV are malignant classifications with increasing proliferation, diffusivity and poorer prognosis (Thakkar et al., 2014). GBM has been traditionally described as a grade 4 Astrocytoma that can present clinically as a primary tumour or a secondary progression from a lower grade glioma (Aldape et al., 2015, Brandner and von Deimling, 2015). In addition to histological features of malignancy, disease stratifying molecular markers are being increasingly incorporated to improve glioma diagnostics. The Isocitrate dehydrogenase 1 or 2 (IDH1/2) genes are the most commonly mutated across all gliomas where mutations generally convey reduced malignancy and improved prognosis (Aldape et al., 2015, Yan et al., 2009). In the recently updated 2021 WHO classification of CNS tumours, IDH wildtype GBM (IDH-wt GBM) has been introduced to denote tumours traditionally described as primary GBM. While grade 4 IDH-mutant Astrocytoma is now used to describe what was traditionally a secondary GBM (Louis et al., 2021). Further legacy GBM tumour types have been reclassified and are outlined in 1.2.4. However, most literature discussed in this thesis was published before the 2021 guidelines and reference to GBM throughout this thesis refers to this outdated nomenclature unless otherwise indicated. Importantly, the experiments described in Chapter 5 of this thesis are performed on cell lines derived from IDH-wt GBM patients.

1.2.1 Epidemiology, clinical features

IDH-wt GBM is the most frequent adult malignant brain tumour, accounting for 15% of all intracranial tumours and nearly half of all primary malignant brain tumours. The tumour occurs preferentially in middle-aged adults (peak incidence 55-85 years), with a slight predominance in males (Ostrom et al., 2014). The IDH-wt GBM is generally found in the subcortical white matter and deep grey matter of either hemisphere and can occur in all lobes of the brain (Ostrom et al., 2019). The clinical presentation depends on tumour location and often manifests with focal neurological deficits such as hemiparesis, aphasia, visual defects and seizures. These symptoms are usually progressive and the time from symptom onset to diagnosis is shorter than 6 month in 80% of patients (Ohgaki and Kleihues, 2007). Through diagnostic imaging (usually magnetic resonance imaging or computed tomography), GBMs appear irregularly shaped with a ring-enhancing component around a central area of necrosis, accompanied by oedema. However, the different molecular subtypes of GBM cannot be readily discriminated by standard imaging studies (Ogawa et al., 2020) (Swaidan et al., 2012), but advanced radiomics studies can to some extent predict the gene expression profiles of newly diagnosed GBM (Macyszyn et al., 2016) (Nandu et al., 2018, Choi et al., 2020) and this can lead to improved diagnostic and prognostic accuracy (Artzi et al., 2019).

1.2.2 Prognosis and treatment

Most patients with IDH-wt GBM die within 15-18 months after therapy with chemoradiation. The 5-year survival rate has been reported as low as 7% (Ostrom et al., 2019) and 10% (Stupp et al., 2009) in the context of clinical trials where patients with slightly more favourable performance status were selected. Favourable prognostic factors are younger age (50 years and below), good performance status, complete tumour resection, and a methylated O-6-methylguanine-DNA methyltransferase (MGMT) promoter which leads to a reduced activity of the MGMT enzyme which antagonises the effects of the standard of care chemotherapeutic Temozolomide (TMZ)(Stupp et al., 2009). In contrast, the IDH-mutant astrocytoma WHO grade 4 (previously known as IDH-mutant GBM, WHO grade IV, or in the more

distant past also termed “secondary GBM”). Some patients with GBM respond well to current treatments (maximal safe surgery, radiation, TMZ chemotherapy, as well as experimental and immunological interventions, but there is a significant heterogeneity in treatment response, which most likely reflects the biological heterogeneity of the disease. For most treatments the basic molecular mechanism for primary or acquired resistance are only incompletely understood.

1.2.3 Histological Hallmarks

The primary approach in diagnosis of brain tumours is histological assessment. Typically, the IDH-wt GBM presents as diffusely growing, cellular glial tumour with astrocytic tumour cells. The cell size and shape can be uniform in some tumours and much more heterogeneous in others. Most tumours show mitotic figures that are easy to identify. Classical additional features are so-called microvascular proliferations (tumour vessels with excessively activated endothelial linings, often with duplicate or multiple layers of endothelium), pseudopalisading cell arrangements and necrosis (Figure 1.1). GBM is one of the most heterogenous neoplasms with a high degree of variability of morphological features between patients and often within individual tumours (Burger and Kleihues, 1989) .

1.2.4 Molecular diagnosis

In the 2016 and 2021 WHO classification of CNS tumours, molecular and genetic markers were incorporated to complement histology-based diagnosis (Louis et al., 2016, Louis et al., 2021). Molecular indicators have been introduced with the aim of minimising interobserver variability, and will be a first important step to direct patients into clinical trials stratified by molecular profiles (Figure 1.1) (Louis et al., 2016b). The IDH-wt GBM is defined by the absence of mutations in the isocitrate dehydrogenase genes 1 and 2, and by the frequent presence of a telomere reverse transcriptase (TERT) promoter mutation, EGFR gene amplification, gain on chromosomes 7 and loss on chromosome 10 (Stichel et al., 2018). Although diagnostic characteristics of IDH-wt GBM, these alterations present with considerable variation between patients and the presence of only one of these alterations is sufficient for diagnosis of an IDH-wt GBM, even in the absence of high-

grade histological features such as pseudopalisades necrosis or microvascular proliferation (Brat et al., 2018) (Louis et al., 2020). This interpatient variability extends beyond diagnostic mutations and each patient will present with an almost unique collection of mutations and large-scale chromosomal aberrations. This variability is a considerable hurdle to successful treatment and underpins the drive for a more patient specific approach to prescribing treatment. A further molecular marker with prognostic (but not diagnostic) relevance in IDH-wt GBM is methylation at the MGMT promoter. MGMT promoter methylation predicts patient responses to the standard of care chemotherapeutic TMZ, with methylation and gene silencing predicting favourable outcomes (Hegi et al., 2005) (Wick et al., 2014).

Several other forms of GBM are described in WHO 2015 classification with updated nomenclature in 2021 classification (Louis et al., 2016, Louis et al., 2021). The IDH mutant GBM (Grade 4 IDH-mut Astrocytoma, WHO 2021) is characterised by mutations in the alpha-thalassemia/mental retardation, x-linked (ATRX) gene, often resulting in the loss of ATRX protein expression, a diagnostically useful marker. The Histone3 K27M mutant GBM (Diffuse midline glioma, WHO 2021) is in midline structures and is typically seen in children (termed diffuse infantile pontine glioma) but more recently with more systematic diagnostic tests also increasingly identified in adults. A very poorly differentiated form of GBM, termed diffuse hemispheric glioma in WHO 2021 classification, carries the histone H3 G34R or V mutation with ATRX mutations also common.

In 2010, Verhaak *et al*., classified GBM entities based on mRNA expression data collected by the Cancer Genome Atlas (TCGA). Hierarchal clustering of 1740 genes from 202 GBM samples revealed 4 distinct GBM subclasses named proneural, neural, classical and mesenchymal (Verhaak et al., 2010). Classifications had relevance to survival data with the proneural tumour types showing statistically significant shorter survival times but was subsequently determined to be largely congruent with the IDH-wt GBM classification. A further sequencing approach by which GBM has been classified is the methylation profile where researchers have proposed 8 sub-classes within the WHO 2016 GBM patient base (Capper et al., 2018). Beyond the goal of improved stratification of GBM entities, expression and methylation classifiers build

substantial repositories of molecular information that can indicate specific therapeutic vulnerabilities associated to different tumour subclasses. However, these analyses are generally limited to bulk samples of the end-stage disease, missing any underlying cellular heterogeneity in the measures used for classification and the dynamic cellular processes that precede and are ongoing during clinical presentation. Therefore, in addition to accurate stratification of the disease, an understanding of tumour evolution and the cellular heterogeneity produced during this process will be vital for advancing therapy. In the following sections factors known to influence these evolutionary dynamics will be discussed.

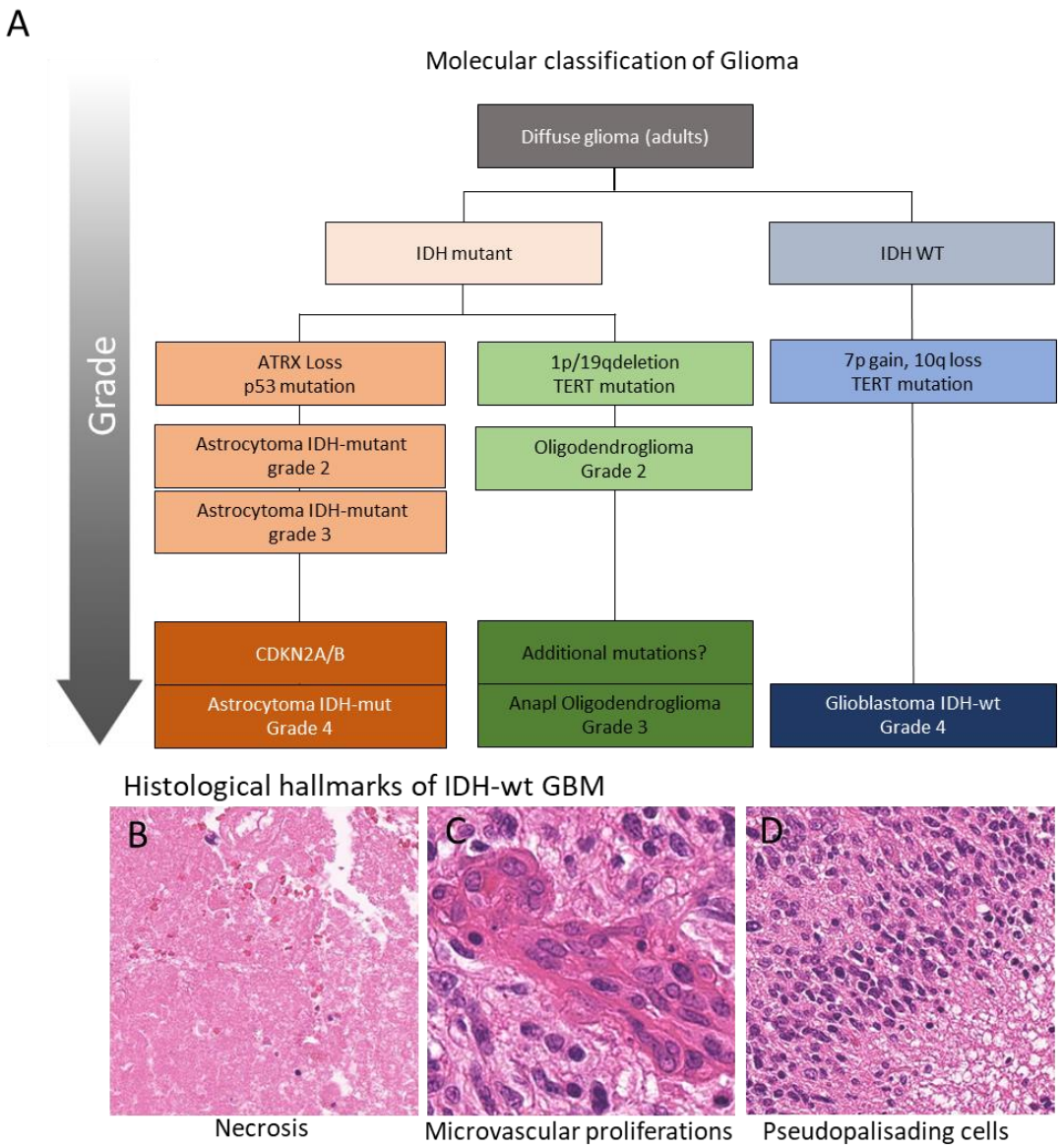


Figure 1.1: Molecular classification of adult glioma and histological hallmarks of IDH-WT GBM: (A) IDH1 is the most commonly mutated gene in gliomas where IDH-mutant gliomas comprise lower grade subtypes compared with IDH-wt gliomas. However, mutations in CDKN2A/B can lead to higher grade IDH-mut tumours. In addition to specific mutations, specific chromosomal aberrations assessed through copy number profiling are commonly found in certain glioma subtypes. 1p/19q co-deletion is a characteristic of Oligodendroglioma and 7p gain/10q loss are common aberations found in IDH-wt GBM. (B-D) Representative histological features associated with IDH-wt GBM which include large necrotic regions void of living cells (B), formation of naïve vasculature structures termed microvascular proliferations (C) and pseudopalisading cell structures where cells appear to be migrating away from regions of necrosis (images taken from in house histological samples).

1.2.5 Cell of origin

The progenitor populations of the adult brain are the most widely accepted candidates of the cell of origin in glioma. In the adult brain, these progenitor populations are located in the sub-ventricular zone (SVZ) lining the ventricles, the sub-granular zone (SGZ) within the dentate gyrus of the hippocampus, and the white matter of the cerebellum (Figure 1.2) (Ming and Song, 2011) . Studies in mice revealed that newly produced neuroblasts in the SVZ migrate along the rostral migratory stream toward the olfactory bulb to replace neurons and encode new olfaction sensory information (Lim and Alvarez-Buylla, 2016). Interestingly, it is also well established that spatially distinct regions of the SVZ give rise to neuroblasts which mature into functionally distinct types of olfactory neurons (Sequerra, 2014).

While SVZ neurogenesis occurs in humans, its function remains debated as there is limited evidence for SVZ neurogenesis contributing to new olfactory neurons in humans (Lim and Alvarez-Buylla, 2016). In the SGZ, neurogenesis is postulated to have functions in encoding new memories and other complex cognitive processes such as pattern separation, which are likely to be conserved in humans (Ming and Song, 2011). In addition to the previously mentioned neurogenic areas, there are additional regions, identified in recent studies, such as hypothalamus, striatum, substantia nigra, cortex and amygdala (Jurkowski et al., 2020). Whether neurogenesis in these areas is dependent on migration from the SVZ and SGZ, or each structure contains its own residual neural stem cell (NSC) pool is yet to be resolved. In addition to sites of neurogenesis, it's also known that oligodendrocyte precursor cells (OPC) and astrocytic precursor cells (APC) exist distributed throughout the brain (Windrem et al., 2020). OPCs can replace oligodendrocytes and contribute to plasticity by enhancing myelination between brain regions (Martins-Macedo et al., 2021). APCs actively give rise to further mature astrocytes and play important roles during inflammatory responses to lesions within the CNS (Martins-Macedo et al., 2021)

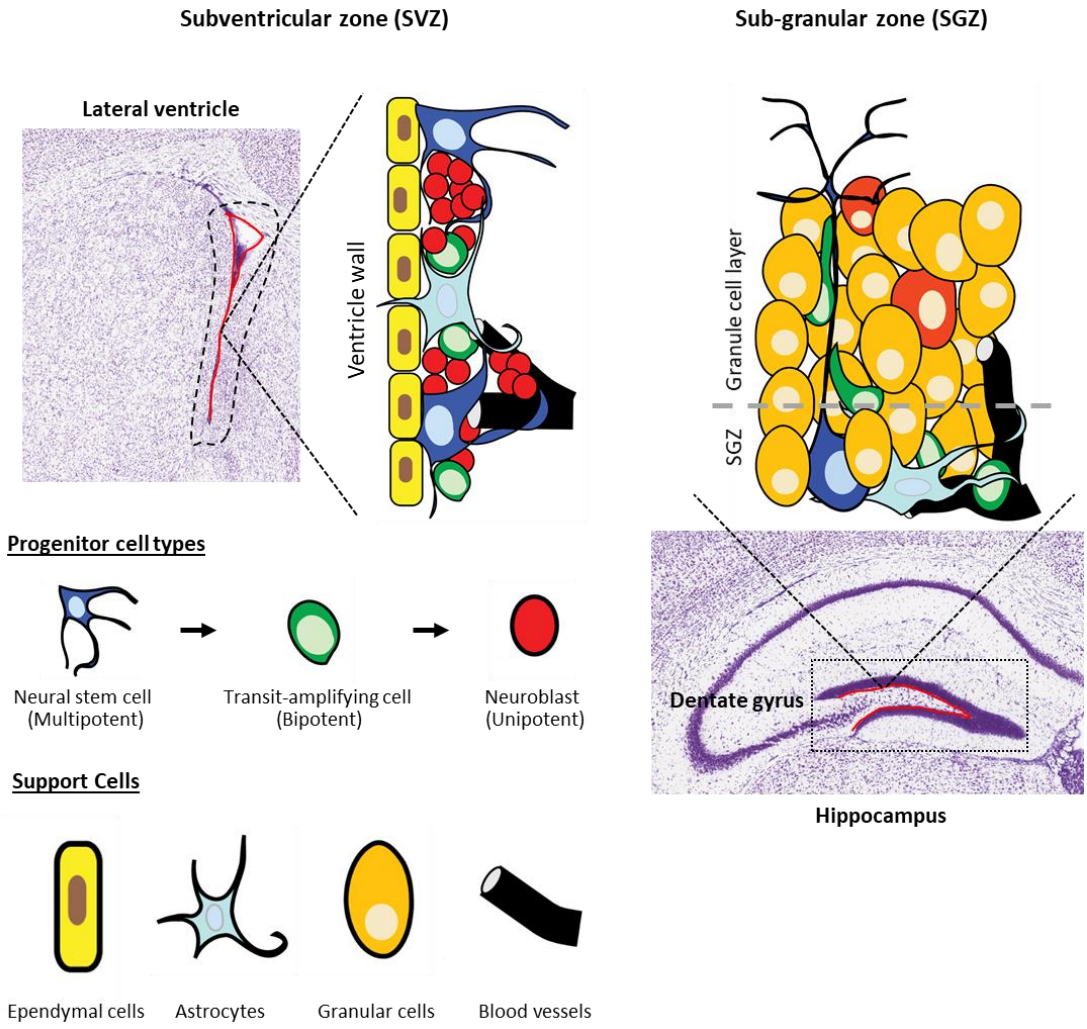


Figure 1.2: Neurogenic niches of the adult brain and the cell types associated. Neurogenesis occurs in the lateral ventricles and hippocampus of adult animals. These neurogenic niches are comprised of NSCs which give rise to restricted potential progenitors which go on to produce new mature neuronal cells. Niches also contain a number of support cells important for regulating stem cell proliferation and differentiation trajectory. (figure adapted from (Kazanis et al., 2008))

The fact that these cell types are actively dividing and evidence of stem cell compartments giving rise to tumours in other organs have placed these progenitors as the most likely candidates as cell of origin of glial or glioneuronal tumours. In both the SVZ and SGZ it is established that quiescent and slowly dividing NSCs, capable of producing all cell types of the brain other than microglia, give rise to multi-potent and more rapidly dividing transiently amplifying cells (Figure 1.2) (Alcantara Llaguno and Parada, 2016). These cells intern give rise to intermediate bipotent cell types which produce unipotent progenitors only capable of differentiating into a single

terminal cell type. As such, APCs and OPCs distributed throughout the brain can be considered unipotent progenitors. The different cell types of this hierarchy from stem-like to terminal progeny are well characterised with established molecular markers associated with each cell type (Jurkowski et al., 2020). Experiments using conditional Cre expression to target deletion of tumour suppressor genes in stem cell or limited progenitor compartments suggest a wide-range of cells may have tumour forming potential. For example, Nestin and GFAP are markers associated with NSCs and targeted inactivation of tumour suppressor genes such as Neurofibromatosis 1 (NF1), Transformation related protein 53 (P53), Phosphatase and tensin homologue (PTEN) and Retinoblastoma transcriptional corepressor 1 (RB1) in cells expressing these markers leads to brian tumour resembling GBM (Alcantara Llaguno et al., 2015). Furthermore, Ascl1 is a marker of a bipotent cell population capable of giving rise to mature neurons and oligodendrocytes and targeted deletion of tumour suppressors in these Ascl1 expressing cells also leads formation of GBM-like tumours (Alcantara Llaguno and Parada, 2016, Azzarelli et al., 2018). In another study, authors targeted deletion of P53 and NF1 in adult OPCs using inducible cre expression under control of Neuron-glial antigen 2 (NG2) promoter sequence. In this context, OPCs went through a protracted multi-stage transformation process but were also capable of malignant progression to produce tumours resembling glioma (Galvao et al., 2014). This finding was also supported by a lineage-tracing study using mosaic analysis with double markers (MADM) to achieve tumour-inducing mutations in each lineage compartment of the adult brain; neurons, astrocytes and oligodendrocytes (Kegelman et al., 2014). This study showed that that OPCs were the cell type that most readily underwent neoplastic transformation during early stages of tumour development and that mature tumours harboured salient OPC expression programmes (Liu et al., 2011).

The experiments outlined above suggest a considerable number of different cell types can undergo malignant transformation and form tumours. Moreover, this array of stem cell and limited progenitors capable of forming tumours may be reflected in the considerable inter-patient heterogeneity observed within glioma sub-classes. Where patient age, presumed cell of origin and underlying driver mutations all

contribute to shaping a tumours molecular and histological characteristics. However, owing to their ability for indefinite self-renewal, naïve stem cell compartments in the brain opposed to limited progenitor populations are the most likely-candidates for glioma cell of origin. Since limited progenitors are capable of only a finite number of divisions, the likelihood of further neoplastic mutations arising after an initial mutation is limited compared to a stem cell which may divide and copy its DNA many more times throughout an organisms life-time.

1.2.6 Intra-tumour heterogeneity: Clonal evolution

Working from the assumption that GBM arises from progenitor cells in the brain, the aetiology of the disease is thought to begin with a mutation in a progenitor cell compartment. Neurogenic compartments of the adult brain are specialised niches composed of various regulatory cell types in close proximity to vessels and cerebrospinal fluid (CSF) for integrating endocrine signals (Ohlstein et al., 2004, Walker et al., 2009). Each of these components play an important role in finely tuning proliferation and directing progenitors toward the correct route of differentiation. It is thought that mutational events affecting cell-cycle regulatory pathways such as RAS/ERK and PI3K/AKT, allow cells within these niches to escape this regulatory framework (Azzarelli et al., 2018). It is believed neoplastic growth occurs in stem cell compartments throughout the lifetime of an organism but are usually kept in check by cell intrinsic fail-safe mechanisms or recognised by the immune system and targeted for destruction (Sever and Brugge, 2015). In the right conditions, a neoplasm can escape these fail-safe mechanisms and develop into a malignancy. These initial mutations are referred to as truncal or driver mutations and occur on tumour suppressor genes, oncogenes, DNA repair-associated genes and genes important for epigenetic regulation (Sanai et al., 2005).

Tumour suppressor genes commonly mutated in GBM include P53, PTEN and Cyclin dependent kinase inhibitor 2A/B (CDKN2A/B). Tumour suppressor gene mutations generally convey a loss of function for their coded proteins which limit the inhibitory regulation they exert on molecular pathways influencing proliferation, survival and cell cycle progression. P53 is a transcription factor affected by many upstream signals primarily influencing cell cycle progression and genome integrity

with further functions related to cell metabolism and stemness (Bieging et al., 2014, Zhang et al., 2018). Notably, P53 is vital for triggering cell cycle arrest and apoptosis in cells with damaged DNA. PTEN is a lipid phosphatase that negatively regulates the phosphatidylinositol 3-kinase (PI3K) pathway by dephosphorylation of phosphatidylinostiol 3, 4, 5-triphosphate (PIP3), to phosphatidylinositol 4,5-biphosphate (PIP2) (Chalhoub and Baker, 2009). PIP3 interacts with AKT which acts on downstream components to promote cell survival and growth (Worby and Dixon, 2014). Therefore, loss of function in tumour repressor genes leads to loss of inhibitory regulation on pathways that can promote development of neoplasias.

Oncogene regulated pathways overlap with tumour suppressor gene signalling but play a promotive function in pathway activity. Common examples are mutations in receptor tyrosine kinases where PDGFR and particularly EGFR are found mutated in GBM (Brandner and von Deimling, 2015). Activation of these RTKs promotes downstream activation of effector molecules RAS and AKT to promote cell proliferation and survival (Cantanhede and de Oliveira, 2017, Nazarenko et al., 2012). In GBM, mutations lead to overexpression and constitutively active forms of these receptors that override inhibitory regulation from tumour suppressors and other regulatory elements (Oprita et al., 2021).

DNA damage response (DDR) pathways form an intricate network of proteins which identify and repair DNA damage in response to genotoxic stress from tumour-suppressor and oncogenic dysfunction or environmental factors (Sousa et al., 2020). The most common DDR associated alteration in GBM is MGMT promoter methylation which has clinical importance in conferring susceptibility to treatment with TMZ (Dunn et al., 2009). TMZ is a DNA alkylating agent and its damaging effects on DNA are repaired by MGMT protein action, therefore, methylation of the promoter and reduced protein expression support positive responses to TMZ treatment (Brandner and von Deimling, 2015, Koschmann et al., 2016). ATRX is a further DNA repair protein which is responsible for carrying out non-homologous end joining and ATRX loss is associated with paediatric GBM and lower grade gliomas ((Koschmann et al., 2016)). While not as well established in GBM, loss of effective DNA repair is a step in the early progression of many other tumours where DNA instability increases the

likelihood of acquiring further tumour promoting mutations (Sousa et al., 2020). Conversely, in later stages of GBM, DDR molecular machinery is found over expressed conveying a greater ability for cells to correct the high rate of newly induced mutations. As discussed, spontaneous mutations can drive tumour progression but an overly severe tumour burdened can also lead to tumour cell death, a fate avoided by over expression of DDR components (Bartkova et al., 2010). Therefore, given the important role for oncogenes and tumour suppressors in integrating signals for cell-cycle progression and survival, mutations affecting their function can cause normal progenitors to escape the tightly controlled extra-cellular regulation of their niche (Azzarelli et al., 2018, Cha and Yim, 2013, Sanai et al., 2005). Furthermore, mutations to DNA-repair machinery can precede or compound these alterations by increasing the chance of further harmful mutations arising. As a neoplasm grows and forms a tumour, further truncal mutations are introduced alongside passenger mutations which can be detected through sequencing but do not have strong tumour promoting effects. DNA instability and truncal driver mutations are important concepts underlying clonal evolution where heightened mutagenesis and Darwinian selection create a tumour composed of different clonal populations with distinct underlying DNA aberrations (Nowell, 1988).

Clonal evolution can occur through a spontaneous mutational event, for example, a mutation producing a constitutively active RTK conveying greater growth and survival promoting effects than previous mutations could act as a selective pressure promoting predominance of cells harbouring this new alteration. This process can build with the introduction of further mutations promoting growth and survival creating new clonal lineages with advantageous growth properties. Alternatively, as the tumour evolves, environmental conditions can change, for example hypoxia in regions of necrosis, inflammation or infiltration of immune cells such as macrophages or T-cells (Clevers, 2011). These can create microenvironmental and immunological selection pressures where a clone with the ability to switch to anaerobic metabolism or the highest immunosuppressive expression signature will escape immune detection most effectively and outgrow other clones. Various models for clonal evolution have been put forward, notably these include linear, branching, neutral

and punctuated forms of clonal predominance which are outlined in **Error! Reference source not found.** (Davis et al., 2017). Early molecular evidence of clonal evolution giving rise to intra-tumour heterogeneity in GBM was the identification of intermingled cells with differential amplification of platelet derived growth factor receptor (PDGFR) and epidermal growth factor receptor (EGFR), demonstrated by fluorescence *in situ* hybridisation (FISH) or quantitative PCR (Snuderl et al., 2011). Suggesting that copy number profiles of a tumour bulk reporting amplification of two RTKs may reflect the presence cells with different RTK amplifications, rather than a generic dual amplification shared by all cells.

To explore the functional basis of this observation, human cell lines were established from tumours showing co-amplification of PDGFRα and EGFR. Interestingly, established lines from these tumours also contained cells with either co-amplified PDGFRα and EGFR or cells with amplification of just one of these RTKs (Szerlip et al., 2012). Interestingly, for all cell lines assessed, cells with co-amplified RTKs were far less prevalent than cells with a single EGFR or PDGFR amplification. It was also demonstrated that selection of a single RTK-amplified population could be achieved through pharmacological stimulation or inhibition of either PDGFRα or EGFR, suggesting *in vivo* these cells are functionally different in terms of their response to growth factor signalling (Szerlip et al., 2012). Furthermore, enrichment for each of the single-amplified populations and DNA sequencing suggested all populations shared a common clonal origin. Suggesting that populations with differential RTK amplifications may arise through more complex mechanisms than mutational acquisition and clonal outgrowth. It is possible that undefined mechanisms of genetic inheritance during cell division are altered in GBM cells allowing ongoing production of cells harbouring diverse gene amplifications which can be selected for by environmental pressures.

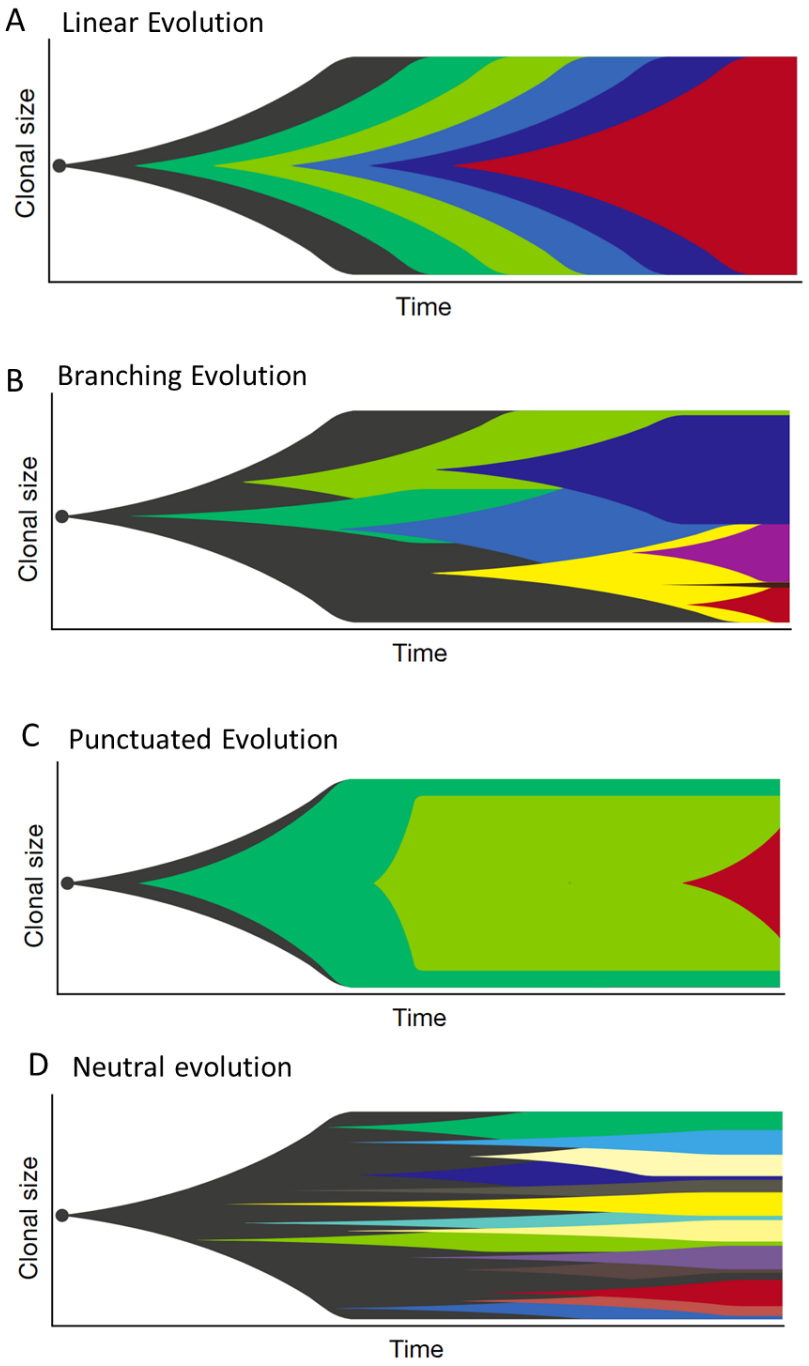


Figure 1.3: Different forms of genetic clonal evolution in cancer: (A) Linear evolution is where tumour cell populations advance continuously with new "fitter" clones successively outgrowing legacy clones. This mode of evolution is often observed in liquid tumours and is less common in solid tumours. (B) Branching evolution is where new clones arise from previously predominant populations, are selected for and co-inhabit the tumour. Evolution in this fashion is more often found in solid tumours perhaps influenced by a variable spatial microenvironment. (C) In punctuated evolution, rather than mutational events being an ongoing process, tumour and clonal evolution is characterised by bursts or genetic alteration giving sequential rise to new clonal populations. (D) Neutral evolution proposes that as new mutations arise in clones, there is no selection of "fitness" and all separate clones retain the same propensity to grow regardless of the mutations they carry. Figure adapted from (Vendramin et al., 2021)

Several seminal studies investigated tumour heterogeneity and genetic evolution by spatial biopsy sampling. Through intraoperative sampling and sequencing of up to 6 biopsies per tumour, authors found CNVs that were common in all biopsies, shared amongst some biopsies and in each biopsy unique variations were also present (Sottoriva et al., 2013). Additionally, it was shown that in 6 of 10 tumours, that there were regionally distinct expression profiles, corresponding to 1 of the transcriptomic GBM classes (proneural, neural, mesenchymal and classic) (Verhaak et al., 2010). Authors postulated that common alteration shared across all regionally distinct biopsies were early mutational events in tumour evolution with shared CNVs occurring at intermediate stages and alterations unique to each biopsy were acquired latest in tumour progression. With some theoretical assumptions, authors use this information to reconstruct and model the time course of a tumours clonal evolution. Although, evidence of clonal cells carrying variable CNVs, discussed in the previous paragraph, cautions the validity of lineage reconstruction from bulk sequencing techniques in this fashion (Szerlip et al., 2012), their modelling suggested that RTK amplifications are often truncal events occurring early in a founder clone during tumour evolution. It is estimated GBM can be in development for a median of 13 years before diagnosis, therefore, insights to the patient specific dynamics of this evolutionary process will be important for furthering disease understanding and aid in developing more targeted combinatorial treatments (Wang et al., 2016).

A further influential study focussing on the dynamic nature of GBM genomic profiles performed mutational analysis on matched low-grade primary IDH-mut glioma and recurrent GBM (Johnson et al., 2014). Similar to results with serial spatial sampling, authors find shared and unique mutational events in the patient matched primary and recurrent tumours. Results were highly variable between patients, with some recurrences sharing similar mutational profiles suggesting linear clonal evolution and others showing a low degree of similarity in mutational profile indicating branched evolution (Figure 1.3B). In one case, mutational profile of a first, second and third recurrence of the same tumour suggested branched and linear evolution taking place at different stages of the disease (Johnson et al., 2014). Suggesting type of clonal evolution pattern is not an intrinsic property of the tumour but perhaps a process

dependent on stage of development and environmental contexts (**Error! Reference s ource not found.**).

Through various sequencing approaches, these studies identified intra-cellular factors affecting cellular heterogeneity in GBM. Pointing to the likely circumstance that at the time of treatment, tumours are composed of distinct genetic clones of different predominance and functional potential where mutational analysis of the bulk tumour mass will overlook functional mutations present in subsidiary genetic clones (Sottoriva et al., 2013). Upon a change of environment such as resection and treatment exposure, the varied functional properties of subsidiary clones leave the tumour well poised to adapt and overcome this new challenge (Wang et al., 2016). Combining these findings with the clear heterogeneity between patients conveys cautionary implications on the likely success of treatments targeting a single molecular pathway in GBM (Ghosh et al., 2018). A combinatorial, patient specific approach to treatment may harbour more success clinically and a further understanding of cell heterogeneity will be critical to accurately inform such a strategy.

1.2.7 Intra-tumour heterogeneity: Cancer stem cells

In 1994, acute myeloid leukaemia cells (AML) were identified that had a distinct immunophenotype compared to other cellular fractions, and they showed greater cancer-inducing potential upon transplantation into immunocompromised mice (Lapidot et al., 1994). This study was a critical step in support of a long-standing hypothesis that cancer cells divide in a hierarchical manner, with cancer stem-like cells (CSC) at the apex of this hierarchy (Batlle and Clevers, 2017, Lathia et al., 2011, Vescovi et al., 2006). Using expression patterns of the two surface markers CD34 and CD38, it was shown this AML inducing potential was restricted to cells with a $CD34^{+}/CD38^{-}$ phenotype, where $CD34^{+}/CD38^{+}$ and $CD34^{-}$ cellular fractions were incapable of inducing AML after transplantation(Lapidot et al., 1994). Critically, these $CD34^{+}/CD38^{-}$ cells represented about a 1000-fold smaller fraction of AML cells compared to a previously identified colony-forming cell population (Lapidot et al., 1994). These unique properties and the uneven distribution supported the idea that $CD34^{+}/CD38^{-}$ AML-inducing cells constituted a less mature population towards the

apex of a hierarchy. This was a finding with profound implications for understanding tumour development and treatment resistance across many cancer types.

The CSC hypothesis proposes that cancer cell populations follow the general rules of differentiation hierarchies found in normal tissues. Stem cell hierarchies are well-illustrated by the Waddington landscape model of balls rolling down a hill, with branching valleys representing routes of differentiation (Figure 1.4A) (Wang et al., 2011). In this analogy, balls start at the top (stem cell) of a hill and can only roll down the valleys as differentiation is a linear process, and once a ball (cell) enters a valley (i.e., corresponding to a cellular state) it is now restricted to subsidiary valley's it can nter. In other words, potential for differentiation and cell proliferation is lost as cell's move down from the apex of a hierarchy. These concepts are distinct from those assumed in the clonal evolution hypothesis where all cells are equal and capable of indefinite self-renewal (Nowell, 1988). In this model, CSCs are the only cells capable of indefinite self-renewal adding another dimension to the network of cell diversity of glioma (Lathia et al., 2015). Within a tumour, CSCs divide infrequently and are often in a state of quiescence, occasionally giving rise to more rapidly dividing restricted potential cell populations which make up the bulk of the tumour mass which subsequently produce non-proliferative, terminally differentiated cells (Figure 1.4B) (Lathia et al., 2011). The implications of this concept on treatment were clear, i.e. if the CSC populations can be "eradicated" at the top of the hierarchy, all cells in the tumour could in theory be eradicated (Yu et al., 2012). This hypothesis also offered an attractive explanation for tumour recurrence, where a gross total resection removes the bulk of tumour cells which are predominantly rapidly dividing but restricted potential progenitors, leaving behind pockets of invasive CSCs, even in very small residual quantities. Quiescent, non-dividing CSCs which have an established *in* vitro resistance to TMZ which may contribute to their ability in escaping eradication during treatment, a behaviour also considered possible through overexpression of DNA-repair enzymes (Sousa et al., 2020, Liu et al., 2006). Once a treatment regimen has been completed, residual CSCs could re-enter the cell cycle and repopulate the tumour. Therefore, in GBM, as in other cancers, there has been

a large effort to characterise and isolate stem-like cell fractions for stem cell targeting treatments.

Cells displaying *in* vitro properties similar to NSCs have been successfully isolated from a variety of different tumour types including GBM. These presumed CSCs have the ability to form spheres, show indefinite self-renewal, multipotent differentiation potential and could form tumours upon transplantation into immunocompromised mice (Lathia et al., 2015). Furthermore, immunophenotyping revealed they also show expression of markers for NSCs and other brain cell prohenitors, such as Nestin, Nanog, Musashi, OLIG2, Pax6, SOX2, SOX1 and GFAP (Hemmati et al., 2003, Ben-Porath et al., 2008, Ligon et al., 2007, Kim et al., 2010, Tunici et al., 2004). However, many of these markers are nuclear or cytoplasmic and can only be enriched for through transgenic fluorescent reporter approaches. Flow cytometry staining and FACS sorting is rapid technique for segregating cells by expression of surface proteins and as such, surface markers of different potential glioma CSC populations were sought after.

Promonin-1 (CD133) was the first protein put-forward as a candidate surface marker of glioma CSCs (Singh et al., 2004, Singh et al., 2003). CD133 was an early identified marker of neural stem and progenitor cells and was utilised to label identify potentially similar cell populations in glioma (Uchida et al., 2000). Early studies showed $CD133^+$ enriched in cells with higher *in vitro* self-renewal, proliferation, differentiation and *in vivo* tumour forming potential. As few as 300 $CD133^+$ cells could form tumours while as many as 100,000 $CD133^-$ cells were incapable of seeding and forming a tumour after xenotransplantation. Furthermore, isolated $CD133^+$ glioma CSCs were shown to harbour *in vitro* resistance for TMZ and radiotherapy (Singh et al., 2004). However, subsequent studies demonstrated tumour forming potential of $CD133^-$ cells which interestingly gave rise to tumours containing $CD133^+$ cell populations (Wang et al., 2008, Singh et al., 2004). Although, a possible explanation for this finding is fluctuations in CD133 expression associated with cell cycle progression (Barrantes-Freer et al., 2015). Further studies isolating this $CD133^+$ cell compartment of glioma lead to the discovery of other enrichment markers for glioma cells with stem-like properties. $CD133^+$ cell populations showed increased expression

of CD44 compared with $CD133^-$ populations (Liu et al., 2006). CD44 is a membrane glycoprotein upregulated in many cancers and is associated with motility, proliferation, apoptosis and angiogenesis (Chen et al., 2018). Further evidence suggests CD44 also confers resistance to the cytotoxic agent TMZ through supressing the Hippo apoptosis signalling pathway (Xu et al., 2010). CD44 is particularly highly expressed in GBMs with the "mesenchymal" transcriptional and epigenetic profile, and previous studies claimed that CD44 expression was significantly associated with worse patient survival (Krishna et al., 2013, Jin et al., 2017).

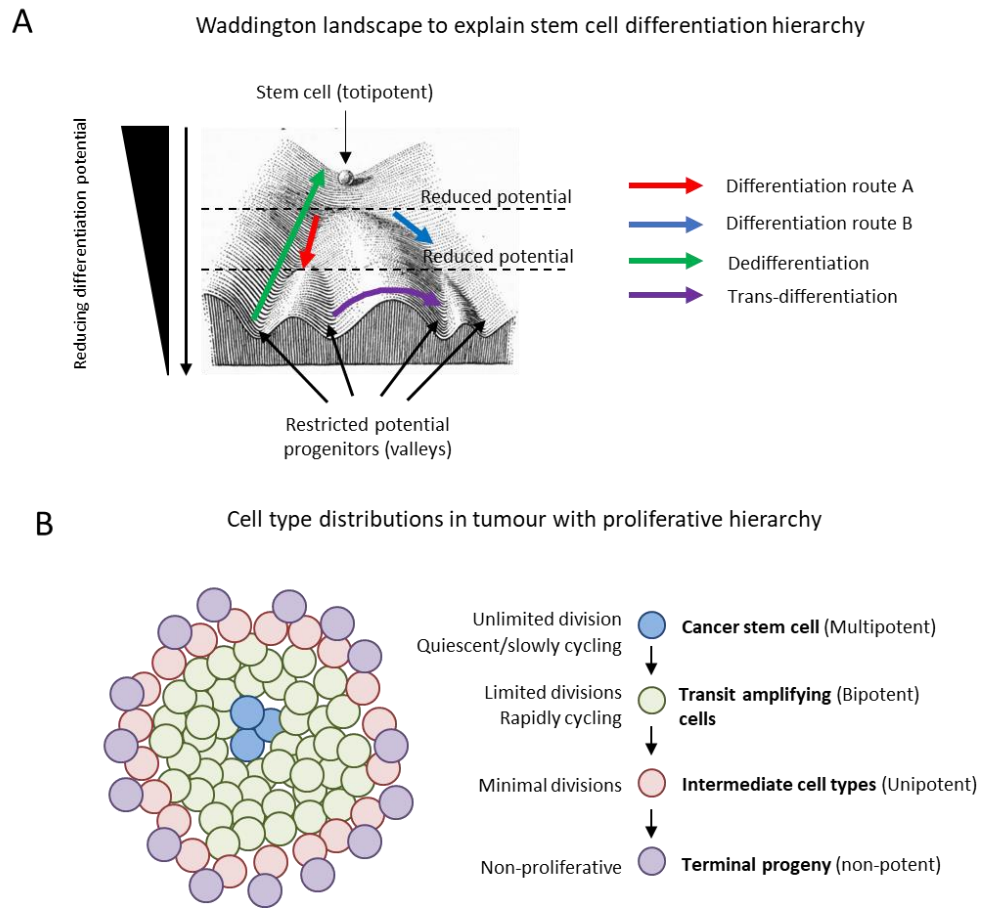


Figure 1.4: CSC proliferative and differentiation hierarchies: (A) The Waddington landscape displaying the proposed process of cell differentiation in a rigid hierarchy. Characterised by assymetrical divisions giving rise to progeny with restricted differentiation potential and the ability to only produce certain cell types (red and blue arrows). In normal developmental cellular hierarchies this is a linear process, however in cancers it is likely that dedifferentiation (green arrow) and transdifferentiation (Purple arrow) can also occur. (B) The presence of hierarchies has implications on the proportional cell type composition of tumours. CSCs divide slowly but indefinitely, comprising a rare portion of tumour cells. Occasionally, a CSC gives rise to a transit amplifying cell which are the most abundant cell type of the tumour and divide rapidly but for only a limited number of cell cycles. Transit amplifying cells give rise to intermediate and terminal progeny which have a highly limited number of divisions and comprise a limited portion of the tumour. Waddintongton landscape in an adapted from (Waddington, 2014).

A further glioma CSC candidate marker identified through the study of $CD133^+$ was the glial progenitor marker A2B5 found to be expressed by approximate 60% of cells across a number of primary human glioma lines (Ogden et al., 2008). Indeed, $A2B5^+CD133^-$ cells were capable of forming tumours after xenotransplantation, further suggesting tumour forming potential is not the exclusive property of $CD133^+$ cells. Interestingly, it was found that CD133 and A2B5 are variably expressed across samples of different patient, suggesting markers expressed by glioma progenitors are not conserved between patients but rather tumour dependent. $A2B5^+$ cells were shown to have a migratory and proliferative phenotype with the propensity to differentiate into more astrocytic and oligodendrocyte cell types (Tchoghandjian et al., 2010). A2B5 is an antibody epitope targeting antigens found on certain trisialogangliosides found expressed during specific periods of early brain development and also expressed in a pocket of glial progenitor cells in mature sub-cortical white matter (Baeza-Kallee et al., 2019).

Stage-specific embryonic antigen 1 (SSEA-1/CD15) was put forward as a further marker of glioma CSCs, a marker often found expressed in the nearly 40% of glioma cell lines containing no $CD133^+$ cells (Son et al., 2009). $CD15^+$ cells were found in a number of $CD133^-$ tumour lines and showed the *in vitro* characteristics of self-renewal and differentiation while also showing potency in tumour forming potential (Son et al., 2009). CD15 has subsequently been linked with cells of the proneural GBM subtype as well as found expressed in medulloblastoma (Krishna et al., 2013, Read et al., 2009).

A recent study offered a new perspective on the dynamic expression of these CSC surface markers in glioma (Dirkse et al., 2019b). Dirkse., et al performed "surface marker phenotyping" across a number of GBM cell lines, co-staining for the 4 surface markers CD133, CD44, A2B5 and CD15 and quantifying 16 different expression phenotypes. Across cell lines they found substantial variability in the expression of surface markers and interestingly isolation and characterisation of 16 different phenotypes (derived from combinations of the four individual surface markers) showed no difference in proliferation or self-renewal even in the phenotype negative

for all 4 markers (Dirkse et al.,). This challenges the validity of these surface proteins as markers of stem-like cells at the apex of a hierarchy. Purified populations of each of the 16 different phenotypes had indeed the capacity to give rise to all other 15 surface marker phenotypes. This is compelling evidence (in the context of cancer) that these markers may not represent binary indicators of stem-like properties as has been shown in normal tissues during organ development and differentiation (Prager et al., 2020). Finally, a cell lines marker expression can vary in different environmental contexts, whereby cultured cells enrich for the expression of certain markers depending on oxygen pressure, nutrients, differentiation conditions, and after xenografting. A phenomena that had already been speculated (Scott et al., 2019, Scott et al., 2014, Enderling, 2015) and also demonstrated in a more simple experimental setting investigating CD133 and CD44 expression only (Brown et al., 2017). This offers an attractive explanation and integration of seemingly contradictory evidence across previous studies, that intrinsic cell lines variability and experimental conditions for example culture media can profoundly impact on the reproducibility of glioma GSC marker profiles. Thus, suggesting that cellular phenotypes in cancer are more plastic and can exist in a spectrum of states with many possible transitions between different states. These findings also caution the functional validation of markers in cancer suggesting there is apparent redundancy in associated stemness pathways and that apparent plasticity in marker expression may impede the likelihood of success in targeting these markers for therapeutic intervention.

Beyond surface marker staining and cell isolation, single cell RNA sequencing (scRNA-seq) has emerged as probably one of the most powerful tools for investigating intra-tumour heterogeneity and CSC characteristics (Wagner and Klein, 2020, Sun et al., 2021). In the first published GBM sc-RNAseq study, authors demonstrate heterogeneity in single cell copy number alterations within a single tumour and also show that cells within a single tumour can correspond to different GBM expression classifications (Patel et al., 2014). Furthermore, cell expression signatures ranging from stem-like to more differentiated cell types were observed supporting the idea of hierarchical divisions from glioma CSCs (Patel et al., 2014). Several subsequent

scRNA-seq studies have been published, using different parameters and modelling to classify gene expression states of cells in GBM (Couturier et al., 2020, Wang et al., 2019, Neftel et al., 2019). One prominent scRNA-seq study of GBM, posited that cells exist across a spectrum of expressional states characterised by their resemblance to neural, astrocytic, oligodendrocytic and mesenchymal progenitors with many cells found in intermediate states between these more defined expressional programmes (Neftel et al., 2019). In this study, authors also demonstrate how the underlying genetic aberrations of a tumour shape the distribution of cell expression states. In a further study, using a different dataset and analysis pipeline, authors concluded that cell states are arranged along a single axis ranging from mesenchymal to proneural, with mesenchymal cell types representing the less mature more stem-like population giving rise to proneural cells (Wang et al., 2019). Similarly, in the 4-progenitor model, proliferation is enriched within the oligodendrocyte and neural precursor states (Neftel et al., 2019). In both models, many cells are found in intermediate states, expressing genes associated with multiple progenitor cell types, seemingly supporting findings of complex state transitions suggested through surface marker phenotyping of glioma cell lines (Dirkse et al., 2019b)

In conclusion, these studies suggest a scenario where GBM cells do exist in some hierarchical arrangements, but lack the more rigid linear differentiation rules observed in normal tissues. Potential for dedifferentiation and trans differentiation appear to be a likely characteristic of GBM where under steady conditions cells may divide in a hierarchical manner, but mutations or a change to the tumour microenvironment can shift the balance of this hierarchy, with effects on clonal predominance or differentiation trajectory. This is the basis of the attractor states hypothesis (Figure 1.5) of glioma CSCs which combines aspects of clonal evolution, stem cell hierarchies and cellular plasticity to describe tumour evolution and the establishment of intra-tumour heterogeneity in GBM (Prager et al., 2020). This attractor states hypothesis will be further explained in the following chapter.

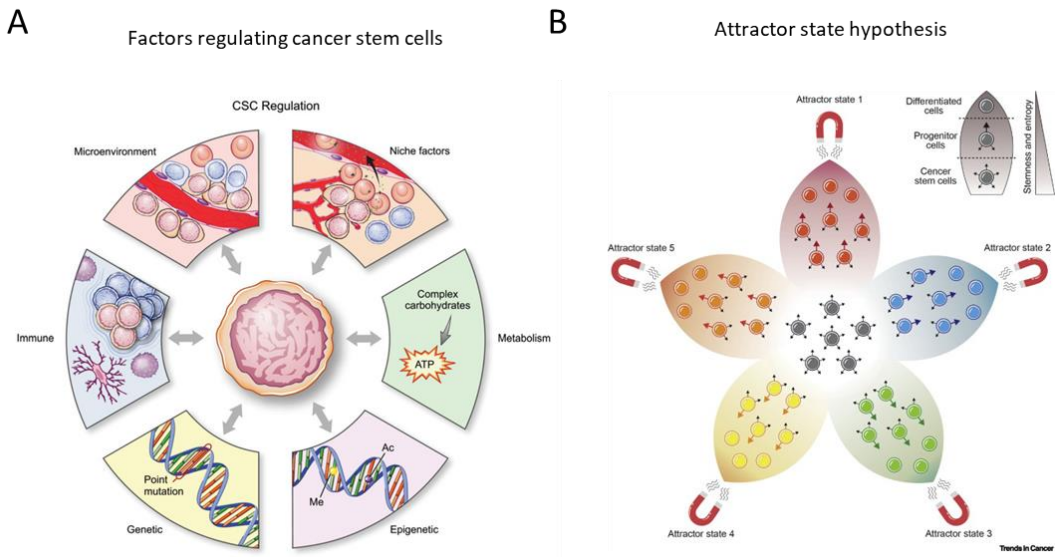


Figure 1.5: Factors influencing cancer stem cell regulation and the attractor state hypothesis: (A) Stem cell proliferation, quiescence, differentiation and migratory behaviours are all governed by wide range of cell intrinsic (genetic, epigenetic and metabolic) and cell-extrinsic (microenvironment, niche componants and immune system) factors. (B) Such factors interact to create tumour conditions which direct and attract cancer stem cells and their progeny toward on certain differentiation trajectories or other behaviours. Unlike rigid hierarchies, cells attracted to certain states can revert back toward other progenitor types or even immature cancer stem cells upon new environmental or cell intrinsic triggers. Schematic in panel A taken from (Lathia et al., 2015) and schematic in panel B taken from (Prager et al., 2020)

1.2.8 Intra-tumour heterogeneity: Microenvironment

In the attractor state hypothesis, signals from variable microenvironmental components of GBM shape clonal selection and programme cells toward particular differentiation states (Figure 1.5) (Lathia, 2021, Prager et al., 2020). Histological assessment of GBM had outlined the presence of three well characterised environmental niches; perivascular stem cell niche, the peritumoral invading edge and perinecrotic tumour core (Schiffer et al., 2018). Tumour cells are found expressing different markers in each of these niches as well as different behavioural phenotypes (Schiffer et al., 2018). In the perivascular niche, supposed glioma CSCs localise on arterioles and capillaries in close contact with vascular endothelial cells while other parenchymal brain cells such as pericytes, reactive astrocytes and immune cells are also found (Filatova et al., 2013, Charles et al., 2012). In the peritumoral invading edge, cells migrate along vessels and the basal lamina secreting factors which modulate the ECM (Kim et al., 2011). This invasive niche is thought to

be a large reason for recurrence in GBM where resection will never remove the entirety of malignant cells. Invading cells will recruit components of vasculature such as endothelial cells and pericytes to divide and form vessels for establishing further perivascular niche structures (Hira et al., 2018). Finally, in the peri necrotic core of the tumour the microenvironment is defined by hypoxia and acidic pH where cells are observed arranged in pseudopalisading structures reminiscent of layers formed in neural rosettes (Martínez-González et al., 2012).

Further work has established that beyond these three histologically identifiable niches, glioma cells interact considerably with non-tumour cells such as neurons, reactive and non-reactive astrocytes, infiltrating immune cells (macrophages, t-cells and natural killer cells), resident brain microglia and fibroblasts (Bougnaud et al., 2016). Such interactions have been shown to have both inhibitory and promotive effects on tumour cell progression through a variety of mechanism (Silver and Lathia, 2018). scRNA-seq is well suited for studying these interactions as transcriptome information is collected for all non-tumour cell types also present in the tumour (Caruso et al., 2020), for example transcriptome heterogeneity of infiltrating macrophages (Ochocka et al., 2021). However, scRNA-seq is not without its limitations, it can only detect RNA from the most highly expressed genes in a cell and is limited to inferences of cell intrinsic properties, overlooking effects of non-cellular microenvironmental factors. Furthermore, proteins and other biomolecules laid down by non-tumour cells such as neurotransmitters and polysaccharides can influence tumour cell states and drive tumour progression. Other chemical factors such immune cytokines, pH and hypoxia further affect tumour cells where in the case of hypoxia, cells are forced to switch into a glycolytic rather than oxidative form of metabolism and immune cytokines transition glioma cells into immunosuppressive expression programme (Scott et al., 2019, Scott et al., 2014). Therefore, various components of the tumour microenvironment have wide-reaching effects on glioma cell biology shaping cellular plasticity and clonal evolution further promoting tumour heterogeneity which drives progression and promotes resilience.

1.3 Lineage-Tracing

Lineage-tracing encompasses a wide range of approaches which fundamentally aim to follow the fate of cells accross divisions, migration and differentiation (Kretzschmar and Watt, 2012). As such, it is a suitable approach to investigate aspects of tumour heterogeneity such as clonal evolution and stem cell plasticity. Firstly, it is important to define the terms "*clone"* and *"lineage"* which are used in this thesis. The terms "clone" or "clonal" define a population of cells that are all descendent from the same "ancestor" cell, while "lineages" are the identities or expressional cellular states that can be found within a bulk or clonal population. Generally speaking, clonal populations are related in terms of their shared genetic inheritance and separate lineages are created through epigenetic changes within a clone (Valor and Hervás-Corpión, 2020).

1.3.1 Early concepts and techniques

Lineage-tracing has fundamentally transformed our understanding of the finely tuned and regulated processes of organism development. The nematode worm Caenorhabdtis elegans (C. elegans) and Xenopus clawed frogs were the subject of early pioneering work in cellular lineage-tracing. Compared to other organisms, Xenopus embryos are proportionally large, allowing experimenters to perform dissection and transplantation experiments more readily. Early studies by Spemann and Mangold identified organiser structures which were vital for laying down the dorsoventral axis of Xenopus development (Elinson and Holowacz, 1995). Ultimately, this work lay the seed for the expansive studies into morphogenic gradients that drive the spatial and temporal organisation of cell fate decisions in the development of all organisms. Developing C. elegans embryos are completely translucent allowing for direct observations of cell divisions through light microscopy. Their invariant pattern of cell division could be fully characterised showing adult males produce 1031 somatic cells arising in the same place and order. Researchers could irradiate individual cells with a laser microbeam or disrupt cellular pathways with drug inhibitors to assess how their well-characterised development was affected (Nigon and Félix, 2017). Critically these studies showed the developmental trajectory of

certain early progenitors is pre-defined at preliminary stages of development, which represented an early clue of linear differentiation hierarchies.

1.3.2 Modern Applications

Today, how cell heritage is related to establishment of cell populations with distinct functions and developmental trajectories remains a central question in the study of tissue development homeostasis, regeneration and disease. Advances in transgenics and gene editing technologies such as cre recombinase, viral gene delivery and CRISPR have underpinned the development of new lineage-tracing applications in complex mammalian systems (Kretzschmar and Watt, 2012). For example, cloning of the gene for green fluorescent protein (GFP) (Chalfie et al., 1994, Shimomura et al., 1962) and establishing the promoter sequences for a number of developmentally important genes allowed the establishment of reporter mouse models (Li et al., 2018). In these models, investigators could interrogate gene function, as any cells expressing their gene of interest will be labelled with a selected fluorescent reporter. However, these transgenic mouse models only labelled cells while they were actively expressing a gene of interest. If researchers were interested in the developmental fate a marker defined progenitor population, such a model would not allow tracking of terminal progeny produced from these progenitors. To overcome this caveat, similar models incorporating Cre-lox recombination were introduced, where the coding sequence of a reporter gene (such as GFP) – under the control of a constitutively expressed promoter (e.g. ROSA26) - is preceded by a LoxP flanked STOP cassette (STOP-GFP) (Soriano, 1999, Mao et al., 2001). A further transgene is introduced which permits Cre expression under the control of a specific promoter for the gene of interest. In this setting, once a cell population begins to express the gene of interest, Cre expression will be triggered which in turn excises the STOP cassette that precedes the reading frame of GFP and leads to active transcription of the reporter construct. As GFP is under a constitutive promoter, any progeny of cells that undergo this initial recombination event will also express GFP, allowing identification of downstream cell types produced by early progenitor populations. This strategy was highly influential in characterising progenitor populations of developing

mammalian systems but in some circumstances lacked specificity as developmental genes often serve different functions at different stages of development.

The generation of conditional Cre systems such as the modified oestrogen receptor (CreERT) were introduced to overcome this problem (Littlewood et al., 1995, Feil et al., 1997, Feil et al., 1996). In this setting, Cre and modified estrogen receptor genes are actively transcribed under the control of a promoter linked with a gene of interest in animals also containing a STOP-GFP cassette. The oestrogen receptor is modified such that it sequesters Cre-recombinase proteins, preventing removal of the stop cassette and active transcription of a reporter gene. Only when the oestrogen receptor is exposed to a synthetic ligand (4-OH Tamoxifen) will Cre be released and excise the stop cassette leading to GFP expression. Critically, 4-OH tamoxifen can be applied to in vivo systems as its pre-metabolite Tamoxifen allowing to control both the type of cell expressing the reporter and the time at which they begin to express the reporter gene. As tamoxifen can be delivered to adult animals, conditional cre models were particularly useful to investigate gene function in mature tissue stem cells during homeostasis or damage repair(Kim et al., 2018).

A further approach which often combined many of the transgenic and Cre based strategies above was the production of chimeric mice, in particular before the availability of more sophisticated methods of transgenesis, conditional Cre-lox systems and the more recent CRISPR Cas9 systems (Kretzschmar and Watt, 2012, Snippert and Clevers, 2011). In this context, embryonic stem cells (ESCs) were harvested from animals with a desired transgenic manipulation or, WT ESCs were isolated in culture and genetically modified. This would involved either deletion or overexpression of a gene of interest alongside a fluorescent reporter for identification (Giangreco et al., 2009). These modified ESCs were then injected into a syngeneic wild-type developing embryo such that the modified cells and WT cells had the potential to contribute to every germ layer of the resultant organism. Introduced cells could be identified through histological stains or a fluorescent reporter and allowed direct comparisons of how cells descendent from the modified ESCs compared with that of the WT host cells. Specifically, these studies were well suited

to answering questions around whether a genes function had a cell intrinsic or extrinsic effect.

One limitation of fluorescent reporter-based tracing approaches in investigating high multi-cellular developmental processes is the restrictive light spectrum available for such labels. This does not allow the degree of sensitivity required to follow or reconstruct all lineages and clonal populations present in such complex systems. However, recent sequencing-based approaches for lineage reconstruction offer much greater sensitivity and precision in tracking of complex multi-clonal settings (Zhang et al., 2020). One example is viral barcoding where investigators can create a library of random plasmid sequences that can be delivered to an *in vivo* or *in vitro* proliferative population. This approach can produce as many as 2 x 10^5 distinct barcode sequences (Nguyen et al., 2014b, Nguyen et al., 2015, Nguyen et al., 2014a). Quantities of certain barcodes within a population after a tracing period can also be determined to provide information about clonal sizes. Such approaches are particularly useful in cancer models where large numbers of cell lineages are actively dividing.

Further advances to these sequence-based approaches are being driven by the rapid development of single-cell "-omics" (Wagner and Klein, 2020). Viral barcode quantification requires PCR amplification of barcode sequences isolated from homogenised whole cell populations to provide a relative quantification of barcodes and clonal prevalence. Single-cell sequencing can reconstruct both clonal and lineage identities of individual cells providing an absolute measure of clonal prevalence and cell types comprising these clones. For example, copy number alterations in single cells of cancer have been used to reconstruct tumour clonal evolution owing to shared and unique CNVs (Yu et al., 2020). Furthermore, heritable somatic mutations introduced passively during development and throughout adult tissue homeostasis can be detected through sequencing and used to identify clonal cell populations (Wagner et al., 2018). Finally, of target effects Cas-9 based genome editing introduce heritable point mutations that have been used to identify the clonal relatedness of cells (Zafar et al., 2020). Nevertheless, sc-seq is still hampered by technical limitations such as poor gene coverage and sequencing only a limited number of cells in a

population, which may not provide the full picture of clonal prevalence's and lineage identities.

1.3.3 Brainbow Toolkit

As mentioned above (Modern Applications 1.3.2) one aspect limiting tracing approaches with fluorescent proteins is the restricted number of simultaneously detectable colours. In recent years, significant effort has been directed at expanding the array of emission spectra produced by recombinant fluorophores (Shcherbo et al., 2010, Lambert et al., 2020). A further step toward maximising the potential of lineage inference through fluorescent labelling were new Cre recombinase strategies to label cells with different fluorophores, or and combinations of fluorophores (Rodriguez et al., 2017, Livet et al., 2007). Figure 1.6 shows the design of vector constructs containing multiple fluorophore-expressing genes with comibnations of incompatible and inverted LoxP sites (Livet et al., 2007, Cai et al., 2013). These were termed brainbow constructs and while the initial publication framed them as tools for studying fine cellular structures and neuronal connectivity, their real potential lay in lineage tracing applications. These two constructs (Brainbow 2.1 and 3.2), were also used in our study (Figure 1.6). Brainbow 2.1 makes use of parallel and inverted loxP sites along with inverted fluorophore sequences that upon cre exposure randomly express one of 4 possible fluorophores (Figure 1.6). Critically, the inverted LoxP sites and fluorophore sequences allow for changing of a cells colour upon secondary exposure to Cre recombinase. The brainbow 3.2 construct makes use of incompatible LoxP sites spaced between three different fluorophore sequences such that upon cre exposure, the cassette will be randomly edited to express one of the three fluorophores (Figure 1.6). This construct was used to produce the brainbow mouse where three of these constructs have been inserted into the mouse lines genome. Upon exposure to the Cre protein, the recombination outcomes at these three transgenes will lead to overlapping expression of fluorophores in different combinations expanding the number of possible fluorescent labels (Figure 1.6). These "multiplex" fluorescent labelling approaches offered new potential for investigating dynamics within progenitor populations, superseding early single fluorophore reporter models which essentially homogenise the population of

interest. However, some early limitations of brainbow constructs were poor fluorophore expression levels which required immuno-amplification for detection and an initially limited number of promoters which restricted labelling to brain cell populations ((Livet et al., 2007, Cai et al., 2013, Roossien and Cai, 2017)).

An important step in realising the potential of these constructs in lineage-tracing was the development of the Rosa26-confetti mouse (Figure 1.6), by cloning the Brianbow2.1 construct into the strong constitutive promoter locus reverse orientation splice acceptor 26 (Rosa26) (Snippert et al., 2010). In this study, performed in the context of intestinal crypt homeostasis, Rosa26-confetti was combined with a second generation conditional Cre recombinase fused to a mutant estrogen ligand-binding domain (CreERT2) transgene under the control of a promoter associated with the intestinal stem cell marker Leucine rich repeat containing G protein-coupled receptor 5 (Lgr5). This was the first time that the proliferative dynamics within this population could be studied and a model of neutral competition was postulated in which individual intestinal crypts repeatedly tend toward clonality throughout a lifetime (Snippert et al., 2010). In follow-up studies, $Lgr5^{+}$ confetti tracing were combined with models of tumorigenesis, establishing a link between $Lgr5^{+}$ stem cells in the origin and maintenance of intestinal adenomas, a striking finding which supported a model of tumour biology where stem cells were on top of hierarchies of cell proliferation (Schepers et al., 2012).

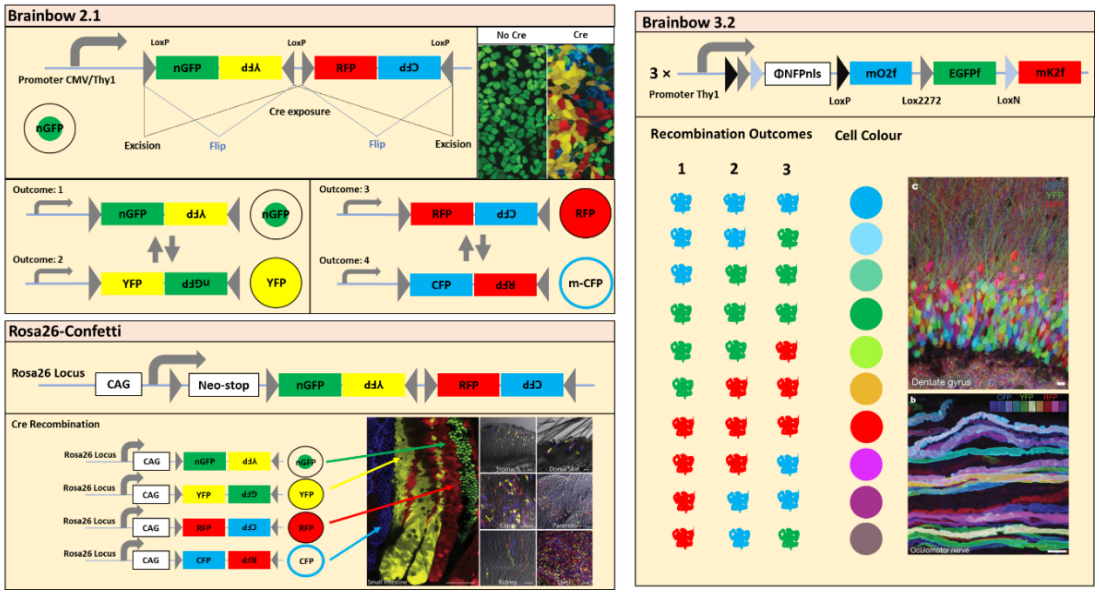


Figure 1.6: Transgenic constructs for multi-colour labelling and lineage-tracing: (A) Structure of the brainbow 2.1 construct containing with possible expression of CFP (membranous), GFP (nuclear), YFP (cytoplasmic) and RFP (cytoplasmic). In the unrecombined formation all cells express GFP. LoxP sites are arranged such that an initial cre exposure will recombine the construct into 1 of 4 possible conformations outlined in the lower panel with cellular expression of one possible fluorophore. Inverted LoxP sites permit flipping of the cassette and changing of cell fluorophore expression (YFP ↔ GFP or RFP ↔ CFP). (B) Rosa26-confetti construct involved inserting brainbow 2.1 in the ROSA26 locus along with a neo-stop cassette for cre-inducible fluorescent tracing in many tissue types. (C) Brainbow3.2 locus comprises just three fluorophores (mO2f, EGFP and mK2F) and the arrangement of incompatible cre recognition sites (LoxP, Lox2272 and LoxN) mean upon recombination there is stochastic expression of just a single fluorophore. Before recombination all cells express the same non fluorescent but immunodetectable fluorophore (ϕNFPnls). Diversification of detectible labels is achieved through insertion of multiple Brainbow 3.2 cassettes, where differing recombination outcomes lead to overlapping expression of different fluorophores, increasing the variety of label hues produced. Examples of labelling are displayed in mouse dentate gyrus neurons and oculomotor nerves.

1.3.4 Lentiviral Gene ontology (LeGO-vectors)

Brainbow constructs were the first of such multiplex and overlapping fluorophore strategies for cellular labelling. In the initial publications, brainbow was described as a method for studying neural circuitry and connectivity with its applications in lineage and clonal tracking not emphasised. However, the technique triggered a cascade of new clonal tracking approaches using fluorophore co-expression to expand the number of unique fluorescent tracking labels that can be introduced to a cell population (Loulier et al., 2014, Sakaguchi et al., 2018, Weber et al., 2012, Hadjieconomou et al., 2011, Xiong et al., 2015). These approaches have been extensively reviewed with each strategy working through different molecular and genetic mechanisms designed to fit certain experimental niches (Weissman and Pan,

2015). One such approach and the basis of the techniques applied in human GBM here, were *Lentiviral gene-ontology vectors* (LeGO-vectors) (Weber et al., 2008). These were initially described as a panel of modular lentiviral vectors primarily for applications *in vitro* populations. The base constructs contained sequences to express a variety of fluorophores (one fluorophore per construct) that were coupled to modular cassettes for downregulating or overexpressing genes of interest (Weber et al., 2008). After the publication of brainbow methodologies, application of these vectors was re-imagined, conceiving RGB marking where three accurately titred LeGO-vectors, each with a different expressed fluorophore are evenly co-transduced into a desired cell population (Figure 1.7A) (Weber et al., 2012, Weber et al., 2011). The result is a mixture of cells comprising seven different labelling outcomes with either single, double or triple fluorophore expression. A critical principle in maximising the number of identifiable lineages is that of regional gene expression, where chromosomal location of lentivector transgenesis dictates the strength of fluorophore expression (Caron et al., 2001). Therefore, depending on insertion site, fluorophore expression can theoretically be high, medium or low, meaning that even cells expressing the same combination of fluorophores will appear in different hues. This methodology is not limited to tracing in vitro cells as in vivo delivery of lentivirus' permits tracking of cells in the developing brain (Figure 1.7B) (Gomez-Nicola et al., 2015). While a powerful technique for clonal tracking, RGB marking is limited in its ability to trace specific lineage-identities. Unlike transgenic brainbow techniques where fluorescence expression can be coupled more easily with cell type specific promoters or timed during development through co-application with inducible CreERT transgenes.

Various iterative improvements of RGB marking have been described previously such as the combination of marking and viral barcoding for more precise clonal tracking (Cornils et al., 2014). A further study performed serial passaging of RGB marked cell lines for up to 38 passages, assessing fluorescent labels by flow cytometry every 4 passages (Brenière-Letuffe et al., 2018). In addition to long term retention of label expression, serial passaging led to outgrowth of clones across all cell lines, indicated by a loss of label diversity in the traced population. Most cultures became dominated

by clonal populations indicated by expansions of cell detections corresponding to a particular

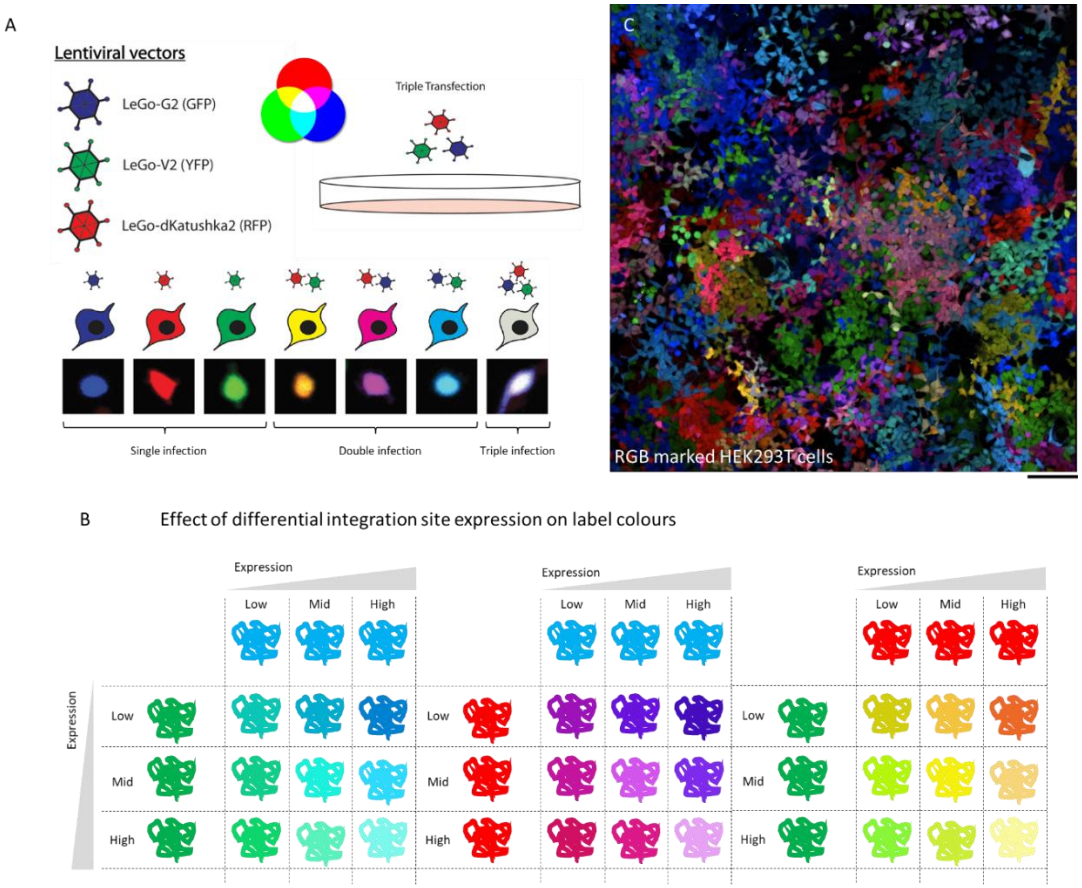


Figure 1.7: Principles of LeGO-vector based RGB marking: (A) Principle of delivering three different fluorophore harbouring Lentivectors simultaneously giving rise to different cell colours depending on transduction outcome (Single, Double or Triple). (B) Demonstrating how variable fluorophore expression depending on site of chromosomal integration can give rise to further colour hues. (C) Example of RGB marking in HEK293T (All personal data).

colour group (Brenière-Letuffe et al., 2018). Studies outlining methods for defining clonal population fluorescence have also been described, enhancing the applicability of accurately tracing multiplex fluorescent cell clones through imaging and flow cytometry, an application particularly desirable in studies of cancer which are less expensive than genetic barcoding (Wu et al., 2016, Gambera et al., 2018, Coffey et al., 2013). A further advancement of LeGO-vector based marking was the publication of "optical barcoding" where authors expanded the panel of lentiviral fluorophores up to seven (Mohme et al., 2017). With seven detectable fluorophores and up to two or three fluorophores allowed per cell, 28 or 63 unique label combinations

respectively, could be created. However, these approaches involved establishing clones from single cells with parallel maintenance of all differently coloured clones before mixing for tracing assays (Mohme et al., 2017, Wu et al., 2016, Gambera et al., 2018, Coffey et al., 2013). When assessing multiple cell lines such an approach does not scale well and parallel maintenance of many clones from many cell lines is unfeasible. Nevertheless, this technique has been applied in glioma studies looking at the effect of immunodeficient and immunocompetent *in vivo* models on tumour clone selection (Maire et al., 2020).

While the array of colour hues produced through these multiplexed fluorescent approaches is expansive, the number of unique colour hues is still limited compared to sequencing and genetic barcoding approaches (Wagner and Klein, 2020, Zafar et al., 2020). However, fluorescent tracing conveys a number of advantages over single cell sequencing approaches. Fluorescent tracing permits *in situ* assessment of lineage which can even be performed on live cells as described in organoids grown in this report. Sequencing approaches, generally require homogenisation of tissue to access the genetic labels for clonal quantification. Interactions between cells and their environment are vital for many cancer processes which can be more easily assessed when tissue integrity is retained for analysis (Charles et al., 2012). However, *in situ* spatial transcriptomics a method for identify specific DNA fragment sequences on histological slides is a technically demanding technique currently in its infancy but has recently been combined to read genetic barcodes on tissue (Hilscher et al., 2020, Askary et al., 2020) Nevertheless, fluorescent labelling approaches remain an inexpensive and experimentally adaptable alternative to viral barcoding with *in situ* sequencing. Beyond imaging, fluorescent lentiviral labels can be assessed through flow cytometry for highly accurate quantification of labels in a traced population which also permits sorting of different populations for genomic applications or continued culture as FACS retains cell viability (Maire et al., 2020, Mohme et al., 2017). Something not been reported yet is the combination of fluorescent lineage-tracing and surface marker phenotyping (Dirkse et al., 2019b) an approach explored considerably in this report.

1.4 Experimental glioma models

Early murine models of glioma included spontaneous mutation models and carcinogen mediated tumour induction. The VM and BRVR mutant mouse lines and their interbred progeny would develop tumours resembling astrocytoma's at incidences of lower than 2% (Fraser and Mcconnell, 1975). Owing to this low tumour penetrance these models were never widely adopted in neuro-oncology. The first attempt at inducing mouse glioma model was to implant pellets loaded with carcinogens such methylcholoanthrene and benzyprene which could induce tumours histologically resembling a range of glioma types with a typical latency of 10 months and a penetrance of around 50% (Seligman et al., 1939, Zimmerman and Arnold, 1943). Interestingly, the form of tumour arising through this method depended on the site of pellet implantation with, for example, medulloblastoma-like tumours forming after cerebellar implantation and ependymoma-like tumours forming after implantation to the ventricular wall. More recently, genetically engineered mouse models (GEMM) were developed combining cre recombination and CRISPR gene editing to introduce disease relevant mutations to more accurately model the human disease.

Early examples of GEMMs included cre mediated deletion of P53 alongside other tumour suppressor genes such as *Nf1, Pten* or *Rb* (Zheng et al., 2008, Chow et al., 2011). Where cre delivery is achieved by either focal viral injections (Jacques et al., 2010) to appropriate brain regions or through transgenic cre expression casettes under control of promoters of specific brain progenitor such as Nestin, GFAP and GLAST (Kwon et al., 2008, Llaguno et al., 2009, Zhu et al., 2005). More refined and temporally controlled models incorporated the tamoxifen inducible CreERT2 trasngene under control of these promoters where experimenters could dictate the onset of cre expression (Benedykcinska et al., 2016b) which was also an advantage of viral cre delivery methods. A notable addition to these models was the development of the IDH1 R132H cre inducible mutation which mimics the most common mutation in human glioma (Sasaki et al., 2012). Further accurate models of the human disease are being introduced with CRISPR methods where human specific mutations can be introduced to progenitor cell compartments of the brain through

viral injections (Robertson et al., 2019). In this study we incorporate the PDGFβ-Ires-Cre (PIC) retroviral model where intraventricular retroviral injections to $P53^{flox/flox}/Pten^{flox/flox}$ mice induce loss of tumour suppressor alleles and overexpression of the potent mitogen PDGFβ in SVZ neural progenitors (Lei et al., 2011, Sonabend et al., 2013, Sonabend et al., 2014). This model reports near 100% tumour penetrance and produces tumour histologically resembling IDH-WT GBM.

As discussed in the previous section about CSCs (Section: 1.2.7), protocols are now available for the derivation of immortal cells from human brain tumours (Galli et al., 2004, Singh et al., 2003). Importantly, these cell types can capture aspects patient specific differences in tumour biology which is considered an important stumbling block in the development of therapies using pre-clinical models (Reifenberger et al., 2017). Beyond *in vitro* cultivation, these cell lines can induce tumour formation when xenografted into the brain of Non-obese diabetic/severe combined immunodeficient mice (NOD/SCID) animals where they recreate histological aspects of the original tumour and aspects of intra-tumor heterogeneity (Singh et al., 2004, Neftel et al., 2019). However, these models are limited in their potential to mimic interactions between tumour cells and the immune system.

Isolated GBM cell lines could be maintained either in adherent culture or as small three-dimensional floating spheroids. Recently, larger scale hydrogel matrix based three-dimensional and organoid culture systems which permit longer-term culture of the cells without passaging have been developed (Hubert et al., 2016). Three-dimensional Matrigel® models have been shown to recreate hypoxic environmental niches while displaying regional heterogeneity in glioma CSC marker expression (Hubert et al., 2016). A further rapid organoid model termed GBM organoids (GBO) omitted the use of Matrigel®, instead, minced tissue of GBM biopsies were cultured on an orbital shaker in normal glioma CSC culture media supplemented with the growth factors FGF and EGF (Jacob et al., 2020). Biopsy fragments grew into near spherical structures with striking histological resemblance to the original tumour and show promise in establishing patient specific tumour vulnerabilities. Further models incorporated methods of cerebral organoid production from human induced pluripotent stem cells (iPSCs) (Lancaster et al., 2013). One approach was the

development of neoplastic cerebral organoids (NeoCORs) where CRISPR editing strategies were employed to induce oncogenic mutations in cells of mature organoids (Ogawa et al., 2018, Bian et al., 2018). A further approach was developed around the observation that co-cultured glioma spheroids could spontaneously attach and infiltrate developing cerebral organoids (da Silva et al., 2018). This approach was refined and the term cerebral organoid glioma (GLICO) was coined showing cells phenocopy patient tumours and show tumour microtube formation which preceded tumour cell invasion (Linkous et al., 2019). In this model, iPSCs could be derived from a patient's dermal fibroblasts and used to culture cerebral organoids which could then be co-cultured with patient matched glioma CSCs to mimic any interactions between tumour cells and parenchymal neuronal cells.

The three-dimensional glioma models are yet to be widely employed in glioma research and still lack large components of the tumour microenvironment that can only be modelled through *in vivo* xenografting. As mentioned, xenografting also omits influences of the immune system on tumour development and therefore syngeneic mouse allografts of transgenic inducible models are especially well suited to studies of immune effects on glioma progression. Nevertheless, these *in vitro* models offer a higher through-put and cheaper alternative to animal models where experimenters can perform drug screens on a patient derived model that more accurately mimics the *in vivo* disease.

1.5 Outline of this study

Given the increasing awareness of GBM intra-tumour heterogeneity and its influences on tumour progression, recurrence and treatment resistance, we aimed to develop fluorescent lineage-tracing approaches that would permit interrogation of clonal heterogeneity within a bulk cell population. In Chapter 3, we aim to develop an *in vivo* based approach combining a retroviral Cre-inducible GEMM with rosa26-confetti based labelling (Lei et al., 2011, Snippert et al., 2010). Hypothesising that confetti labelling will allow identification of different clonal populations produced during tumorigenesis. The GEMM used is a rapid and highly penetrant system which

will be used for exploring suitable laboratory techniques for readouts of clonal behaviour.

In chapter 4, we aim to use LeGO-vectors to establish a clonal tracking technique in primary human cell lines. Hypothesising that LeGO-vector based labelling will allow detection and parallel tracking of a greater number of mixed clonal populations compared with rosa26-confetti labelling. Informed by initial findings in Chapter 3, flow cytometry and fluorescence histology of adherent and three-dimensional Matrigel® spheroids are used to establish an experimental readout combining clonal tracking with CSC surface marker labelling.

Finally, in chapter five we aim to use our developed pipeline of fluorescent tracking and surface marker labelling to understand the relationship between clonal outgrowth and CSC surface marker plasticity in response to changes of culture environment. Testing the hypothesis that clonal expansion drives CSC surface marker plasticity in response to a change in culture environment.

Chapter 2: Methods

2.1 Animal models and experiments

2.1.1 Animal management

All procedures performed on mice were according to Institutional and UK Home Office guidelines (Project license PA79953C0). The ARRIVE guidelines were followed as part of the institutional policy and licensing of all experiments. All mice were kept at the Biological Services Facility, UCL. Mice were housed in racks with a laminar-filtered airflow system under pathogen-free conditions at constant temperature (22°C ± 2°C), with relative humidity (55% ±5%), 12-h dark/light cycles and *ad libitum* access to food and water.

2.1.2 Transgenic mouse lines and genotyping

R26-confetti animals were obtained from Hans Clevers, Hubrecht Institute, Utrecht, Netherlands (Snippert et al., 2010). Mice carrying $P53^{loxP/loxP}$ (Marino et al., 2000) and $Pten^{loxP/loxP}$ (Marino et al., 2002) transgenes were inter-crossed resulting in co-deletion of Pten and P53 upon cre exposure (Jacques et al., 2010, Zhang et al., 2019). GLAST-cre ER(T2) mice were obtained from Magdalena Götz, Helmholtz Centre Munich, Germany. These three genotypes were crossed to produce $GLAST^{CreERT2/CreERT2}$/$Pten^{lox/lox}$/$P53^{lox/lox}$/$ROSA^{confetti/WT}$ herein termed confetti-GBM and used throughout chapter 3 of this thesis. NOD/SCID animals were used for all xenograft experiments using U87 and other primary human cell lines DNA was extracted from mouse earclips and genotyping performed using Phire II polymerase (Thermofisher, F125L) all primers shown in Table 1.

Table 1: Sequences of primers used for genotyping

Gene	Primer sequence	Band Size (BP)
P53^{flox}	Forward - AAGGGGTATGAGGGACAAGG Reverse - GAAGACAGAAAAGGGGAGGG	584
Pten^{flox}	Forward - GGCAAAGAATCTTGGTGTTAC Reverse - GCCTTACCTAGTAAAGCAAG	300
Rosa26-confetti	Forward - GAATTAATTCCGGTATAACTTCG Reverse - CCAGATGACTACCTATCCTC	300
GlastCreERT2	Forward - ATTTGCCTGCATTACCGGTC Reverse - ATCAACGTTTTGTTTTCGGA	400

All sequences shown here are from 5' to 3'

2.1.3 Stereotaxic injections

Intra-cranial/intraventricular injections of PIC retrovirus into neonatal mice were performed at postnatal day (P) 0 – P4 (Lei et al., 2011, Sonabend et al., 2013). Pups were anesthetised with 3 – 4% isoflurane (Isoflurane-Vet, Merial Animal Health Ltd) and oxygen flow rate of 4 L/min during injections while maintained on a heatpad at approximately 25°C during the whole procedure. A 10 µl 26 guage Hamilton syringe with 51mm needle was mounted then washed with 70% EtOH and PBS (Sigma, 84100). Prior to injections retrovirus was mixed with polybrene (Sigma, TR-1003) to a final concentration of 8 µg/ml to facilitate infection. Injections were targeted at 0.5 – 1mm left of the sagittal suture and in line with the centre of the left eye. Once the needle had visibly penetrated the skull, the syringe was inserted a further 1mm and 1ul of virus was injected into the left ventricle. After 5 seconds the needle was removed gradually and the mouse left to recover on a heat pad. Pups were returned once capable of righting themselves or showing visible intent.

Xenografts of U87 and primary human GBM cell lines were performed on NOD/SCID animals at 3 – 9 months of age using a 25 guage Hamilton syringe with a 51mm needle. Cells were prepared as a single cell suspension in sterile PBS (Gibco, 14-190-235) at 1 x 10^7 or 5 x 10^6 cells/ml for primary cell lines or U87 cells respectively. Animals were anaesthetised using a constant flow of isoflurane and 10 µl of cell suspension was injected to the same region as intraventricular injections described above. Since xenografts were performed on adult mice, after penetrating the skull, the needle was inserted a further 2 mm and cells expelled. Animals were given the analgesic carprofen (1:50, 5µl/g)(Fisher scientific: 53716-49-7) after injections were complete and monitored for recovery throughout the rest of the day then daily checks for emergence of neurological symptoms associated with tumour formation.

2.2 Histology

2.2.1 Tissue cryosectioning

Whole brain and organoid tissue were fixed overnight in 4% PFA at 4°C. The following day tissue was washed in PBS (sigma, 84100) and then cryopreserved by 24h incubation in PBS with 15% sucrose followed by approx. 2h incubation in PBS with 15% and 7.5% bovine gelatine (sigma, G9391) at 37°C. Samples were embedded into cryo-moulds on ice then stored at 4°C overnight. Finally, samples were frozen in dry ice cooled isopentane and stored at -80°C. Using a Leica 3050M cryostat, brain tissue was cut to a thickness of 30 µm and organoid tissue to a thickness of 20 µm.

2.2.2 Fixed tissue vibratome sectioning

Whole brain and organoid tissues were fixed overnight in 4% PFA (Thermo-fisher: AAJ19943K2) at 4°C then washed in PBS (Sigma, 84100). For whole brain tissue, samples were mounted by removing cerebellum and gluing tissue directly to vibratome mounts. For organoid tissue, sample were embedded in 1% agarose (sigma A9539) which was allowed to set then cut into a cube for gluing to vibratome mounts. All sectioning was performed on a Vibratome 1500 sectioning system.

2.2.3 Live organoid vibratome sectioning

2% agarose (sigma A9539) in sterile PBS (Gibco, 14-190-235) was prepared by heating agarose/PBS mixture in microwave until all powder had fully dissolved, evaporated PBS was replaced, and mixture was cooled to 45°C (just above gelling temperature) in a water bath. Living mature tumour spheroids were placed into single wells of 24-well plate in 1 ml of freshly warmed 37°C culture media. Using Pasteur pipette, 1ml of 45°C 2% agarose/PBS was rapidly mixed with organoid containing media. Organoids became embedded as mixture solidifies with well serving as mould. Solidified moulds were removed from well, dried and glued to vibratome chucks. Organoids were then cut at 200 – 500 µm on Vibratome 1500 sectioning system and collected back into media.

2.2.4 Clarity tissue clearing

Passive clarity procedure was employed for this aspect of the project (Tomer et al., 2014, Yang et al., 2014), reagents described in table 2. Whole brain tissue was collected and fixed overnight in 4% PFA at 4°C. Brains were then sectioned with vibratome to the desired thickness and stored overnight at 4°C in hydrogel polymer solution (HPS). The following day, 7ml tubes containing sections and HPS were degassed by gently bubbling NO_2 though the solution for 15mins. Bottles were immediately sealed with parafilm and samples polymerised at 37°C for 4 hours. After 4 hours, samples were washed thoroughly in PBS then transferred to a 50ml falcon tube and incubated at 37°C in 8% SDS solution until all lipid had diffused from the tissue. Once cleared, sections were washed 4 x in PBS for 20 mins at 37°C. Finally, samples were incubated in refractive index matching solution (RIMS) overnight and imaged first thing in the morning.

Table 2: Reagents for Passive clarity	
Hydrogel monomer solution	4% Acrylamide (Bio-rad, 1610140) 0.25% Azo-initiator (Wako, VA-044) Dissolved in 1 x PBS (Sigma, P4417)
Detergent solution	8 % Sodium dodecyl Sulphate (ACROS, 230425000) Dissolved in 1 x PBS (Sigma, P4417)
Refractive Index Matching Solution (RIMS)	80% Histodenz™ (Sigma, D2158) 0.1% tween-20 (Sigma, P1379 0.01% Sodium Azide(Sigma S200-2)

2.3 Cell and tissue culture

2.3.1 Cell types and maintenance

The use of HEK 293 platinum E, SVZ isolated mouse NSCs, neurosphere derived mouse brain tumour cells and human glioma CSCs are described in this project. Mouse NSCs and mouse cancer cells were maintained on poly-lysine (Sigma, P6282) and laminin (Sigma, L2020) coated plates. While human glioma CSCs were maintained on laminin coated plates (Sigma, L2020). All cells were maintained at 37°C, 5% CO2, 90% humidity and in appropriate media outlined in table 3. Medium was changed every 48h unless stated otherwise. Passaging was performed when cells were ~90% confluent and were detached using Accutase™ (Sigma, A6964).

Two patient cell lines, designated G19 and G61, were used extensively throughout Chapter 5 of this study. Cell line G19 is derived from the GBM of a 65 year old female. The tumour is wild-type on IDH1 and IDH2 genes and carries a TERT promoter mutation (C228T). The copy number profile shows MYCN and PDGFRA amplification, and chromosome 7 gain. Methylation profiling yielded the methylation class of GBM, IDH-wt, subclass RTK I. Cell line G61 is derived from the GBM of a 66 year old female. The parent tumour showed no mutations on the IDH1 and IDH2 genes, and a TERT

promoter mutation present (C228T). The copy number profile shows a CDKN2A/B homozygous deletion, and chromosome 7 and 10 loss. Methylation profiling yields the methylation class GBM, IDH-wt, subclass RTK II.

2.3.2 Neurosphere derivation and tissue dissociation

For neural stem cell isolation, mice at postnatal age 7 days (P7) were culled by removal of the head and whole brain was sampled using autoclaved dissection instruments then stored on ice. For collecting PIC retroviral induced cancer cells from adult mice, animals were culled by rising CO2 concentrations, followed by dislocation of the neck before removing the whole brain. The remaining procedure was carried out in a laminar flow tissue culture hood with autoclaved sterile dissection instruments. Sampled whole brains were submerged in 70% EtOH for 5 – 10s then washed in PBS. When collecting P7 NSCs, non-forebrain regions were then removed. When collecting PIC induced brain tumour cells for *in vitro* culture or FACS analysis, non-tumour regions were discarded. Forebrain or tumour regions were then dissociated using the Worthington Papain tissue dissociation kit (Cat No. LK003150) as per the manufacturer's protocol. Dissociated tissue was passed through a 70 µm filter before plating onto an uncoated 10cm dish. After 5 – 7 days, NSCs form large floating spheroid structures while differentiated cells are dead or adherent. Spheroids were collected by extracting medium and dissociation in Accutase solution® (Sigma, A6964). Dissociated spheroids were stored in liquid nitrogen after suspension in serum free freezing medium (Millipore, SCM014).

Table 3: Tissue culture media

Table 3: Base formulas for cell culture media	
Cells	Media
Platinum E HEK293	Dulbueco's Modified Eagle Medium (Lonza, 12-604F) 10% Foetal Bovine Serum (Sigma, F2442) 1% Penicillin and Streptomycin (Sigma P4333)
Neural Stem Cells &	Dulbueco's Modified Eagle Medium/F12 (Lonza, 12-719F)

Cancer stem cells	B27 Supplement (Gibco, 17504-044)
	20 ng/ml Epidermal Growth Factor (PeproTech, 315-09)
	20 ng/ml Fibroblast Growth Factor (PeproTech, 100-18B)
	1 % Penicillin and Streptomycin (Sigma, P4333)

2.3.3 Retro and lentiviral production

The PIC retroviral construct was kindly provided by Prof. Peter Canoll (Columbia University Medical Center, USA,) and the use of this retrovirus to generate gliomas have been described previously (Lei et al., 2011, Sonabend et al., 2013). Fluorescent lentiviral gene ontology vectors (LeGO)-vectors were described previously and available through addgene® (Weber et al., 2008). Retrovirus production was carried out in Platinum E cells where transduction mixtures containing plasmid DNA, Fugene® and opti-mem were added to a 60% confluent plate and left overnight at 37°C. For fluorophore and shRNA expressing lentivirus production, 2nd generation packaging plasmid psPAX2 (addgene: 12260) and envelope expressing plasmid pMD2.G (addgene: 12259) were used. Briefly, packaging and envelope plasmids were mixed with fluorescent expressing plasmids and Lipofectamine 2000 ® reagent before being added to a 60% confluent plate of HEK293T cells and left overnight at 37°C. For both virus preparations, transduction media were replaced with fresh culture medium and cells incubated for a further 24h at 37°C. After 24h, culture medium was extracted into a syringe, passed through 0.45ul filter (Millipore: HVLP09050) and mixed with Retro X concentrator (Clontech, PT5063-2) at a 1:3 (concentrator : Medium). This mix was incubated for 16 – 20 h at 4^{o}C and the centrifuged at 1500g and 4^{o}C for 1h and 30 mins. The supernatant was extracted, and the pellet resuspended in opti-mem®.

2.3.4 PDGFβ-IRES-Cre retrovirus titre estimation

0.1 x 10^{6} unrecombined ROSA26-confetti NSCs were seeded into coated wells of a 24-well plate and cultured for 24h (n = 3 repeats per test). After 24h, a mixture containing 5 µl of retrovirus, polybrene (Sigma, TR-1003) (concentration 8 µg/ml) and 1 ml of appropriate culture medium was added to each well. Control wells received

culture medium with 8 µg/ml polybrene (Sigma, TR-1003) and no virus. Cells were incubated in this mixture for 24h then washed with PBS before adding regular culture media. After a further 48h culture, each test well was washed with PBS and cells were fixed by 10 mins incubation in 4% PFA. Cells were then incubated in 1 µg/ml Hoechst 33342 (Invitrogen, H3570) for 20 mins. Finally, cells were washed in PBS then imaged on an inverted Zeiss LSM710 confocal microscope. For each test well, 10 regions were imaged (10x magnification) and the proportion of cells expressing confetti fluorophores was calculated. Adjudged by the recombination efficiency of a previous viral preparation achieving near 100% tumour penetrance, viral preparations with > 2.5% of cells recombined were considered of sufficient titre for successful *in vivo* injections.

2.3.5 Fluorescent LeGO vector titre estimation

Fluorescent lentivirus titres show variability with different cell lines and were performed in the same fashion on either HEK293T cells or primary human glioma CSCs. 5 x infection mixtures were produced containing 8µg/ml polybrene in 500 µl of culture media with either 1, 1 x 10^{-1}, 1 x 10^{-2}, 1 x 10^{-3} or 1 x 10^{-4} µl of concentrated virus. These mixtures were added to separate wells of a 24-well plate containing 5 x 10^4 cells seeded 24h earlier. Cells were incubated with virus mixtures for 4h and were then washed 1x with PBS and replenished with appropriate media. Proportion of infected cells was then determined by measuring fluorescence on flow cytometer (cytoflex-s, Beckmann coulter). Flow readings were taken at 72h post infection for HEk293T cells and 120h post infection for primary human glioma CSCs. Reactions that resulted in 5 – 30% of cells infected were used to calculate titre as these values fall within the linear range of infection outcomes. Titre in PFU was then calculated using the following formula.

$T = N \times P/V$

Where T is titre (PFU), N is number of cells seeded (5 x 10^4), P is proportion of transduced cells (e.g 18% = 0.18) and V is volume of virus added (ml, e.g 1 x 10^{-3} µl = 1 x 10^{-6} ml)

2.3.6 NFIX shRNA expressing lentiviral titre estimation

Initial knock-down (KD) validation of NFIX shRNA lentivectors were carried out with a construct containing expression for a fluorescent reporter. As such, these titre estimations were carried out as described for fluorescent lentiviruses above. For use with fluorescent tracing a vector containing no fluorescent reporter was used. Briefly, a range of viruses preparation volumes (10, 1, 1 x 10^{-1}, 1 x 10^{-2} and 1 x 10^{-3} µl) were added to 5 x 10^{4} human glioma CSCs with 8µg/ml polybrene in appropriate culture media. Virus was incubated with cells for 4h, cells were washed 1 x with PBS and fresh media added. After 120h, cells were immunostained for NFIX (1:1000)(NOVUSbio, NBP2-58904) allowing detection of cells with KD which showed dimmer staining. The proportion of dim cells was used to estimate the number of cells infected and titres were calculated using the equations described in the section above.

2.3.7 Adeno-cre and 4-OH tamoxifen *in vitro* recombination

4-OH tamoxifen (Sigma, H7904) was prepared as a 5 mM stock solution dissolved in ethanol and stored at -20°C. Stock was dissolved in media to 5 µM, 2.5 µM, 1.25 µM and 6.25 x 10^{-1} µM. As above, unrecombined R26-confetti NSCs were seeded in a 24-well plate and cultured for 24h. Media was replaced with 4-OH tamoxifen supplemented media which was replaced every 24h for 72h. After 72h incubation in 4-OH tamoxifen, cells were washed, fixed, stained and imaged as described for the retroviral titre test. Adeno-cre virus was prepared in HEK293 cells. Unrecombined R26-confetti NSCs were plated in a 24-well plate and cultured for 24h as described above. Adeno-cre virus (10^{10} IFU/ml) was diluted to 5 x 10^{-3}, 1 x 10^{-4}, 2 x 10^{-4}, 4 x 10^{-4} and 8 x 10^{-4} in culture medium. Cells were incubated with virus supplemented media for 2h then washed and replaced with normal medium. After 72h, cells were washed, fixed, stained and imaged as described for the retroviral titre test.

2.3.8 Fluorescent labelling, RGB marking and dual-labelling

Selection of fluorophore combinations is dependent on experimental application and is described in results for specific experiments. In the simplest labelling regime three

un-titred LeGO-vectors were added to cell culture media at equal dilution with 8 µg/ml polybrene and incubated with cells for 4h before washing cells and replacing media. Labelling was inspected by confocal microscopy at 72 – 120h post infection. For RGB marking, volumes required to add equal amounts of three separate virus preparations were calculated using the following equation based on mathematical set theory.

$V = 0.7 \times N / T$

Where, *V* is virus preparation volume (ml), *N* is number of seeded cells and *T* is the virus titre (PFU). 0.7 denotes 0.7 virus particles per seeded cell which according to poisson distribution is required for a transduction rate of 50%. This is a theoretical value and actually obtained transduction rates will be lower. To achieve optimal label distribution a transduction rate of 60 – 70% per vector is required. Multiple infections adding 2x, 4x and 8x the volume calculated for a theoretical transduction outcome of 50% were added and flow cytometry used to determine which infection produced the optimal transduction outcome per each of the three vectors used.

For dual-labelling, four (section: 5.2.1) or five (Section: 4.2.6) fluorescent LeGO-vectors were used to stain a cell population. Separate dual-lentiviral infections were carried out, comprising all possible combinations of applied lentiviral labels. Using five vectors required a total of 10 dual-infections and 4 vectors a total of 6 dual-infections. Volumes of concentrated virus added to each infection were calculated using the equation described for RGB marking. To achieve equal mixtures of single and double positive cells, virus volumes were calculated with the aim of achieving ~75% transduction rate in each separate reaction. 120h after viral infection separately labelled populations were mixed and colour compositions assessed using flow cytometry. Alternatively, in latter experiments flow cytometry was used to collect only the cells expressing two fluorophores. In this context, achieving a 75% infection per vector added was not required as single positives were removed during sorting process.

2.3.9 Matrigel® tumour organoid culture

Organoids were cultured as previously described using either murine confetti-GBM cells or primary human glioma CSCs. Aliquoted Matrigel® (Corning, 356234) was thawed on ice overnight and pipette tips were stored at -80°C for handling Matrigel®. For seeding 5000 cells per organoid a 1 x 10^6 cells/ml suspension of cells was prepared using an automated cell counter (Millipore, PHCC0000). Organoid moulds were created by pressing a piece of parafilm between two autoclaved PCR plates. Using cooled pipette tips cell suspension was thoroughly mixed with Matrigel at a ratio of 1 : 4 (cell suspension : Matrigel). Still using cooled pipette tips, 20 µl drops of the cell/Matrigel mixture were added to the moulds. Moulds were covered and incubated for 2h at 37°C to facilitate matrix gelling. Seeded organoids were removed from moulds then cultured individually in 12-well plates in appropriate cell culture media with orbital shaking at 80RPM.

2.4 Plasmid cloning

All plasmids used in this study are outlined in table 4. Experimental plasmids for PIC retrovirus, lentiviral LeGO-vectors, NFIX shRNA, viral capsid and packaging vectors were acquired either as pre-transformed bacterial stabs or purified plasmid. XL10-Gold® Ultra competent cells (Stratagene, 200314) were transformed and selected for with appropriate antibiotic after innoculation on LB agar (Sigma: L7025-100TAB) plates. A single clonal colony was selected and transferred to mini culture in 5ml Lb broth (Sigma: L3522) with appropriate antibiotic selection. When vectors were supplied in a pre-transformed bacteria stab, they were added straight into a 5ml LB broth mini-culture. 100 µl of mini-prep cultures were then transferred to a 100ml midi-prep. After 18h of culture, bacteria pellets were collected and plasmid DNA purified with QIAGEN® CompactPrep Plasmid Midi Kit (QIAGEN, 12843) as per the manufacturer's instructions and purity assessed on a ND-800 spectrophotometer (Nanodrop, Delaware, United states).

Table 4: Summary of all plasmids used in this study

Vector	Source	Bacteria selection	Mamallian selection	Purpose	Promoter	Fluorescent reporter
PIC	Professor Peter Canoll, Columbia University	Ampicillin	NA	Retroviral PDGFβ and Cre expression		NA
PMD2.G	Addgen: 12259	Ampicillin	NA	Lentivrial capsid		NA
psPAX2	Addgene: 12260	Ampicillin	NA	Lentiviral packaging		NA
shNFIX-1	Dharmacon SMARTvector®	Ampicillin	Puromycin	shRNA expression	hCMV	TurboRFP
shNFIX-2	Dharmacon SMARTvector®	Ampicillin	Puromycin	shRNA expression	hCMV	TurboRFP
shNFIX-Scramble	Dharmacon SMARTvector®	Ampicillin	Puromycin	shRNA expression	hCMV	TurboRFP
LeGO-EBFP2	Addgene: 12260	Ampicillin	NA	lentiviral fluorophore expression	SFFV	EBFP2
LeGO-S2	Addgene: 85211	Ampicillin	NA	lentiviral fluorophore expression	SFFV	T-sapphire2
LeGO-G2	Addgene: 25917	Ampicillin	NA	lentiviral fluorophore exression	SFFV	EGFP
LeGO-V2	Addgene: 27340	Ampicillin	NA	lentiviral fluorophore expression	SFFV	Venus
LeGO-mOrange2	Addgene: 85212	Ampicillin	NA	lentiviral fluorophore expression	SFFV	mOrange2
LeGO-dKatushka2	Addgene:	Ampicillin	NA	lentiviral fluorophore expression	SFFV	dKatushka2

2.5 Flow cytometry

2.5.1 Adherent cell preparation and marker staining

Adherent cells were soaked in appropriate volume of Accutase® solution (Sigma, A6964 until all cells had visibly detached but no longer than 10 minutes. Accutase® solution was deactivated using 5ml of sterile PBS and a single cell suspension achieved through repeated pipetting. Cells were then pelleted by centrifugation at 200 RCF

then washed by resuspending in flowcytometry staining buffer (FSB) with vortex followed by centrifugation (hereby: washed through centrifugation). Cells were split into experimental reaction mixtures in 1.5 ml Eppendorf's, re-pelleted and supernatant aspirated. Pellets were broken apart using a vortex at 1,500 RPM and resuspended cells incubated for 30 mins at 4 °C in 50 µl of Fc receptor blocking solution (10 µl FC receptor block + 40 µl FSB). After 30 mins, samples were resuspended using vortex and 50ul of appropriate volumes of antibody mixture was added and reactions incubated for a further 30 mins at 4 °C. For each antibody and cell line used a single stain control sample was carried out. After 30 mins, reaction mixtures were flushed with 1 ml of FSB and washed by centrifugation. This process was repeated two further times and final sample resuspended in 200 µl of FSB. 200 µl of each reaction were loaded into individual wells of round bottom 96-well plates and samples assessed on a cytoflex-s cell analyser (Beckman coulter, High Wycombe, Buckinghamshire, UK). Owing to incomplete aspiration of supernatant after final wash, each reaction contained 10 – 20 µl of residual cell suspension after loading 200 µl for analysis. These residual cells were pooled together from each sample and used as a calibration sample to set laser gains for each experiment. A calibration sample was required for each cell line being acquired because fluorophore brightness would vary between cell lines.

2.5.2 Three-dimensional tumour spheroid preparation and marker staining

Structural integrity of whole or sectioned tumour spheroids was first disrupted through mechanical dissociation with a P1000 or P200 pipette tip. Spheroid tissue was then enzymatically dissociated by incubation in 750 µl of Accumax® solution for 20 mins at room temperature. After 20 mins, dissociation was aided by repeated pipetting of tissue and tissue left for a further 10 mins. Quality of dissociation was assessed under a light microscope and generally one further pipetting step and 10 mins incubation yielded adequate dissociation. If a small number of single cells had been achieved after 40 mins Accumax® incubation, tissue would be incubated for longer but never beyond a total incubation time of 1h. To remove any residual tissue clumps dissociated tissue was passed through a 40 µm cell strainer into a 1.5 ml

Eppendorf. Strained dissociated cells were then washed 1x by centrifugation. Fc receptor blocking and surface marker staining was then carried out as described for adherent cells.

2.5.3 Surface marker phenotyping cytometer setup and compensation

All surface marker phenotyping and flow cytometry-based lineage-tracing was performed on a Beckman-coulter cytoflex-S cell analyser. All endogenous fluorophores, fluorescent stains, antibodies, conjugated fluorophores, dilutions, lasers and bandpass filters used for analysis are shown in Table 5. To set laser parameters, a calibration sample containing a mixture of all lentiviral labels and surface marker stains for a given experiment was first analysed. To achieve optimal laser parameters, a calibration sample was required for each cell line being assessed as fluorescent signal intensity varied between cell lines. Firstly, the calibration sample was plotted on FSC-A/SSC-A dot-plots to remove debris and dead cells from analysis. Single cells were subsequently separated from cell doublets by plotting detections on FSC-W/FSC-H dot-plots. Once single cells were isolated, histograms were used to set laser gains for each fluorescent signal being detected. Laser power was set such that the brightest population for each fluorescence signal was within the limit of detection axis and well separated from other dimmer populations. This process of setting laser parameter was repeated for each cell line being analysed and laser gains were then applied to all test samples for that cell line. After setting laser powers, samples were resuspended and ran in batches of five, resuspending the next 5 samples after the first batch was completed. Sample flow speed was set to reach 2000 event/s or an abort rate of below 5% and entire sample (200 µl) is recorded. As stated above, single surface marker stain controls were produced, and a mixture of these single stain controls was used to set compensation. Single surface marker stain controls were produced for every experiment and each different cell line assessed in that experiment. Single lentiviral label controls were only used during preliminary experiments which determined suitable lentiviral label combinations.

Table 5: List of laser and filter sets used for detecting fluorescent probes				
Fluorescent Probe	**Application**	**Epitope**	**Laser (nm)**	**Bandpass (nm)**
ECFP	r26-confetti	NA	405	525/30
EYFP	r26-confetti	NA	488	550/40
GFP	r26-confetti	NA	488	510/30
tdimer2	r26-confetti	NA	561	610/20
EBFP2	LeGO-vector	NA	405	450/45
T-sapphire	LeGO-vector	NA	405	525/40
EGFP	LeGO-vector	NA	488	525/40
Venus	LeGO-vector	NA	488	525/40
mOrange2	LeGO-vector	NA	561	585/42
Katushka2S	LeGO-vector	NA	633	690/50
APC	surface marker	A2B5	633	660/10
APC-vio® 770	surface marker	CD44	633	780/60
PE-vio® 770	surface marker	CD133	561	780/60
PerCP-vio® 700	surface marker	CD15	488	690/50
BV-780	surface marker	CD15	405	780/60
BV-711	surface marker	CD133	405	712/25
Ghost dye 710	Cell viability	NA	633	712/25

2.5.4 Confetti label flow cytometry gating

Lasers and filtersets used for all flow cytometry and FACS experiments are shown in Table 3. Dissociated ROSA26-Confetti labelled tumuor cells we separated from clumps and dead cells through gating on dotplots of side scatter area (SSC-A) and forward scatter area (FSC-A). Singe cells were then separated from doublets using forward scatter width (FSC-W) and forward scatter height (FSC-H). Isolated single cells were then plotted in two separate plot across four collection parameters;

1. Blue laser 510/530 (GFP) and blue laser 550/40 (EYFP)
2. Yellow laser 610/20 (RFP) and Violet laser 450/45 (eCFP)

After applying compensation, automatic gating was performed to identify populations corresponding to one of the 4 markign fluorophores (eCFP, EGFP, eYFP and RFP)

2.5.5 RGB marking fluorescent barcode flow cytometry gating

Lasers and filter-sets used for all flow cytometry and FACS experiments are shown in Table 3. Multiple Fluorophore combinations are used and are indicated in results, here, we use Fluorophore A, B, C, etc to describe gating strategy for each labelling approach. Cells we separated from clumps and dead cells through gating on dot-plots of side-scatter area (SSC-A) and forward scatter area (SSC-A). Single cells were then separated from doublets using forward scatter width (FSC-W) and forward scatter height (FSC-H). Flow analysis of RGB marked cells was achieved by testing effective separation of underlying fluorescent signals in control cells stained for a single fluorophore. Single colour controls were also used to setup laser values and compensation values for effective separation of single, double and triple labelled cells in a mixed population. Important to note that cells exposed to higher virus loads will be labelled at a higher average brightness. Gates were applied to results for mixtures of singly labelled cells and boolean logic used to identify all 8 different label outcomes in an RGB marked population. Briefly, avoiding cell detections for single positive cells, gates were drawn to isolate the three dual labelled populations from the single cells on plots with axis combinations:

0. X-axis/Y-axis
1. Fluorophore A/Fluorophore B (Dual label A + B)
2. Fluorophore A/Fluorophore C (Dual label A + C)
3. Fluorophore B/Fluorophore C (Dual label B + C)

Triple label cells were excluded form dual label counts by excluding cells that occur in more than one dual label gate. Triple label cells were then counted by detecting cells that occur in all three of these dual label gates. All cells not captured by the three gates above were potted on:

1. SSC-A/Fluorophore A
2. SSC-A/Fluorophore B
3. SSC-A/Fluorophore C

Gates were drawn around cells showing expression of a single tracing label. Single gate drawn around label negative population on plot showing SSC-A/Fluorophore A.

Cells in this gate, excluding those in positive gates for Fluorophore B and Fluorophore C were considered the isolated negative cells.

For dual-label optical barcoding with five fluorophores (A. B, C D and E), single colour controls were produced by labelling cells with a single fluorophore at the same viral load as the dual label reactions to achieve comparable brightness of fluorophores in controls and target cells. Isolation of single cells was performed in the same fashion was done for RGB marking. Gates for isolation and quantification of all 15 label groups and unlabelled cells was performed by initially plotting gates on single colour control data on the following 10 dotplots. With a gate drawn around quadrant capturing cells positive for both markers.

0. X-axis/Y-axis
1. Fluorophore A/Fluorophore B
2. Fluorophore A/Fluorophore C
3. Fluorophore A/Fluorophore D
4. Fluorophore A/Fluorophore E
5. Fluorophore B/Fluorophore C
6. Fluorophore B/Fluorophore D
7. Fluorophore B/Fluorophore E
8. Fluorophore C/Fluorophore D
9. Fluorophore C/Fluorophore E
10. Fluorophore D/Fluorophore E

Cells not captured by any of these gates were then plotted on the following dot-plots.

1. X-axis/Y-axis
2. SSC-A/Fluorophore A
3. SSC-A/Fluorophore B
4. SSC-A/Fluorophore C
5. SSC-A/Fluorophore D
6. SSC-A/Fluorophore E

Remaining cells positive for any of these markers were gated and single gate for negative cells drawn on the SSC-A/Fluorophore A plot. As with RGB marking, Boolean

logic used to isolates negative cells by excluding any cells found in the other four single positive gates. Any cells found to be positive for 3 or more fluorophores were not included for analysis as these were likely cell doublets carrying two different labels. Colour proportions described in results are determined as a proportion of all identified labelled cell populations.

2.5.6 Dual-label barcoding and adjunct surface marker staining flow analysis

Dual-label barcoded cells (Figure 4.13) were labelled with antibody conjugates targeting CD44 (APC780), CD133 (BV711), CD15 (BV780), A2B5 (APC) all together and as single stain controls as described in 2.5.1 and 2.5.2. Single stain controls were mixed and used to set laser power and compensation for removal of bleed through between channels. Once compensation was completed, single cells were identified by gating as described in 2.5.5 and 2.5.4. In some experiments (indicated in results) the viability dye Ghost Dye™ Red 710 (Cell signalling) was used to remove non-viable cells from further analysis. This is a free amine binding dye and non-viable cells with compromised membranes will readily react with this dye appearing brighter in the detection channel. After gating for single viable cells our dual label cell populations were first identified by plotting single viable cells on the following 6 dotplots also displayed in Figure 2.1.

(X-axis/y-axis)

1. EBFP2/T-sapphire (BS)
2. EBFP2/Venus (BV)
3. EBFP2/mOrange2 (BO)
4. T-sapphire/Venus (SV)
5. T-sapphire/mOrange2 (SO)
6. Venus/mOrange2 (VO)

These plots are shown regularly throughout results section and are labelled with the codes BS, BV, BO, SV, SO, and VO where the first letter indicates the fluorophore detected on x-axis and second letter the fluorophore detected on y-axis Figure 2.1. These plots reveal the six dual-labelled trace populations which are boxed for as shown in Figure 2.1C. At this stage some detections are reported as positive for more

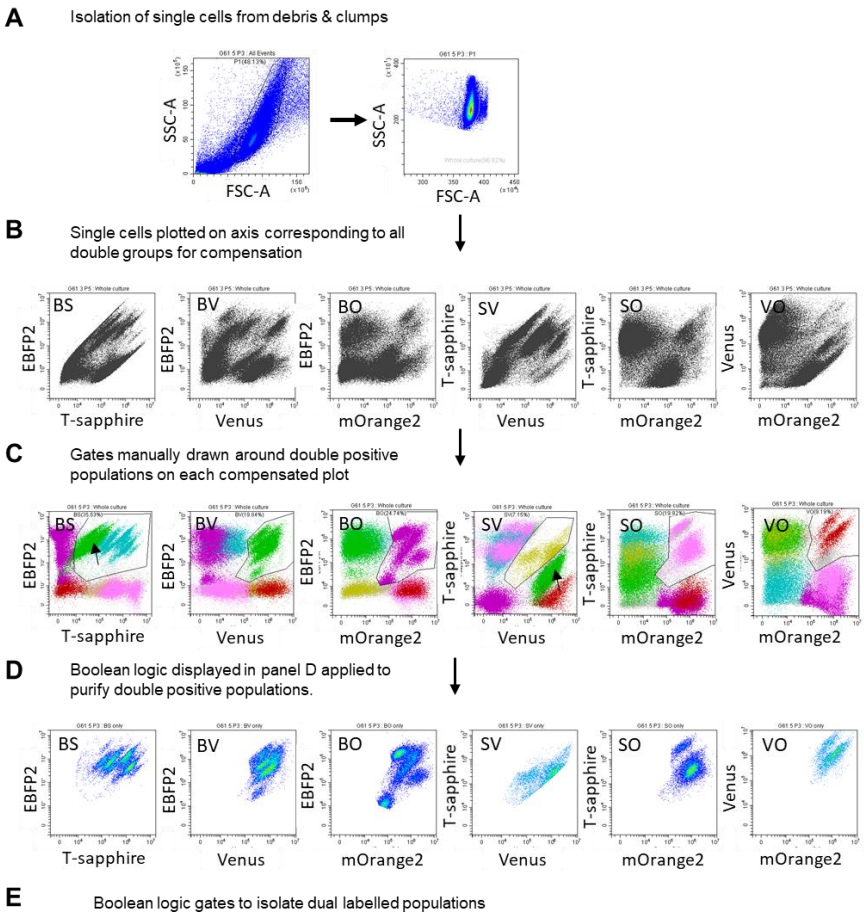
than two fluorophore. These cells were removed from quantification by applying the Boolean logic shown in Figure 2.1E. Furthermore, the EBFP2/Venus (label code BV and displayed as green in dot-plots/bargraphs) dual labelled cells produce an incorrectly positive signal in the t-sapphire detection channel (405 laser & 525/40 filter) that could not be corrected through compensation. We reasoned this was caused by a Förster resonance energy transfer (FRET) between EBFP and Venus proteins. Briefly, the 405nm laser (used in t-sapphire detection channel) will also excite EBFP2 which emits photons with wavelengths in the range of Venus excitation. Venus and t-sapphire have highly similar emission spectra and this Venus excitation caused through FRET will be detected in the t-sapphire channel as indicated by arrows in Figure 2.1C. This issue was incoorperated into our analysis pipeline and was largely circumvented when applying Boolean logic is to purify the populations.

Once all dual label populations are identified, gating for delineating all 16 cell surface marker phenotypes was performed. Initially, taking the gated single cell population, cells are plotted on the following three dot-plots as shown in Figure 2.2B-D and gates setup to detect the six populations outlined below.

x-axis/y-axis – (gates drawn)

1. CD44/A2B5 – quad gate
    a. Lower left gate: $CD44^-/A2B5^-$
    b. Upper left gate: $A2B5^+$
    c. Upper right gate: $CD44^+/A2B5^+$
    d. Lower right gate: $CD44^+$
2. CD44/CD133 – single box gate
    a. Box gate: $CD133^+$
3. CD44/CD15 – single box gate
    a. Box gate: $CD15^+$

Once gating is completed on single stain controls these gates are then applied to all test samples of the corresponding cell line. To detect all 16 different surface marker phenotypes from the above 6 gated populations, Boolean logic was again used as shown in Figure 2.2E. Finally, once these sixteen surface marker populations were achieved marker analysis of any dual label population (BS, BV, BO, SV, SO or VO) or



BS only= **BS** AND NOT (**BV**) AND NOT (**SV**) AND NOT (**BO**) AND NOT (**SO**) AND NOT (**VO**)

BV only = NOT (**VO**) AND (**BV**) AND NOT (**SO**) AND NOT (**SV**) AND NOT (**VO**)

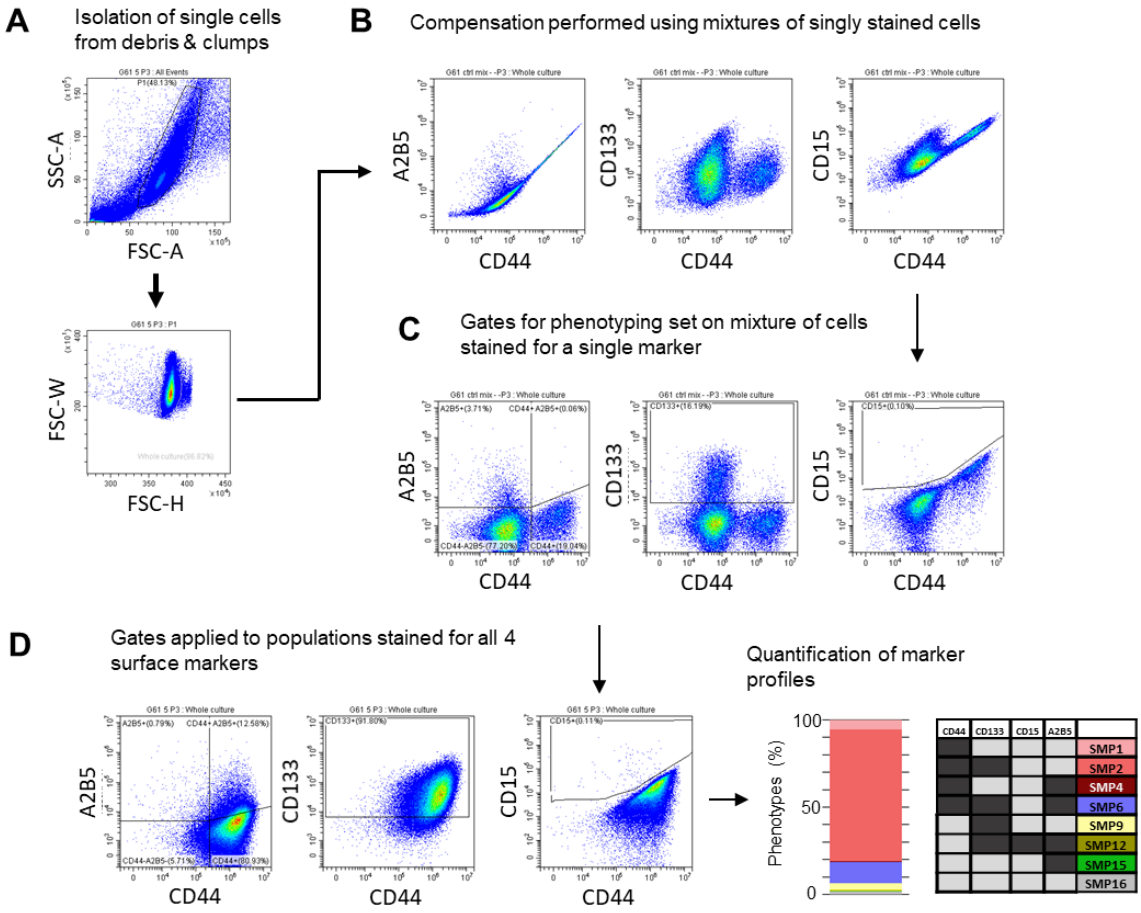
BO only = NOT (**BS**) AND NOT (**BV**) AND **BO** AND NOT (**SV**) AND NOT (**SO**) AND NOT (**VO**)

SV only = NOT (**BS**) AND NOT (**BV**) AND NOT (**PO**) AND **SV** AND NOT (**SO**) AND NOT (**VO**)

SO only = NOT (**BS**) AND NOT (**BV**) AND NOT (**PO**) AND NOT (**SV**) AND **SO** AND NOT (**VO**)

VO only = NOT (**BS**) AND NOT (**BV**) AND NOT (**PO**) AND NOT (**SV**) AND NOT (**SO**) AND **VO**

Figure 2.1: Gating to isolate cell detections corresponding to double labelled cells. A, side scatter area (SSC-A) and forward scatter area (FSC-A) were used to remove dead cells and debris in the preparation. This P1 population was then plotted on FSC-height (H) and FSC-width (W) to remove any possible doublet cells. B, Purified cell detections were then plotted on flow plot axis permitting detection of each of the six double positive populations (BS, BV, BO, SV, SO, VO) and compensation applied to the cells. Pilot experiments using single colour control cells (B, S, V, O) were performed to determine sufficient separation of fluorescent signals. C, After compensation, gates were drawn to isolate the dual labelled populations that contain clonal streaks we aim to trace. Arrows in BS and SV plots show that even after compensation, the double positive population BV (green) bleeds through considerably into the t-sapphire channel. This is caused by Fluorescence Resonance Energy Transfer (FRET) whereby 405 laser exciting t-sapphire will also excite EBFP2 in the BV population. EBFP2 emission in turn excites Venus which can also be detected in the 525/40 nm filter used for T-sapphire. Effects of this on accurate detection of double labelled populations have been mitigated by stringent gating and separation through Boolean logic gates D. Finalised double positive populations achieved through Boolean logic to remove any events that are detected across more than one double positive gate. This population is used to pull events belonging to streaks/clusters. D. Boolean logic gates to achieve final populations for tracing. Note, for BV the gate “not BS” is omitted because of FRET described earlier.



E Boolean logic to isolate 16 different phenotypic populations applied to samples stained for all 4 fluorophores

CD44+ only =	NOT(CD44+ A2B5+) AND NOT(A2B5+) AND NOT(CD44-A2B5-) AND CD44+ AND NOT(CD133+) AND NOT(CD15+)
CD44+ CD133+ =	NOT(CD44+ A2B5+) AND NOT(A2B5+) AND NOT(CD44-A2B5-) AND CD44+ AND CD133+ AND NOT(CD15+)
CD44+ CD15+ =	NOT(CD44+ A2B5+) AND NOT(A2B5+) AND NOT(CD44-A2BS-) AND CD44+ AND NOT(CD133+) AND CD15+
CD44+ A2B5+ =	NOT(A2B5+) AND CD44+ A2B5+ AND NOT(CD44-A2B5-) AND NOT(CD44+) AND NOT(CD133+) AND NOT(CD15+)
CD44+ CD133+ CD15+ =	NOT(CD44+ A2B5+) AND NOT(A2B5+) AND NOT(CD44-A2B5-) AND CD44+ AND CD133+ AND CD15+
CD44+ CD133+ A2B5+CD44+ =	A2B5+ AND NOT(A2B5+) AND NOT(CD44-A2B5-) AND NOT(CD44+) AND CD133+ AND NOT(CD15+)
CD44+ CD15+ A2B5+ =	CD15+ AND NOT(CD133+) AND NOT(CD44-A2B5-) AND NOT(A2B5+) AND CD44+ A2B5+ AND NOT(CD44+)
CD44+ CD133+ CD15+ A2B5+ =	NOT(CD44-A2B5-) AND NOT(CD44+) AND CD133+ AND CD15+ AND NOT(A2B5+) AND CD44+ A2B5+
CD133+ only =	NOT(CD15+) AND CD44-A2B5- AND NOT(A2B5+) AND NOT(CD44+ A2B5+) AND NOT(CD44+) AND CD133+
CD133+ CD15+ =	CD133+ AND CD15+ AND NOT(CD44+) AND CD44-A2B5- AND NOT(A2B5+) AND NOT(CD44+ A2B5+)
CD133+ A2B5+ =	A2B5+ AND NOT(CD44+ A2B5+) AND NOT(CD44-A2B5-) AND NOT(CD44+) AND CD133+ AND NOT(CD15+)
CD133+ CD15+ A2B5+ =	NOT(CD44+ A2B5+) AND A2B5+ AND NOT(CD44-A2135-) AND NOT(CD44+) AND CD133+ AND CD15+
CD15+ only =	NOT(CD44+ A2B5+) AND NOT(A2B5+) AND NOT(CD44+) AND CD44-A2B5- AND NOT(CD133+) AND CD15+
CD15+ A2B5+ =	NOT(CD44+ A2B5+) AND A2B5+ AND NOT(CD44-A2B5-) AND NOT(CD44+) AND CD15+ AND NOT(CD133+)
A2B5+ only =	NOT(CD44+ A2B5+) AND NOT(CD44-A2B5-) AND A2B5+ AND NOT(CD44+) AND NOT(CD133+) AND NOT(CD15+)
negative =	NOT(CD44+ A2B5+) AND NOT(A2B5+) AND NOT(CD44+) AND CD44-A2B5- AND NOT(CD133+) AND NOT(CD15+)

Figure 2.2: Gating to phenotype of stained cell populations. A, For all samples, SSC-A and FSC-A were used to remove dead cells and debris in the preparation. This P1 population was then plotted on FSC-H and FSC-W to remove any possible attached cell. B, A mixture of cells either unstained or stained for one of the four phenotyping antibodies (CD44, CD133, CD15 or A2B5) were used for compensation. Uncompensated cells are shown in panel B. C, Single-stain mixture after compensation, this population was used to designate 6 gated populations from which Boolean logic gates could be used to calculate proportions of all 16 surface marker profiles. D, Quantification of marker profiles for this data is shown in graph on right. The legend shows the surface marker profiles (SMP) represented in the graph, please refer to Fig. 2 for a table with all marker profiles. E. All 16 Boolean logic gates used to separate cells belonging to different surface marker profiles.

any clonal streak formed within a dual label population can be assessed by plotting detections that fall within their gates on the three surface marker dot-plots described above.

2.6 Imaging

2.6.1 Laser Scanning Microscopy

All fluorescent confocal imaging was performed on either a Zeiss lsm 710 or lsm 880 microscope (Carl Zeiss, Cambridge, UK). Imaging of cryosectioned and cleared confetti tumours, RGB marked U87 tumours, all cultured organoids and adherent cultures were performed using separate tracks for acquisition. Laser excitation and emission filters was setup according to supplementary table 5. In the case of ROSA26-confetti imaging, ECFP and tDimer2 were acquired on the same track owing, to sufficient spectral separation. EYFP and EGFP emission spectra overlap considerably and were acquired on separate excitation and emission tracks for better signal separation. Laser powers and detector gains were set appropriately for the brightness of fluorescence within the sample which could vary between cell lines and affected by fixation. For acquisition of dual labelled U87 and primary cells, spectral unmixing experiments were performed. Spectral unmixing requires simultaneous excitation with all required laser wavelengths and laser powers set to achieve a balance between intensity of each fluorescent signals and avoidance saturated pixels. Individual emission spectra for each fluorophore (EBFP2, T-sapphire, EGFP, Venus, mOrange2 & Katushka2S) and the far-red nuclear dye DRAQ5™ (thermofisher, 62251) were acquired using the Zen black Lamda stack acquisition function. Acquisition was setup such that emitted light intensities were detected at 10 nm increments from 400 to 720nm, producing a lambda stack comprised of 32 images. Once reference spectra were acquired by selecting pixels positive for a single fluorescent singla, the Zeiss black online fingerprinting function could be used to unmix overlapping spectral signals in real-time during subsequent acquisitions.

2.6.2 Light-sheet Microscopy

A light-sheet Z1 (Carl Zeiss, Cambridge, UK) microscope was used for rapid volumetric imaging of cleared confetti tumours and whole confetti tumour organoids. Detection of ECFP and tDimer2 was possible, however, EGFP and EYFP were collected as the same signal because the microscope was not fitted with appropriate emission filters to separate this signal. Filter sets were setup according to the values in table 5. Cleared brain sections and organoids were mounted by gluing samples to a low refraction glass pipette and placing them suspended within the imaging chamber. Laser powers and detector gains were set appropriately for the brightness of fluorescence within the sample.

2.6.3 Image Analysis

For *in vitro* confetti recombination assays comparing adeno-virus and 4-OH tamoxifen, quantification and analysis recombination indicated by fluorescence expression were performed on FIJI using binary thresholding (Schindelin et al., 2012). Briefly, .czi zeiss image files were separated into their individual acquisition channels (ECFP, EGFP, EYFP and tDimer2) and exported as .tif files. Each image was then converted to 8-bit greyscale and subject to binary thresholding for detecting either fluorescent cells or labelled nuclei with the FIJI particle analyser and analysis of double labelling was performed using the co-localisation threshold plugin.

For imaging based quantification of labels in samples transduced with LeGO-vectors, images were collected using spectral unmixing and the open source image analysis software quPath (Bankhead et al., 2017) was used for quantifying label distributions. Samples were labelled with the nuclear dye DRAQ5™ and the nuclear segmentation algorithm StarDist (Schmidt et al., 2018) was used to detect DRAQ5™ labelled cells. After cell segmentation, fluorescence intensity data for each labelled fluorophore detected are exported to a data frame. Subsequently, distributions of cellular fluorescence intensity for each detected fluorophore were visualised to manually assign intensity threshold for designating cells positive or negative for a given fluorophore. This analysis reported a number of triple positive cells which were likely

incorrect as cells were exposed to just two fluorescent lentivirus'. As such, cell detections reporting expression of three fluorophores were removed from analysis.

2.7 Statistical analysis

2.7.1 Recombination efficiencies and NFIX KD

Comparisons of ROSA26-confetti recombination efficiencies were performed in Graphpad Prism 5 using 2-way Anova assuming guassian distribution with Bonferroni correction for multiple comparisons. Exact statistical test used is indicated in text and figure legends. With the same analysis performed on westernblot data investigating NFIX KD in primary cell lines. All bar graphs and error bars indicate mean of sample data and standard error of the mean respectively.

2.7.2 Comparison of tracking label groups frequencies in Organoids

Tracking label group composition and changes in colour group frequencies were analysed by Anova with a for correction for multiple comparisons. Parameters of statistical test applied to each experiment are outlined in text and figure legends.

2.7.3 Cosine Similarity analysis of clonal surface marker data

Strategy of cosine similarity analysis was formulated between James Innes and Andrew Lowe. Design and implementation of the analysis pipeline was largely performed by Andrew Lowe with input from James Innes refining process. Quantification of surface marker phenotypes was achieved through binarizing signal intensities of each surface markers. Assignment of a cell as positive or negative for a given marker was achieved by cut off thresholds determined through analysis of single label controls. From 4 markers, a total of 16 different surface marker combinations could be determined. Thus, the combination of these 16 different phenotypes for a given cell line or cell clone make up their phenotypic distributions. Estimations of the degree of similarity between and two phenotypic distributions was determined using Cosine of Similarity.

$$\cos(\theta) = \frac{\mathbf{A} \cdot \mathbf{B}}{\|\mathbf{A}\|\|\mathbf{B}\|} = \frac{\sum_{i=1}^{n} A_i B_i}{\sqrt{\sum_{i=1}^{n} A_i^2} \sqrt{\sum_{i=1}^{n} B_i^2}}$$

In the above equation A and B represent the inner vector product of the 16-dimensional phenotypic distributions of a given pair-wise clonal or cell line comparison. The output of this equation is a value between 0 – 1 where a comparison yielding a cosine similarity of 1 indicates the two samples have the same phenotypic distributions with similarity decreasing as cosine similarity values tend toward 0. Assessments of the similarity of more than two subsets have been achieved by determining the mean cosine of similarity of all pair-wise combinations.

To determine the influence of sampling variation in our cosine similarity par-wise comparisons we undertook a Monte Carlo simulation; where an *in silico* population hypothetically comprised of 16-different proportions of surface marker phenotypes was repeatedly sub-sampled. Repeatedly choosing (x100,000) sub-samples of cells (i.e. of the order empirically seen for single clones) without replacement enabled probability distribution curves for expected deviations in similarity between the phenotype of the population and the sub-samples. Under these conditions we determined a cut value of CoS = 0.98, where any difference observed between two matched phenotypic distributions could not be explained by sampling error ($P < 0.001$, indicated by dashed lines in plots of Chapter 5)

Chapter 3: Exploring fluorescent Lineage tracing applications in murine glioma models

3.1 Introduction

The Rosa-26confetti mouse was first described by Clevers *et al.,* and is now a well-established transgenic model for fluorescent lineage-tracing (Clevers, 2011, Schepers et al., 2012, Snippert et al., 2010) (Figure 3.1). Upon Cre exposure the confetti construct recombines into 4 possible orientations corresponding to expression of either m-cereulean (membrane, blue), eGFP (nuclear, green), Venus (cytoplasmic, yellow) or tdtomato (cytoplasmic, red). To gain technical experience in fundamental methodologies and explore techniques which could complement lineage-tracing in primary human GBM cells, we first sought to develop a confetti mouse tumour model. Confetti animals were bred into a background of tumour inducible GLAST$^{CreERT2/CreERT2}$/*Pten*$^{lox/lox}$/P53$^{lox/lox}$ to create the genotype GLAST$^{CreERT2/CreERT2}$/*Pten*$^{lox/lox}$/P53$^{lox/lox}$/ROSA$^{confetti/WT}$ referred to as confetti-GBM. We firstly trialled *in vivo* and *in vitro* method for inducing recombination of tumour suppressor alleles and the confetti locus. For *in vitro* recombination, Cre delivery to derived adherent confetti-GBM NSCs was mediated by either adenovirus (Adeno-cre) or 4-OH-tamoxifen (Benedykcinska et al., 2016a). *In vivo,* intraventricular injections of PDGFβ IRES Cre (PIC) or IP tamoxifen injections were trialled. Owing to its speed and penetrance, PIC induced confetti tumours were used for exploring techniques suitable for use with fluorescent clonal tracking.

Compared to other cellular tracking techniques, fluorescent tracking offers the unique ability of identifying clones *in situ.* In human GBM, intra-tumour heterogeneity is a well described feature and has been identified by sequencing of spatially distinct tumour biopsies (Sottoriva et al., 2013, Yu et al., 2020). Therefore, we aimed to develop an approach utilising tissue clearing in combination with light-sheet microscopy for rapid fluorescent imaging of large three-dimensional tumour

regions. Additionally, live *in vivo* and *in vitro* imaging experiments were trialled with the aim of assessing proliferation and motility of different tumour clones. FACS separation of human CSC clones and comparisons of gene expression would be a powerful tool for investigating how clonality is linked to intra-tumour heterogeneity. Therefore, we aimed to optimise approaches for tissue dissociation and sorting of labelled cells. Matrigel® tumour spheres recreate aspects of intra-tumour heterogeneity and can be cultured for longer without intermittent passaging. These models are higher-throughput, cheaper and more ethical prelude to *in vivo* experiments, reducing the need to sacrifice animals while optimising experiments. Therefore, we sought to develop an organoid model to use in combination with confetti cell tracking. Through trialling a number of approaches with confetti fluorescent tracking we aim to inform development of improved approaches in primary human GBM CSCs.

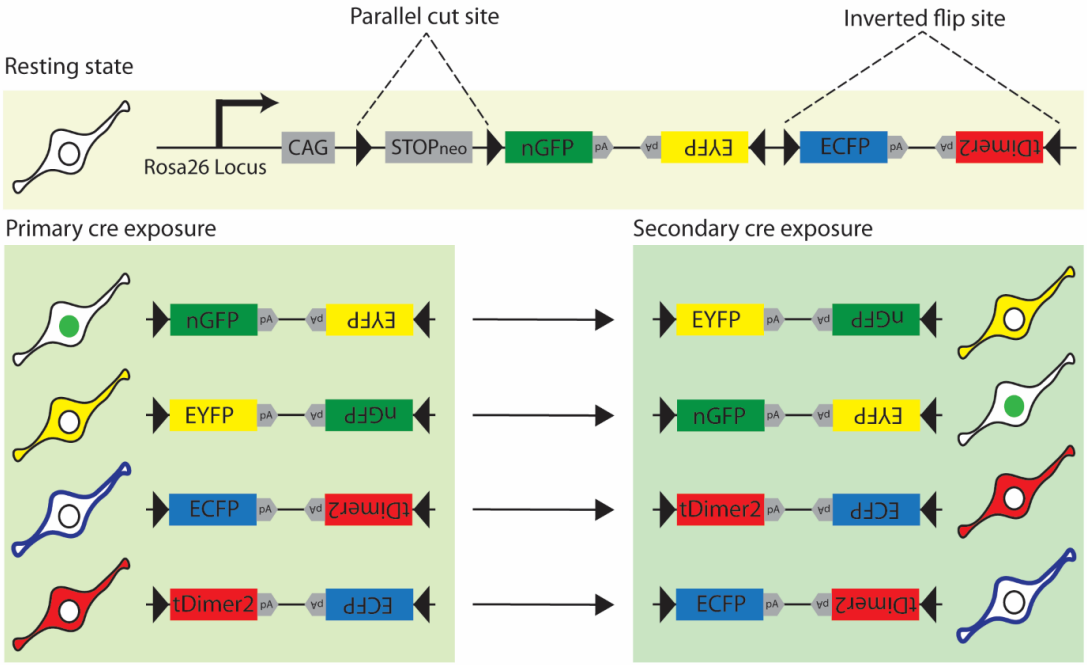


Figure 3.1: Schematic of r26-confetti genetic construct. The construct is inserted to the rosa26 locus for considerable protein expression. At resting state (before cre exposure) a neo-stop cassette prevents gene expression. The LoxP site design involves parallel and inverted orientations which allow for induction of fluorophore expression and a subsequent flipping of the fluorophore coding regions to change expression after a secondary Cre exposure. Upon a primary Cre exposure, cells express either nuclear EGFP, cytoplasmic EYFP, membranous ECFP or cytoplasmic tdimer2. Upon a secondary cre exposure EGFP switches to EYFP, EYFP switches to EGFP, ECFP switches to tdimer2 and tdimer2 switches to EXFP expression.

3.2 Results

3.2.1 *In vitro recombination of GLASTCreERT2/confetti/Pten/P53* murine NSCs *in vitro with* Adenovirus-cre or 4-OH tamoxifen

To explore methods for recombining the confetti locus, NSC populations were established from the forebrain of P7 confetti-GBM mice using the neurosphere culture method. These cells carry the GLAST CreERT2 transgene for an inducible, transient Cre expression upon exposure to 4-OH tamoxifen, rendering any recombination event permanent. To test the effectiveness of this approach in recombining the confetti-locus we performed a serial dilution of 4-OH tamoxifen and applied these to isolated confetti-GBM NSCs (Figure 3.2 A-E). To determine optimal working concentration that would balance cellular toxicity and recombination effectiveness we exposed cells to 6.25 x 10^{-1}, 1.25, 2.5, and 5 µM. For each concentration three repeat culture wells were tested and five regions were imaged for each repeat. 6.25 x 10^{-1}, 1.25 and 2.5 µM 4-OH tamoxifen concentrations were all successful in achieving recombination and showed similar confluence to control conditions (**Error! Reference source not found.**A-E). However, 5.0 µM 4-OH t amoxifen led to excessive cell death and for this reason, recombination efficiency was not calculated for this condition (Figure 3.2 E, F & G). For all three concentrations there was no significant difference in recombination efficiency (One-way ANOVA, Bonferroni correction, 95% CI), suggesting a lower concentration of 4-OH tamoxifen would be sufficient to induce maximal recombination in culture (Figure 3.2 E). Furthermore, across all conditions it was very rare to observe a recombination outcome resulting in EGFP expression while m-Cerulean could not be imaged as its excitation and emission spectra overlaps with the Hoechst 33342 nuclear stain.

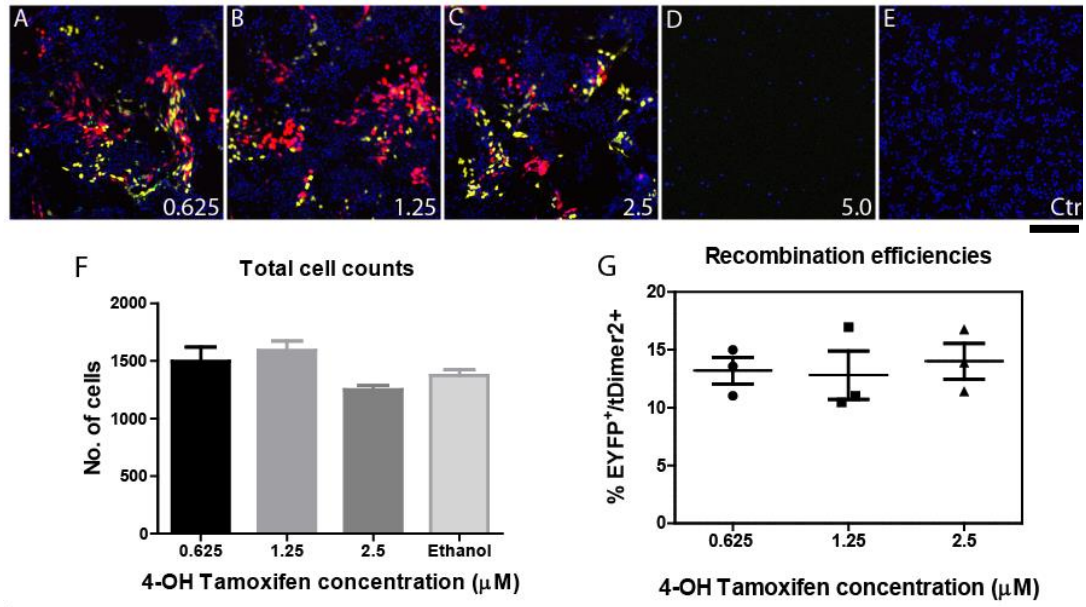


Figure 3.2: In vitro recombination efficiencies of GLASTCreERT2/R26-confetti neural stem cells. (A – E) Representatives of Hoechst stained r26-confetti neural stem cells after exposure to 4-OH tamoxifen; 0.625, 1.25, 2.5 and 5.0 denote the 4-OH tamoxifen concentration. 5.0 µM was not further analysed owing to cell death (scale bar = 200µm). (F) Shows average number of Hoechst stained nuclei per field of view sampled after recombination, data represents the mean of n = 3 repeat wells, each determined through mean count of n = 5 ROIs. Repeated measures ANOVA with Bonferroni correction revealed no significant difference between test conditions and ethanol control (CI = 95%). (H) Shows percentage of $EYFP^+/tdimer2^+$ cells counted after 4-OH tamoxifen exposure in each condition. Each point represents a separate test well (n = 3) which is a mean count of $EYFP^+/tdimer2^+$ cells across n = 5 ROIs. Bars indicate mean and SEM. Repeated measures ANOVA with Bonferroni correction revealed no significant difference between conditions (CI = 95%).

To further explore the confetti locus expression dynamics, we passaged the recombined confetti-GBM NSCs exposed to 1.25 µM of 4-OH tamoxifen assessed cell colour frequencies over 120h (n = 3 repeat cultures and n = 1 x $5mm^2$ ROI per time point) after removing 4-OH tamoxifen (Figure 3.3A-L). Using the binary threshold method outlined in methods (Section: 2.6.3), the proportion of labelled cells expressing each fluorophore was calculated. At 24h after passaging, 24% ± 1% SEM of all coloured cells were $EYFP^+$ while 40% ± 1% SEM and 36% ± 1% SEM were $tdimer2^+$ and $ECFP^+$ respectively (Figure 3.3M). $EGFP^+$ cells were observed at very low frequency (< 0.01) and were often co-expressed with EYFP and thus were not included in the quantification. Interestingly, these outcomes are not equal as described in the original Brainbow publication (Livet et al., 2007) but quantifications were not reported and similar distributions in expression outcomes after

recombination of the confetti locus have been reported since (Reeves et al., 2018). Over the following 120h, the proportion of tdimer2^{+} cells increased and the proportion of ECFP2^{+} cells decreased, although these trends did not reach statistical significance (Figure 3.3M). At all time points, a number of cells co-expressing ECFP and tdimer2^{+} were observed in imaging (Figure 3.3A', E', & I'). It is likely that prolonged Cre-exposure results in multiple flipping events of the confetti construct, resulting in expression of both fluorophores. However, after Cre has dissipated and the construct "settles" in a single orientation, cells will eventually adopt a single colour. In support of this, we observed a reduction in the proportion of ECFP^{+}/tdimer2^{+} cells across the 120h which reached statistical significance by 120h post infection ($p < 0.0001$).

As an alternative to CreERT2 mediated recombination, we also tested *in vitro* recombination through infections with cre expressing adenovirus (adeno-cre). Five dilutions (5×10^{-3}, 1×10^{-4}, 2×10^{-4}, 4×10^{-4} or 8×10^{-4}) of Adeno-cre were applied to naïve confetti-GBM NSCs (n = 3 repeats per viral dilution and n = 5 ROIs per repeat) to determine its effectiveness in recombining the confetti locus (Figure 3.4 A). A dilution $5x10^{-3}$ resulted in nearly 100% recombination. However, accurate quantification of these infections was not possible using binary thresholding, so this group was not quantified. One-way ANOVA with 95% CI found significant difference between means of total cell counts across conditions ($p = 0.37$) but post-hoc Dunnet's multiple comparisons with control condition did not reach significance for any viral concentration (Figure 3.4 H). After quantification, recombination efficiencies were found to be 0.86 ± 0.02 SEM, 0.59 ± 0.02 SEM, 0.37 ± 0.01 SEM and 0.22 ± 0.01 SEM for $5x10^{-3}$, $1x10^{-4}$, $2x10^{-4}$, $4x10^{-4}$ and $8x10^{-4}$ dilutions respectively, showing a near linear reduction in recombination efficiencies with increasing viral dilution (Figure 3.4 I).

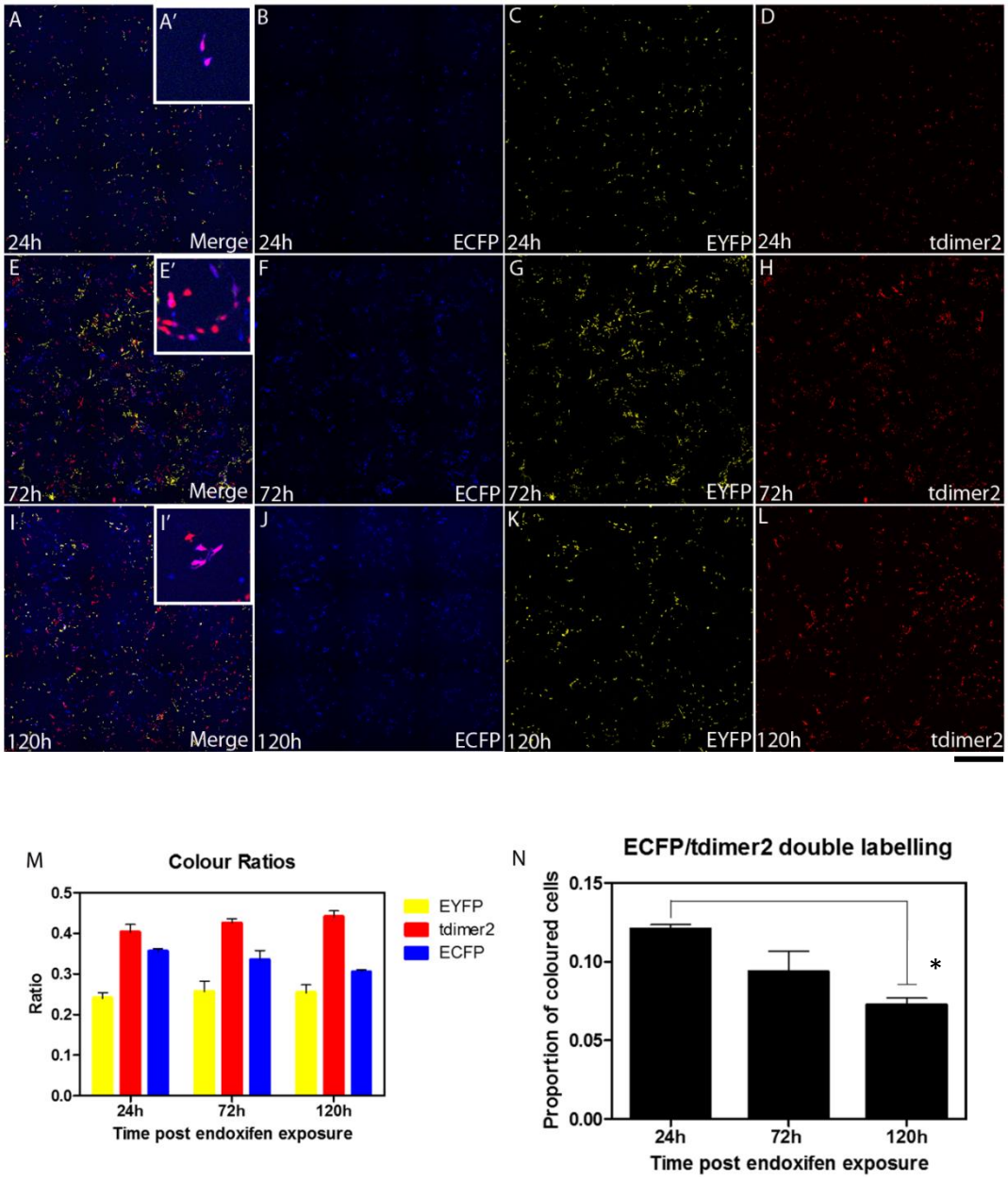


Figure 3.3: Passaging of 4-OH Tamoxifen induced GLASTCreERT2/R26-confetti cells reveals fluorophore expression dynamics. (A- L) Merged and single channel fluorescent images of 1.25 µM 4-OH tamoxifen recombined cells at 24h, 72h and 120h post recombination (Scale bar = 500 µm). (A', E', I') zoomed examples of cells at each time showing expression of both ECFP and tdimer2. Panels to the right of A, E and I show single channel images of ECFP, EYFP and tdimer2 at each time point (n = 3 repeats and n = 1 x 5mm^2 ROI per repeat)(Scale bar = 1mm). (M) Quantification of cells expressing each fluorophore over the time course of the experiment. 24h; 0.24 ± 0.01 SEM, 0.40 ± SEM and 0.36 ± 0.01 SEM were found to be the proportions of EYFP, tdimer2 and ECFP expressing cells respectively. Over the time course of the experiment, the proportion of tdimer2$^+$ cells increased and ECFP$^+$ cells decreased but this effect did not reach significance (One-way ANOVA, Bonferroni correction with 95% CI). (N) Quantification of time of cells double positive for ECFP and tdimer2 at 24h (mean = 0.120 ± 0.002 SEM), 72h (mean = 0.094 ± 0.012 SEM) and 120h (mean = 0.073 ± 0.004 SEM). One-way ANOVA with Bonferroni correction and 95% CI found significant difference between dual labelling at 24h and 120h (P = 0.0264)

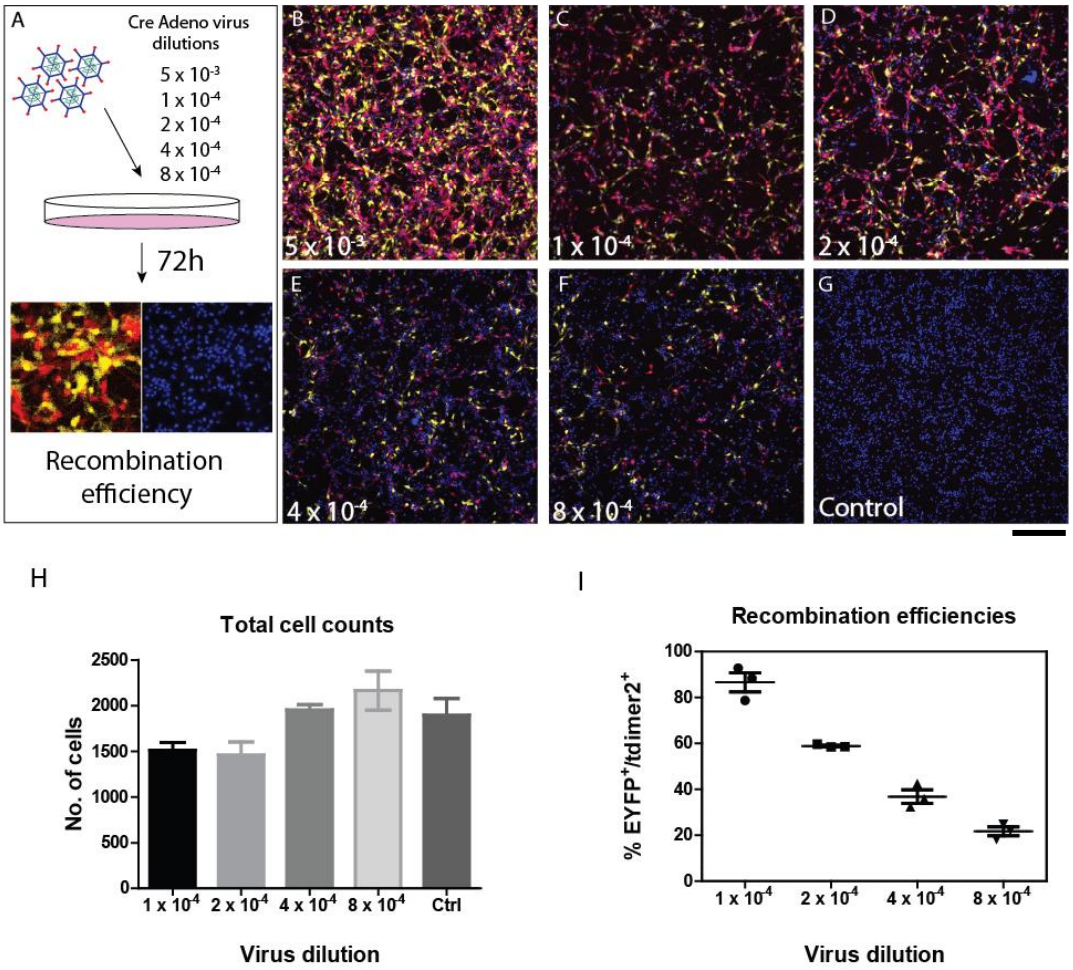


Figure 3.4: in vitro recombination of Adeno-cre induced r26-confetti neural stem cells. (A) schematic of experimental design. (B – G) Representatives of Hoechst 33342 stained sample regions from r26-confetti neural stem cells stained with ad-cre dilutions of 5 x 10^{-3}, 1 x 10^{-4}, 2 x 10^{-4}, 4 x 10^{-4} or 8 x 10^{-4} and a no virus control (n = 3 repeats per viral dilution and n = 5 ROIs per repeat)(Scale bar = 200 µm). (H) Shows quantification of Hoechst 33342 stained cells per field of view during analysis. One-way ANOVA with 95% CI revealed statistical difference between means (p = 0.37) but post-hoc Dunnets multiple comparison test comparing each condition to control did not reach significance for any condition. (I) Recombination efficiencies of as-cre infected cells across 4 concentrations (n = 3 repeats per viral dilution and n = 5 ROIs per repeat). Recombination efficiencies showed a near linear reduction with increasing virus dilution.

3.2.2 Intraperitoneal Tamoxifen injections lead to extensive gene recombination within the subventricular zone and cortex

One possibility for *in vivo* activation of the confetti locus and subsequent tumour formation is to induce Cre expression within the SVZ using the GLASTCreERT2 Transgene (Mori et al., 2006). GLAST is a marker specific for SVZ NSCs which are the current leading candidate for glioma's cell of origin and unlabelled tumorigenesis in this fashion has already been established in our lab (Benedykcinska et al., 2016a). To test the effectiveness of recombination via this method animals (n = 6) were given a 3-day course of tamoxifen injections and culled 7 days later (Figure 3.5A). In all animals, cryosectioning and confocal imaging revealed extensive recombination throughout the upper and lower regions of the SVZ where cells expressing each of the four confetti fluorophores could be observed (Figure 3.5B & C). However, a considerable number of recombined cells were also visible scattered throughout the cortex (Figure 3.5D). GLAST is also a marker of activated astrocytes (Preston et al., 2019) which judging by cellular morphology is likely the identity of the labelled cells found in the cortex. These results suggest tamoxifen mediated recombination targeting GLAST expressing cells is a viable strategy for *in vivo* confetti labelled tumour induction. However, our previous work with this technique reported protracted periods of tumour development and low tumour penetrance (Benedykcinska et al., 2016b). Therefore, we next attempted to combine confetti labelling with the rapid and high penetrance PIC retrovirus glioma model.

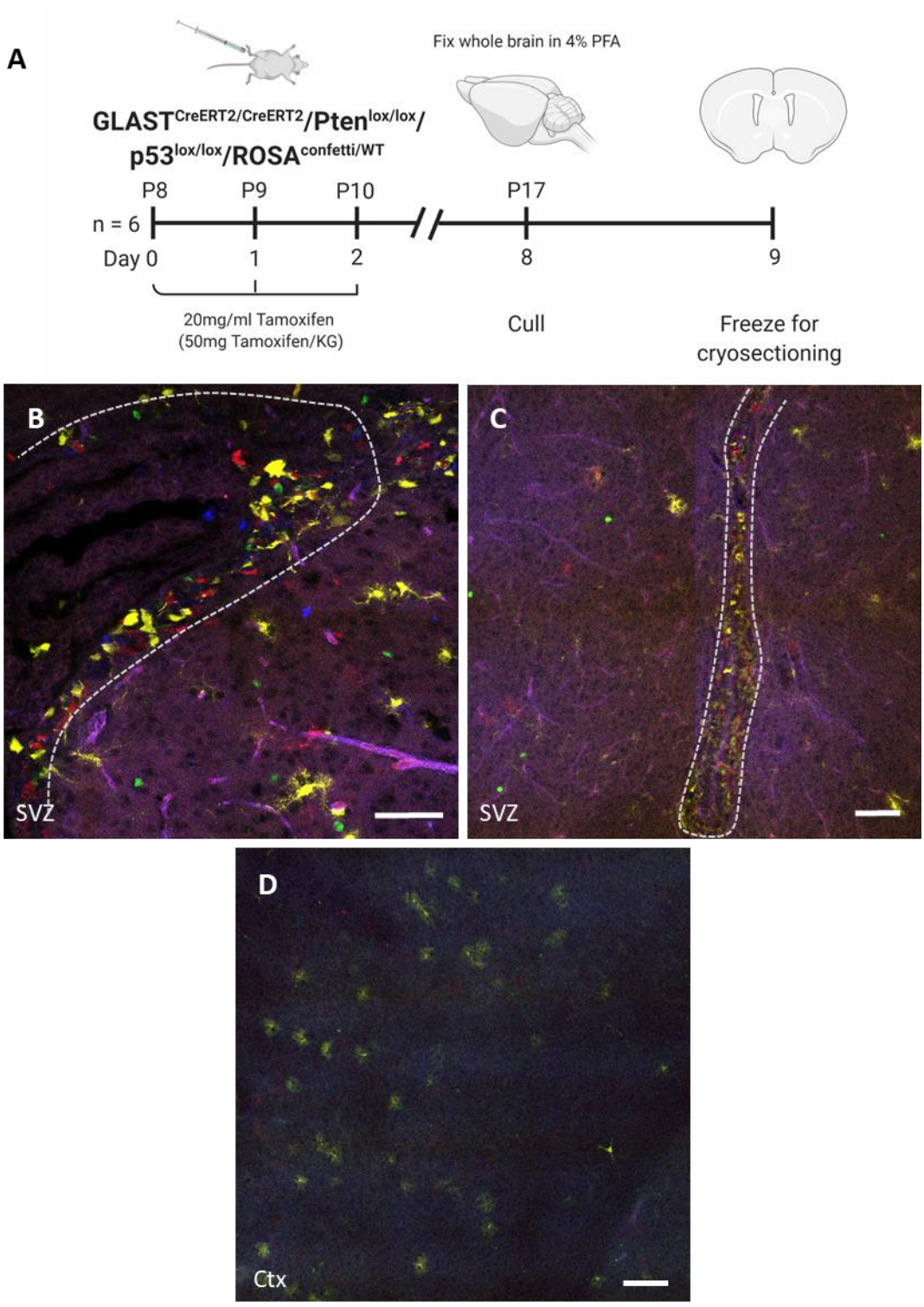


Figure 3.5: Tamoxifen delivery effectively recombines the confetti locus in GLAST CreERT2 transgenic animals (A) schematic showing time course of experiment and genotype of animals used (n = 6). (B & C) Confocal imaging of subventricular zone (SVZ) in cryo-sectioned mouse brains after course of tamoxifen injections. Extensive expression of confetti fluorophores can be observed suggesting effective activation of GLAST CreERT2 in target cell populations (A scale bar = 100 µm, B scale bar = 200 µm). (D) Example of further confetti fluorescence expression in astrocytes of the cortex which like SVZ NSCs are also GLAST expressing cells (Scale bar = 200 µm).

3.2.3 Relative titre calculations increase PIC retrovirus *in* vivo tumour penetrance

Neonatal intraventricular injections of PIC retrovirus into PTEN$^{lox/lox}$/P53$^{lox/lox}$ mice produce rapidly growing tumours with nearly 100% penetrance (Lei et al., 2011). The resulting fast experimental turnover makes this an excellent model for developing our lineage tracing methodologies. In our experiments, this model were hampered by low tumour penetrance thought to be due to insufficient virus titre during virus preparation. To address this we increased viral titre, by concentrating the viral preparations. Furthermore, to ensure new virus preparations were of satisfactory titre and capable of achieving the expected high tumour penetrance, they were subject to a relative titre test. Virus preparations were incubated with neurosphere-derived (Galli et al., 2004) neural stem cells from a Confetti-GBM mouse (n = 3 repeats). 72h later, the number of recombined cells present was calculated across 10 sample regions (Figure 3.6B - E). Previously, virus preparation #1 demonstrated high tumour penetrance (Figure 3.6F), therefore, we reasoned virus preparations with a greater recombination efficiency than preparation #1 should yield high tumour penetrance. For virus preparation #1, the mean fraction of *in vitro* recombined cells was found to be 2.16% (± 0.24% SEM) (Figure 3.6G). For new concentrated virus preparations, these values were 6.47% (±1.02% SEM) and 4.21% (±0.73% SEM) for preparation 4 and 5 respectively (Figure 3.6B). Both new preparations reported an increased recombination efficiency compared to preparation 1 but only preparation 4 reported a statistically significant increase (One-way ANOVA 95% CI, Bonferroni correction) (Figure 3.6G). Virus prep 4 was used for subsequent injections and reported a 100% tumour penetrance across three separate litters (Figure 3.6H). Mice injected with virus prep 4 reported a median survival rate of 27.5 days (Figure 7.15B) and shared histological features consistent with human GBM (Figure 7.15C-F)

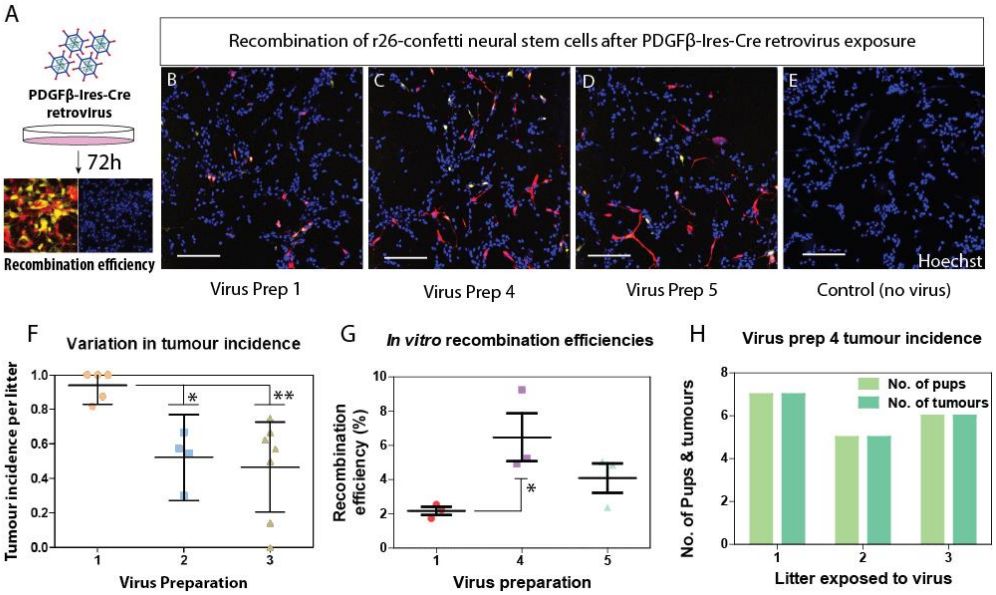


Figure 3.6: Relative titre calculations ensures consistent tumour formation in PIC retroviral model. (A) Schematic showing PIC virus application, 72h hour wait until imaging of confetti fluorescent reporter and Hoechst 33342 nuclear stain for calculating recombination efficiencies. (B – E) Example fluorescent imaging of viral titre test ROIs for PIC retroviral prep 1 (B), prep 4 (C), prep 5 (D) and no no virus control (E)(Scale bar = 200 µm). Viral preps 2 & 3 were not subject to *in vitro* recombination assay. (F) Displays difference in tumour penetrance between different virus preparations, each point represents tumour incidence in a litter of injected animals. Virus preps 2 (n = 4 litters) and 3 (n = 5 litters) showed a statistically significant reduction in tumour penetrance compared to prep 1 (n = 5 litters) (One-way ANOVA, Bonferroni correction, 95% confidence intervals, p = 0.006). (G) Displays quantification of *in vitro* recombination results of virus prep 1 and newly concentrated virus preps 4 and 5 (n = 3 technical repeats per virus prep). One way ANOVA with 95% confidence intervals didn't reach significance but reported p = 0.051. Bonferroni correction of multiple comparisons did find $p < 0.05$ for prep 1 vs prep 4 but not prep 1 vs prep 5. (H) Tumour incidence across 3 litters injected with virus prep 4 was 100%.

3.2.4 Confocal Imaging of PIC injected mice reveals tumours of variable organisation, location and clonal content

Confocal imaging experiments of PIC induced tumours were limited by low tumour penetrance and as such only a small proportion of injected animals were eventually imaged. In total 160 neonatal mice were injected for confocal imaging experiments. Of these, 106 animals were processed for cryosectioning while 54 were processed for vibratome sectioning. In total 15 animals from the cryostat group and 13 animals from the vibratome group were imaged by laser scanning confocal microscopy. Cryosectioning was found to be a more suitable approach for fluorescent imaging of

confetti labelled tumours. This section presents the best examples of cryosectioned tumours to illustrate the variable location, structure, development and clonal content of tumours produced after PIC retrovirus injections.

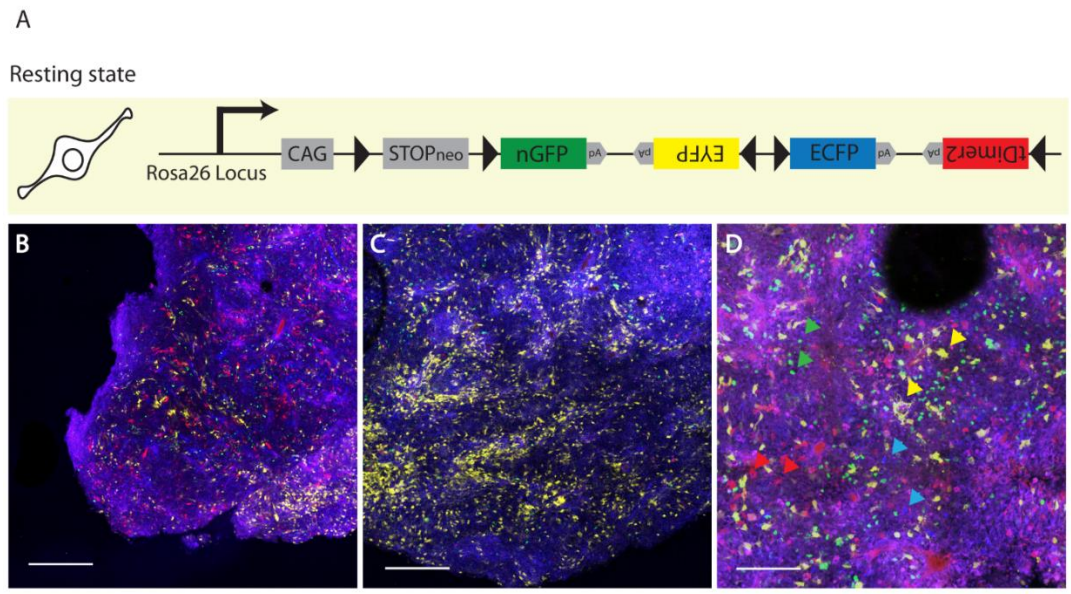


Figure 3.7: R26-confetti reveals distinct tumour lineages after intraventricular PIC retroviral injections. (A) r26-confetti locus design allows for expression of four different fluorophores. (B and C) Distinct regions of the same tumour (n = 1) demonstrating spatial heterogeneity in clonal composition (scale bar = 100 µm). (D) High power image highlighting presence of all rosa6-confetti fluorophores designated by correspondingly coloured arrows (scale bar = 50 µm).

Figure 3.7 shows an example of cryosections from a confetti labelled tumour in the cortex. Cells expressing each of the 4 confetti fluorophores (ECFP, nGFP, EYFP and tdimer2) were identified. However, cellular discrimination was much easier for cells expressing EGFP, EYFP or tdimer2, while ECFP expressing cells generally appear as a continuous region of blue fluorescence where single cells cannot be easily identified. Interestingly, tumour label composition was found to vary across different tumour regions; a region containing many labels is show in Figure 3.7B while a region predominated by EYFP expressing cells is shown in Figure 3.7C. This would indicate that discrete tumour regions can contain a variety of different clones while some regions are comprised primarily of cells from a single clone. This would be congruent with human disease where spatially discrete biopsies of a single tumour were found to have different clonal contents.

Figure 3.8 displays images of an apparent choroid plexus tumour seven weeks after PIC injections. In Figure 3.8A we can see the rostral portion of this tumour retained within the left ventricle which would suggest it is a tumour of choroid plexus origin. Cells are predominately labelled with EGFP or EYFP and are densely clustered. In a more caudal region of this tumour we can identify two histological distinct tumour regions; a densely populated region proximal to the left ventricle (Figure 3.8B.1 & B.2) and a leading edge of the tumour where cells adopt a much more dispersed distribution (Figure 3.8B.3) perhaps indicating cell migration. Between these two regions, there is a portion of the tumour with apparently few labelled cells. It is possible this is a region of necrosis with few living cells or alternatively this region of the tumour is comprised of cells labelled with ECFP that we rarely detected during histological assessment of PIC induced tumours.

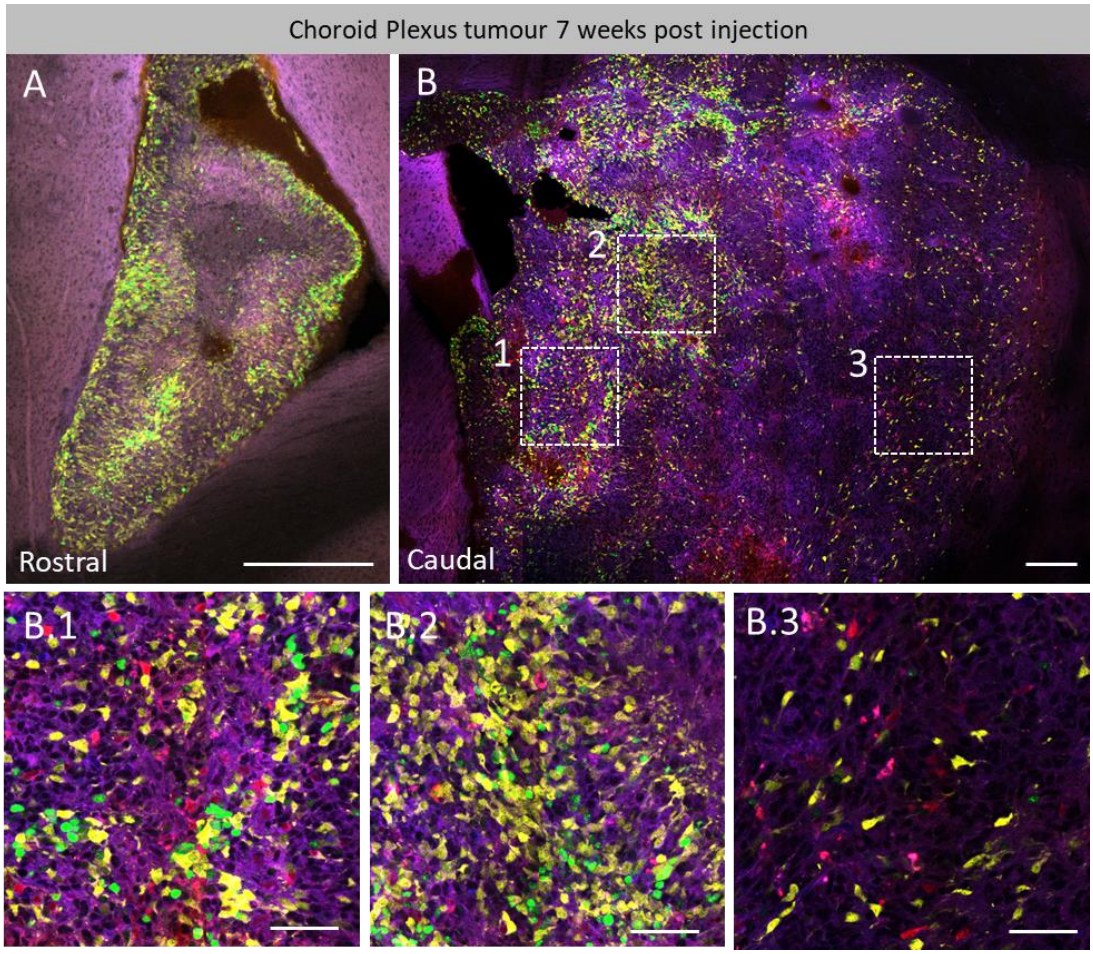


Figure 3.8: R26-confetti labelled choroid plexus tumour after PIC retroviral injections: (A) Confocal imaging of rostral tumour region contained within ventricle (scale bar 200 µm). (B) Caudal tumour region showing invasion in brain parenchyma (scale bar = 200 µm). (B.1 & B.2) Proliferative regions

proximal to the ventricle showing densely packed cells in palisading structures (scale bar = 100 µm). (B.3) Tumour region distal to the ventricle at the invading edge of the tumour showing cells less densely packed (scale bar = 100 µm).

Figure 3.9 shows three examples of another common tumour location and structure found after PIC injections. In these examples, tumours presented within the subcortical white matter and each one is comprised of just a single label suggesting these tumours comprised primarily of a single clone. Generally speaking, these tumours were confined within the subcortical white matter with cells primarily migrating laterally with a small proportion exiting the white matter dorsally or ventrally. In Figure 3.9A & B, cells can also be seen migrating dorsally, perhaps along cortically projecting axonal tracts while in Figure 3.9C cells can be seen exiting the subcortical white matter ventrally between the lateral ventricles demonstrates three examples of tumours that formed in confetti animals but presented containing very few or no labelled cells. Figure 3.10A contains a caudally located tumour that formed beneath the lateral ventricles while tumours displayed in Figure 3.10B & C were found lateral to the ventricles. Similar to what was observed in the tumour shown in Figure 3.8, it is unclear if these are regions of necrosis or packed with poorly detected ECFP expressing cells. Alternatively, it is possible that constitutive cre activity is leading to persistent flipping of the confetti locus and eventual loss of fluorescence expression through damage to the locus.

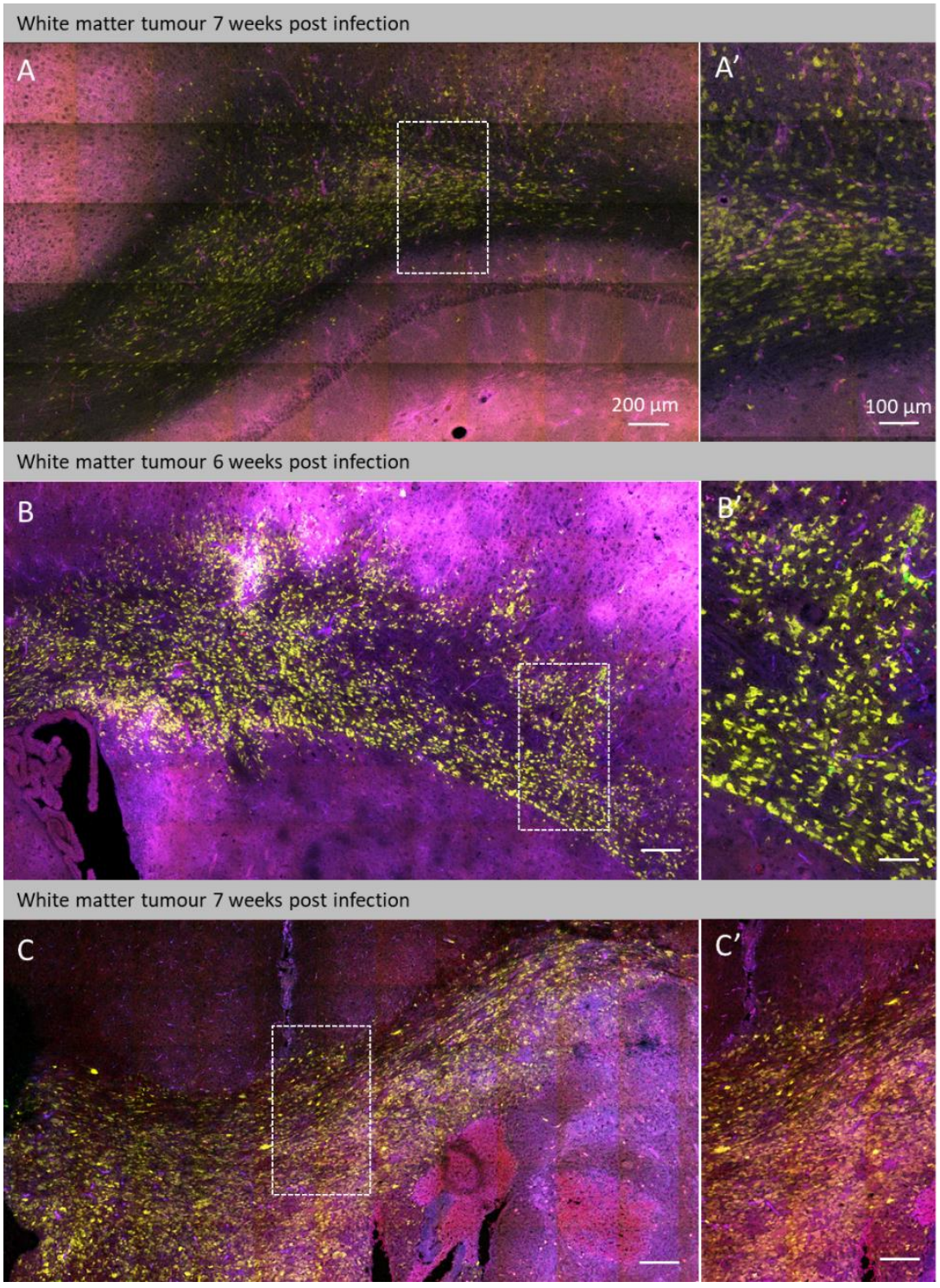


Figure 3.9: Examples of tumours forming in sub-cortical white matter after PIC retroviral injections. (A – C) Three separate tumours forming with cells restricted largely in the sub-cortical white matter (scale bar = 200 µm). (A' & B') Cells exiting the white matter and migrating dorsally into the cortex (scale bar = 100 µm). (C') Cells exiting the white matter into brain regions ventral to the sub-cortical white matter (scale bar = 100 µm).

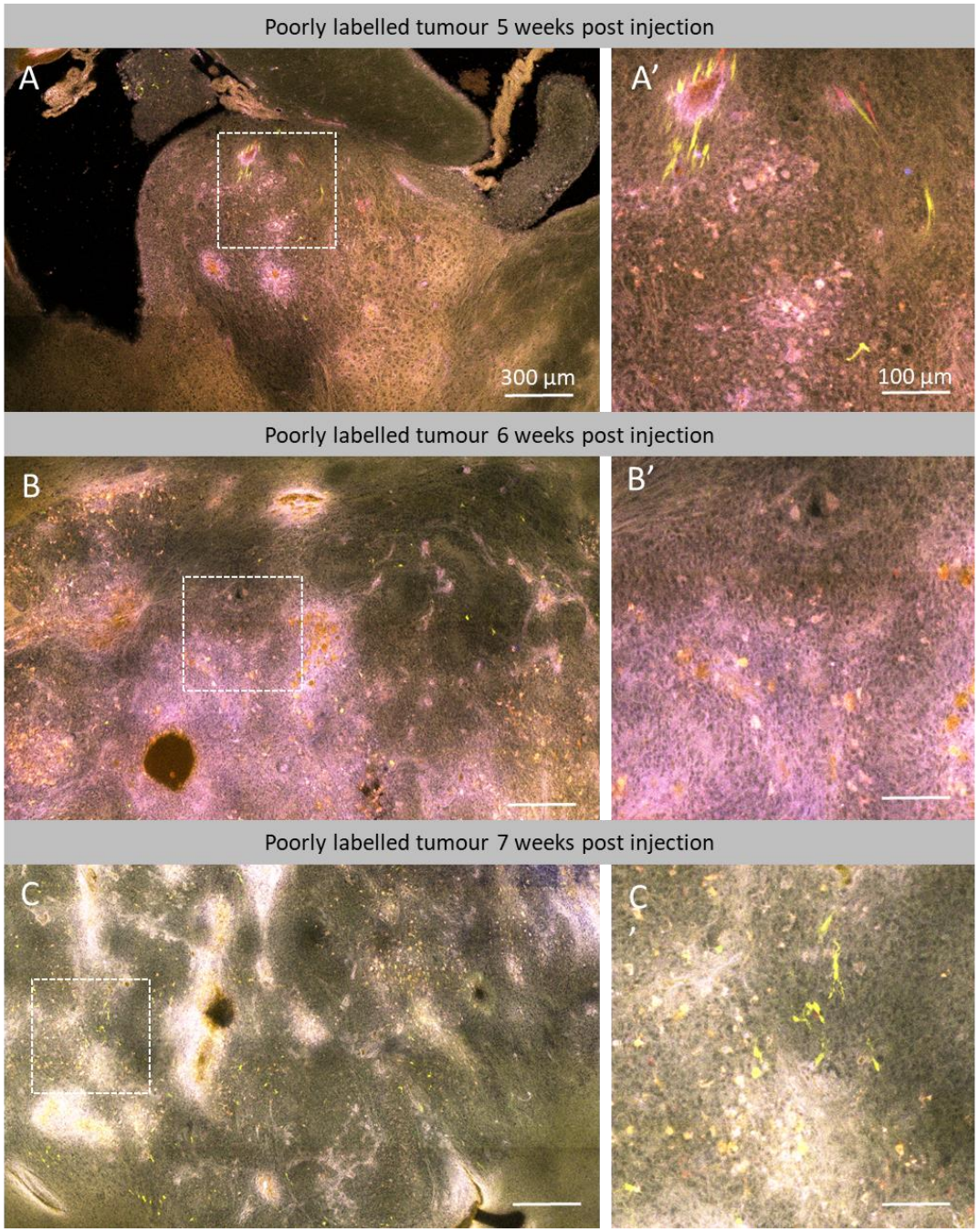


Figure 3.10: Examples of tumours forming in central brain regions containing many unlabelled cells. (A-C) Three separate tumours where the bulk of observed cells were found not expressing and R26-confetti labels (scale bar = 300 µm). (A' – C') boxed regions in A – C demonstrating the sparse presence of labelled tumour cells but extensive areas of unlabelled cells (scale bar = 100 µm).

3.2.5 PIC induced confetti tumour fluorescence is preserved after clarity tissue clearing

Despite the issues discussed above, We were able to use r26-confetti mice injected with the PIC retrovirus to optimise our clearing protocol. This method will allow us to obtain large scale spatial information on the migration and localisation of differently labelled clones. The passive clarity procedure involves overnight tissue fixation in 4% PFA, then embedding of sample into a polyacrylamide hydrogel (Figure 3.11A). Once embedded, samples are then passively cleared of lipids by bathing in a detergent solution (8% SDS). Finally, to minimise heterogeneous light scattering events samples are incubated overnight in refractive index matching solution (RIMS) (50).

Early trials of using passive clarity to image r26-confetti tumours were unsuccessful. Tissue would visibly clear but when tumour regions were investigated little or no endogenous fluorescence was observed. We were initially using large pieces of brain tissue which required up to 1 weeks incubation in SDS for delipidation, we therefore modified our approach to clear 1mm thick coronal slices (Figure 3.11: Passive clarity tissue clearing permits deep imaging and preservation of R26-confetti fluorophores. (A) Schematic of theory behind clarity tissue clearing. Tissue is fixed then embedded in acrylamide hydrogel to fix tissue macromolecules. Lipids are then diffused from the tissue by incubation in detergent solution (8% SDS).(B-E) Examples of brain tissue before and after clarity procedure demonstrating excellent optical clearing (n = 2)(Scale bar = 500 µm). (F & G) Regions containing cells with preserved endogenous fluorescence (scale bar = 100 µm). (H) Orthogonal view of a 600 µm z-stack of Hoechst-stained clarified brain tissue. Hoechst staining was visible at up to 300 µm of depth after which signal intensity was very low (n = 2).). These sections could complete delipidation in around 16h incubation in SDS. Cleared sections in Figure 3.11D & E were imaged after overnight SDS incubation and 2h in RIMS. Critically, when incubating 1mm sections in 8% SDS, endogenous fluorescence was preserved (Figure 3.11F & G). Hoechst 33342 could be added to the RIMS solution at 1 µg/ml to examine imaging depths. The orthogonal view of a 600 µm z-stack is shown in Fig 6H, hoechst stained nuclei were easily observable up to ~300 µm of depth (Figure 3.11H). Nuclei were still visible below this point, but intensities were very low.

Figure 3.12A-H shows images of a R26-confetti PIC tumour taken at 50 µm increments in the z-axis. The mean corrected fluorescence intensity of ECFP, EYFP and tdimer2 for the image is shown in in Figure 3.12I in arbitrary units (AU), EGFP was not measured. EYFP+ and tidimer2+ cells were still discernible at 300 µm of depth although fluorescent intensity had reduced drastically for both fluorophores. ECFP signal appeared to reduce faster than EYFP and tdimer2 with most signal disappearing by around 200 µm in depth. Shorter wavelengths are subject to more scattering events than longer wavelengths which may explain apparent lower ECFP+ signal at greater imaging depths

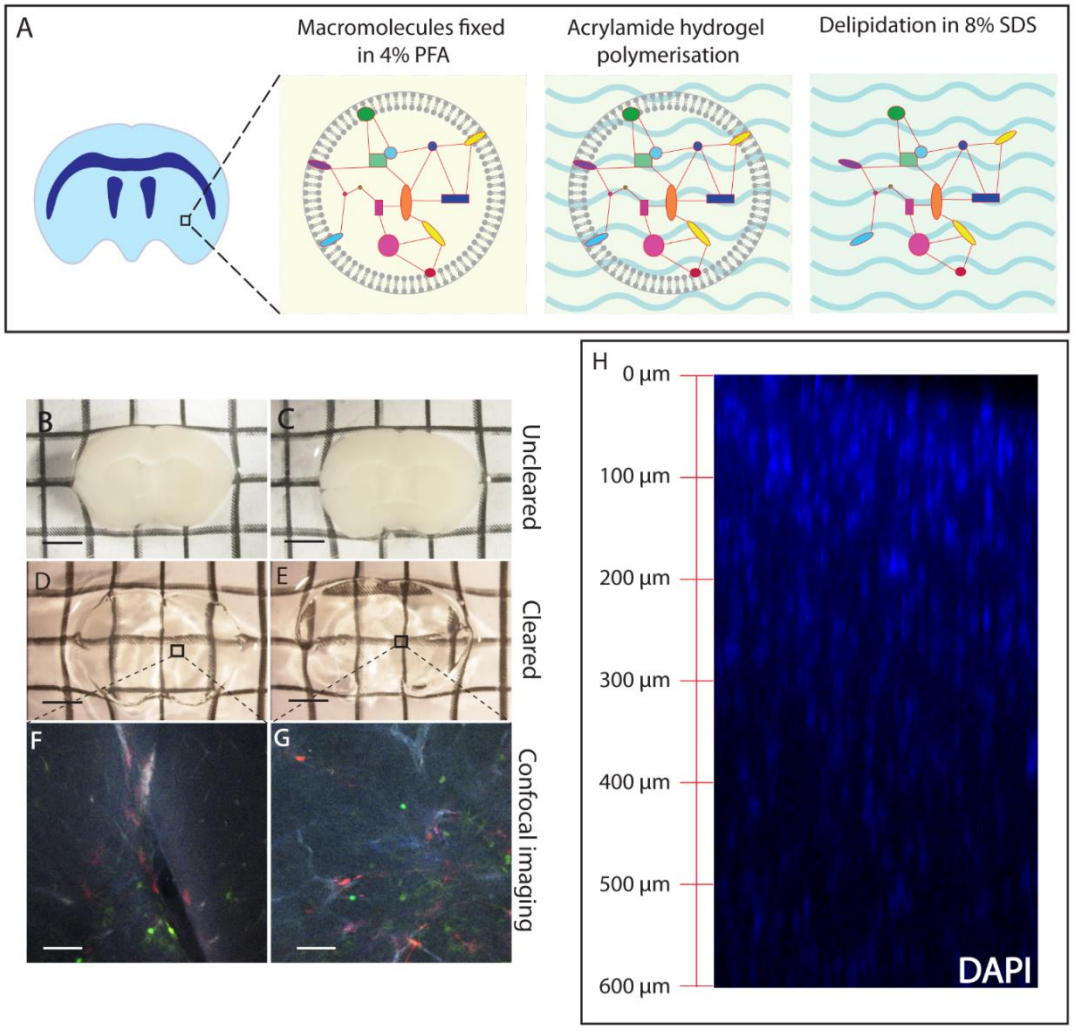


Figure 3.11: Passive clarity tissue clearing permits deep imaging and preservation of R26-confetti fluorophores. (A) Schematic of theory behind clarity tissue clearing. Tissue is fixed then embedded in acrylamide hydrogel to fix tissue macromolecules. Lipids are then diffused from the tissue by incubation in detergent solution (8% SDS).(B-E) Examples of brain tissue before and after clarity procedure demonstrating excellent optical clearing (n = 2)(Scale bar = 500 µm). (F & G) Regions containing cells with preserved endogenous fluorescence (scale bar = 100 µm). (H) Orthogonal view

of a 600 µm z-stack of Hoechst-stained clarified brain tissue. Hoechst staining was visible at up to 300 µm of depth after which signal intensity was very low (n = 2).

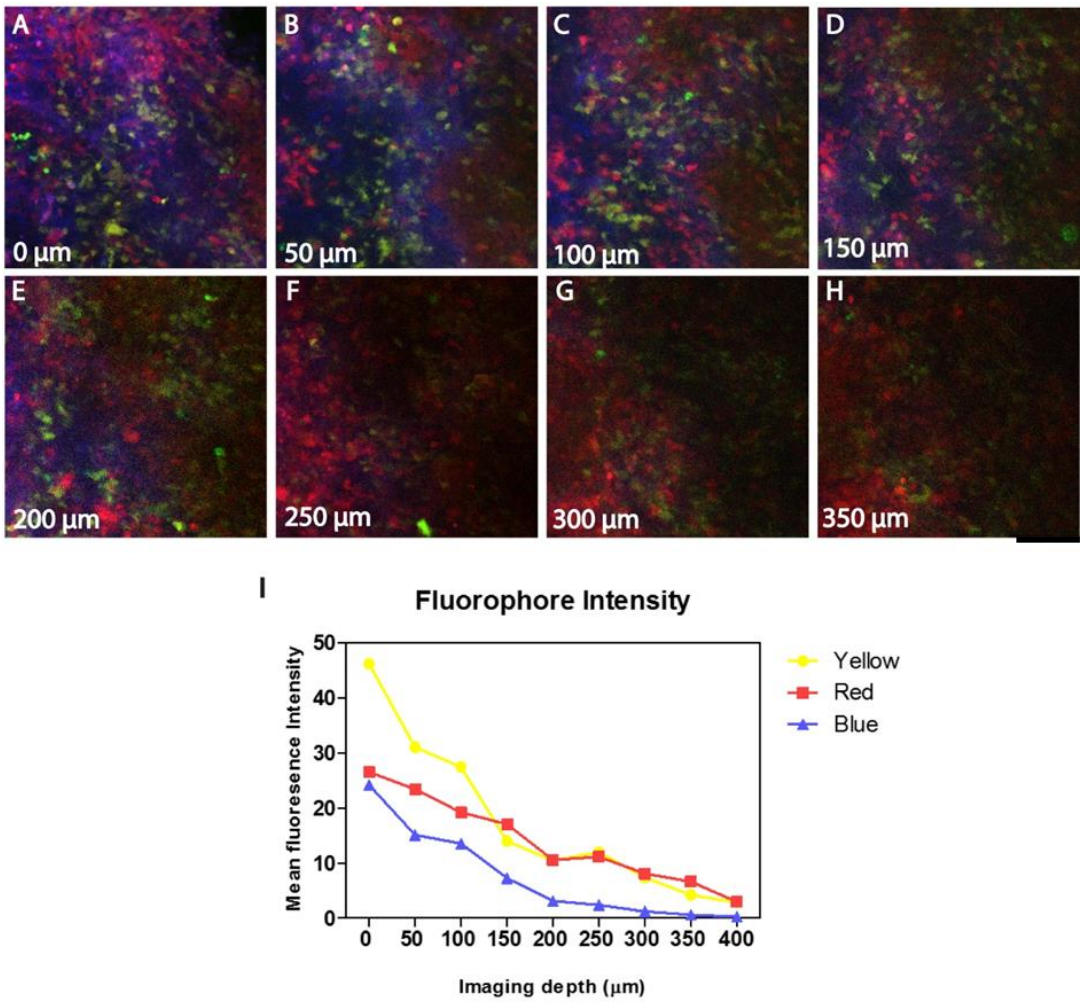


Figure 3.12: Clarity is best suited to imaging longer wavelength fluorophores. (A – H) Representative merged images of rose26-confetti tumours containing ECFP, EGDP, EYFP and tdmer2 cells imaged at 50 µm increments in the z-plane (scale bar 100 µm). (I) Mean corrected fluoresence in arbitrary units for each ECFP, EYFP and tdimer2. EYFP and tdimer2 fluorescence is still detectable up ~300 µm while ECFP detection is lost at almost 200 µm. N = 1 cleared tumour sample.

3.2.6 FACS sorting of PIC induced confetti tumours reveals recurrent flipping of the confetti locus

As an alternative to investigating cells in situ, we planned to isolate and propagate the four genetically labelled lineages from the same tumour. We would then use in vitro assays and organoid culture to see if the behaviours of separate lineages diverged during tumour propagation. To explore this method, tumours were again induced using PIC retroviral injections and propagated until animals developed clinical signs of intracranial tumours (4 – 5 weeks). Fresh brain tissue was sampled, and tumour regions extracted then dissociated using the Worthington papain dissociation kit® (Section: 2.3.2). We then attempted to separate the ECFP+, EGFP+, EYFP+ and tdimer2+ cells and propagate them in culture. The cell sorting was set up with a gated system where cells were initially gated for EGFP or EYFP expression (Figure 3.13A & B). Cells negative for both were then passed onto a further gate where they were tested for tdimer2 or ECFP expression (Figure 3.13A & C). With this setup we effectively isolated tdimer2+ and ECFP+ cell populations (Figure 3.13D & G). However, there was an issue deviating between EGFP+ and EYFP+ cells as our EGFP collection was composed primarily of EYFP+ cells (Figure 3.13E). Surprisingly, most cells in the EYFP collection were double positive for ECFP and tdimer2. A further complication arose when isolated populations were imaged 2 days after sorting. Many cells within the initially pure ECFP and tdimer2 collections were positive for both markers (Figure 3.13H' & K'). By 5 days after FACS, the ECFP collection contained cells expressing tDimer2 only (Figure 3.13L'). While the tdimer2 collection contained a large portion of cells expressing ECFP only (Figure 3.13O').

This observation suggested a considerable number of tumour cells was constantly changing colour during tumour propagation. Retroviral genes integrate into the host genome so PIC retroviral induced tumour cells constitutively express Cre recombinase. As a result, the confetti locus would be constantly flipping throughout the lifetime of these cells. It is reported that genomic regions flanked by inverted LoxP sites can become unstable or damaged by long-term cre exposure (Meinke et al., 2016). This may explain why some cells are not changing colour. It is conceivable that the construct has become "damaged" and is locked in a single orientation.

Additionally, the observation that some PIC induced tumours contained a large portion of unlabelled cells could also be attributed to this effect. Perhaps the construct does not correctly re-insert during flipping and fluorophore expression ceases.

The suggestion that cells are constantly changing colour during tumour propagation further limits the PIC model for Lineage tracing. It would only be possible to compare two lineages; ECFP+/tdimer2+ and EGFP+/EYFP+ as it is likely that the colour of a given cell continuously flips during tumour propagation. Given these results and factors discussed previously, we reasoned that the PIC retroviral model was not a suitable system for confetti lineage tracing.

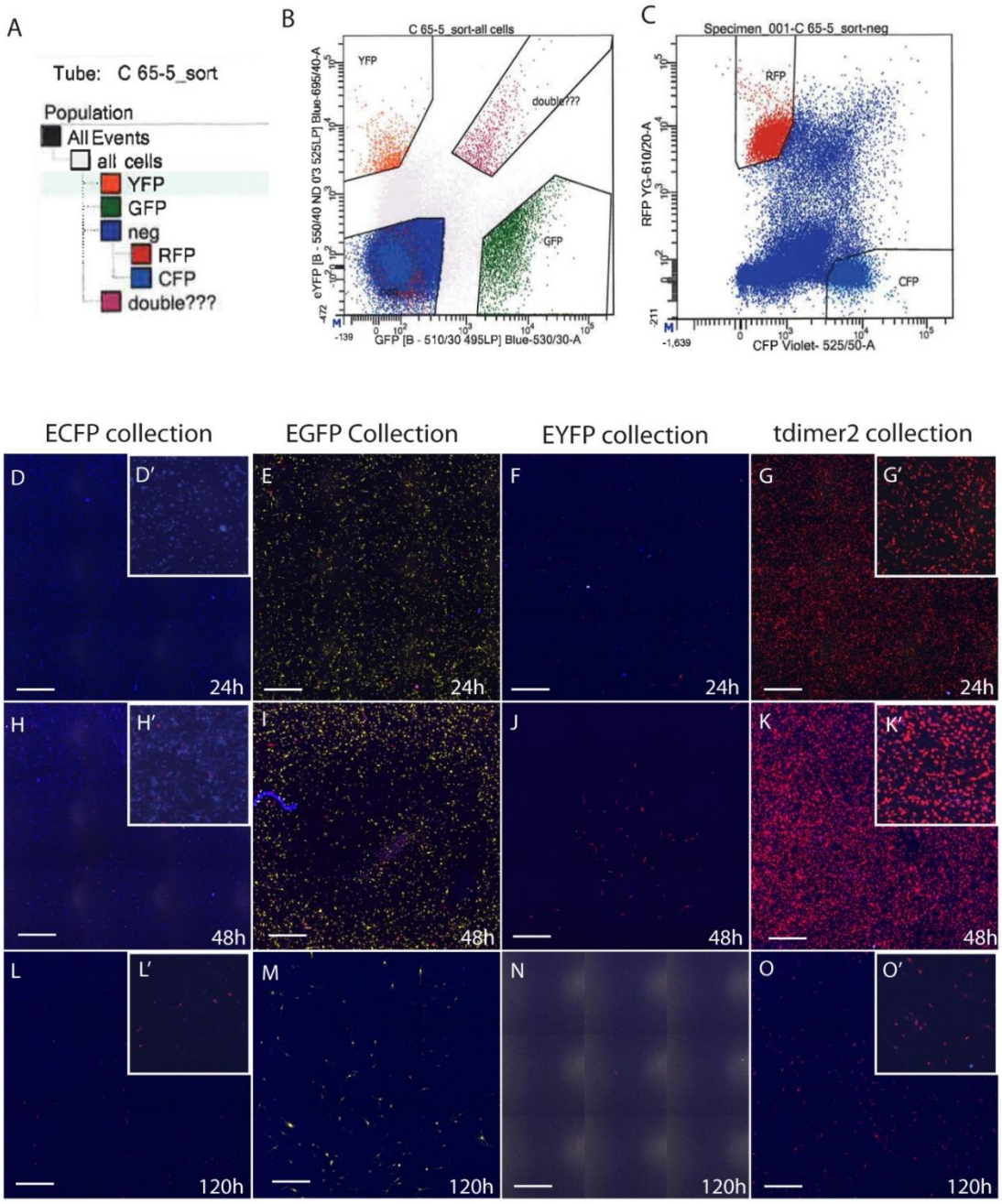


Figure 3.13: FACS analysis of PIC induced r26-confetti tumour cells reveals autonomous flipping of the confetti locus. (A) Demonstrating the gated setup of our sort (n = 1 PIC induced tumour). (B) Fluorescent cells plotted for EGFP/EYFP expression; regions highlighted demonstrate which cells were collected. Double positive cells were detected at this stage (C) Cells that were negative for EYFP and EGFP (bottom left in B) plotted for ECFP/tdimer2 expression. Regions where ECFP^{+} and tdimer2^{+} cells were collected from are designated a CFP and RFP respectively. Spread would suggest presence of double labelled cells. Lower panel (D-O) demonstrate adherent culture of r26-confetti tumour cells for each collection at 24h, 48h and 120

h post sort (Scale bars 400 µm). (D', H', L') Show emergence of tdimer 2 expressing cells in the ECFP collection, (G', K', O') Shows emergence of ECFP expressing cells in the tdimer2 collection.

3.2.7 Sectioning of living tumour tissue and ex vivo culture permits serial imaging of tumour tissue capturing gradual cell death

To achieve serial imaging of confetti labelled cell populations, we explored sectioning of freshly sampled confetti-GBM brain samples with PIC induced neoplams. PIC induced confetti-GBM tumour tissue (n = 6) was sectioned on a tissue chopper and transferred to wells containing raised inserts such that sections were not submerged in media. After sectioning, neoplasms were found generally within cortical regions (Figure 3.14B-F) but in one case we observed an outgrowth in the sub-cortical white matter (Figure 3.14G). Imaging of sections revealed apparent clonal neoplasms (Figure 3.14C & D) as well as neoplasms composed of multiple clones (Figure 3.14B, D-F). Sections were maintained in the presence of 1 µM 4-OH Tamoxifen with the aim of flipping the confetti cassette and visualise cells changing colour. Unfortunately, no apparent changes in cell colour were observed through the period of culture which in most cases lasted 4 days. By this point, much of the tissue had become necrotic and cell fluorescence was diminishing, limiting the detection cells. Sections produced by the tissue chopper were quite thick which rendered focussing on neoplasms in the same plane on consecutive imaging days difficult. Furthermore, the thickness of sections may have compounded the tissue necrosis that was observed. Nevertheless, the detection of labelled cells suggests that with some improvements, slice culture could be a useful technique for assessing confetti labelled tumours. Furthermore, culturing and imaging slices of tumour organoids may be a better approach for achieving serial imaging of labelled tumour cells.

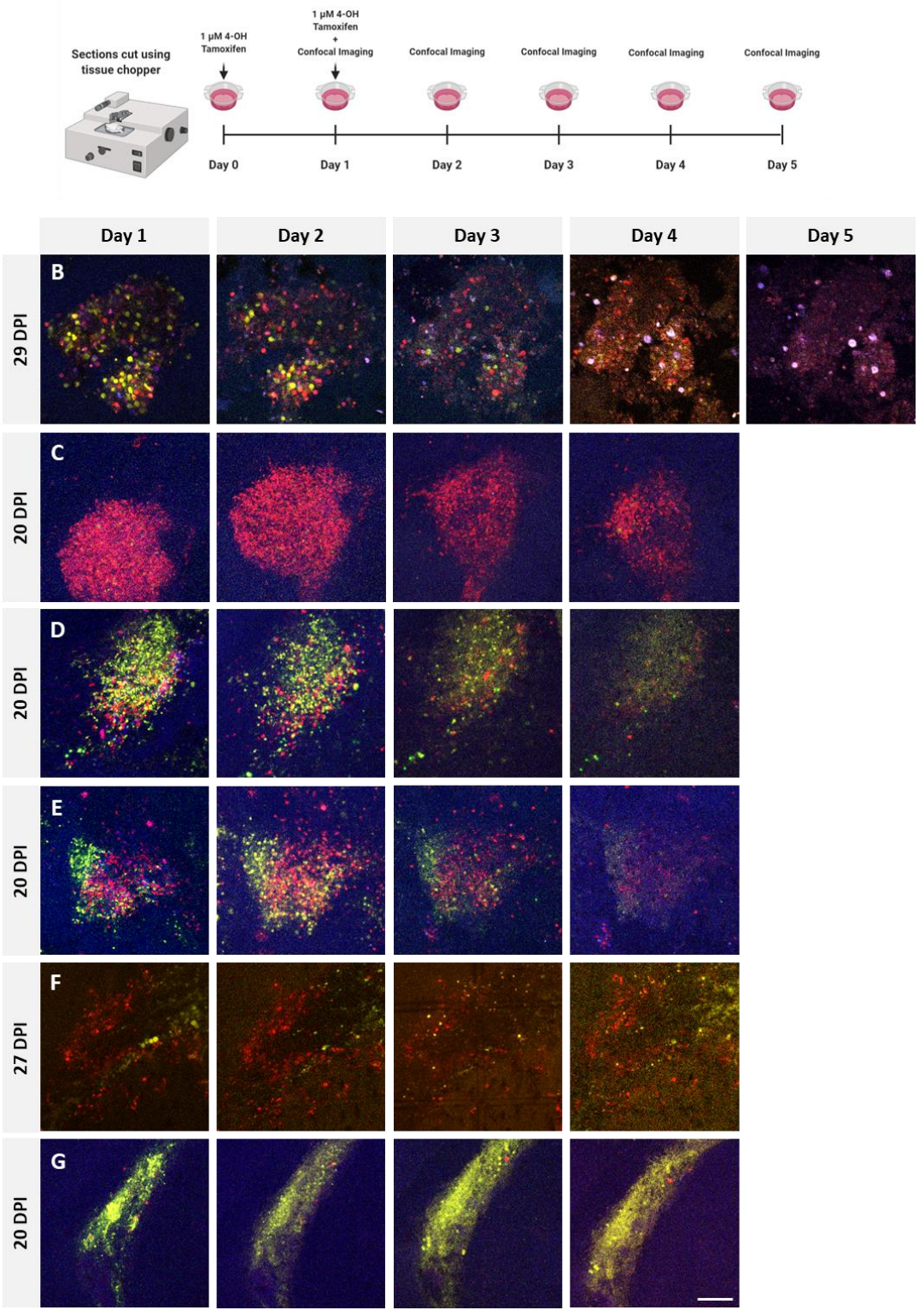


Figure 3.14: Sectioning of living tumour tissue and ex vivo culture permits serial imaging of confetti labelled tumours but shows gradual tissue necrosis by 4 or 5 days. (A) Schematic of experimental pipeline where tissue sections are maintained in media with 4-OH tamoxifen. (B – G) Examples of cellular neoplasms imaged over sequential days in animals injected with PIC retrovirus. Number of days since retroviral injection is displayed on the left-hand side.(n = 12, 6 displayed) (Scale bar = 200 µm)

3.2.8 Successful isolation and propagation of confetti labelled glioma cells in three-dimensional cultures

To assess intra-tumour heterogeneity, we seeded cells from different tumour lineages into three-dimensional organoid culture. When cultured in this fashion, cancer cells can be cultivated for protracted time period, potentially over months, without the need for passaging. Over such long culture periods, cells diversify and differentiate recreating aspects of heterogeneity observed *in vivo* (Hubert et al., 2016). Furthermore, the availability of nutrients at the core of a 3D culture is limited and as cells divide within the 3D scaffold, central regions become hypoxic and necrotic (see also discussion in section 5.3). Essentially, separate niches are created within the culture, fluorescent tracking could be used to explore how clonality relates to a cells propensity for proliferation in the internal hypoxic niche and/or the nutrient-rich surface.

Figure 3.15B-P shows cryosections of organoids seeded with a mixture of R26-confetti cancer cells isolated from a PIC induced tumour. Representative images of organoids frozen after 6, 8, 10, 12 and 14 days of culture and stained with Haematoxylin and Eosin (H&E) to compare their structures. Up to 8 days of culture cells appear to occupy organoid regions evenly (Figure 3.15B&C). By 12 days of culture central regions begin to show a much lower density of cells likely due to nutrient restriction within the organoid core (Figure 3.15E&F). By day 10, cells are much more densely located around the periphery and have formed an interconnected network of processes containing dispersed nuclei (Figure 3.15N). By day 14, densely packed nuclei line almost all regions of this network (Figure 3.15P). Interestingly, at day 14 more cells were observed in the organoid core than at 12 days (Figure 3.15T & U). Suggesting these cells possess a heightened ability to proliferate within this hypoxic nutrient restrictive niche.

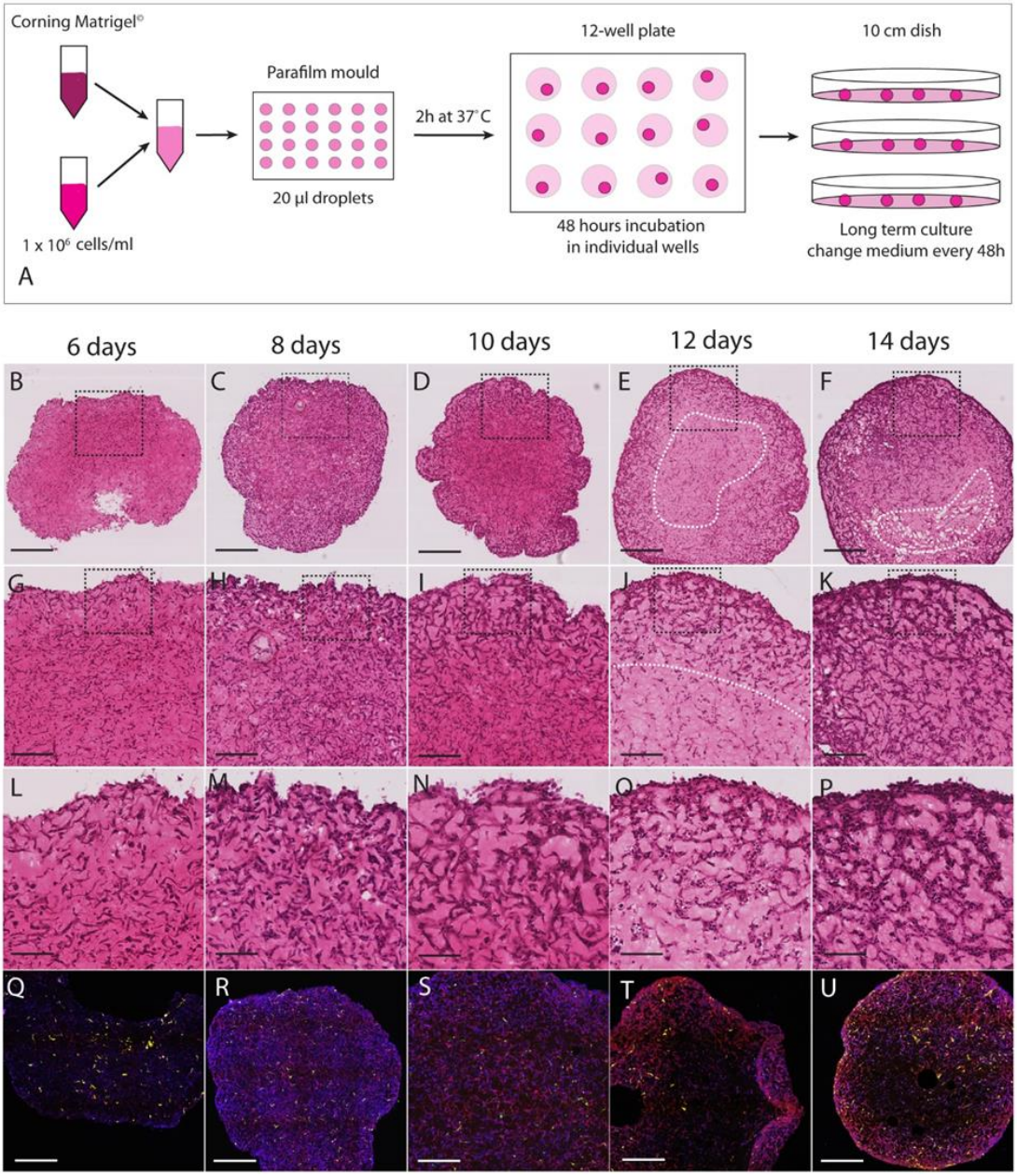


Figure 3.15: H&E histology and fluorescent imaging from r26-confetti organoid experiments. (A) Schematic of procedure for GBM organoid culture. Cells are mixed with Matrigel then 20 µl droplets are placed into parafilm moulds. Droplets are polymerised for 2h at 37°C the cultured in isolation for 2 days before being placed in 10 cm culture dishes. (B-F) Macro view of H&E stained cryosectioned organoids at 6, 8, 10, 12 and 14 days of culture (Scale bar = 500 µm). After 12 days (E & F), cells sparsely occupy the core region suggested to be caused by nutrient restriction and hypoxia. (C-P) increasingly higher magnification images showing structural changes as culture progresses. G – K scale bar = 200 µm, L – P scale bar = 50 µm. (Q-U) Fluorescent images of cultured organoids demonstrating the ability to visualise all r26-confetti fluorophores (Scale bar = 500 µm).

Figure 3.15Q-R displays fluorescent images of the samples in Figure 12B-P. It was possible to distinguish cells of all the different colours, $ECFP^+$ cells were much easier to distinguish compared to imaging in vivo. Fluorescent images also demonstrate the reduced cell density within core regions observed at 12 days of culture (Figure

3.15T&U). Three-dimensional imaging with Light-sheet microscopy was also performed (n = 1 organoid) revealing spatial heterogeneity in clonal composition (Figure 3.16) Therefore, tumour organoids represent a good model system for assessing tumour heterogeneity and comparing clonal dynamics of cells from a single tumour.

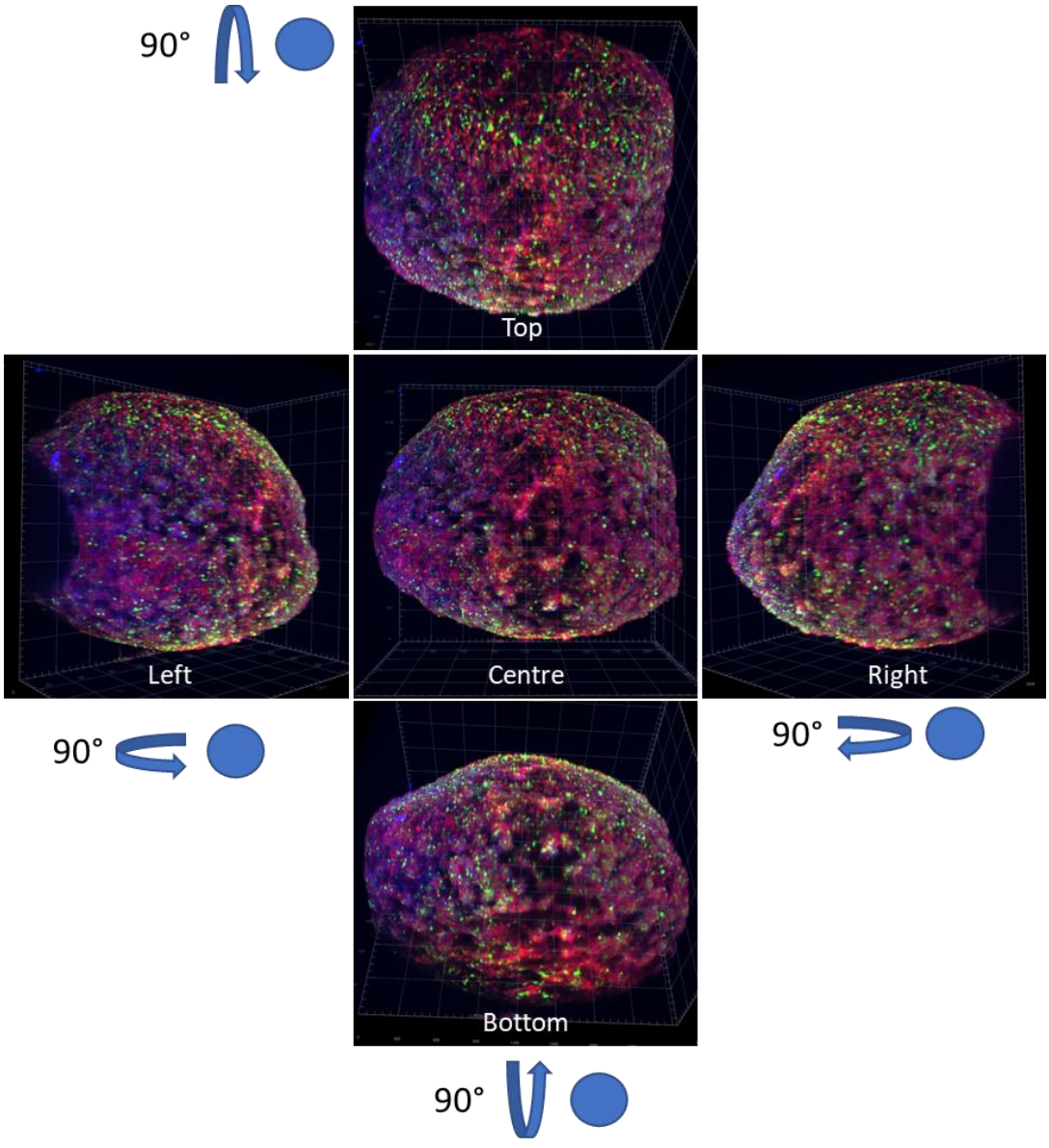


Figure 3.16: Light-sheet imaging reveals regional clonal heterogeneity of Matrigel glioma organoids: Three-dimensional reconstruction of confetti labelled organoid surface (n = 1). **Arrows and blue dot which represents Matrigel sphere**, indicate the degrees by which the central view has been rotated to give the top, bottom, left and right views of the organoid surface (Grid in 400 µm units).

3.3 Discussion

In this chapter we explored *in vitro* and *in vivo* applications for lineage-tracing in murine models of GBM. By crossing PTEN$^{flox/flox}$; P53$^{flox/flox}$ animals with ROSA26-confetti animals we generated a transgenic line where cellular Cre exposure induces expression of combinations of fluorophores as well neoplastic transformation in the same cells. Using the PIC retroviral glioma model (Zhang et al., 2019, Lei et al., 2011, Sonabend et al., 2013, Sonabend et al., 2014) we successfully generated tumours harbouring cells labelled with ROSA26-confetti fluorophores (Figure 3.7). Confocal imaging of confetti-labelled tumours revealed varied outcomes in tumour formation and clonal composition. Whether this heterogeneity represents stochastic variations in the model system or inherent heterogeneity in the evolution of tumours induced through PIC injections remains to be seen. With this model system we successfully developed approaches for tissue clearing and three-dimensional imaging, FACS label detection and separation, live cell imaging, serial imaging of tumour tissue sections and three-dimensional culture systems.

Firstly, through FACS sorting and serial passaging of distinctly labelled populations we identified persistent flipping of the confetti cassette in tumour cells produced through PIC injections. This was attributed to the design of the confetti locus in terms of parallel and inverted loxP sites. After the initial Cre exposure the confetti locus is excised to leave a cassette containing two fluorophores (ECFP and RFP or EGFP and EYFP) flanked by inverted loxP sites for flipping the cassette and changing fluorophore expression. In the PIC model, Cre expression remains constitutive due to integration of the retroviral transgene into the genome of target cells. It is likely that this leads to persistent flipping of the confetti locus and explains why red cells appeared in the blue sorted population and blue cells appeared in the red sorted population after a few days of culture. This represents a major disadvantage for lineage tracing as cells keep changing colour combinations and render it difficult or impossible to identify related progenies. Therefore, models using transient cre delivery for tumour induction would work better in combination with confetti lineage-tracing. For example, we demonstrated extensive recombination and expression of the confetti cassette after a course of *in vivo* tamoxifen injections to

cells expressing the GlastCreERT2 transgene. However, tumours induced in this fashion have the technical disadvantage of low penetrance and incubation times of up to 10 months compared with 30 -40 days using PIC-induced tumours.

A further option was to establish NSC cultures from PTEN$^{flox/flox}$; P53$^{flox/flox}$ animals and perform the Cre recombination step *in vitro.* We achieved this by incubating cells with Adenovirus-cre or, in cell lines expressing the GLASTCReERT2 transgene, incubation with 4-OH Tamoxifen. By incubating with adeno-cre, the highest dose achieved a recombination and thus fluorescence expression in nearly 100% of cells, while 4-OH tamoxifen delivery achieved a maximum of ~14% recombination efficiency across all concentrations tested. Suggesting that only a subset of murine NSCs in culture are expressing GLAST and susceptible to recombination in this fashion. Nevertheless, both adeno-cre and 4-OH tamoxifen approaches successfully produced recombined labelled cells that would be suitable for subsequent allografting. One would suspect that successfully recombined murine NSCs with PTEN and P53 loss would outgrow unrecombined cells within a few passages. However, allografting cells is not as accurate a model for tumour development compared with *in vivo* recombination approaches where cells exit their canonical *in vivo* regulation to from neoplasms. One advantage of *in vitro* recombination for subsequent allografting would be the ability to quantify the starting label compositions in the transplanted cell populations which is advantageous for identifying the underlying label population from which cell expansions have occurred. Many PIC-induced tumour were composed almost completely of cells with the same label and whether this was due to expansion of a single clone or a large representation of cells with that colour after initial recombination was unclear. As shown previously, we find that in the setup with *in vitro* recombined confetti cells, fluorophore expression is not evenly distributed throughout the recombined populations. Therefore, it is likely that similar distributions of fluorophore expression in the recombined cell population occur during *in vivo* recombination.

In tumours where multiple confetti labels and therefore multiple clones were present, clonal populations did not show any distinct, or reproducible organisational or structural pattern. Imaging applications of confetti lineage-tracing have great

power in revealing the tightly controlled and ordered patterns of clonal behaviour in tissue development and regeneration. While it is possible that persistent flipping of the confetti locus was masking any underlying order to tumour development, unsurprisingly, it appears PIC induced tumours show a highly disordered and variable clonal development. Suggesting that little understanding of clonal tumour development can be achieved through direct observation of end-stage confetti labelled tumours. Alternatively, investigating the early stages of tumour development where smaller neoplasms are forming and clonality can be more easily identified, may be a better application of confetti lineage-tracing in this model. Moreover, looking at time series of tumour development may also provide a more informative readout of tumour development. For example, where end stage tumours form containing cells primarily of an individual label, time series analysis would indicate if these tumours generally began with greater clonal diversity which was lost over time.

Clonal diversity is part of the multi-faceted concept of tumour heterogeneity where cell intrinsic and cell extrinsic factors shape tumour cell diversity on genetic and epigenetic level, which affects cellular phenotypes. Since we observed development of both clonally diverse and near-monoclonal tumours it would be interesting to see how clonal diversity influences tumour biology. For example, classification of tumour clonality through flow cytometry combined with RNA sequencing would reveal any underlying transcriptomic differences between clonally diverse and clonally uniform tumours. Furthermore, FACS could be used to isolate individual clones for subsequent RNA-sequencing or even single-cell RNA sequencing to assess differences within clones. Such experiments would address the contentious questions of how cell heritage contributes to shaping tumour cell biology and bulk tumour heterogeneity. A recent study combined viral barcoding and single-cell RNAseq with xenografting of human glioma CSCs (Neftel et al., 2019), and showed that cells of an individual clone exist across multiple characterised expression states resembling parenchymal brain cells suggesting cellular heritage has little influence over the expression states of individual cells. However, this study was limited to just a few cells per clones (<10). FACS sorting and subsequent RNA sequencing of fluorescently traced cells would

permit sequencing of many more cells per clones. Such an approach would be valuable for interrogating cellular hierarchies between clones within an individual tumour (Lan et al., 2017). A possibility to explore this further would be the combination of fluorescent tracing and surface marker expression analysis through flow cytometry to investigate the relationship between clonality and CSC surface marker expression (Dirkse et al., 2019b). Our experimental setup exploring this option is described in Chapter 5. Alternatively, surface marker staining could be used to characterise immune responses to monoclonal or clonally diverse tumours to investigate how clonal architecture influences immune responses and vice versa. Finally, with the increasing availability of *in situ* sequencing techniques further development of fluorescent tracing applications could permit interrogation of spatial and clonal gene expression. Such an approach would have great value in investigating how microenvironmental factors influence cellular states while providing a readout of clonal proliferation.

Here, we successfully achieved optical clearing and high-depth (>300 µm) imaging of confetti tumours using passive clarity and confocal microscopy (Tomer et al., 2014, Yang et al., 2014). Early experiments found considerable bleaching of expressed fluorophores when labelled tumour tissue was incubated in SDS for prolonged periods. To mitigate this, fixed brain tissue was sectioned to 1mm thick slices, thus enabling complete delipidation of tissue after a single overnight incubation in SDS. With this method, endogenous fluorescence remained well-preserved and imaging of differentially coloured cells was achieved at larger imaging depths. In keeping with previous studies (Richardson and Lichtman, 2015) we observed that the depth of fluorescence detection depends on the wavelength of emitted light. Thus, EYFP and RFP were detected at greater depths than eCFP with its shorter emission wavelength. This effect is well characterised and is associated with greater refraction and scattering of shorter wavelengths as they pass through the cleared tissue (Richardson and Lichtman, 2015).

Effective three-dimensional imaging of large tumour areas using confocal microscopy was unfeasible, due to the long image acquisition times. Even the use or an advanced Laser scanning confocal microscope Zeiss LSM880 with its faster image acquisition

compared with the LSM710 was still insufficient for imaging of many of the larger tissue samples. Issues of protracted image acquisition could be resolved through the use of light sheet microscopy where we were able to image a ~3mm^2 region at up to 1mm in depth in around 5 minutes. However, limited laser lines and emission filter sets on the Zeiss light-sheet Z1 microscope limited the fluorophore acquisition to EYFP and RFP only. Nevertheless, use of a light-sheet microscope with appropriate lasers and filters could have resulted in effective three-dimensional imaging of fluorescently traced tumours. Here, we did not explore the potential for immunostaining on cleared tissue for potential interrogation of clonality and expression of prognostic markers. It is likely that prolonged incubations in detergent solutions for permeabilization required when immunostaining would affect the detection of endogenously expressed confetti fluorophores. As such, antibodies targeting expressed fluorophores would also be required to enable detection of the fluorescent labels but lead to highly complex immunostaining protocol.

In conclusion, this chapter describes the approaches to establish a lineage tracing model, combining the induction of flow for expression in combination with Inactivation of tumour suppressor genes, using cream mediated recombination. Whilst in principle providing a useful model, there were specific technical limitations, such as the continued recombination events, rendering this specific model system is suitable. The model was then modified using a transient Cre expression (GLAST-cre ER(T) promoter responding to tamoxifen exposure). This elegantly resolved the issues of continued Cre activity, but was fraught with low efficiency rates and long incubation times to develop tumours. A further approach was the derivation of neural stem cells from mice expressing a combination of the brainbow and tumour suppressor constructs (PTEN and P53) flanked by loxP sites, and the combination with either Adeno-Cre or tamoxifen incubation of the cultured spheres. However, subsequent imaging of thick slices derived from allografted tumours was met with technical limitations, either slices that could not be penetrated by the lasers of the confocal microscopy systems, or by the limited availability of filter sets in the light sheet microscopy setup.

Chapter 4: Optimising labelling assays for lineage-tracing primary human GBM cell lines

4.1 Introduction

For fluorescent lineage tracing in human cells, we adopted RGB ("red green and blue") marking. This technique involves lentiviral delivery of 3 fluorophores to produce up to 8 different distinct labels. An accurate virus titre and calculations based on mathematical set theory (Weber et al., 2012) achieve nearly even gene transfer rates for each delivered fluorophore. This technique is adaptable and can be expanded to include more fluorophores achieving a greater number of distinguishable labels (Mohme et al., 2017).

In the most elaborate application to date, an optical barcoding producing 41 unambiguously labelled U87 cell clones using six different fluorophores, was achieved (Mohme et al., 2017). Subsequently, barcoding and in vivo tracing of 21 distinctly labelled GL261 mouse glioma cells was achieved (Mohme et al., 2017). This approach relies on single cell FACS sorting to obtain homogenously labelled clonal populations. In previous work, primary cells have been selected for the neural stem cell marker CD133 to extract cells with the potential for indefinite self-renewal (Lan et al., 2017). Furthermore, achieving adequate yields of clonal cells through this approach can take upwards of 2 - 3 months which is likely to introduce changes in the epigenetic identity of clonal cells. Clonal populations established from single cells have undergone a substantial number of cell divisions in a context that is far removed from the GBM microenvironment. Therefore, in our work we aim to establish clones from mixtures of cells, bypassing the need for single-cell sorting.

In 2018, another study was published demonstrating clonal emergence after long-term culture of RGB marked cells (Brenière-Letuffe et al., 2018). This was demonstrated for iPSC lines as well as HEK293 and two fibroblast lines for which

clonal populations were detected by 4 passages after labelling. Clonal populations of dual or triple labelled cells were easily recognised as they created streaky clusters on flow cytometry dot-plots. Once a clone has expanded substantially, these cells can be FACS enriched for further manipulation or sorted straight into cell lysing buffer for extracting nucleic acids. Such clones arise after competition and/or cooperation with other lineages rather than their ability to withstand experimental stress and grow from isolation. Although fluorescence expression for cell tracing is widely used, it is important to concede that lentiviral infection and fluorophore expression may also affect the outgrowth of clones.

Nonetheless, we reasoned clonal emergence achieved with this approach was more suitable to study human glioma CSCs. To improve the procedure, we aimed to grow labelled cells as tumour organoids rather than serially passage them in adherent culture. In 2016 Rich et al (Hubert et al., 2016)showed a three-dimensional tumour organoid culture which emulate elements of the glioma microenvironment and cellular heterogeneity. Cells cultured in organoids show better levels of engraftment and are more invasive in xenografts when compared to adherently cultured human glioma CSCs. It Is not established if the same clonal emergence occurs in organoids but since these can be seeded with as few as 1000 cells we reasoned it was likely clonal populations would arise and could be collected after 4 – 6 weeks, representing an excellent tool for studying gene expression in human glioma CSC clones.

To validate labelling results of previous studies, we first trialled a co-transduction of untitred concentrated LeGO-vectors to U87 cells which were subsequently xenografted or grown as organoids. In parallel, virus titre were established and controlled RGB labelling techniques were optimised in HEK293T cells and subsequently applied to primary GBM cells G61. Primary cells were then subject to serial passaging and organoid culture using flow cytometry to assess the emergence of clones. Informed by experimental feedback form these assays we further optimise LeGO-vector based labelling for clonal detection. In addition to establishing a clonal tracking technique we also explored options for functional assessment of clones. One possibility explored was KD of a novel gene associated with glioma progression or alternatively combining fluorescent tracing with

phenotyping of glioma CSC marker expression to investigate clonal dynamics and marker plasticity.

4.2 Results

4.2.1 RGB-marked U87 glioma xenografts, three-dimensional culturing and FACS sorting

In the first instance we attempted to establish RGB marking using three different LeGO-vectors in adherent U87 glioma cultures. First, we aimed at producing a population carrying all 8 possible colour combinations from un-titred virus preparations. We delivered virally packaged LeGO-vectors carrying EGFP, Venus or Katushka2S at equal dilutions in three separate transductions (Figure 4.1). Dilutions ranged from 1:2000 and 1:8000 and this fluorophore combination was chosen to permit detection of the nuclear stain Hoechst 33342 alongside tracing labels. We achieved expression of all three fluorophores in each of the three transductions. As expected, cells exposed to the highest concentration of virus particles showed the highest proportion of cells expressing the three delivered fluorophores (Figure 4.1B). Furthermore, in 1:4000 dilution transduction more cells showed complex colours as a result of expressing two or three different fluorophores. Across all three transductions EGFP was expressed most widely, followed by KAtushka2S with Venus expressed least abundantly. This suggested that harvested viral preparations carried variable concentrations of virus particles and as such, titering of the viral preps would be required to deliver equal amounts of viral particles when performing co-transductions in this manner. Nevertheless, with the current set of label cells we were also able to optimise FACS parameters to separate the 7 different label combinations (Figure 4.2). We next set out to assess the prolonged expression of these vectors through culture of labelled U87 cells in three-dimensional organoids and through orthotopic xenografting of labelled cells into immunocompromised mice.

The culture exposed to a viral dilution of 1:2000 was selected for further culturing as it contained the greatest diversity of labelling. These cells were expanded to sufficient numbers for seeding into organoids and xenografting. We seeded 5000 U87

cells into 12 organoid cultures harvesting (n = 4) organoids at 2, 3 and 4 weeks of culture for cryo-sectioning (. Examples of serial sections from individually harvested organoids at the three time points are shown in Figure 4.3B - D. Histological assessment of labels revealed good preservation of fluorescence expression at all

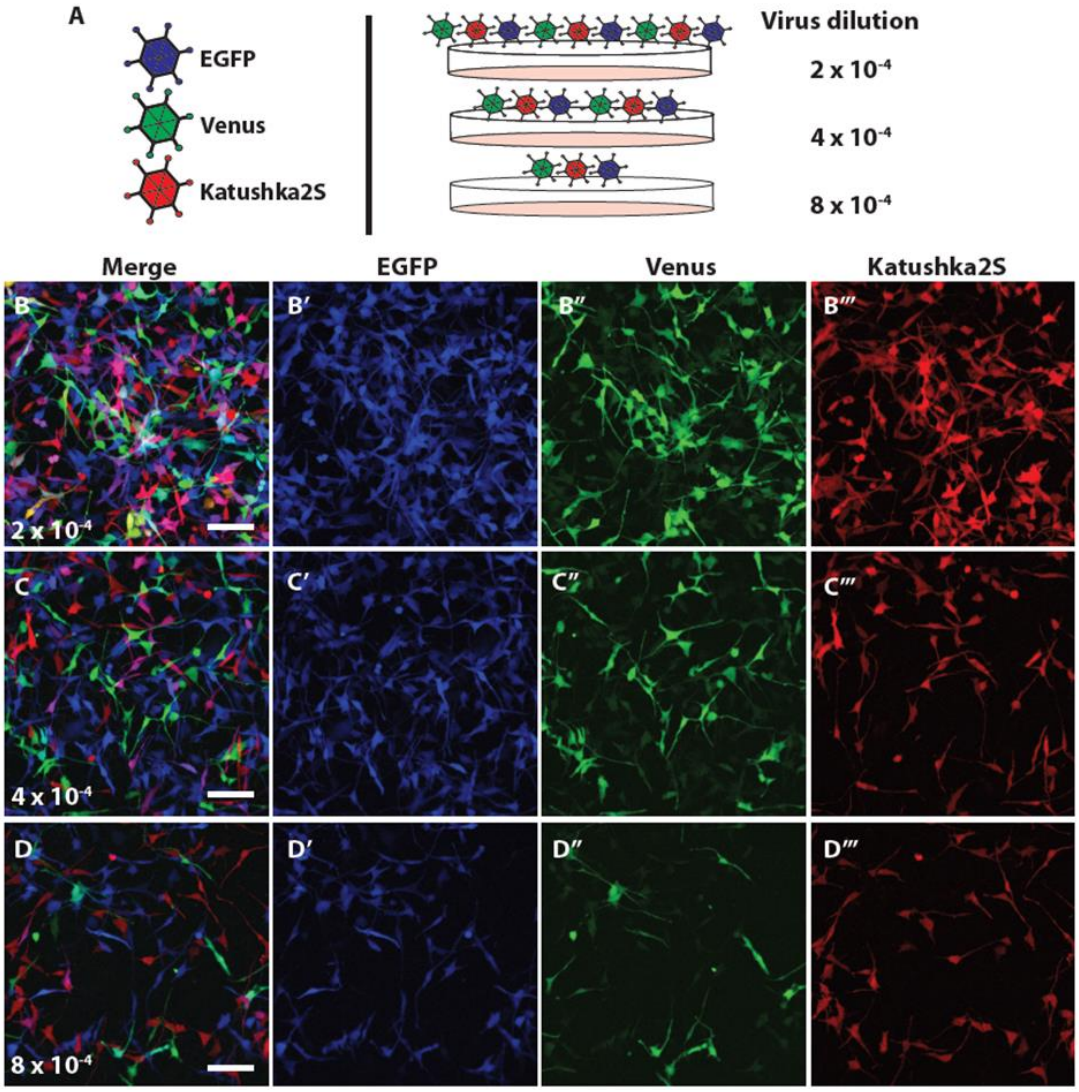


Figure 4.1: LeGO-vector transductions in U87 produce cells expressing three fluorophores in 7 combinations. (A) Schematic showing the three viral encoded fluorophores (EGFP, Venus and Katushka2S) and the corresponding dilutions of the three viruses in the different transductions (n = 1 infection per dilution). (B) 2 x 10^{-4} viral dilution produces cells of all 7 fluorophore combinations through overlapping expression. (C) 4 x 10^{-4} viral dilution achieves lower levels of viral infection and less effective at producing cells labelled with 2 or more fluorophores. (D) 8 x 10^{-4} viral dilution achieves lowest level of infection achieving mainly single label infections (red, blue and green). Single channel images show EGFP has the highest level of infectivity across the three dilutions and Venus has the lowest

time points of organoid culture. Furthermore, from time points 2 to 4 weeks we also observed a reduction in the label variability observed in harvested organoids. In the 2 week organoid (Figure 4.3B) cells expressed a variety of fluorescent labels spatially distributed throughout the sections. However, in our sections of organoids harvested at 3 weeks and 4 weeks we can see that certain labels are beginning to achieve prominence in the cultures. By 3 weeks (Figure 4.3C) we can see an organoid

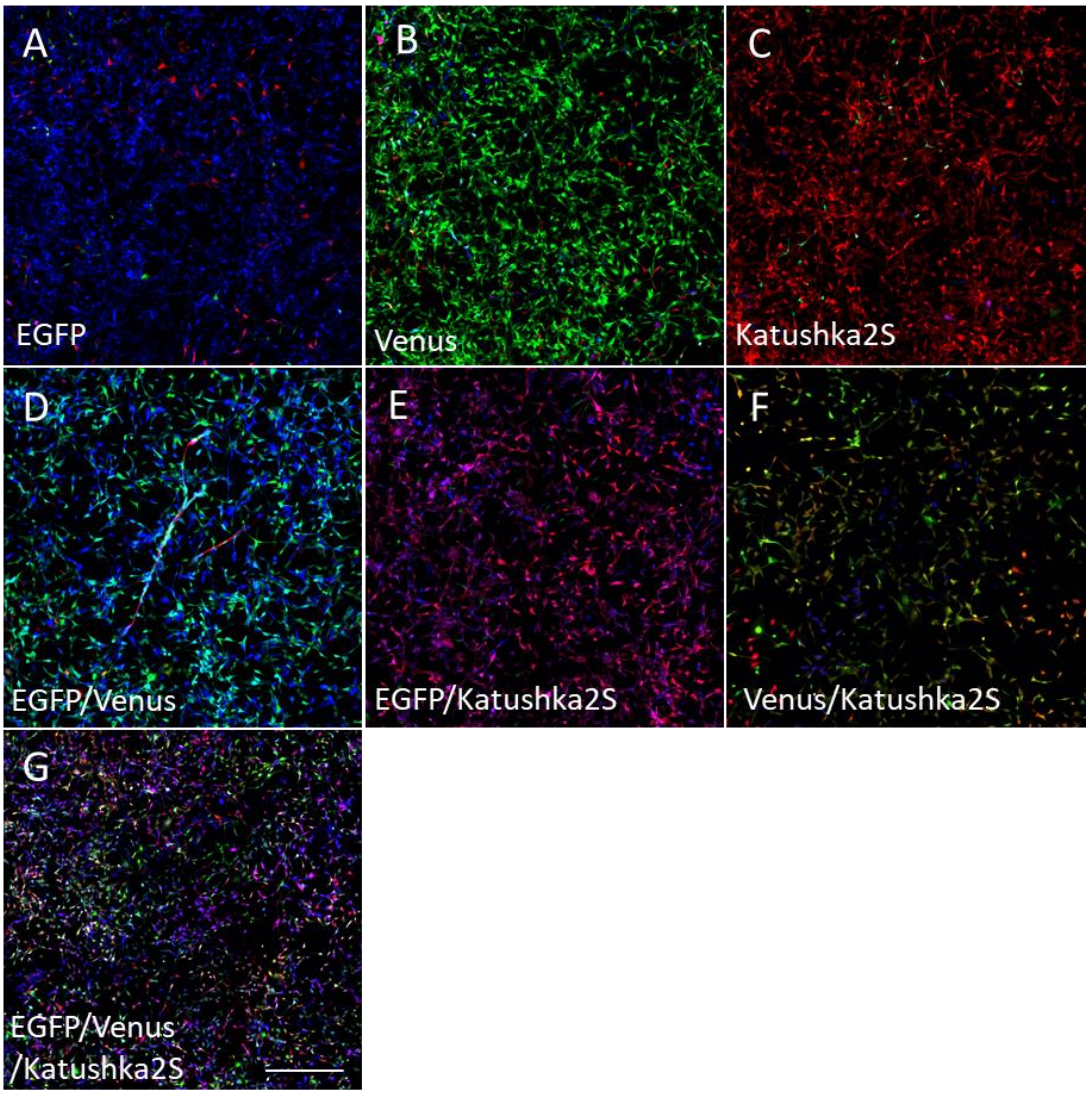


Figure 4.2: Fluorescent activated cell sorting can purify uniquely labelled populations with only minimal contamination from other label. (A – G) Merged confocal images of cell populations after FACS experiment to isolate cells carrying the same label combination. Corresponding fluorophore of each collection shown in bottom left of image panel. (D – E) Double positive and (G) triple positive cells are not homogenously labelled with many different hues produced from the same underlying label combinations (scale bar = 200 µm)

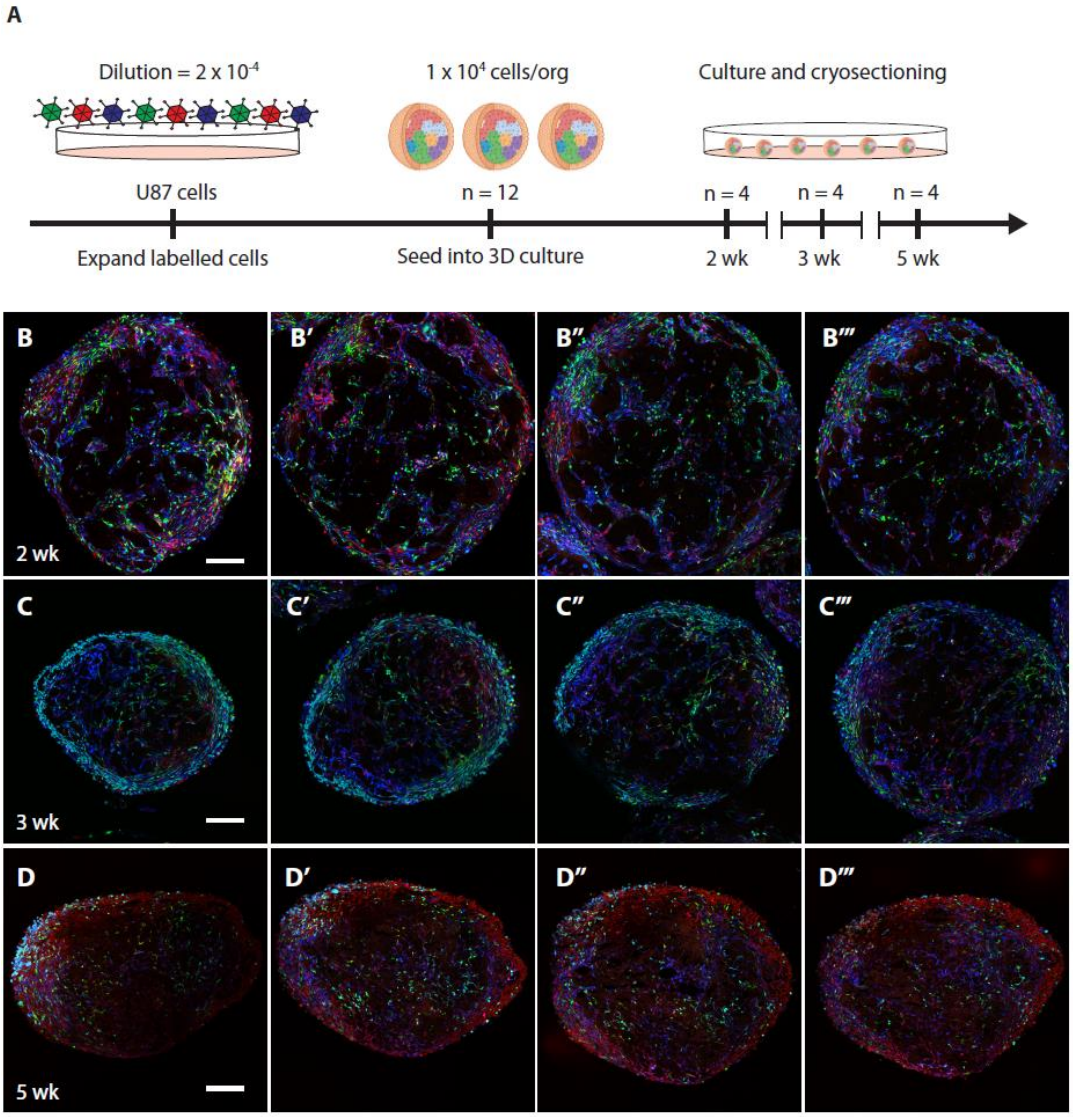


Figure 4.3: Organoids grown from triple transduced U87 cells show expansions of cells carrying the same label. (A) Experimental schematic, cells from the 2 x 10^{-4} dilution triple infection were expanded and 10,000 cells seeded into (n = 12) organoids with 4x organoids harvested after 2, 3 and 4 weeks of culture. (B) Serial sections of an organoid harvested after 2 weeks of culture, showing a large degree of label variability throughout the organoid with cells of all 7 underlying fluorophore combinations visible. Cells forming a network internally and also forming a thicker layer around the periphery. (C) Serial sections of an organoid harvested at 3 weeks showing a reduction in label variability amongst the cells. A population of cyan cells can be seen extensively around the periphery across all the sections. (D) Serial sections of an organoid harvested at 5 weeks showing prevalence of cells harbour Katushka2S expression (red) across all sections. Cyan (EGFP + Katushka2S), green (Venus) and blue (EGFP) are the only other labels visible throughout the organoid. Internal network structures visible at 2 weeks not as apparent (scale bars = 300 µm).

primarily composed of cells coloured in cyan (EGFP + Venus) or Blue (EGFP) with a smaller number of cells represented by green (Venus) and purple (EGFP + Katushka).

Interestingly the colours appeared to show spatial localisation with cyan coloured cell localising to the surface and blue cells more prominent centrally. Reduced label variability indicative of clonal outgrowth was even more pronounced in organoids imaged at 5 weeks, the example sections (Figure 4.3D) show an organoid with red (katushka2S) coloured cells growing throughout and smaller clusters of green, blue and cyan coloured cells.

Figure 4.4 and Figure 4.5 show examples of tissue harvested form immunocompromise mice after xenografting of our labelled U87 cells. We injected 50,000 labelled cells in mice (n = 6) and harvested tissue after 5 weeks of incubation. Labelled cells were detected in 3 out of 6 injected animals. In one animal we found cells growing around the periphery of the brain tissue clustered around blood vessels (Figure 4.4 A, B and C). Additionally, cells were also observed migrating dorsally along blood vessels into more central brain regions (Figure 4.4C). In the other two animals, labelled tumour cells were not found within brain tissue but dense tumour tissue had formed in the meningeal spaces (Figure 4.5A and B). Like organoid culture we found a reduction of label complexity indicative of clonal outgrowth and nuclear labelling with Hoechst 33342 suggests outgrowth of many unlabelled tumour cells. In both animals (Figure 4.5A and B), labelled cells can be observed in spatially discrete locations seemingly migrating and spreading through the tissue from more densely populated regions. Apparent migratory streams (Figure 4.5 B’ and C) and other interesting structural features (Figure 4.5A’ and B’) are observed in densely labelled regions while cells appear to spread and are almost evenly distributed throughout sparsely labelled regions (Figure 4.5A’’ and B’’).

In conclusion, LeGO-vectors can be transduced by lentiviral vectors into glioma cell lines to label cells in more traceable colour hues than the number of delivered of delivered virus. Fluorophores are well expressed and appear to show stable expression over time judged by detection of fluorescent cells after organoid culture (4 weeks) and xenografting (5 weeks). We found that the diversity of label population reduced during organoid culture and xenografting, perhaps as a result of clonal outgrowths within the traced cell population. However, fluorophores were not evenly expressed throughout the initially labelled populations and a large portion of

unlabelled and therefore untraceable cells were observed, particularly evident after xenografting. To achieve even distributions of fluorophore expression, maximise label diversity and minimise the number of unlabelled cells we set out to titre virus preparations for a more controlled labelling approach.

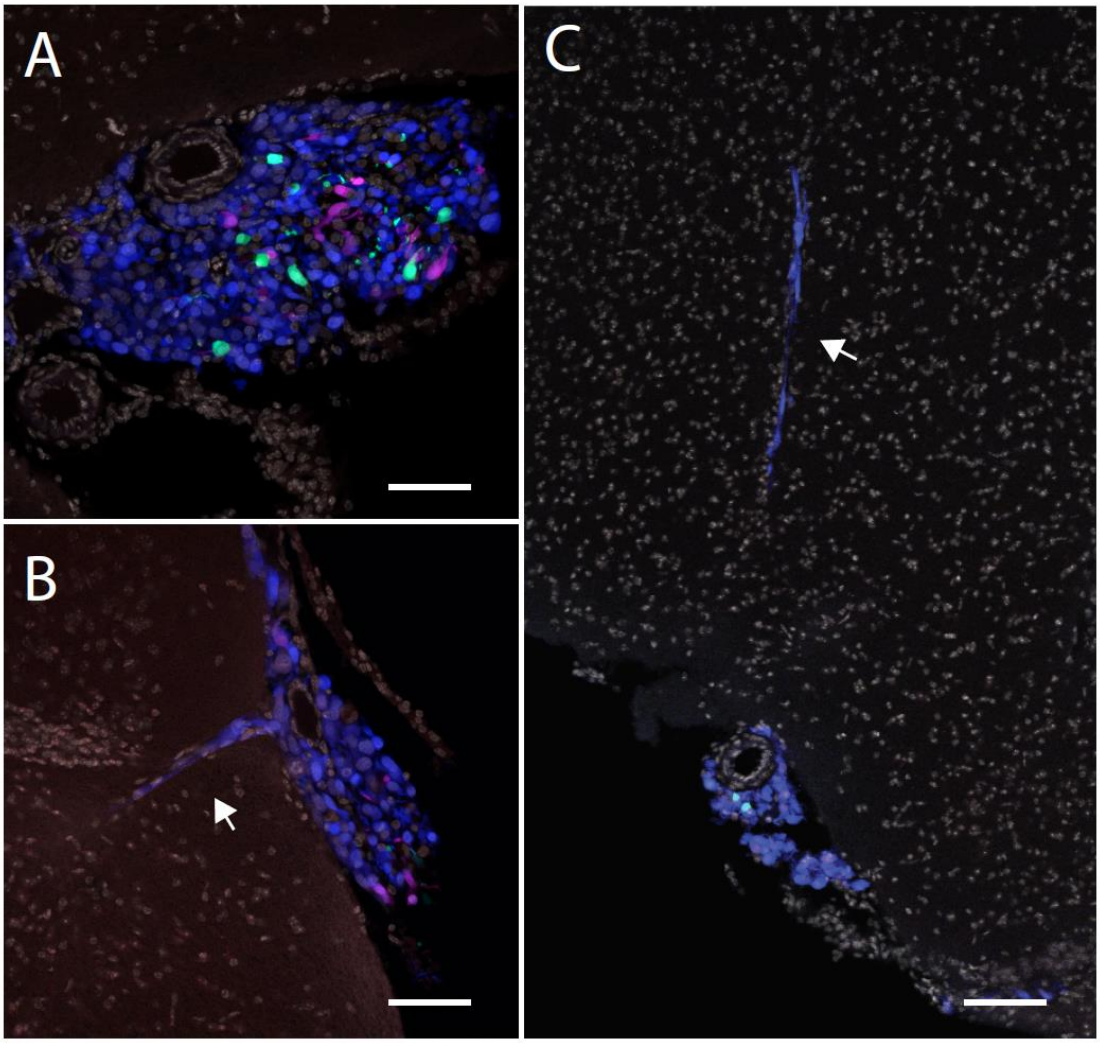


Figure 4.4: Labelled U87 cells grow around blood vessels after xenotransplantation into NOD/SCID immunocompromised mice. (A) U87 cells around blood vessel with primarily blue (EGFP) label with a smaller presence of purple (EGFP and Katushka2S) and cyan (EGFP and Venus) populations (Scale bar = 200 µm). Sections counter stained with Hoechst 33342. (B) Further example of blue cells growing around a blood vessel with some cells migrating int brain along the vessel (Scale bar = 200 µm. (C) A further proliferative cellular colony with another example of cells migrating into the brain along a blood vessel (arrow). (scale bar = 300 µm)

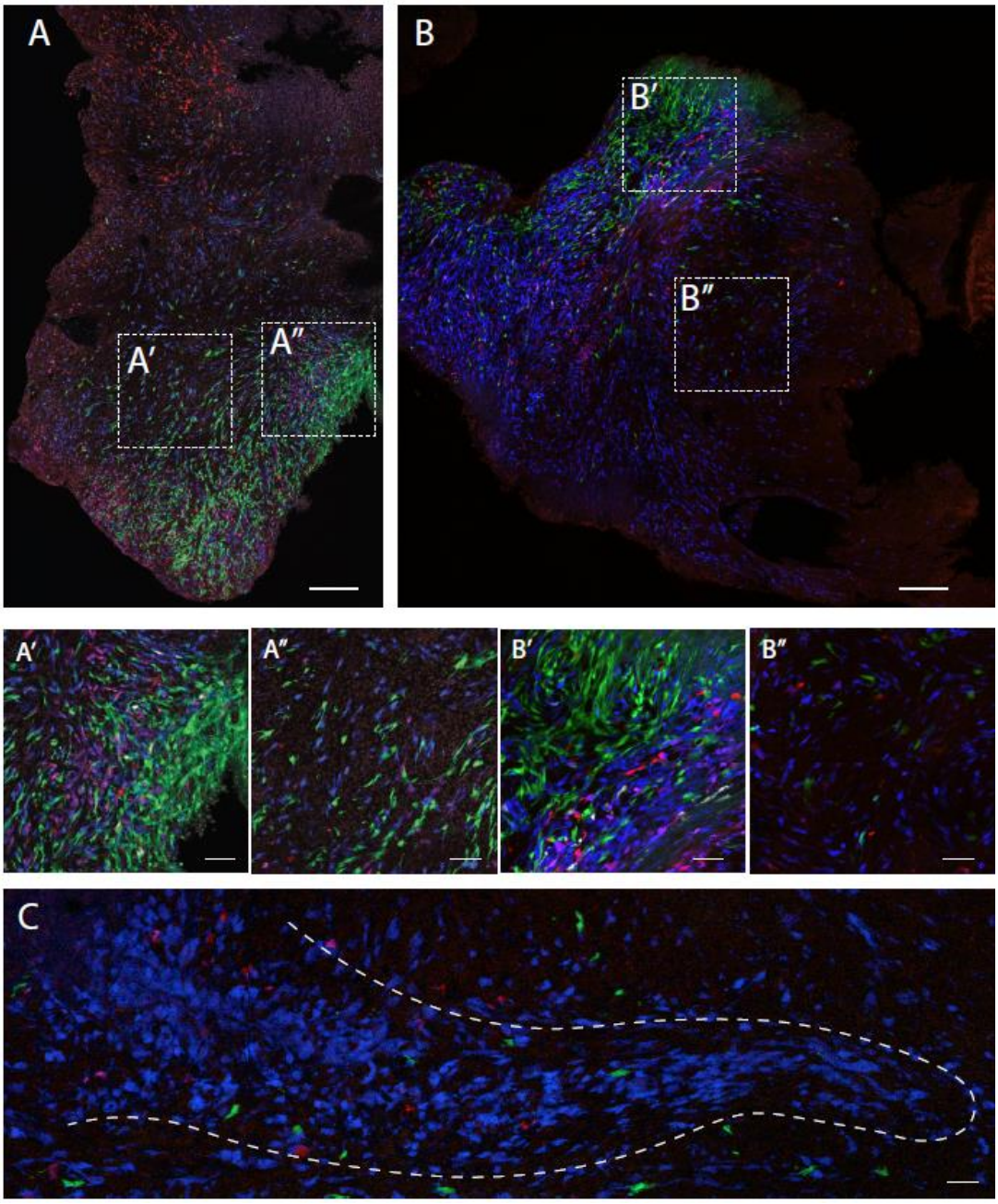
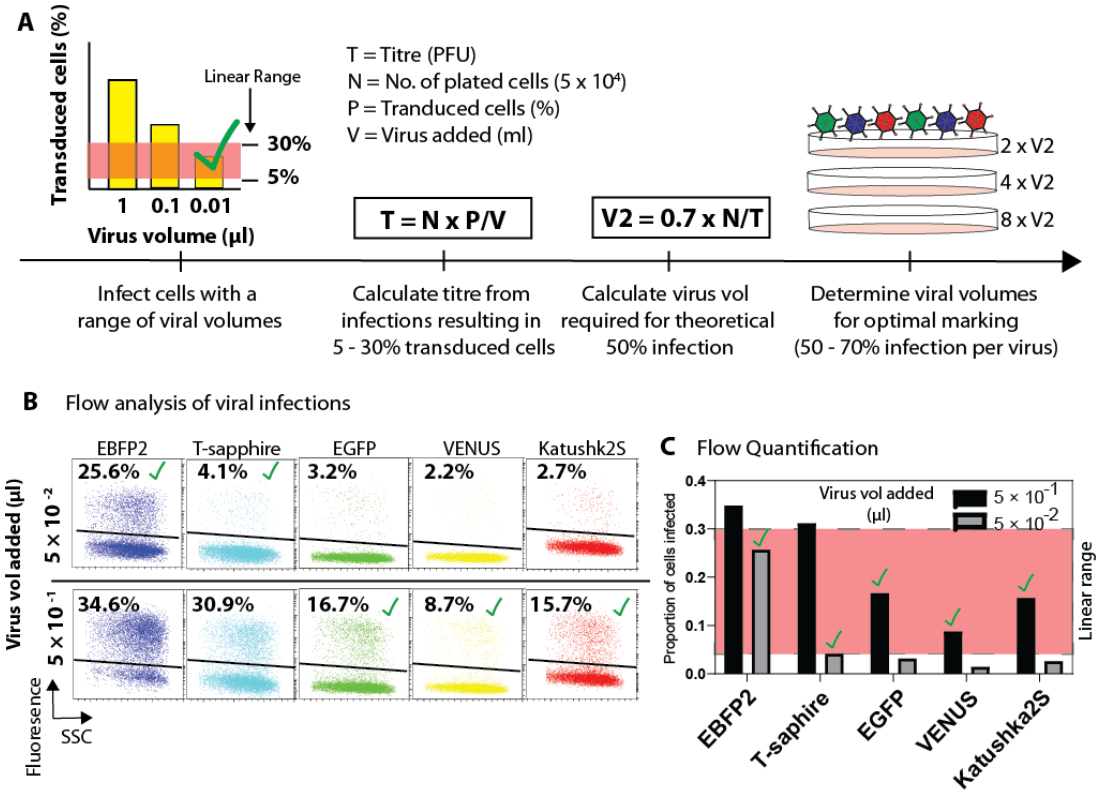


Figure 4.5: Xenotransplanted labelled U87 cells form dense tumour tissue comprising multiple labels and many unlabelled cells. (A) Tumour tissue (animal ID:G61.2) forming in the meninges containing a variety of labels (red, green blue and purple) as well as large regions containing unlabelled cells. (A') region of densely packed labelled cells with cells of three label combinations present (blue, green and purple). (A'') Separate region showing coloured cells less densely packed and the presence of many nuclei corresponding to cells without a fluorescent label. (B) Tumour tissue from (C61.4) growing in the meninges and containing cells with a variety of labels; primarily blue (EGFP) with smaller purple (EGFP and Katushka2S) and green (Venus) populations. (B') Labelled cells again forming regions where they are densely packed (B'') and regions where they are more disperse. (C) Population of blue (EGFP) cells forming a migratory structure in another region of tumour shown in (B).

4.2.2 Optimal RGB labelling in HEK293T and G61 primary glioma

We first optimised titred delivery of viral fluorophores in HEK293T cells. This involved exposing HEK293T cells to 10-fold serial dilutions of our virus in isolation as described previously (Weber et al., 2012). For each virus, the dilution that resulted in a transduction efficiency of 5 – 30% was used to calculate the viral titre in plaque forming unites (PFU) (Figure 4.6 A). This represents the linear range of viral infection where higher values wil include cells with two-plus transgene integrations and result in an under estimation of virus concentration (Fehse et al., 2004, Weber et al., 2012) Raw flow cytometry data and quantification of two 10-fold dilution transduction outcomes can be seen for five vectors in Figure 4.6B and C. Using calculated titres, the volume of each virus preparation required for a 50% infection was determined, this was termed the V2 volume. Optimal RGB labelling will produce a cell population carrying a near even distribution of single, double and triple transduction outcomes which is achieved at a transduction level of 50 – 70% per delivered viral fluorophore. To achieve this, multiple co-transductions were performed where either 2x, 4x or 8x the calculated V2 volume of each virus was delivered (n = 1 per V2 value). Summary of viral titre and virus delivery calculations for optimal RGB marking can be found in Figure 4.6D. Quantification of fluorophores in these three labelled populations are shown in n parallel we set out to achieve similar labelling results in a primary human cell line which are often less amenable to viral infections. . Separation of all 8 different label populations; single, double and triple transduction (2x, 4x and 8x) outcomes are shown in flow cytometry dot-plots (n parallel we set out to achieve similar labelling results in a primary human cell line which are often less amenable to viral infections. A, B and C). Imaging corresponding to the flow cytometry results are shown in Figure 4.7A',B' and C'. n parallel we set out to achieve similar labelling results in a primary human cell line which are often less amenable to viral infections. D – F show quantification of transduction outcomes in our three trialled infections. Across all three infections we achieve relatively equal levels of fluorophore expression for the three co-transduced vectors (n parallel we set out to achieve similar labelling results in a primary human cell line which are often less amenable to viral infections. D). As expected, the proportion of infected cells went up when the



D Results of titre and optimal labelling calculations

Fluorophore	Cells seeded	Viral Vol (ml)	Proportion infected	Viral Titre (PFU)	V2 (µl)	Infection A (2 x V2)(ul)	Infection B (4 x V2)(ul)	Infection C (8 x V2)(ul)
EBFP	5×10^4	5×10^{-5}	0.256	2.57×10^8	0.14	0.28	0.56	1.12
T-sapphire	5×10^4	5×10^{-5}	0.041	4.14×10^7	0.85	0.85	1.7	3.4
EGFP	5×10^4	5×10^{-4}	0.167	1.67×10^7	2.09	2.09	4.18	8.36
Venus	5×10^4	5×10^{-4}	0.087	8.78×10^6	3.98	Not calculated		
Katushka2S	5×10^4	5×10^{-4}	0.157	1.57×10^7	0.233			

Figure 4.6: RGB marking with titrated virus preparations improves label diversity in HEK293T. (A) Experimental schematic for determining virus titres; transductions with 10-fold serially diluted virus preparations looking for infection with 5 – 30% transduction efficiency (n = 1 infection per dilution). Formulas used for calculating titre in PFU from serial dilution transductions and formula for calculating appropriate virus volumes for optimal labelling. Formulas provide a theoretical result so transductions with 1x, 2x and 4x the calculated V2 value are performed to determine optimal volumes. (B) Example flow cytometry plots from 5 x 10^{-3} and 5 x 10^{-2} transductions for EBFP2, T-sapphire, EGFP, Venus and Katushka2S. Negative cells cluster at bottom of plot with positive cells at the upper end of plots. (C) Quantification of data shown in B, with red region indicating the linear range transduction efficiencies required for accurate titre calculations. (D) Table summarising titre calculations for the 5 tested lentivirus preparations, with final three columns showing viral volumes delivered to cells for optimal marking.

volume of delivered virus was increased. Interestingly, the proportion of cells labelled by each of the three fluorophores was consistent across all three separate infections despite the volumes being changed suggesting virus could be accurately

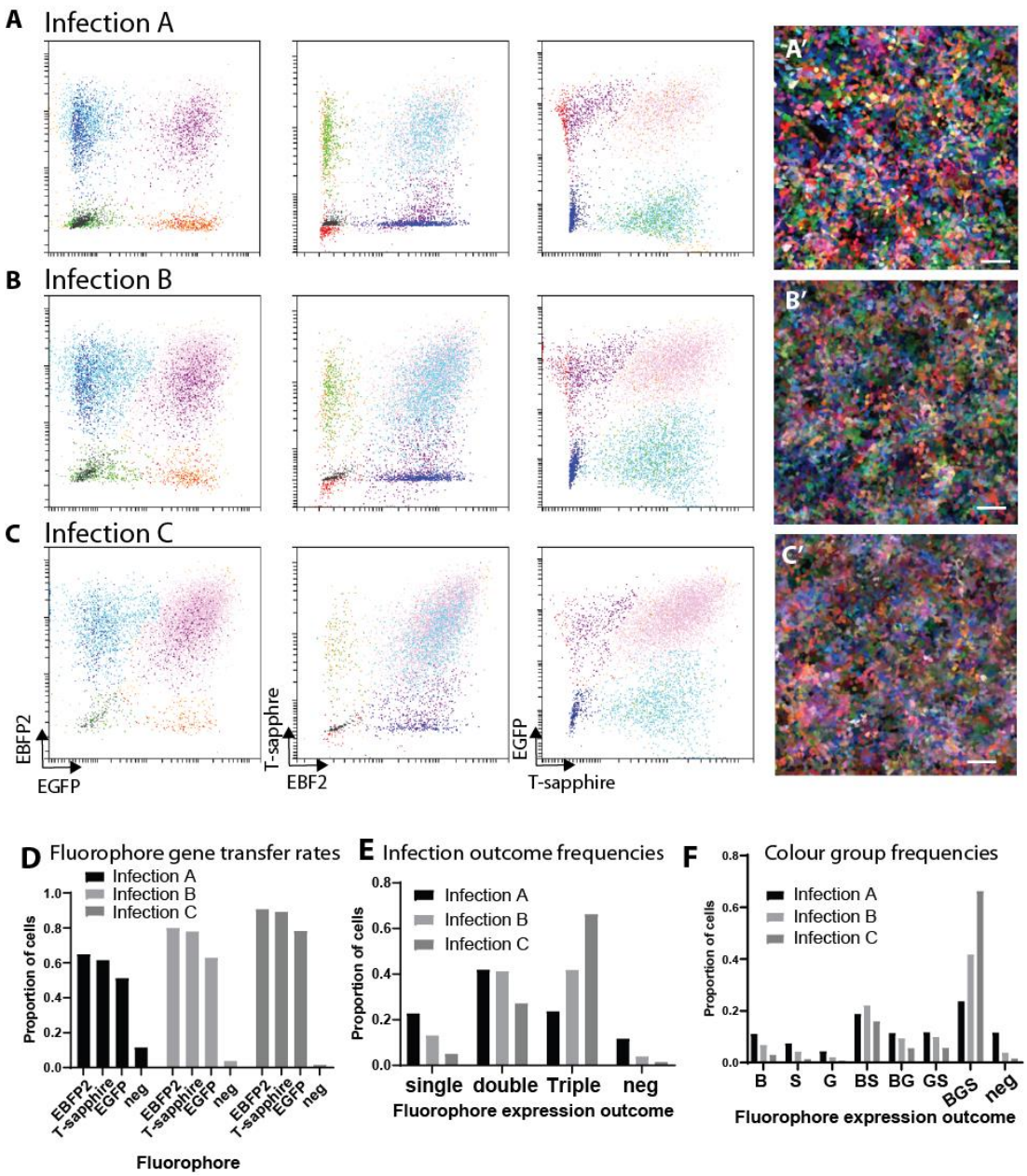


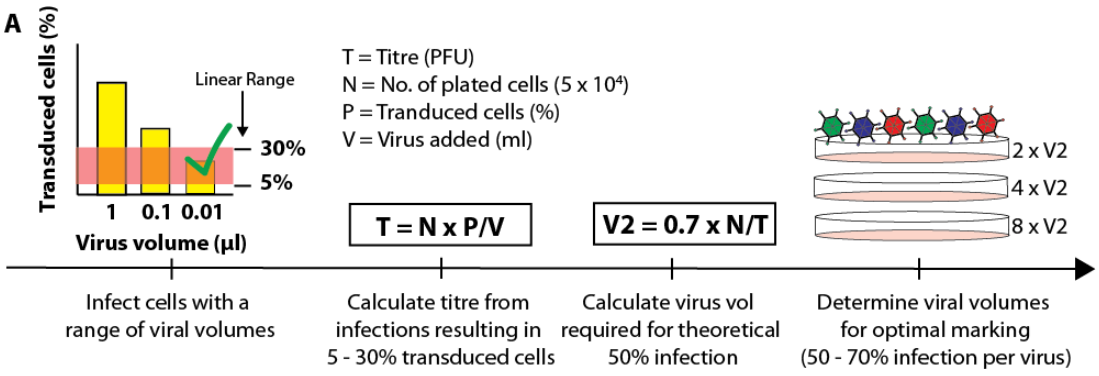
Figure 4.7: Flow cytometry permits accurate quantification of HEK293T RGB marking transduction dynamics. (A – C) Flow cytometry dot plots showing separation of all 7 label outcomes from RGB marking with EBFP2, T-sapphire and EGFP. Blue (EBFP), green (T-sapphire), Red (EGFP), Purple (EBFP2 + EGFP) Cyan (EBFP2 + T-sapphire), Yellow/Orange (EGFP + T-sapphire) and Pink (EBFP2, T-sapphire and Venus)(n = 1 recording per infection**)**. (A' – C') Corresponding imaging of cell populations displayed in A – C flow plots. (D) Gene transfer rates for the three individual fluorophores across the three different co-infections (scale bars = 100 µm). Overall infection level increase with relative proportions of each fluorophore in the labelled populations remaining largely consistent. (E) The proportion of cell carrying either single, double, triple or no label across the three separate co-infections. (F) Full break down of all 7 colour groups and the unlabelled (neg) cellular proportions across the three separate co-infections. Data shown in D – E were derived from a single flow recording of >30,000 cells for infections A, B & C.

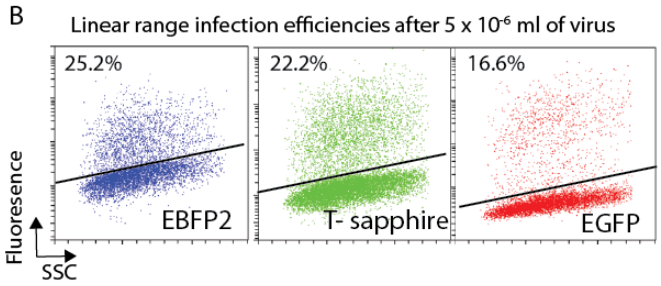
pipetted at different concentrations while keeping the relative proportions even as calculated by the titre data. As delivered viral volume was increased, we found unlabelled, single and double labelled cells decreased in the labelled population while triple labelled cells increased. Complete breakdown of the labelled cell populations across the three separate infections is shown in n parallel we set out to achieve similar labelling results in a primary human cell line which are often less amenable to viral infections.

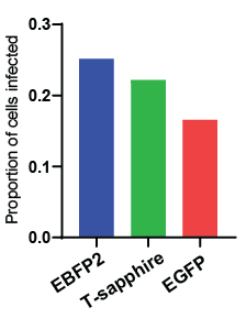
In parallel we set out to achieve similar labelling results in a primary human cell line which are often less amenable to viral infections. Titres were determined for EBFP2, T-sapphire and EGFP harbouring LeGO-vectors in primary cell line G61 (n = 1 transduction per viral dilution). Flow cytometry plots and quantifications of G61 titre data, as described for HEK293, is shown in Figure 4.8A – D. Results indicate that G61 cells required 10 times fewer viral particles to achieve comparable levels of fluorescence expression than HEK293T cells (Figure 4.8D).

The dynamics for viral gene delivery were like those in HEK293T (Figure 4.7). Flow cytometry dot-plots used for quantification of label populations (Figure 4.9A, B & C) and imaging (Figure 4.9A', B' & C') were also comparable to results in HEK293. Since new titres were calculated for G61, the ratio of the three fluorophores in each of the cultures had shifted with t-sapphire most prominently expressed throughout the G61 cultures (Figure 4.9D). Again, the proportion of single/double/unlabelled cells went down and the proportion of triple labelled cells went up as the amount of delivered viral particles increased (Infection A – C, Figure 4.9E & F)

In conclusion, LeGO-vectors can effectively label primary human glioma cells with an array of complex labels discernible through imaging and flowcytometry. This model represents a highly controlled system for labelling where the viral delivery regime can be adapted to include more or less fluorophores, in multiple combinations and delivered at different concentrations to control distributions of delivered labels







D Results of titre and optimal labelling calculations

Fluorophore	Cells seeded	Viral Vol (ml)	Proportion infected	Viral Titre (PFU)	V2 (µl)	Infection A (2 x V2)	Infection B (4 x V2)	Infection C (8 x V2)
EBFP	5×10^4	5×10^{-6}	0.252	2.52×10^9	0.014	**0.028**	**0.056**	**0.112**
T-sapphire	5×10^4	5×10^{-6}	0.22	2.22×10^9	0.016	**0.032**	**0.064**	**0.128**
EGFP	5×10^4	5×10^{-6}	0.166	1.67×10^9	0.021	**0.042**	**0.084**	**0.168**

Figure 4.8: RGB marking in primary line G61 produces comparable labelling results to HEK293T. (A) Schematic of RGB marking experimental pipeline. (B) Flow cytometry plots of viral infection (1 x 10^{-6} ml) producing gene transfer rates within the linear range for EBFP2, T-sapphire and EGFP (n = 1 transduction per viral dilution). (C) Quantification of the flow plots shown in B. (D) Summary of data values required for calculating titre and virus volumes required for optimal RGB marking.

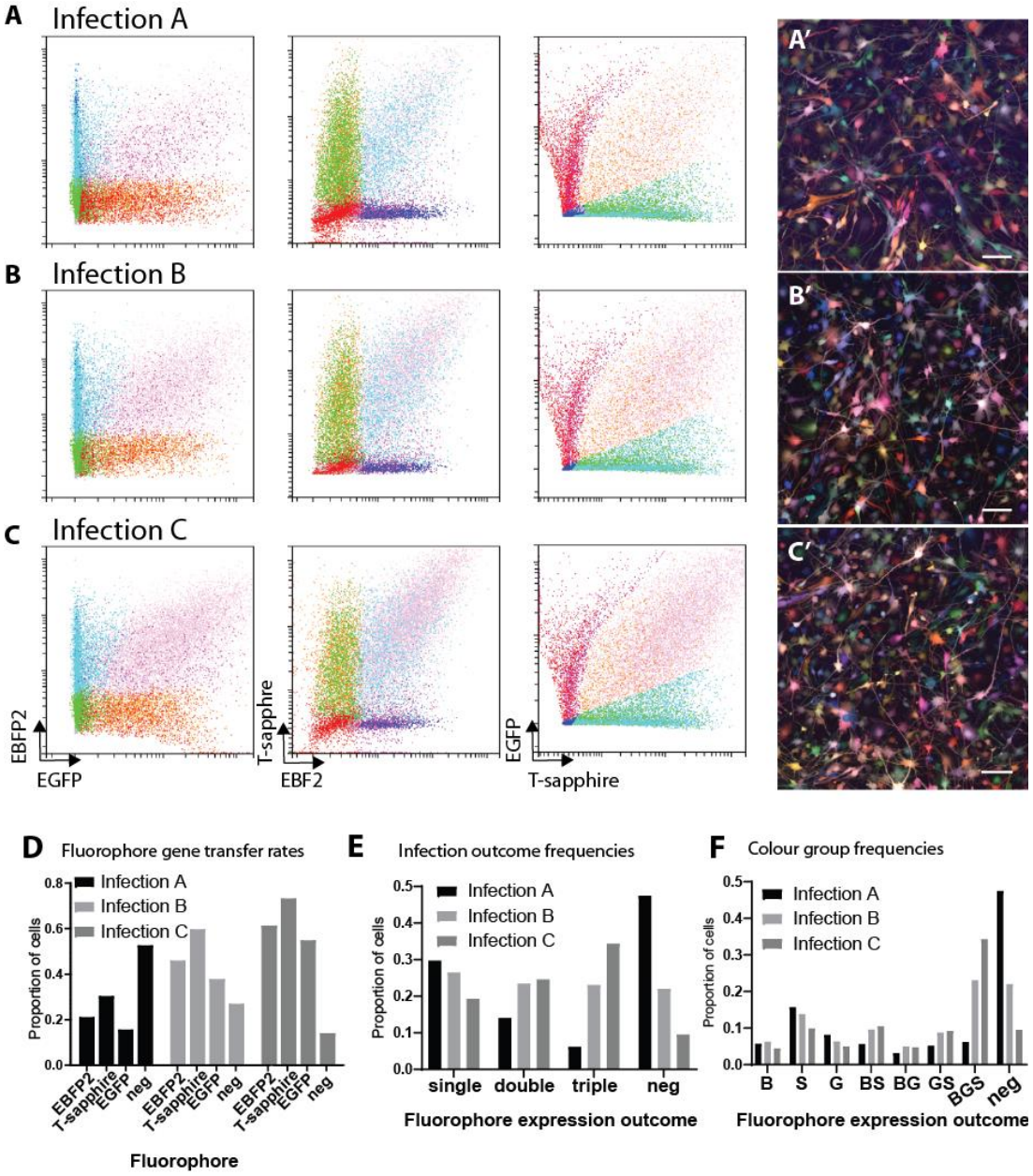


Figure 4.9: Quantification of G61 RGB marking transductions through flow cytometry. (A – C) Flow cytometry dot plots showing separation of all 7 label outcomes from RGB marking with EBFP2, T-sapphire and EGFP. Blue (EBFP), green (T-sapphire), Red (EGFP), Purple (EBFP2 + EGFP) Cyan (EBFP2 + T-sapphire), Yellow/Orange (EGFP + T-sapphire) and Pink (EBFP2, T-sapphire and Venus). (A' – C') Corresponding imaging of cell populations displayed in A – C flow plots. (D) Gene transfer rates for the three individual fluorophores across the three different co-infections (n = 1 technical repeat per infection) (scale bars = 100 µm). Overall infection level increase with relative proportions of each fluorophore in the labelled populations remaining largely consistent. (E) The proportion of cell carrying either single, double, triple or no label across the three separate co-infections. (F) Full break down of all 7 colour groups and the unlabelled (neg) cellular proportions across the three separate co-infections. Data shown in D – E were derived from a single flow recording of >30,000 cells for infections A, B & C.

4.2.3 Serial passaging and flow analysis of RGB marked G61 suggests clonal expansion

The labelling system was developed to trace multiple clones within a mixed population of glioma CSCs. Previous studies using RGB marking had demonstrated that after protracted period of culture (>30 Passages) label complexity in RGB marked cell populations reduces, suggesting outgrowth of some clones and involution of other clones occurred (Brenière-Letuffe et al., 2018). Interestingly, populations labelled with two fluorophores appear as well demarcated streaks (or clusters) on flow cytometry dot plots (Mohme et al., 2017). A population of labelled cells comprised of many streaks would be an excellent starting point for precise tracking of clonal cell populations rather than mixtures of clones with the same label.

To assess the ability of our RGB marked G61 cells to form clonal populations manifesting as streaks on flow cytometry plots, we serially passaged labelled cells. Cells were cultured in 12-well plates and diluted 1:10 at each passage in an attempt to accelerate the emergence of clonal populations (n = 1 passaged culture per infection). Figure 4.10 shows the isolated double positive groups (BG, BS, and GS) for G61 RGB infection B and C (Figure 4.8F & G). Interestingly, already at 4 passages clonal populations manifested as emerging streaks in the BS dual labelled populations for both infection B and C (Figure 4.10). Furthermore, by passage 8, clusters had begun to form in the other two dual-labelled populations for infections B and C. The prevalence of all colour groups was monitored for both infections and is shown in Figure 4.10 B & C. In both infections we saw an overall reduction in the proportion of triple labelled cells and an increase in the number of unlabelled cells. However, there was no particular trend in how the single or dual labelled populations, with some increasing and others decreasing in prevalence.

In conclusion, RGB marked cells labelled with two fluorophores can effectively form streaks on flow cytometry plots over serial passages. However, we also found that triple labelled cells may have a growth disadvantage compared to single, double and unlabelled cells from within the same culture. Performing multiple dual-

transductions and FACS removal of singly labelled cells has been adopted for future labelling assays to maximise the number of dual-labelled cells produced. Dual-labelled cells show sufficient clonal growth and can be accurately detected as clones through flow cytometry.

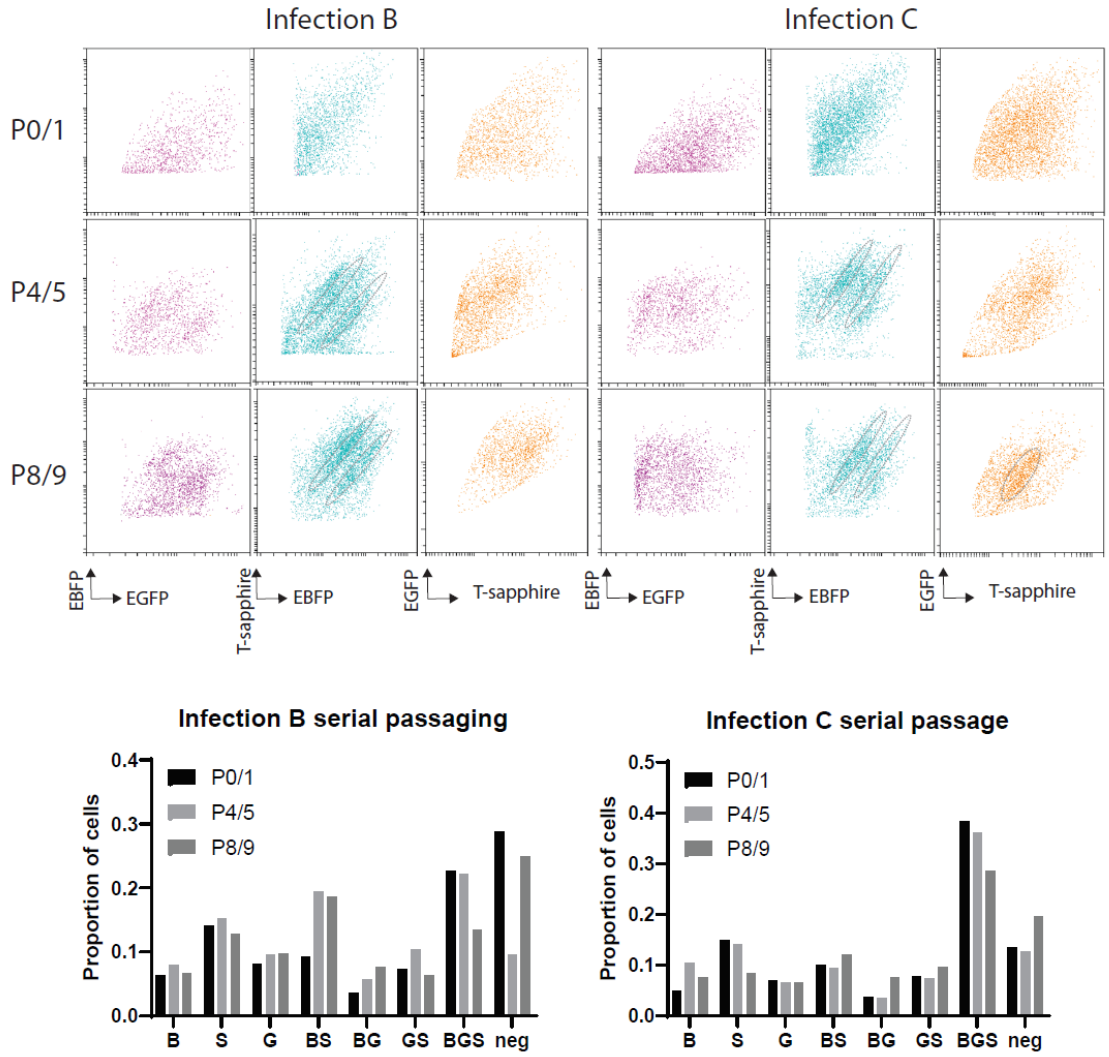


Figure 4.10: Serial passaging of optimally RGB marked G61 cells shows emergence of clonal streaks in dual-labelled cell populations and a steady decline in triple labelled cells. (A) flow cytometry dot plots of G61 cells from RGB marking infections B and C across 3 different passaging timepoints (n = 1 per infection). Ovular dotted boxes on EBFP2 + T-sapphire plots show the emergence of streaks in culture after 4 passages. By P8 clusters are observed in the other dual labelled populations but do not form the pronounced streaks observed in EBFP2 and T-sapphire labelled cells. (B) Quantification of the 7 colour groups and negative cells for Infection B across the three time points. (C) Quantification of the 7 colour groups and negative cells for Infection C across the three time points.

4.2.4 Organoid culture of RGB marked G61 reveals clones with different proliferation rates and regional localisation

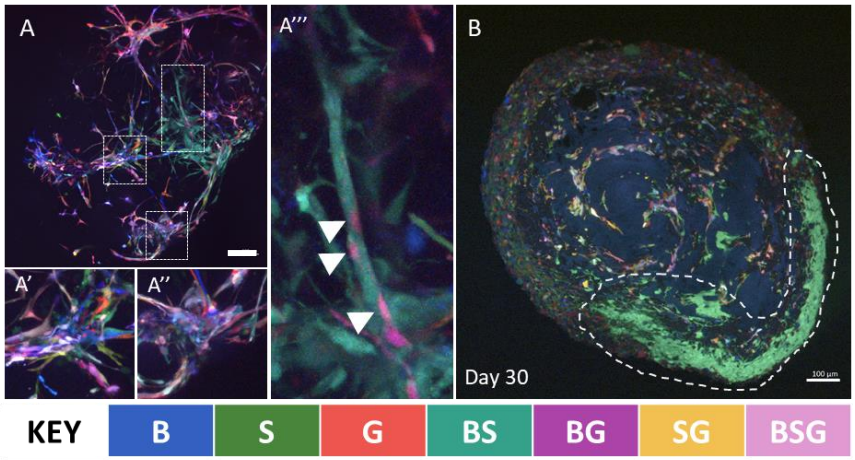
To further explore the effectiveness of RGB marking for investigating clonal dynamics in glioma CSCs we seeded passage 4 RGB marked G61 cells into three-dimensional organoid cultures. Multi-colour confocal imaging of a living whole RGB marked G61 organoid at day 14 (n = 2 imaged, 1 representative organoids displayed) of culture shows cells forming both clonal and mixed clusters that become interconnected through multi-cell migratory tracts (Figure 4.11 A). In this organoid we identify a cyan EBFP2^{+}/T-sapphire^{+} cluster of cells projecting multi-cellular processes away from its central cluster (Figure 4.11A'''). We can observe a further population of purple EBFP2^{+}/EGFP^{+} cells migrating along this tract toward other regions of the organoid. After 30 days of culture organoids develop a surface densely populated with cells resulting in difficulties with imaging deeper regions. Figure 4.11B shows the same organoid displayed in Figure 4.11A after 30 days of culture, freezing and cryo-sectioning. There was a large expansion of cyan EBFP2^{+}/T-sapphire^{+} cells that are likely to have been derived from the similarly coloured cells imaged at day 14. However, we found that while fixation and cryo-sectioning did preserve fluorescence sufficiently well to detect multiple distinctly coloured populations, fluorophore brightness and colour distinctiveness was partially lost through this process with higher laser powers and detector gains required. Furthermore, it was not possible to achieve great depth of imaging and as such structural detail was also limited after cryo-sectioning.

To address this issue, we tested an agarose embedding approach for vibratome sectioning to permit serial imaging of living organoids, at greater z-depth and with better fluorophore detection. Figure 4.11C – E displays imaging of (n = 1) an organoid vibratome sectioned in this fashion at day 30. This technique revealed fine details of individual tumour cells as well as macroscopic detail of clonal organisation internally and externally. Interestingly, at time of sectioning (Figure 4.11C) cells in internal regions of the organoid appear to possess only few long processes. However, by day

32 and beyond (Figure 4.11D– F) many more cells can be seen projecting long processes internally as well as small process emanating internally from cells in the dense surface. We hypothesise that this may be due to the previous hypoxic conditions internally that have been disrupted by cutting and exposing internally regions to greater nutrients and oxygen available in the media. Cells observed at the surface are largely dominated by red $EGFP^{+}$ cells spread extensively around the organoid surface. This contrasts with internal regions where a large variety of colour labels are still present and in fact, very few red cells are observed.

In conclusion, agarose embedding and live imaging of RGB marked G61 organoids produces samples with greater potential for meaningful structural and quantitative inference through multicolour confocal imaging. Here, simple observations reveal differing clonal dynamics at the organoid surface and internal organoid regions. While also highlighting a change in cellular structure in response to vibratome sectioning potential due to a changing microenvironment.

Confocal imaging of the same organoid during growth (A) and after cryo-sectioning (B)



Serial confocal imaging of living RGB marked G61 organoid after vibratome sectioning

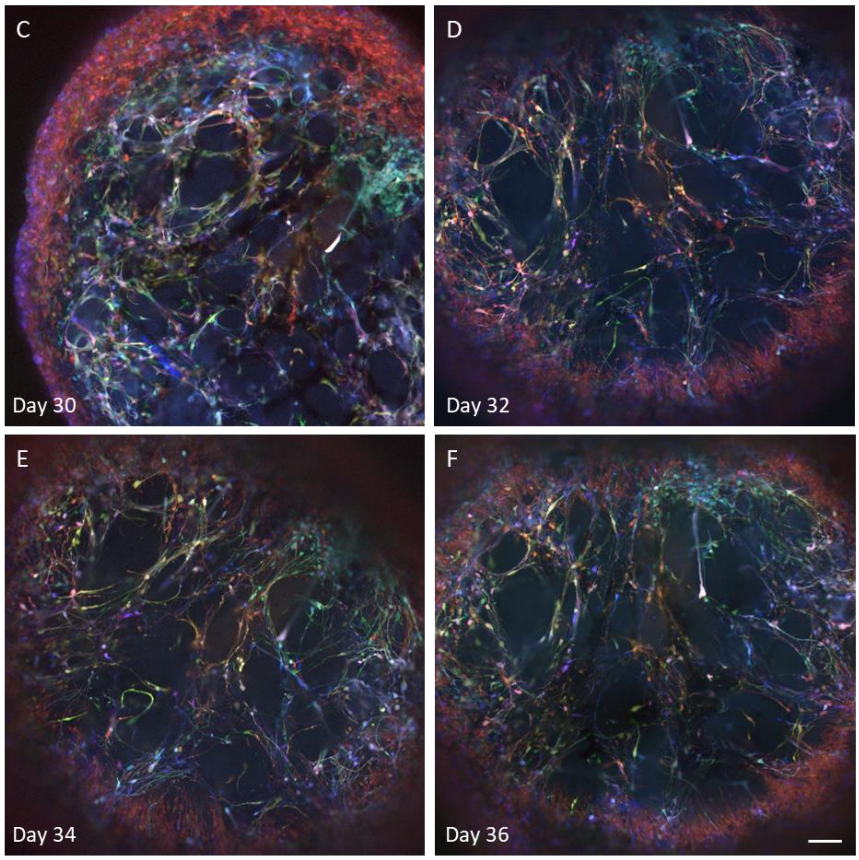


Figure 4.11: Vibratome sectioning and live imaging reveals finer details of cellular structures than fixed frozen tissue. (A) Confocal imaging of three-dimensional RGB marked G61 culture 7 days after seeding cells (scale bar = 150 µm). (A' & A'') Mixtures of variably labelled cells have come together and are forming migratory paths emanating away from the cell cluster. (A''') region showing expansion of cells harbouring the same label, potentially a clonal expansion. Arrows indicate detail of differentially labelled cells migrating along a process emanating from the cluster of cyan cells. (B) Cryo-sectioned image of the organoid shown in A after 30 days of culture. Dotted line highlights large region with cells carrying the same label possibly descendent from the cell cluster shown A''' (scale bar = 100 µm). (KEY) Fluorophore combinations; EBFP2 (B), T-sapphire (S), and EGFP (G) and approximate cell colours. C – E show serial images of a RGB marked G61 organoid after live vibratome sectioning and continued culture. Organoid periphery dominated by red (EGFP) cells with greater label variability in core regions. Cellular structure changes from Day 30 (day of sectioning) to day 36 with cells producing a greater number of processes (scale bar = 100 µm).

4.2.5 Dissociation and flow cytometry reveals large shifts in colour groups after organoid culture which are corroborated through imaging

A further portion (n = 12) of G61 organoids (seeded with cells 4 passages after RGB marking) were dissociated using Accumax® and subject to flow interrogation of tracing labels (EBFP2, t-sapphire and Venus). The population seeded into the organoids was dominated by cells carrying all three fluorophores (BSG, pink), which represented 36% of all cells (Figure 4.12A & B). EBFP2 (B, blue), T-sapphire (S, Green) and unlabelled (neg, grey) cells each represented roughly 10% of the seeding population. With EGFP (G, red), EBFP2/T-sapphire (BS, cyan), EBFP2/EGFP (BG, purple) and EGFP/T-sapphire (GS, orange) positive populations each representing under 10% of the seeding population. After 30 days of organoid culture these lentiviral labels were re-assessed revealing changes in label distributions (Figure 4.12C). Strikingly, BSG which was the majority population at time of seeding were found to be largely absent after organoid culture (Figure 4.12C). Quantification of organoid label distributions (Figure 4.12D) showed EGFP (G) expressing cells and unlabelled cells (neg) showed the highest representation across all organoids (n = 12). One-way ANOVA comparing label group mean prevalences reported a p value of less than 0.0001. Finally, quantification the change in label distributions after organoid culture frther support the fact that EGFP (G) and unlabelled (Neg) cells showed consistent increases across assessed organoids (n = 12). One-way ANOVA comparing label group mean prevalence's reported a p value of less than 0.0001.

In conclusion, these results suggest there is considerable shift in clonal composition after seeding into organoids with some consistent patterns between organoid (prevalence of EGFP) as well as a lot of variation in what cellular colours persist after organoid growth. The loss of all triple positive cells and general persistence of the neg cell fraction suggests that expressing high levels of fluorophore may exert a burdened to cell growth. Furthermore, the density of cell clusters expressing a single fluorophore makes it difficult to assign clonality to a detected group of cells. While in dual positive quadrants of flow plots well demarcated clusters/streaks appear

indicative of clonal expansions. These observations will inform an improved labelling approach for clonal detection.

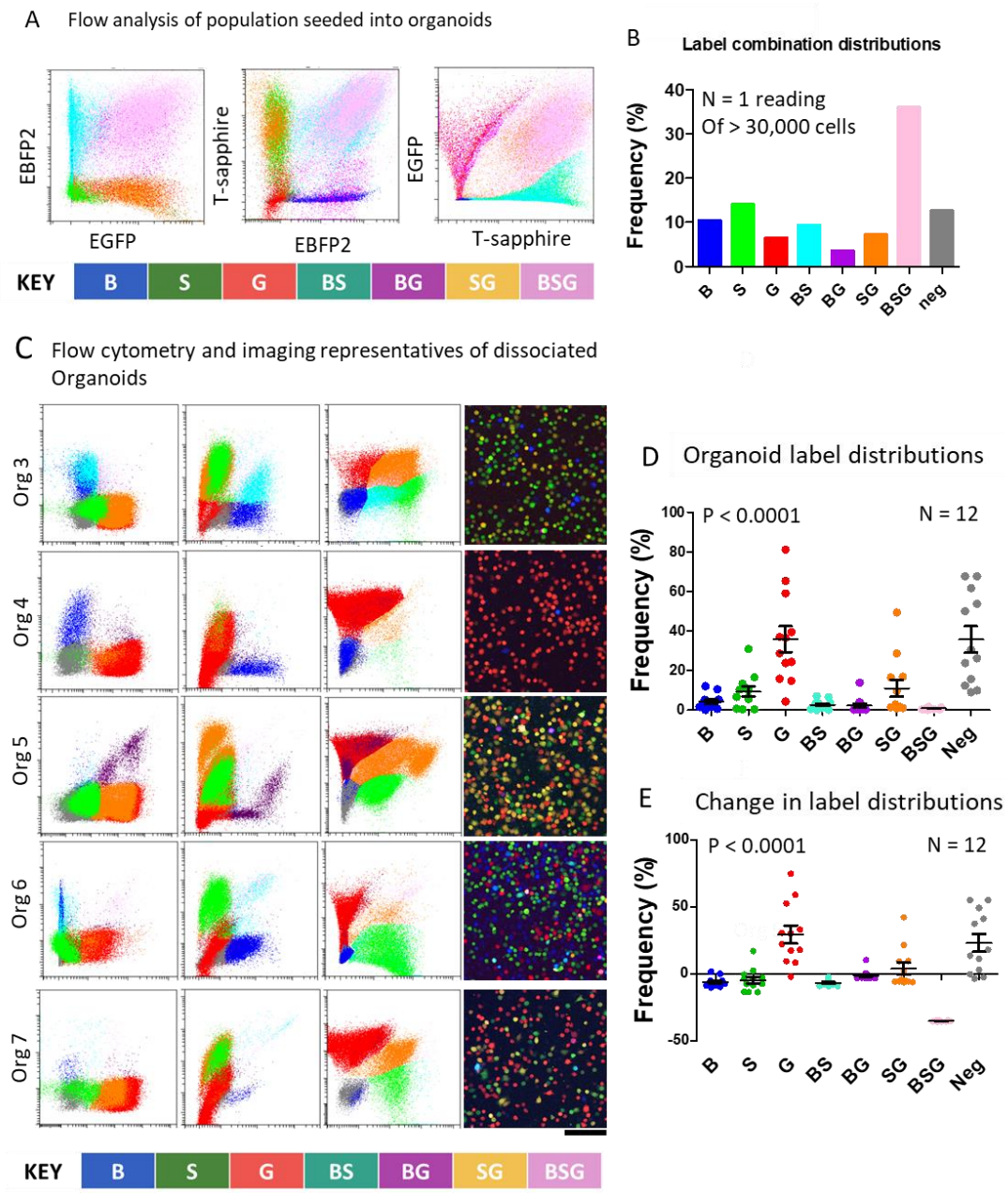


Figure 4.12: Large shifts in colour label composition observed after organoid culture of RGB marked G61 cells. (A) Flow cytometry plots and (B) quantification of RGB label proportions in G61 population seeded into organoids. (C) Representative flow cytometry dot-plots (axis labels same as shown in panel A) and imaging of dissociated organoids after 30 days of culture (n = 5, scale bar = 100 µm). (D) Quantification of label distributions after organoids culture (n = 12) showing EGFP (G) and unlabelled (neg) consistently highly prevalent, one-way ANOVA (CI = 95%), $p < 0.0001$. (E) Quantification of change in label frequency after organoid culture compared to seeding population. EGFP (G), T-sapphire/EGFP (SG) and unlabelled (neg) showed mean increases and all other groups showed a mean reduction in prevalence across n = 12 organoids. One-way ANOVA (CI = 95%), $p < 0.0001$.

4.2.6 Dual-barcoding of U87 cells produces 15 colour groups identifiable through flow cytometry and Imaging.

To overcome issues described in section 4.2.5:

1. Complete loss of triple labelled cells after organoid culture.
2. Prevalence of unlabelled cells that cannot be traced
3. Trend of single label populations outgrowing dual labelled populations

We devised a new labelling regime that would not produce any triple positive cells while maximising the production of dual-labelled cells after transduction. We reasoned that maximising the number of dual labelled cells would increase the ability to identify clonal dynamics (i.e. expansion or attrition) when cells are co-cultured, as these clonal expansions from dual label cells can be identified as streaks on flow cytometry plots following sufficient cellular expansion. To achieve this, two further tracing labels (mOrange2 and Ktushka2S) were added to the panel of labels introduced to cells. In addition, EGFP was replaced by Venus owing to the near identical emission spectra of T-sapphire and EGFP. Instead of applying all virus at once, 10 separate dual transductions of all the possible virus combinations were performed (Figure 4.13A), where - informed by RGB marking results – titred virus particles were applied to achieve approximately 75% infection for each virus. 7 days after virus exposure, cells from the different reactions were pooled together and underlying label populations were assessed using flow cytometry. Figure 4.13B shows detection and separation of all 15 possible fluorophore combinations in this protocol. Barcodes to the right-hand side represent the fluorophore combinations displayed in each set of aligned flow panels (Figure 4.13B). Titre calculations were likely underestimated for mOrange2 (O) and Katushka2S (K) resulting in addition of more virus particles and an overrepresentation of these as a single label and also the dual label group expressing both Katushka2S and mOrange2 (Figure 4.13C). Nonetheless, relative titres proved sufficient for achieving near even proportions of each produced colour group

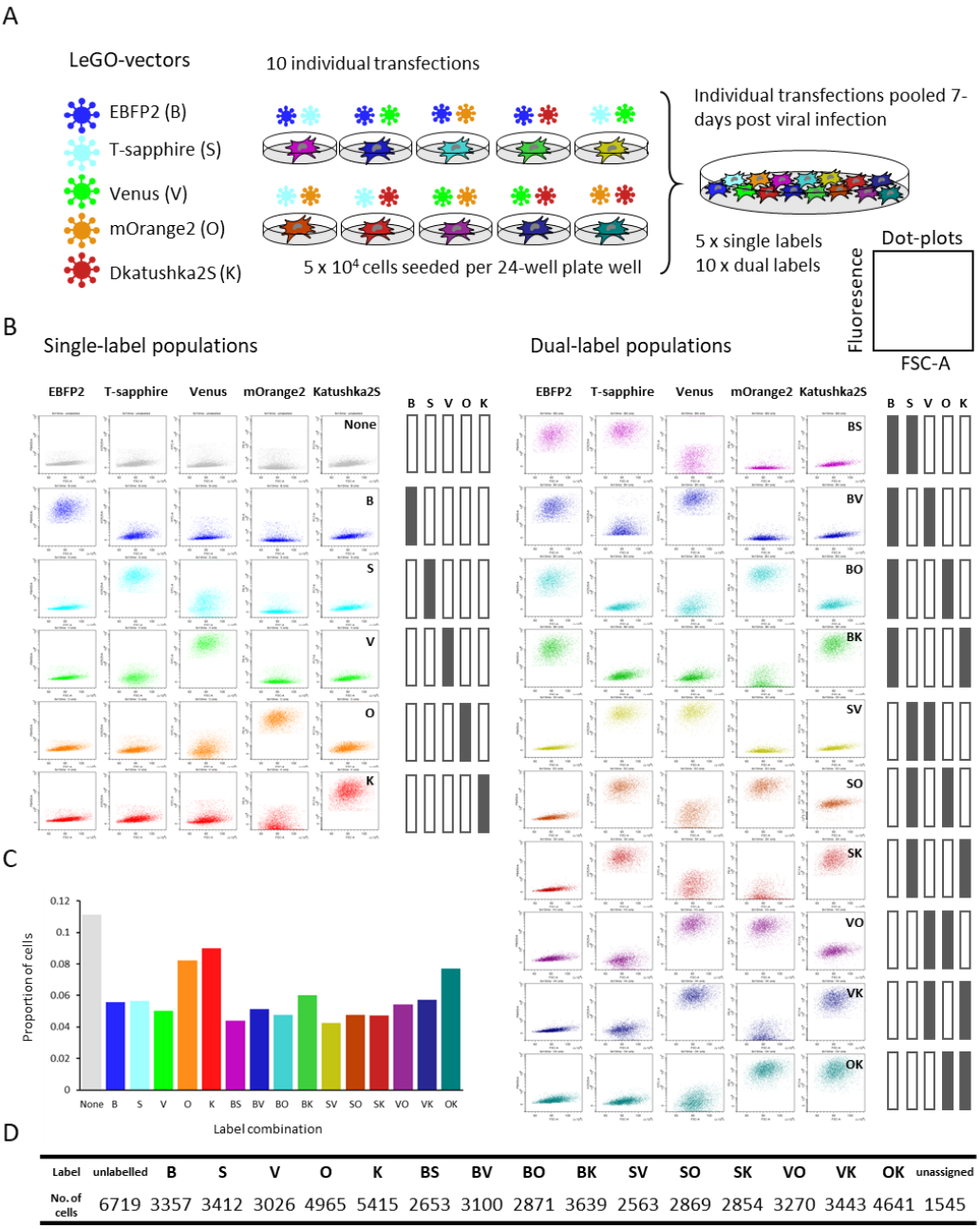


Figure 4.13: Optical barcoding with dual infections to produce 15-distinclty labelled traceable colour groups in U87. (A) Schematic of lentiviral labelling regime for barcoding of cells with 5 different titred fluorophores (EBFP2, T-sapphire, Venus, mOrange2 and Katushka2S. 10 separate dual transductions with each combination of the 5 fluorophores was performed and the 10 populations were pooled 7 days after initial viral infections. (B) Detection and of all 15 colour groups and unlabelled cells in a mixed population by flow cytometry, single label populations on the left and dual-label populations on the right. Each row represent a differentially labelled population. Y-axis shows fluorescence intensity with specific fluorophore indicated at top of panel, x-axis on all plots is side scatter area (SSC-A). (C) Quantification of all 15 colour groups and negative cells in the mixed population. (D) Counting of cells detected in each of the different label groups (n = 1 reading of >60,000 cells. Unassigned group represents excluded cells that analysis rendered positive for three fluorophores.

when the 10-dual transductions were mixed together. Absolute counts for cells assigned to each colour group are shown in Figure 4.13D. A small fraction of cells (2.5%) were unassigned to any group as these appeared in multiple double positive quadrants appearing positive for 3 or more fluorophores (Figure 4.13D).

Flow applications of multi-colour fluorescent lineage tracing have unique potential compared to other lineage-tracing techniques. For example, the ability to separate living cells further downstream culture experiments and assessment of marker expression in different tumour cell populations. However, the most unique aspect of fluorescence based clonal tracking is the ability to investigate cells *in vivo* and look at the histological organisation of clones. To this end, we aimed to develop an image analysis approach for our barcoded cells that automatically identify and classify cells based on their marker expression. Our trace label panel permitted addition of the far-red nuclear dye DRAQ5 that is excited by the red wavelength (633 nm) with an emission spectrum at higher wavelengths than katushka2S. Nonetheless, considerable overlap in excitation and emission spectra of mOrange2, Katushka2S and DRAQ5 warranted the need for unmixing of fluorescent signals to accurately detect overlapping expression of different fluorophores. Emission spectra were acquired by identifying cells expressing a single fluorophore and unlabelled cell labelled with DRAQ5 (Methods). With these spectra it was possible to unmix signals in a population carrying all 6 fluorescent signals including identification of the 10 different dual-labelled cell populations (Figure 4.14).

To test the effectiveness of linear unmixing in identifying and quantifying different label combinations, we developed a pipeline for automatic nuclear segmentation using StarDist script for quPath (Bankhead et al., 2017, Schmidt et al., 2018) and subsequent extraction of pixel intensities for the 6 different unmixed signals (EBFP2, T-sapphire, Venus, mOrange2m Katushka2S and DRAQ5). During the cell passage after flow quantification, 5 large spatial regions were imaged (Figure 4.15A), pixel intensities extracted and the distributions of these intensities (arbitrary units) across the segmented cells were plotted on histograms. Using these distributions, a pixel intensity threshold was determined which served as a cut-off for designating cells as label positive or negative for a given fluorophore. These cut offs were specific for

each fluorophore. Applying thresholds could automatically assign cells to one of the 15 possible label groups or as negative (total cells counted = 8,030) and a comparison of this image quantification with flow cytometry results is shown in figure 4.15B.

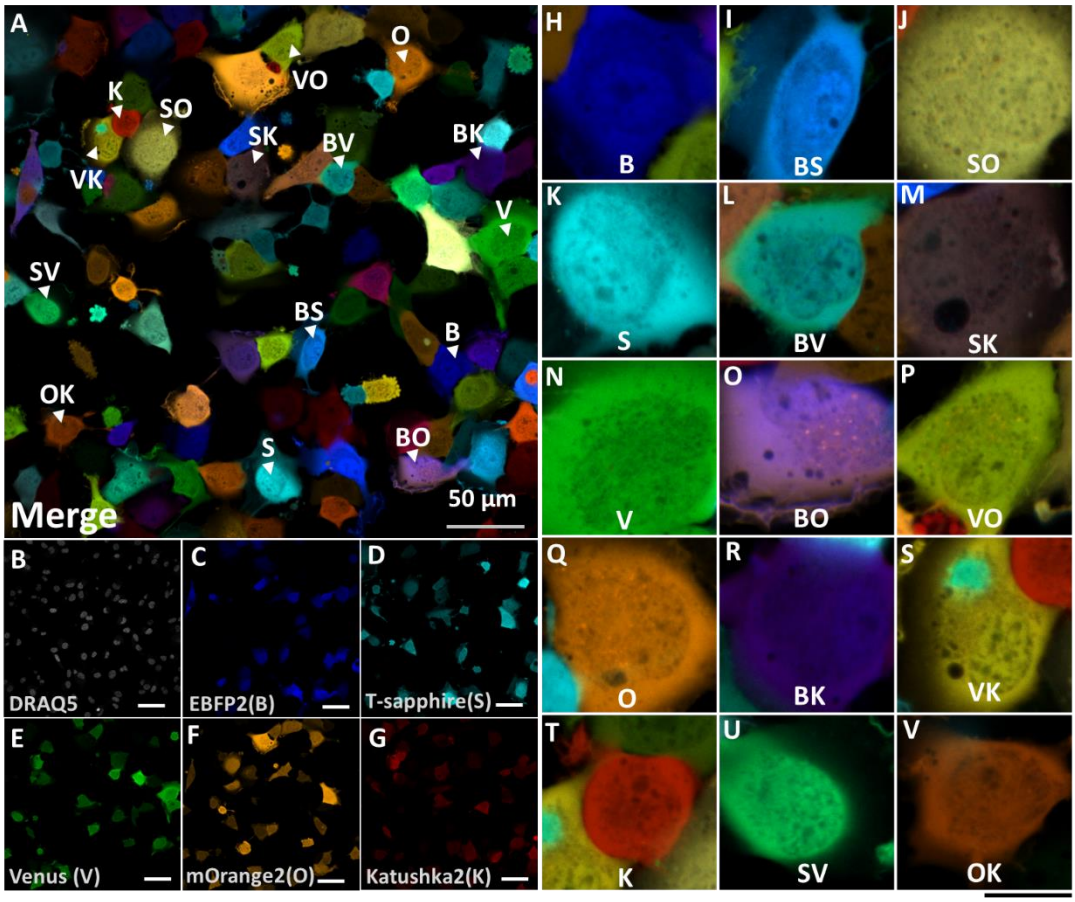


Figure 4.14 Spectral unmixing procedure for accurate detection of barcoded U87 cell populations. (A) Merged channel image of U87 cells in adherent culture labelled with 5 fluorophores, labels and arrows indicate cells of different label combinations shown on right handsize. (EBFP2, T-sapphire, Venus, mOrange2 and Dkatushka2S) (Scale bar = 50 µm). (B – G) Single channel images of merged image shown in panel A (Scale bar = 50 µm). (H – V) expanded view of indicated cells in panel A. Expressed labels indicated in code shown in centre of each panel. (scale bar = 10 µm).

Our imaging analysis pipeline was able to identify cells corresponding to all 15 colour combinations and in particular quantification of single label group sizes (B, S, V, O and K) were comparable to those determined through flow. However, most double label groups were determined at roughly half the prevalence found with flow cytometry and the number of cells designated as unlabelled was 15% larger at 25%. This may reflect insufficiencies of imaging in detecting cells with particularly low levels of fluorophore expression and therefore detection intensity.

In conclusion, we have successfully adapted the labelling procedure to maximise the number dual labelled cells produced without the production of any triple labelled cells. With RGB marking a maximum proportion of 0.25 dual labelled cells with three underlying combinations was achieved. With the adapted labelling regime we were able to achieve a proportion of 0.53 spread across 10 dual labelled populations. We reason in growth assays such as organoids the new labelling regime will be superior for precise identification of clones as dual label cell expansions were previously shown to create well demarcated streaks/clusters on flow plots. Furthermore, our new labelling regime is also effective for imaging and automatic identification of different cellular fractions through linear unmixing of fluorescent signals. However, a significant proportion of unlabelled cells persisted through this technique (11%) which previous assays suggest have superior growth potential over labelled cells. The next step was to culture these cells over serial passages and in organoids to assess growth dynamics.

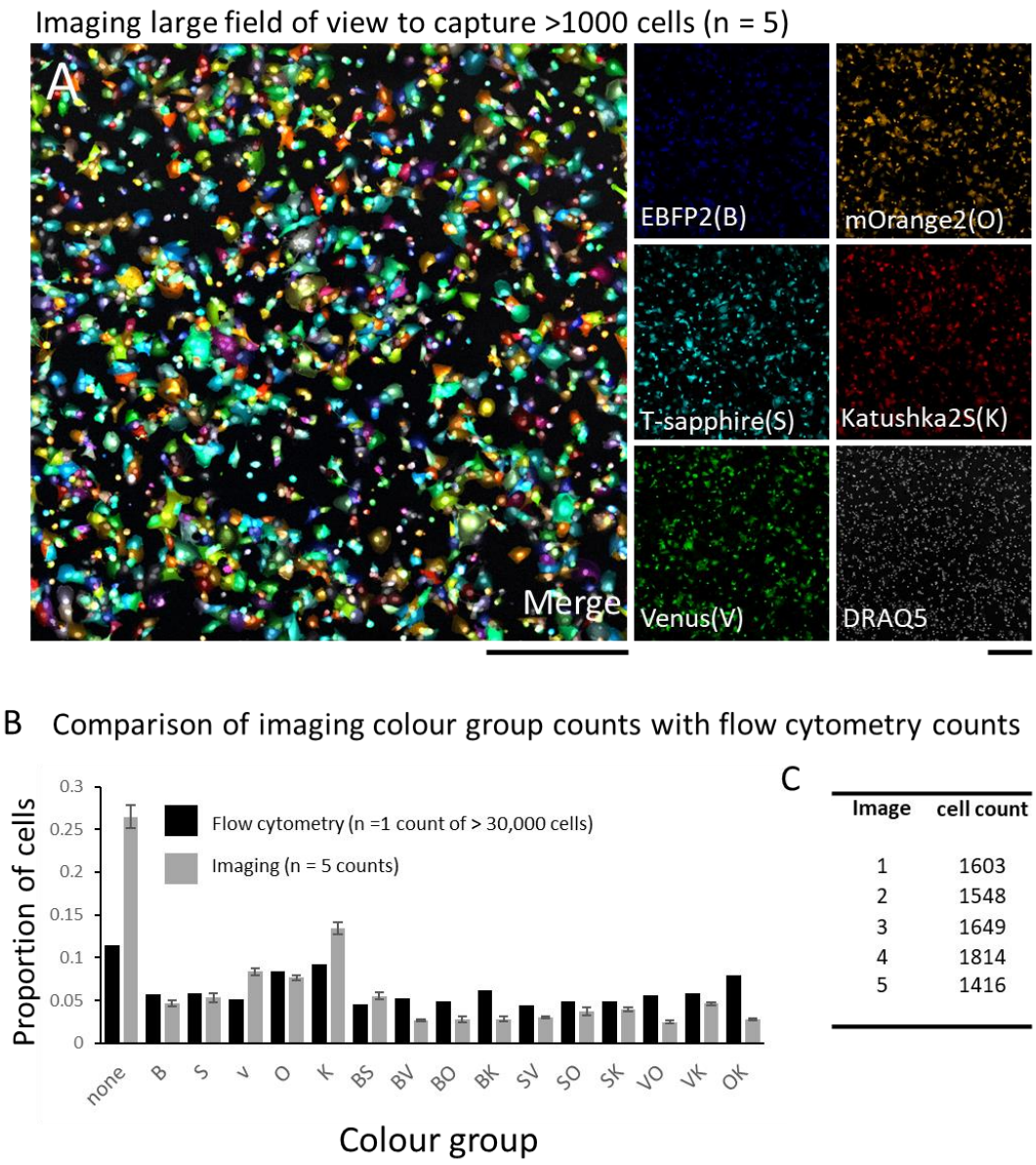


Figure 4.15: Analysis pipeline for effective identification and quantification of colour groups through imaging. (A) Example image (n = 5) of the cell population quantified by flow cytometry in Figure 4.13 (Scale bars = 500 µm). (B) Comparison of colour group quantification by imaging (n = 5 images) compared with flow cytometry (n= 1 reading of > 60,000 cells). Imaging quantification assigned more than twice the number of unlabelled cells and also generally classified a lower number of cells in each dual-label combination. (C) Table displaying the absolute number of single cell detections in each of the images used for quantifications shown in panel B.

4.2.7 Serial passaging and organoid culture of dual-barcoded U87 suggests persistence of more clones than was observed for G61

To test the potential for our dual-barcoding approach in detecting clonal outgrowths as streaks on flow cytometry diagrams, we propagated cells over up to 8 passages in 12-well plate wells (n = 3). Smaller culture vessels were previously showed to enhance the rate that clones grow out *in vitro* (Brenière-Letuffe et al., 2018). We also seeded 5000 cells into organoids (n = 11), to:

1. Dissociate and assess for formation of clonal populations
2. Label with surface marker antibodies to assess marker heterogeneity across label population

Imaging of cryo-sectioned organoids grown from dual-barcoded cells is displayed in figure 4.16. Underlying fluorophores were well preserved and expansions of cells harbouring the same label could be clearly observed throughout the organoids (Figure 4.16A). Unfortunately, attempts to stain dual-barcoded cell nuclei in organoids was not possible as DRAQ5 was found to repeatedly deposit more in the Matrigel® matrix cells were growing in than cellular nuclei. Similar to line G61, by maturity (~30 days culture) cells could be observed in different organisation at the organoid surface and the centre (Figure 4.16A1 & A2). Regional disparities in clonal composition were also observed although no predominant outgrowths covering the whole region of the organoids were observed (Figure 4.16 A3 & A4). Multiple cryosectioned organoids were found to have necrotic cores, largely devoid of labelled cells (Figure 4.16B – M). Cells persisted around the surface where they were still able to get nutrients from the media.

Flow analysis of serial passaging (8-passages) and organoid cultures (~30 days) are displayed in Figure 4.17. In both cases labels expression remained stable with populations of each colour group detected after passaging and organoid culture (Figure 4.17 A - D). However, few clonal streaks were evident in U87 organoid flow-plots, contrary to what was observed for organoid culture of RGB marked G61 cells which did show streaks after culture. We subsequently assessed the shift in label proportions after 8-passages in adherent culture (n = 3) and organoids culture (n =

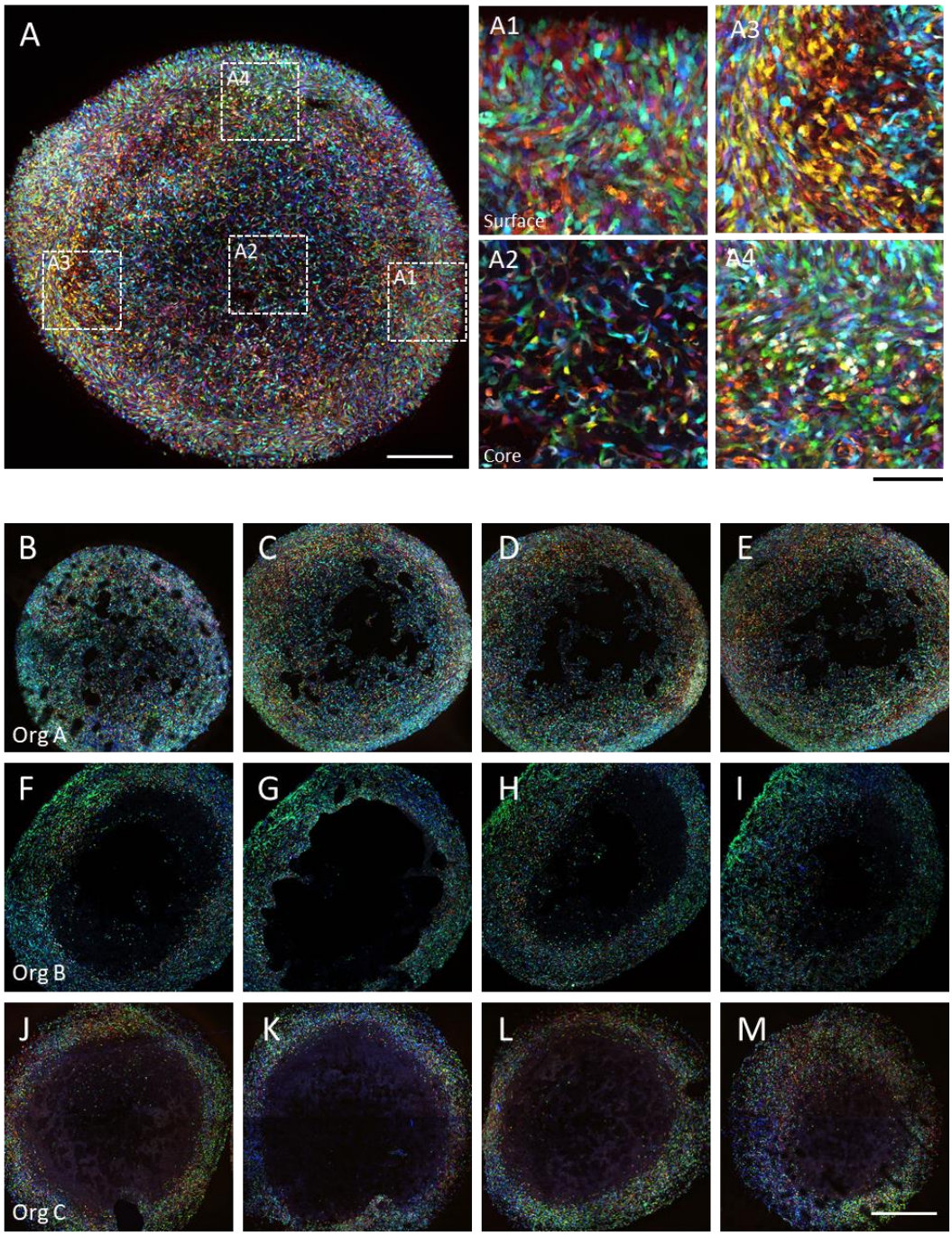


Figure 4.16 Representative imaging of barcoded U87 cells after organoid culture. (A) Example image of organoid after 30 days culture containing cells with a variety of fluorescent labels (scale bar = 500 µm). (A1) Dense cell organisation at the surface. (A2) Dispersed cellular arrangements in the centre. (A3 & A4) Demonstrate regional variation in clonal composition (scale bar = 100 µm). (B – M) serial sectioning of three further organoids; B – E = Org A, F – I = Org B and J to M = Org C (scale bar = 1mm).

5). U87 cells did not show the large proportional shifts observed in primary line G61 with the biggest shift observed in the unlabelled group which was a mean 3% change across organoids. Nevertheless, Two-way ANOVA found statistically significant

difference between means of proportional changes ($P < 0.0001$) but did not find an association with culture condition ($P = 1$) and an overall interaction score of $p = 0.3118$. results did however indicate that dual -label populations were more likely to decrease in their prevalence compared to single label or unlabelled cells irrespective of organoid or adherent culture. Proportional shifts in unlabelled, single labelled and dual-labelled populations were assessed in Figure 4.17F. With unlabelled colour groups ($n = 8$) showing the highest mean increase, single label ($n = 40$) showing a minor degree of mean of increase and double labels ($n = 80$) showing an overall decrease in their proportional representation. One-way ANOVA (CI = 95%) comparing group means suggests differences between means are statistically significant ($P < 0.0001$). Tukeys post-hoc for multiple comparisons also found $P < 0.0001$ for comparisons between each of the means. In conclusion, these results suggest that the number of viral fluorophores expressed may exert a degree of influence over the outgrowth of cells.

To test the potential for combining dual barcoding with staining for multiple surface markers (CD44, CD133, CD15 and A2B5) to discern a total of 16 surface marker expression phenotypes (Dirkse et al., 2019b). As described previously, staining for these 4 surface marker labels revealed marker heterogeneity in the CSC pool of different glioma CSC lines. However, the heterogeneity in marker expression between clonal populations of the same line has not been explored. We planned a panel of surface marker stains that were theoretically discernible from our five underlying lentiviral labels. Fluorophores, antibody targets/conjugated fluorophores, viability dye and their emission spectra are shown in Figure 4.18. Ghost dye 710 viability stain and Percp Vio® 700 (CD15) have near identical emission spectra but are excited by different laser wavelengths, and this also applies to APC-Vio® 770 (CD44) and PE-Vio® 770 (CD133) (Figure 4.18 B).

Combined application of these surface marker antibodies permits detection of up to 16 different single cell marker profiles. All possible marker profiles are displayed in

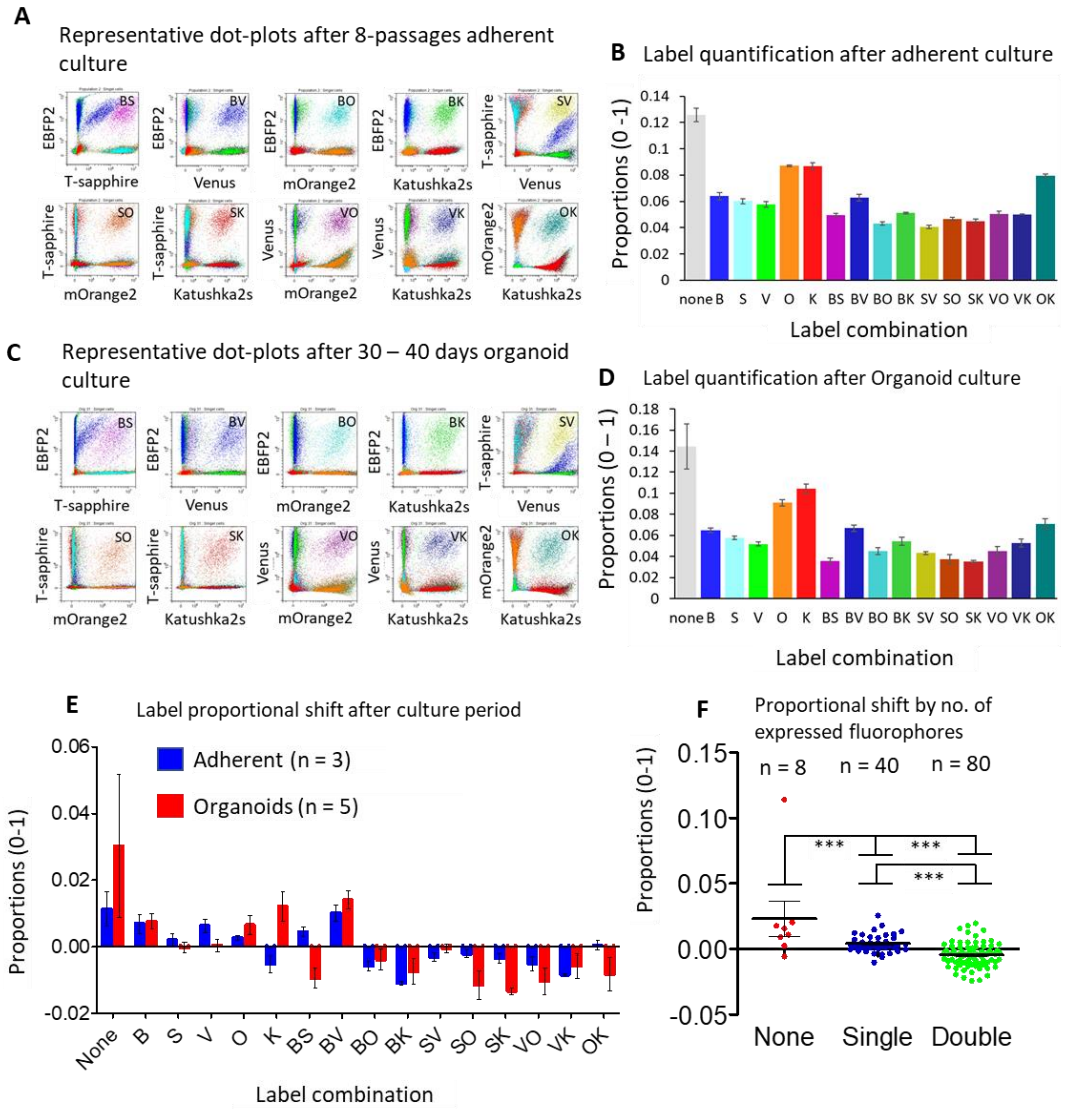
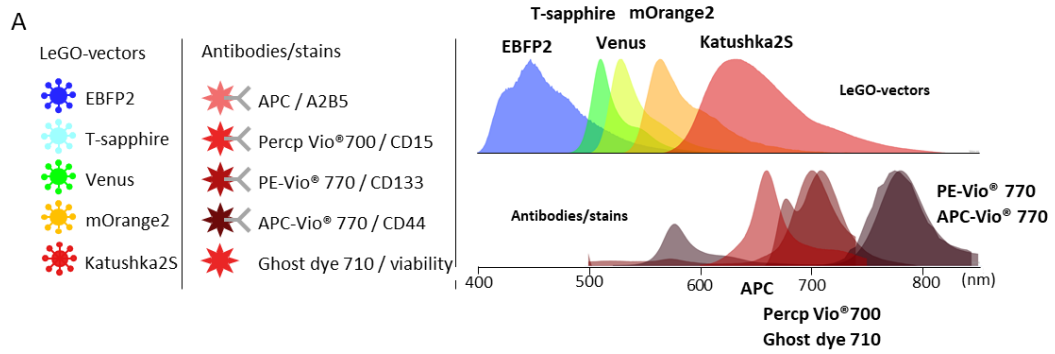


Figure 4.17: Serial passaging and organoid culture suggest single label and unlabelled cells have growth advantage over dual labelled populations. (A) Dot-plots displaying detection of 10 different dual label combinations after 8-passages of barcoded U87 cells in adherent culture. (B) Quantification of dual-label combinations after 8-passages in adherent culture (mean ± SEM). (C) Dot-plots displaying detection of 10 different dual label combinations after organoids culture of barcoded U87 cells. (D) Quantification of dual-label combinations after organoids culture (mean ± SEM). (E) Comparison of label changes across populations cultured over 8-passages (n =3) or in organoids (n = 5) (Mean ± SEM). Results suggest in general dual-label groups are declining more than single or unlabelled. 2-way ANOVA (95% CI) reported a $p < 0.0001$ for variation between colour groups, $p = 1$ for variation between culture condition and $p = 0.312$ for an interaction between these two dependent variables. Indicating that culture condition exerted no statistically significant difference on clonal outgrowth. (F) Quantification of proportional shifts in unlabelled (none), single label (B, S, V, O & K) and double label (BS, BV, BO, BK, SV, SO, SK, VK, VO, OK) from all tested samples in adherent and organoids groups (Mean ± SEM). Statistical analysis suggested significant difference between means of different label groups (one-way ANOVA (95% CI) with $p < 0.0001$, Tukeys multiple comparison test reported $P < 0.0001$ between all label groups).

Figure 4.19A, ranging from single, dual, triple or expression of all four markers. After dissociation, viability and surface marker staining is performed, owing to the

extensive internal cell death we observed after imaging of cryosections, a viability dye would be important for removing dead cells from marker analysis. Isolation of single cells, and removal of non-viable cells is shown in 4.19 B. To establish threshold boundaries for designating cells positive or negative for any of the four markers, control cells stained for a single marker only we required to assess any bleed through into collection channels of the other 3 markers (Figure 4.19C). In this case, we used adherent G61 cells as controls. However, optimal single colour control samples should be from the same conditions and cell line as the sample being profiled with all four markers.

A



B

Fluorophore /antibody /stain	EBFP2	T-sapphire	Venus	mOrange2	Katushka2S	APC	Percp Vio®700	PE-Vio®770	APC-vio®770	Ghost dye 710
Laser (nm)	405	405	488	561	633	633	488	561	633	561
Bandpass Filter (nm)	450/45	525/40	525/40	585/42	690/50	660/10	690/50	780/60	780/60	690/50

Figure 4.18: Experimental design for adjunct surface marker labelling of dual-label Barcoded U87 cells: (A) list of Tracing fluorophores, surface marker targets and antibody labels. Right hand side shows emission spectra of lentiviral labels (LeGo-vectors) and antibodies/viability marker. (B) Excitation laser and bandpass filters used for detection of each fluorescent signal.

Gates placed through reading of single colour controls are then applied to samples being read for all four surface markers (Figure 4.19 D). After staining with all four markers, it can be observed that some cells are being detected in the quadrant associated with co-expression of multiple markers. With the six gates displayed in Figure 4.19D ($CD44^+$, $CD44^+/A2B5^+$, $A2B5^+$, $CD44^-/A2B5^-$, $CD133^+$, $A2B5^+$) As shown in methods, Boolean logic can be implemented to designate all cell detections of the 16 possible marker profiles (Figure 4.19**Error! Reference source not found.**A).

Quantification of these marker profiles is shown a stacked bar graphs for three organoids, with results remarkably consistent between organoids. Between the three organoid five profiles (CD44, CD44/CD133, CD44/CD15, CD44/CD133/CD15 and CD44/CD15/A2B5) are consistently prominent but with subtle fluctuations in their proportions.

Importantly, at this stage we are assessing marker profiles at the whole organoid level, but these cells are also carrying their underlying lentiviral labels that permit each organoid to be split into 15 sub-populations. These 15 populations are shown in Figure 4.20A, comprising 5 x single colour labels and 10 x dual colour labels. However, upon splitting viral label populations and examining surface marker expression an interaction between mOrange2 and dkatshka2S with CD15 detection (Percp Vio® 700). This is demonstrated in Figure 4.20B where histograms showing quantification of CD15 expression in cells of either no viral label, EBFP, T-sapphire, Venus, mOrange2 or Katushka2S demonstrate a shift to the right in cells labelled with either mOrange2 or dKatushka2S. This effect is further exemplified by the proportional expansion of the $CD44^{+}/CD15^{+}$ surface marker phenotype in all viral label combinations containing either mOrange2 or dKatushka2S (Figure 4.20C).

In conclusion, 5-colour dual labelling is a highly effective approach for experimental analysis involving imaging and flow cytometry detection of sub populations. In contrast with glioma CSC line G61, 5-colour dual label U87 cells do not show large shifts in trace label proportions after culture in organoids. Suggesting primary human glioma CSCs may have different clonal dynamics to established and highly passaged cancer cell lines such as U87. Nonetheless, shifts in trace label proportions observed in organoids and in serial passaged adherent culture further support a growth advantage of unlabelled cells compared to labelled cells and potentially even single label over dual label. Finally, we were able to identify marker profiles of organoid cultured 5-colour dual labelled U87 cells but there was cross-talk between underlying lentiviral labels and marker detection which hindered accurate quantification of marker profiles.

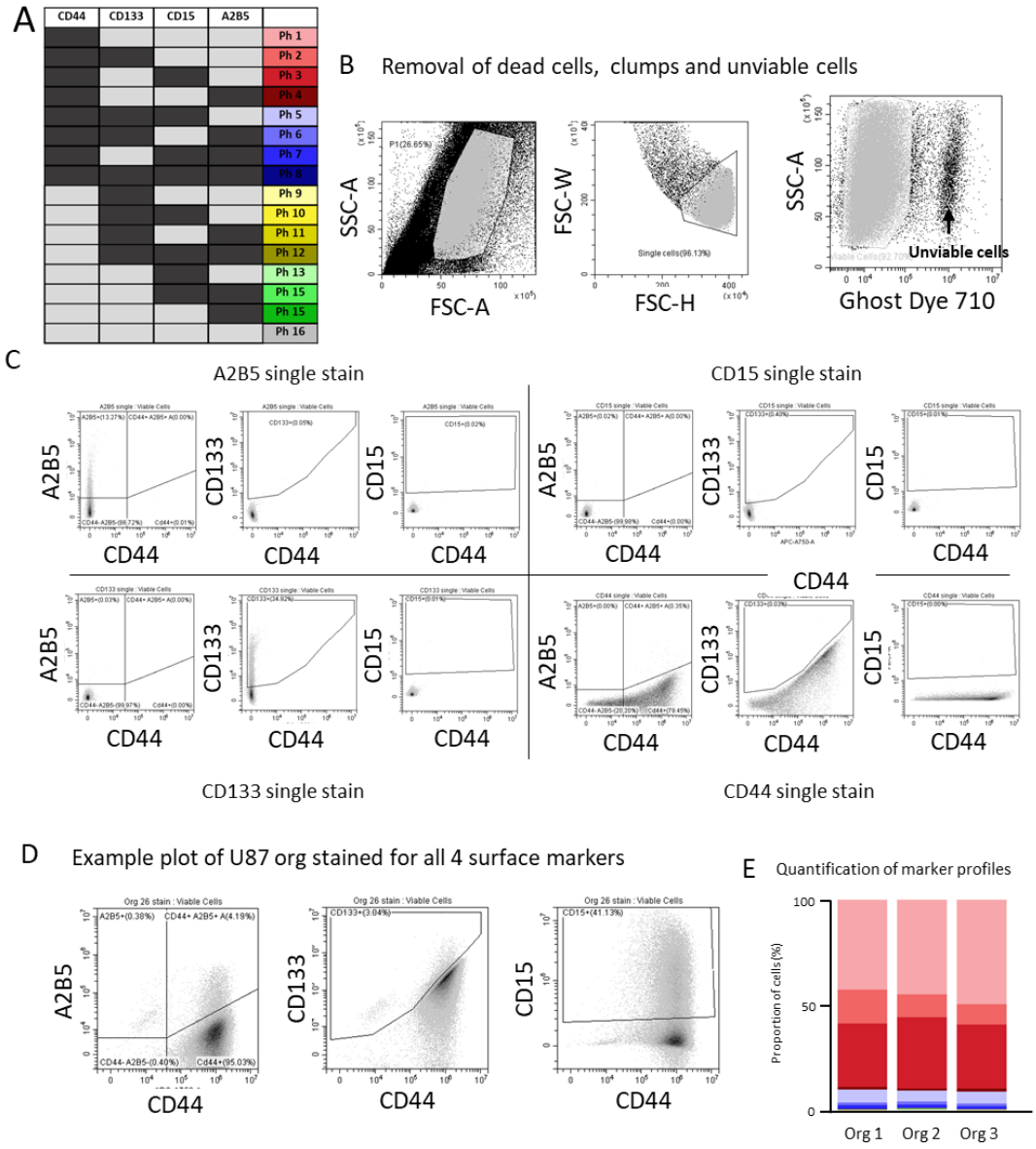


Figure 4.19: Detection of surface marker stains and viability dye in whole 5-colour barcoded U87 organoids. (A) Schematic demonstrating colour scheme used in all subsequent graphs displaying quantification of different surface marker phenotypes (Ph = Phenotype). (B) Flow cytometry pipeline for the removal of dead cells, cell doublets and non-viable cells using viability stain ghost dye 710®. (C) Plots of data acquisition from G61 cells stained for only a single marker; A2B5 (APC-A), (CD15 PerCP-A), CD133 (PC7-A) or CD44 (APC-750-A). These data used to set gates for quantification of marker profiles in samples stained with all 4 surface markers. (D) Representative flow plots of U87 cell organoid stained for all 4 markers simultaneously. (E) Quantification of (n = 3) U87 organoid surface marker profiles.

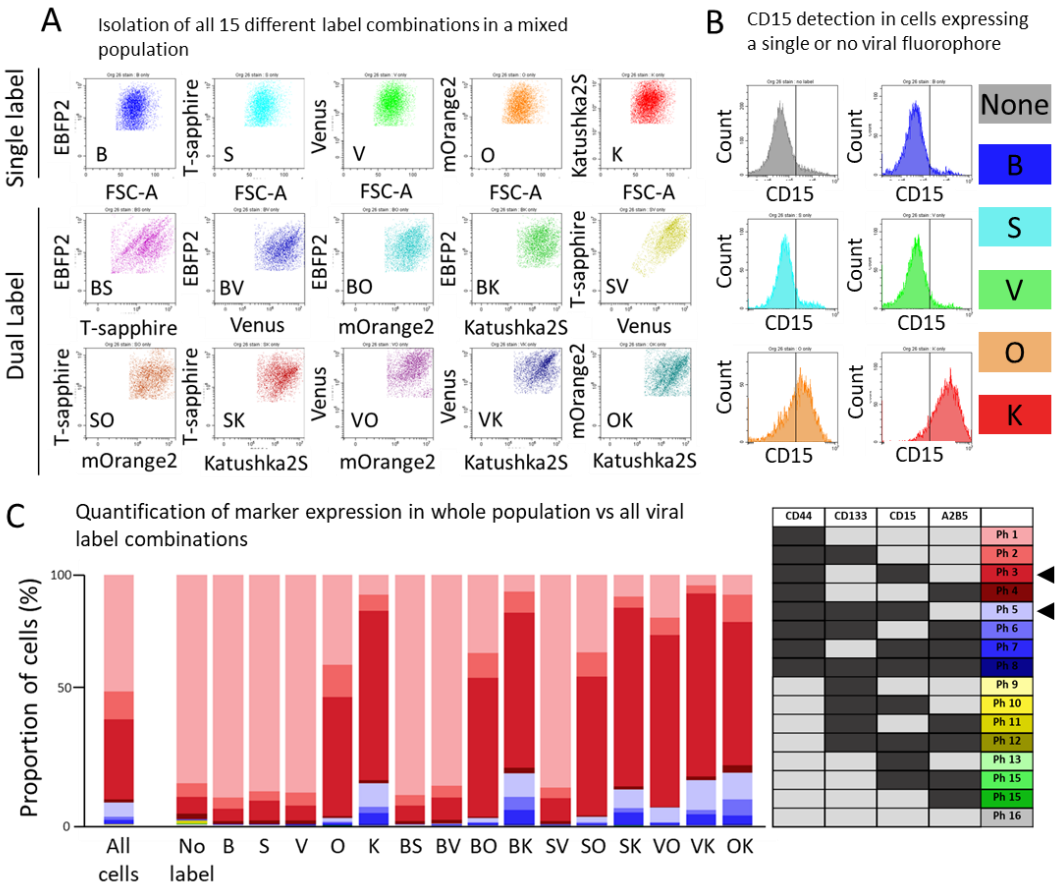


Figure 4.20: Splitting whole organoid into its 15 constituent colour groups reveals errors in marker profiles originating from mOrange2 and Katuhska2S. (A) Isolation of all 15 constituent cellular colour groups from a whole population of U87 organoid cells. (B) Histograms showing number of cells (count) distributed around detection intensity cut-off (vertical line) used to designate a cell as either positive or negative for CD15 expression. Shift cell intensity distributions can be seen in mOrange2 and dKatushka2S labelled cells indicating incomplete separation of the signal for CD15 detection with that for mOrange2 and dKtushka2S. (C) Quantification of surface marker data from 15 constituent groups of n = 3 U87 organoids. Notably, colour groups containing mOrange2 or Katushka2S show aberrant profiles associated with an increase in the $CD44^{+}CD15^{+}$ profile. Suggesting both fluorophores are bleeding into the $CD15^{+}$ collection channel.

4.3 Discussion

4.3.1 Formation of clonal populations and their tracking by barcode labelling

In this chapter, through successful labelling in U87, HEK293T and primary glioma CSC line G61, we demonstrate that LeGO-vectors are an excellent and adaptable tool for labelling cells *in vitro.* The use of LeGO-vectors has a number of advantages compared with confetti labelling. Firstly, through overlapping fluorophore expression, the number of detectable labels introduced can be expanded beyond the number of fluorophores delivered. Secondly, since we have a possible 6 different fluorophores that can be delivered, LeGO-vectors represent a more versatile lineage-tracing tool where label panel design can be adapted for different experimental settings. Preliminary experiments delivering three untitred viral fluorophores to U87 cells produced all seven possible labelling outcomes assessed by confocal microscopy; fluorophore expression persisted well with many labelled cells detected on cryosections of xenografted tissue and organoids while FACS isolation of each labelled group was also achieved. As discussed in chapter 3, the potential for FACS sorting of different clones after xenografting or after organoid culture represents a great opportunity for investigating clonal heterogeneity. Sorted clonal populations could be interrogated genetically looking at their copy number profiles and methylomes or investigating expressional differences through RNA sequencing or qPCR. Furthermore, emergence of clonal populations as streaks on flow cytometry plots further enhances the utility of LeGO-vectors for clonal detection compared with confetti labelling. We observed streak formation after organoid culture and through serial passaging of RGB marked primary G61 cells. A likely explanation for clonal streak on dotplots is that they are a product of lentiviral integration mechanics and regional chromosomal gene expression. Chromosomes are structured such that highly expressed genes are found clustered in co-regulated regions (Ridges) and the same is true for more lowly expressed genes (Anti-ridges)(Caron et al., 2001). As such, lentiviral transgenes carrying the same promoter sequence can vary in their expression as much as 8-fold depending on their site of integration (Gierman et al., 2007). In our context, a dual-labelled cell will have a ratio of expression between its

two fluorophores depending on their site of integration, where protein expression is detected as higher or lower signal intensities during flow cytometry analysis. These integration sites are passed onto cellular progeny leading to cluster/streak formation when a large clonal expansion occurs. In contrast, the cell population produced immediately after labelling has a huge number of transgene integration sites represented within the population and dual-label colour groups appear as spread-out clouds with many different insertion sites represented. In addition to streak formation after serial passaging of RGB marked G61 cells, we also observed a gradual reduction of cells expressing three fluorophores simultaneously. This effect was even more pronounced in organoid cultures of RGB marked G61 where virtually no triple labelled cells persisted. Interestingly, this effect was also observed for serially passaged RGB marked HEK293T but at a slower rate. Furthermore, organoid culture of RGB marked G61 cells seemed to promote outgrowth of single labelled cells at greatest frequency. These observations suggest that the expression of fluorophores impacts cell proliferation and survival. Indeed there are data to suggest that in certain models, GFP expression has cytotoxic effects through facilitating production of reactive oxygen species and impairing myosin cellular transport (Jensen, 2012, Ansari et al., 2016). Further experiments would be required to establish if these effects are influencing the outcome of our clonal assays. Alternatively, a further possibility is that cells were silencing the expression of integrated lentiviral expression cassettes. Promoter methylation on transduced genes is a well-established phenomenon and in our setting, silencing of a single expression cassette in a triple labelled cell would lead to these cells being detected as double or singly labelled. Out with promoter methylation, a further caveat is that with an increasing number of lentiviral integrations, there is an increasing likelihood that an integration will ultimately disrupt a gene or set of genes essential for continued proliferation. As such, cells affected in this way would be outgrown by other populations as was observed for triple positive cell populations in our assays. Another explanation for the prevalence of single positive EGFP cells after organoid culture of RGB marked G61, is that these populations all represent the same dominant clone that was present in the starting bulk population when seeding the organoids. In this case, its outgrowth can be

attributed some intrinsic properties of the clone rather than an association with reduced fluorescence expression.

Nevertheless, to maximise the potential for detecting clones and to address the issue of attrition in triple labelled cells, we modified the labelling method by using five lentiviral fluorophores across 10 separate dual transductions comprising all the dual fluorophore combinations. This strategy did not produce any triple labelled cells and maximised the number of dual labelled starting cells from which discernible clonal streaks could arise. Furthermore, RGB marked G61 cells reported overlapping streaks and we therefore, reasoned that spreading cells across more unambiguous label groups (15 in total; B, S, V, O, K, BS, BV, BO, BK, SV, SO, SK, VO, VK, OK) would reduce the chance of emergent predominant clonal streaks occurring and overlapping in the same unambiguous label group. Unfortunately, U87 cells did not show the same propensity to produce clonal populations as was observed in G61 so we couldn't fully assess the suitability of this labelling strategy in identifying clones. However, quantification of the labelled groups suggested that single label populations were growing at higher rates compared to dual labelled populations. Therefore, to further increase the likelihood of detecting clonal outgrowths as streaks and reduce the effects of potential label growth imbalances, we will introduce a FACS step after the initial labelling to enrich for dual label transduction outcomes only.

4.3.2 Automated identification and quantification of fluorescent labels assessed through imaging

With our barcoded U87 cells we developed a linear unmixing, cellular segmentation and label classification pipeline that was effective at classifying and quantifying cells of all labels. The fluorophore panel of EBFP2, T-sapphire, Venus, mOrange2 and Katushka2S leaves space for far-red nuclear dye, in this case we used DRAQ5. Although imaging produced comparable quantifications of colour groups, generally speaking the number of dual labelled cells was lower than that detected for flow cytometry and the number of negative and unclassified cells was much higher. The use of imaging to classify cells in this fashion has a number of inherent limitations. Firstly, we used nuclear detection to create an ROI for each detected cell from which

we extracted the fluorescence intensity data for each of the five detected fluorophores. With this approach, when detected cells are overlapping, fluorescence of two differentially labelled cells is detected as a single cell causing detections to be incorrectly classified as positive for 3 fluorophores. Because of this, any cells classified as triple positive were removed from the count analysis. LeGO-vector transduction produce dimly labelled cells that are readily detected through flow cytometry. With imaging, many of these dim cells were subsequently mis-classified as label negative as their signal intensity fell below our applied thresholds. Other technical caveats such as background fluorescence and differential focussing planes of fluorophores with different wavelength emissions further hampered the collection of accurate fluorophore intensity readings. Nevertheless, cells of all 15 different label groups were delineated through imaging suggesting, with some further optimisation, histological analysis in this fashion could be a powerful tool for investigating how different clones behave *in vivo.*

Removal of Katushka2S from this label panel would leave free wavelengths for adding immunofluorescence stains with an alexa 594 secondary antibody. Moreover, using the four LeGO-vector labels; EBFP2, T-sapphire, EGFP and Venus would leave free wavelengths for addition of two secondary antibodies (Alexa 561 & 594) for immunofluoresence and a far-red nuclear stain for automated cell detection. Combined immunostaining and clonal detection in xenografted tumours would be particularly well suited to investigating relationships between clonality, tumour heterogeneity and the microenvironment. Comparing clonal propensities for invasion or locations of proliferation could lead to Identification of clones with greater propensity for behaviours which progress tumour development could then be isolated to investigate mechanisms. A further possibility would be to look at whether certain clones have symbiotic effects on one another i.e identifying two clones which grow well when co-cultured could reveal signalling mechanisms which promote tumour progression. Indeed, a recent study using single-cell sorting and clonal production of optically barcoded pediatric GBM cells showed enhanced migratory ability when clones were co-cultured compared to cultured alone (Pericoli et al., 2020)

Allografting of barcoded mouse cells to animals with an intact immune system would permit investigation of how immune responses shape clonality or how different clones shape the immune environment. A study using optically barcoded GL261 marked cells demonstrated that tumours formed in immunocompetent mice had a lower number of persistent clones compared with tumours grown in immunocompromised $Pfp^{-/-}/Rag2^{-/-}$ animals (Maire et al., 2020). Authors also demonstrate that glioma cells responded to immune challenge through activation of the interferon response pathway and the majority of immunosuppressive gene expression signature actually originated from non-tumour stromal cells. Interestingly, the same two clones consistently performed best in overcoming immune challenges after allografting into WT mice suggesting this ability is innate to those two clones (Maire et al., 2020). An interesting follow on study would be too look at how these resistant clones behave in vivo compared to less immune resistant clones by staining for expression of proteins associated with immune escape. For example, investigating clonal expression of intra-cellular adhesion molecule-1 (ICAM-1) known to recruit myeloid derived suppressor cells which exert an array of immunosuppressive effects by modulating the microenvironment. Alternatively, staining for clonal expression of galectin-1 expression, a further adhesion molecule overexpressed in GBM with a characterised mechanism for inducing t-cell death. Such studies could also provide a picture of how the immune environment shapes clonal architecture and vice versa.

4.3.3 Combining optical barcoding with surface marker labelling to simultaneously assess clonal growth and plasticity

With our dual-label barcoded U87 cells we explore the potential for simultaneous flow cytometry reading of lentiviral labels with staining and detection of CSC surface marker expression. Recent work has shown that glioma CSC expression of establish CSC surface markers is highly plastic in different *in vitro* environments but the relationship between marker plasticity and clonal growth was not explored (Dirkse et al.,). We stained three optically barcoded U87 organoids for expression of the

surface markers CD44, CD133, CD15 and A2B5 with the conjugated fluorophores APC-vio®770, PE-vio®770, PercP-Vio®700 and APC, respectively. All 15 colour groups were detectable after surface marker staining, however, colour groups containing mOrange2 and Katushka2S interfered with detection of CD133 and CD15 expression. Based on these results, adapting the LeGO-vector delivered fluorophore combinations and surface marker conjugates, accurate reading of glioma stem-like cell marker expression by clone should be achievable.

Recent work has challenged the consensus that these markers represent indicators of a cell position within a traditional proliferative hierarchy (Scott et al., 2014, Scott et al., 2019, Brown et al., 2017), opening up new possible interpretations of data which seemingly support the presence of cell hierarchies in GBM. Combining RNA-seq with viral genetic barcoding in patient derived xenografts of GBM suggested considerable plasticity in cellular states and that a cells clonal heritage exerted little influence over the expression states a given cell could adopt (Neftel et al., 2019). However, in this study a very small number of cells (<15) were assessed in each of the identified clones. While single-cell RNA-seq is a more precise indicator of a cells state than CSC surface marker expression, our dual barcoding approach could permit a more powerful assessment of clonal plasticity as we can assess marker expression on thousands of cells from a single clone. Moreover, the potential to monitor clonal growth alongside clonal marker expression would permit assessment of how population level marker plasticity is underpinned by clonal outgrowth. Therefore, in the following section we outline changes to the surface marker labelling regime and investigate surface marker plasticity on mixed clones derived from the same primary tumour.

Chapter 5: Combined fluorescent lineage-tracing and surface marker phenotyping reveals clonal dynamics underlying glioma CSC marker heterogeneity

5.1 Introduction

In this final chapter we set out to establish the relationship between clonal dynamics and surface marker expression in primary glioma CSC lines. In the previous chapter we observed cross-talk between the previously utilised lentiviral labels and surface marker stains. Therefore, we first optimised the selection of lentiviral fluorophores and fluorescent probes for effective reading of both tracing labels and surface marker expression. To circumvent the need for single cell sorting to establish clones, we optimised conditions for low cell density seeding to achieve outgrowth of many clones within the same culture. We reasoned this would be a faster approach, yielding a larger number of clones for analysis, while, also mitigating any selective pressures associated with culturing clones from a single isolated cell.

Subsequently, we aim to assess clonal growth and marker plasticity over a number of passages by repeat flow cytometry staining and interrogation of lentiviral labels at each passage. This approach can then potentially provide novel insights of how population level marker expression is underpinned by clonal dynamics and marker heterogeneity. Subsequently, the serially passaged clonal mixtures will be seeded into three-dimensional tumour Matrigel® spheroids to investigate how transition to new microenvironment affects clonal architecture and marker expression. The influence of micro-environmental factors on cellular heterogeneity has implications for treatment development and studying this with respect to clonal populations has not been widely addressed. Furthermore, application of this approach to multiple patient-derived cell lines will provide a picture of how clonal dynamics, marker

heterogeneity and plasticity vary between patients. Consideration of patient specific disease mechanisms is a further aspect of therapy development emerging as an important aspect in achieving more efficacious treatments.

5.2 Results

5.2.1 Dual labelling with FACS purification permits adjunct surface marker labelling and effective detection of clonal expansions

As described above our first step was to achieve a labelled population optimised for clonal tracing and marker assessment. To achieve this, we made a small adaption to our 5-fluorophore labelling regime described in the GBM cell line U87 by removing Katushka2S from the trace label panel and performing only six separate dual transductions to produce 6 different dual labelled populations outlined below.

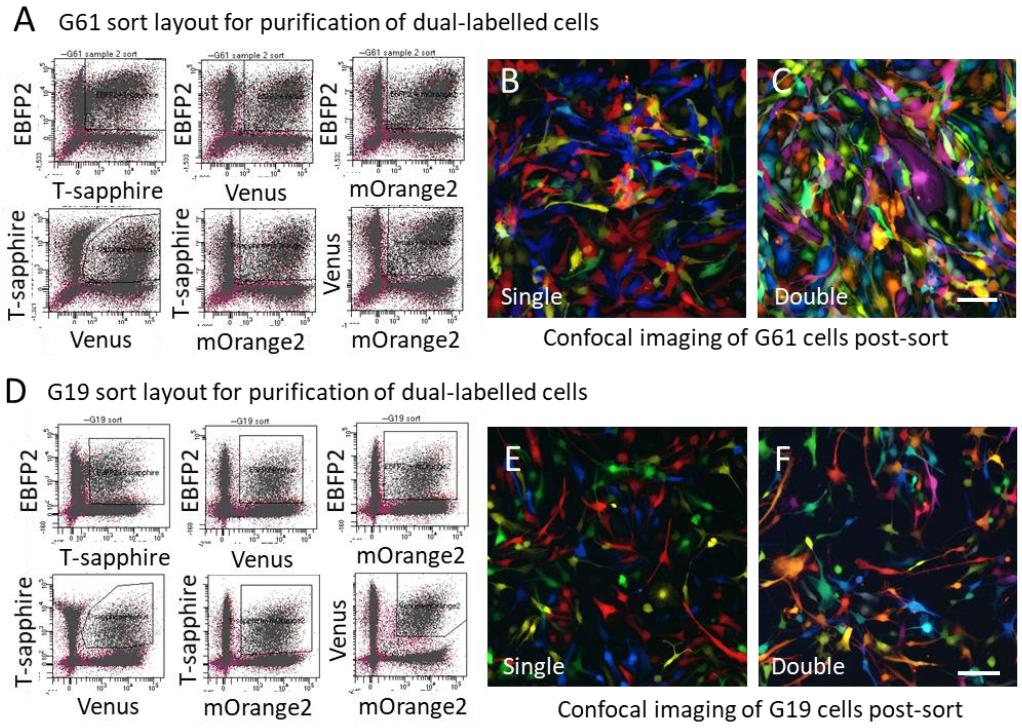
1. EBFP2/T-sapphire (BS)
2. EBFP2/Venus (BV)
3. EBFP2/mOrange2 (BO)
4. T-sapphire/Venus (SV)
5. T-sapphire/mOrange2 (SO)
6. Venus/mOrange2 (VO)

These label populations were established in two primary glioma CSC lines (G61 and G19) which had desirable properties for *in vitro* investigation, i.e. to readily attach to culture plates and rapid propagation. 7-days after exposure to the virus, cells from the 6 separate infections were pooled together and FACS sorted to purify dual labelled cells. As displayed in figure 5.1 A & D, these dual populations were readily identified during the FACS sorting step and gates were manually placed for each cell line for the purification of these six populations. Cells were sorted into two populations;

1. Mixture of the 6 gated dual-labelled populations described above (Double)
2. Mixture of the single label and unlabelled cells falling outside of gates (Single)

The efficacy of this in producing an exclusively dual-labelled populations was demonstrated through confocal imaging (Figure 5.1B, C, E and F). For each of the sorted cell lines the single label collection showed cells nearly exclusively of the 4 primary colours; Blue (EBFP2), Green (T-sapphire), Yellow (Venus) and Red

(mOrange2), while the dual labelled collection displays cells showing labels produced through combinations of these colours. The parameters of this sorting process are displayed in the table in Figure 5.1G, demonstrating the number of dual labelled cells collected for each cell line. In total nearly 600,000 dual labelled cells were collected for G61 suggesting the potential for tracing many different clones. Just under 200,000 dual labelled G19 cells were collected with a high sorting efficiency and few discarded events (96%).



G Parameters of sorts for G61 and G19 viral dual-labelled cell purification

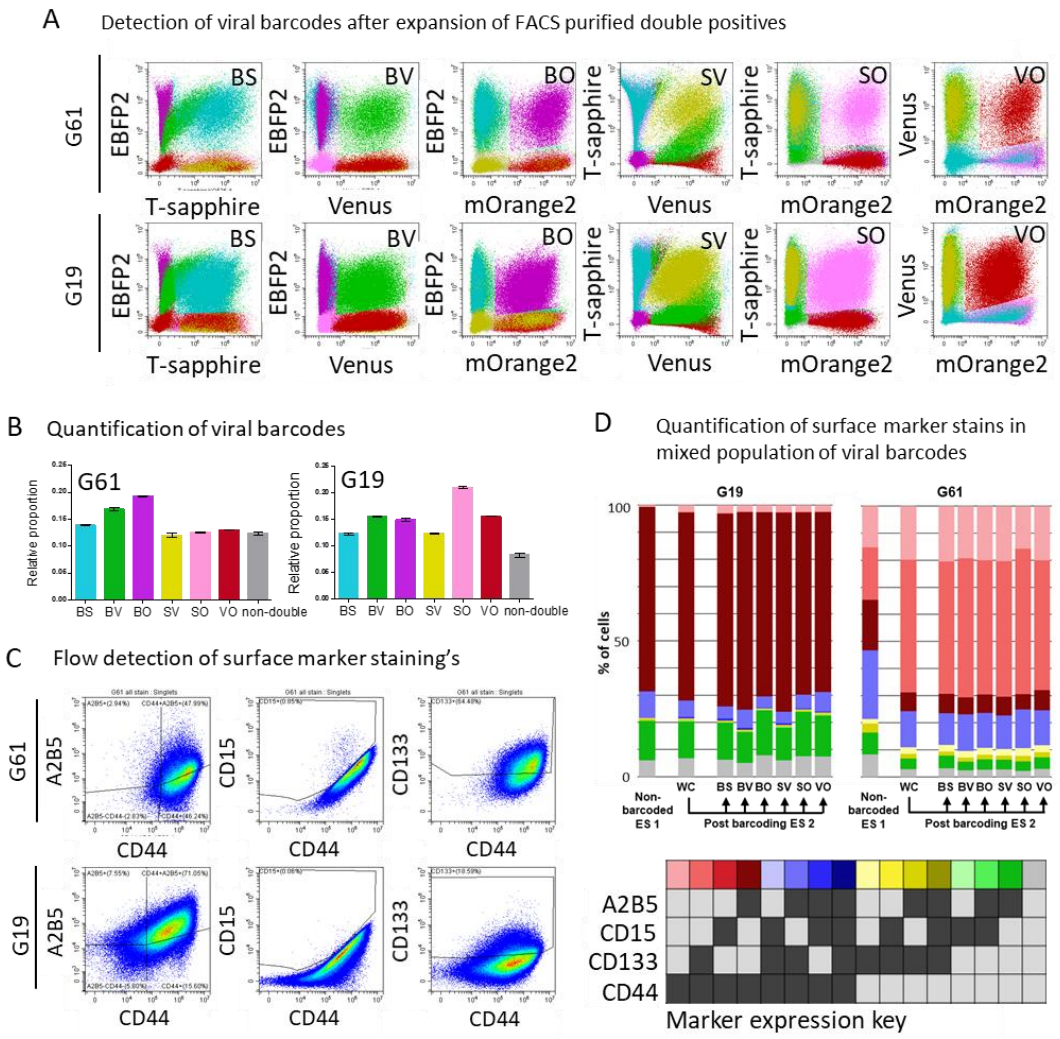
Cell line	Events processed	Aborted events	collected single label events	Collected dual label events	Efficiency (%)
G61	2,343,110	58,795	610,930	598,871	76
G19	711,860	1,377	400,962	195,529	96

*** Efficiency: No. of target cells sorted/total No. of target cells detected

Figure 5.1: FACS mediated isolation of dual-labelled cells to enhance ability for accurate tracing of tumour cell clones. (A) FACS plots and gates used for sorting the double label and single label G61 cells. (B) Confocal imaging of single labelled G61 cells after FACS sorting. (C) Confocal imaging of double labelled G61 cells after FACS sorting (scale bar = 200 µm). (D) FACS plots and gates used for sorting the double label and single label G19 cells. (E) Confocal imaging of single labelled G19 cells after FACS sorting. (F) Confocal imaging of double labelled G19 cells after FACS sorting (scale bar = 200 µm). (G) Parameters and statistics for sorting of 61 and G19 cells. (n = 1 sort for each cell line)

Collected cell populations were expanded as bulk mixtures for a further 7 days to freeze aliquots and test suitability of new probes selected for surface marker staining. Upon testing suitability of surface marker probes, G61 and G19 showed near equal proportions of viral labelled populations which retained their cloud-like distribution which suggested the retention of many labels after the sorting and expansion phase (Figure 5.2A). For both cell lines, a portion of single-positive and unlabelled cells, designated as "non-double" in Figure 5.2B, were observed suggesting these were not completely eradicated during the FACS sorting step. However, all single and unlabelled cells now represented a minor fraction of the whole cell populations that was comparable to a single dual-label group (Figure 5.2C) whereas these were the majority of cells in previous labelling approaches without a FACS step. Using surface marker probes comprising CD44 (APC-750), CD133 (Brilliant Violet 711), A2B5 (APC) and CD15 (Brilliant Violet 780) we were able to distinguish all surface marker signals from one another, and from the underlying viral label signals. In keeping with previous findings and demonstrated here each cell line showed a unique surface marker phenotypic distribution within the bulk sample (Figure 5.2C & D). In both cell lines, CD44 was expressed at the highest level, and A2B5 was also highly expressed in both lines, although, at a higher level in G19. CD133 was expressed on a large portion of cells in the G61 line with lower levels of CD133 expression observed in G19. In each of the middle panels in Figure 5.2C, CD44 (x-axis) is plotted against CD15 (y-axis) and it shows bleed-through from the CD44 probe APC-780 into the Brilliant Violet 780 channel which could not be corrected through compensation. Nevertheless, gating circumvented this issue and permitted the classification of $CD15^{+}$ cells which were found in very low prevalence across the three cell lines. Single stain controls reported almost 0% of cells expressing CD15 across the three cell lines suggesting bleed-through from the CD44 collection channel was not interfering with detection of CD15 expression. When cell line was split into its constituent viral label groups, marginal differences in marker phenotypes were observed for G61 and G19 (Figure 5.2D). Indeed, cosine similarity analysis reported a value of >0.995 when marker phenotypes of each viral label group were compared with those of the overall whole culture. Suggesting bleed-through artefacts observed

during U87 analysis have been mitigated with the adjusted composition of viral labels and surface marker probes.

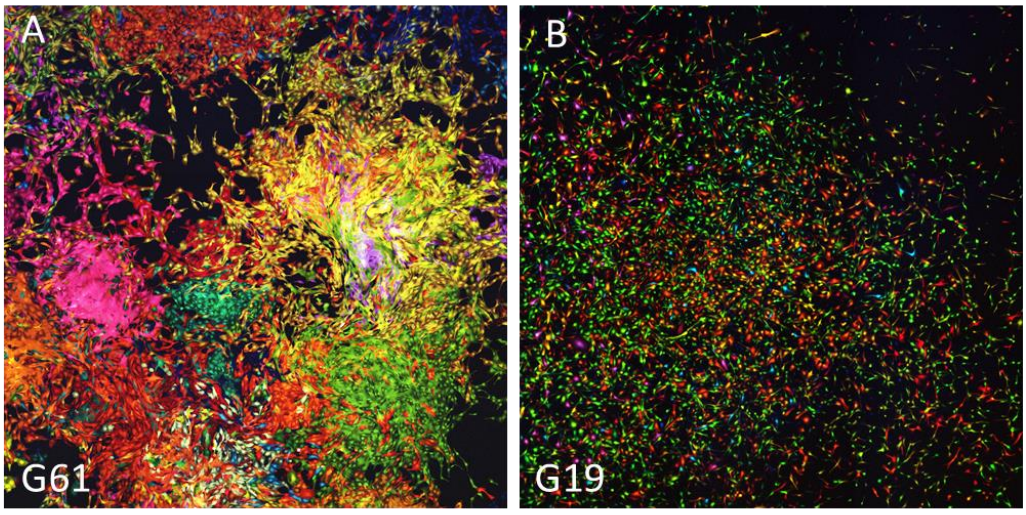


E Cosine similarity of marker expression underlying barcodes compared to whole culture (WC)

	BS	BV	BO	SV	SO	VO
G61	> 0.999	> 0.999	> 0.999	0.995	> 0.999	> 0.999
G19	> 0.999	0.999	0.999	0.999	0.999	0.999

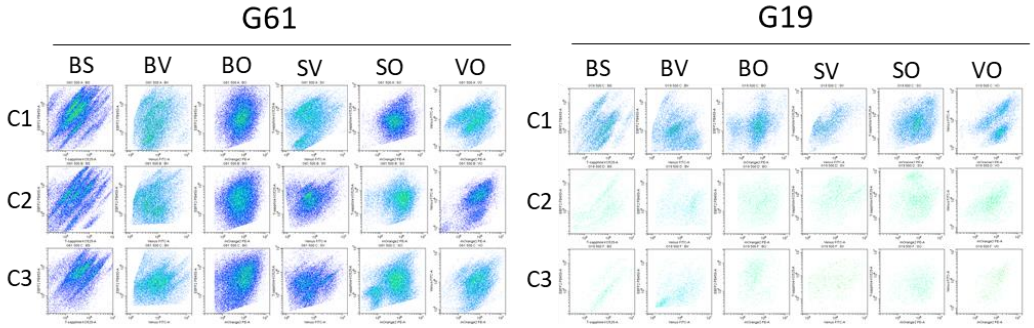
Figure 5.2: Surface marker profiling of differentially labelled glioma CSC populations after FACS purification of dual label cells. (A) Flow cytometry dot-plots showing detection of each dual-label population in a mixed bulk sample of G61 and G19 glioma CSCs after FACS sorting. (B) Quantification of dual barcode labels in G61 and G19 cells after FACS sorting. (C) Dot-plots showing distribution of surface marker expression in G61 and G19 along with gates applied to determine cells positive for each marker. (D) Graph showing distribution of surface marker phenotypes in non-barcoded whole culture, barcoded whole culture (WC) and each of its constituent viral label groups. (E) Summary of cosine similarity analysis comparing viral label group surface marker phenotype distributions to its corresponding whole culture. All values equal or succeed limit of monte carlo (CoS = 0.995) modelling for variation due to sampling error.

In the final step of optimisation for clonal tracking dual barcoded cells were seeded at low density; either 500 or 1000 cells into separate wells of a 48-well plate with the aim of producing traceable clonal mixtures from this starting pool of diversely labelled cells. Confocal imaging of these clonally emergent sub-cultures was performed after 10 days of culture to provide representative images of this process for each cell line (Figure 5.3A & B). These images reveal a phenotypic difference between the two patient-derived lines in terms of clonal colony formation and migration. G61 clonal populations grew in clusters while G19 show apparently greater motility with cells of different coloured clones spread out across the well surface (Figure 5.3A & B). Assessment of dual barcodes in G61 and G19 cells reported formation of streaks and clusters across the six dual labelled groups indicative of expanding dual-labelled clones, with comparable results after seeding 500 or 1000 cells (Figure 5.3D & E). As a result, we reasoned for future assays that trackable clonal populations would be produced by expansion from an initial population of 500 cells.



Confocal imaging of barcodes in cultures 10 days after seeding with 500 cells

C Flow assessment of barcodes after 10 days culture from 500 cells



D Flow assessment of barcodes after 10 days culture from 1000 cells

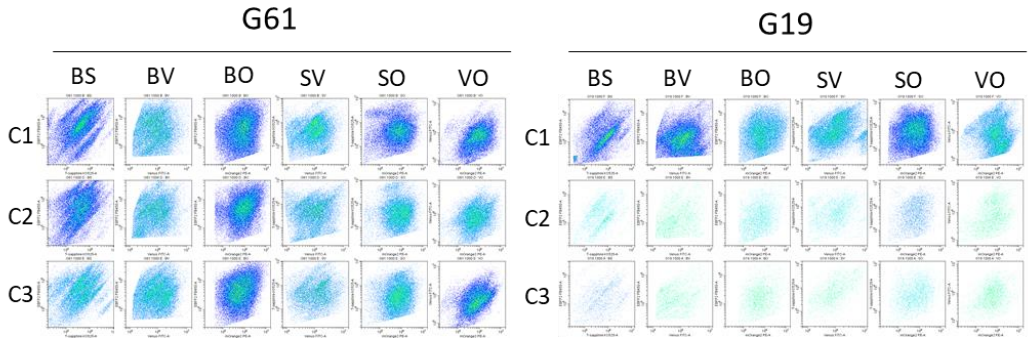
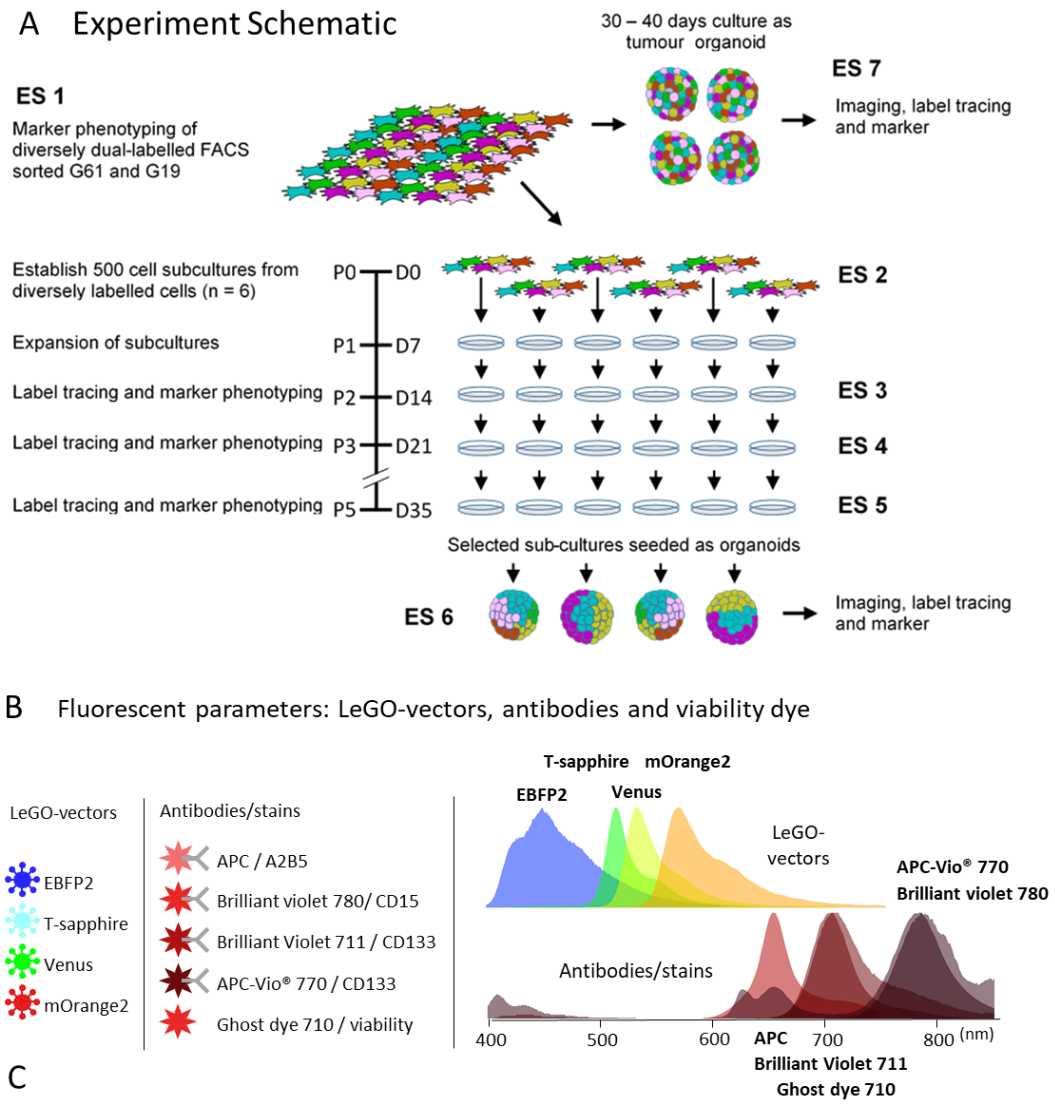


Figure 5.3: Low-density seeding of FACS sorted dual-labelled glioma CSCs can produce mixtures of traceable clones: (A & B) Imaging of 48-well plate wells 10 days after seeding approx. 500 dual-labelled glioma CSCs from patient-derived tumour). Differentially coloured clonal populations can be observed revealing unique behaviours between cell lines. (C) Flow cytometry dot-plots of the six dual-labelled populations (BS, BV, BO, SV, SO and VO) in (n = 3) 500 cell cub-cultures (C1, C2 & C3) after 10-days of growth for G61 and G19. (D) Flow cytometry dot-plots of the six dual-labelled populations (BS, BV, BO, SV, SO and VO) in (n = 3) 1000 cell cub-cultures after 10-days of growth for G61 and G19.

5.2.2 Sub-culturing and clonal tracking of glioma CSC show distinct dynamics in clonal expansion and selection

Combining dual label barcoding, surface marker labelling, and low-density sub-culturing outlined in previous section, we set out to assess the relationship between surface marker expression and clonal dynamics of G61 and G19. To achieve this, we seeded 500-cell sub-cultures (n = 6) for G61 and G19 in 48-well plate wells. Once confluent, sub-cultures were transferred (P1) and expanded in a 12-well plate wells. Sub-cultures were serially passaged in 12-well plates (25% of cells retained at each passage) with clonal proportions and marker expression assessed at passages 2, 3 and 5 (Figure 5.4). In the following section (Section 5.2.4), clonal mixtures established by P5 (n = 3 mixture for G61 and G19) are transferred to Matrigel® spheroids to investigate how clonal predominance and changing culture environment affect CSC surface marker expression (Figure 5.4, ES6). Subsequently (Section 5.2.5), diversely labelled mixtures produced in ES1 are directly cultured in organoids to assess marker expression and clonal expansion in a more complex *in vitro* system (Figure 5.4 ES7).

Quantification of flow cytometry detection of dual barcode and clonal proportions for the 6 sub-cultures across G61 and G19 are shown in Figure 5.5 and Figure 5.6 respectively. In both cell lines, across all 6 sub-cultures, discernible clonal populations arose which constituted the majority of cells in culture, raw dot-plots are shown in supplemental Figure 7.4 & Figure 7.5. And, examples of gating strategy used for detecting clones across passages is shown in Figure 7.6. For G61 we detected 13 – 16 clones per sub-culture and for G19 we detected 14 – 18 clones per sub-culture. Quantification of clonal proportions suggested, in both cell lines, a single or minority of clones were tending toward predominance while the majority were reducing in their proportional representation within the culture. Interestingly, in G61 sub-cultures, the predominant clone at P5 arose form a different barcode label in each culture other than sub-cultures 3 and 6 (C1:BO, C2:BS, C3:SV, C4:BV, C5:SO and C6:SV)(Figure 5.5B-G). However, in G19 sub-cultures 1 – 4, the predominant clone arose in the BV barcode group (Figure 5.6B – E), and. the position of the clonal streak indicated these were all descendent from the same clone (Figure 7.5A – D asterisk). While in G19 sub-cultures 5 and 6, predominant clones arose in barcode label SO.



Fluorophore /antibody /stain	EBFP2	T-sapphire	Venus	mOrange2	APC	Brilliant Violet 780	Brilliant Violet 711	APC-vio®770	Ghost dye 710
Laser (nm)	405	405	488	561	633	405	405	633	561
Bandpass Filter (nm)	450/45	525/40	525/40	585/42	660/10	610/40	780/60	780/60	690/50

Figure 5.4 Scheme for assessing glioma CSC clonal dynamics and marker expression in different environments. Experimental step 1 (ES1), perform marker phenotyping of FACS purified diversely labelled dual-label glioma CSC lines G61 and G19. (ES2) Establish (n = 6) 500 cell sub-cultures to produce traceable clonal mixtures. (ES3) After 14 days and at the first passage 3/4 of the cells are to be harvested to assess surface marker expression. The cells are cultured up to passage 5 with further surface marker profiling at P3 (ES4) and P5 (ES5) after 35 days in adherent culture. (ES6) Portions of cells from 500-cell subcultures will then be placed in organoids to look at how their clonal dynamics and marker expression shift in a new environment. (ES7) Organoids will also be grown from the diversely labelled glioma CSC populations which will be assessed for their clonal content and marker expression once they reach maturity after 30 – 40 days. (B) Summary of viral labels and surface marker probes used for clonal tracking and surface marker phenotyping. (C) Table displaying cell analyser laser and filter setups for each fluorescent signal.

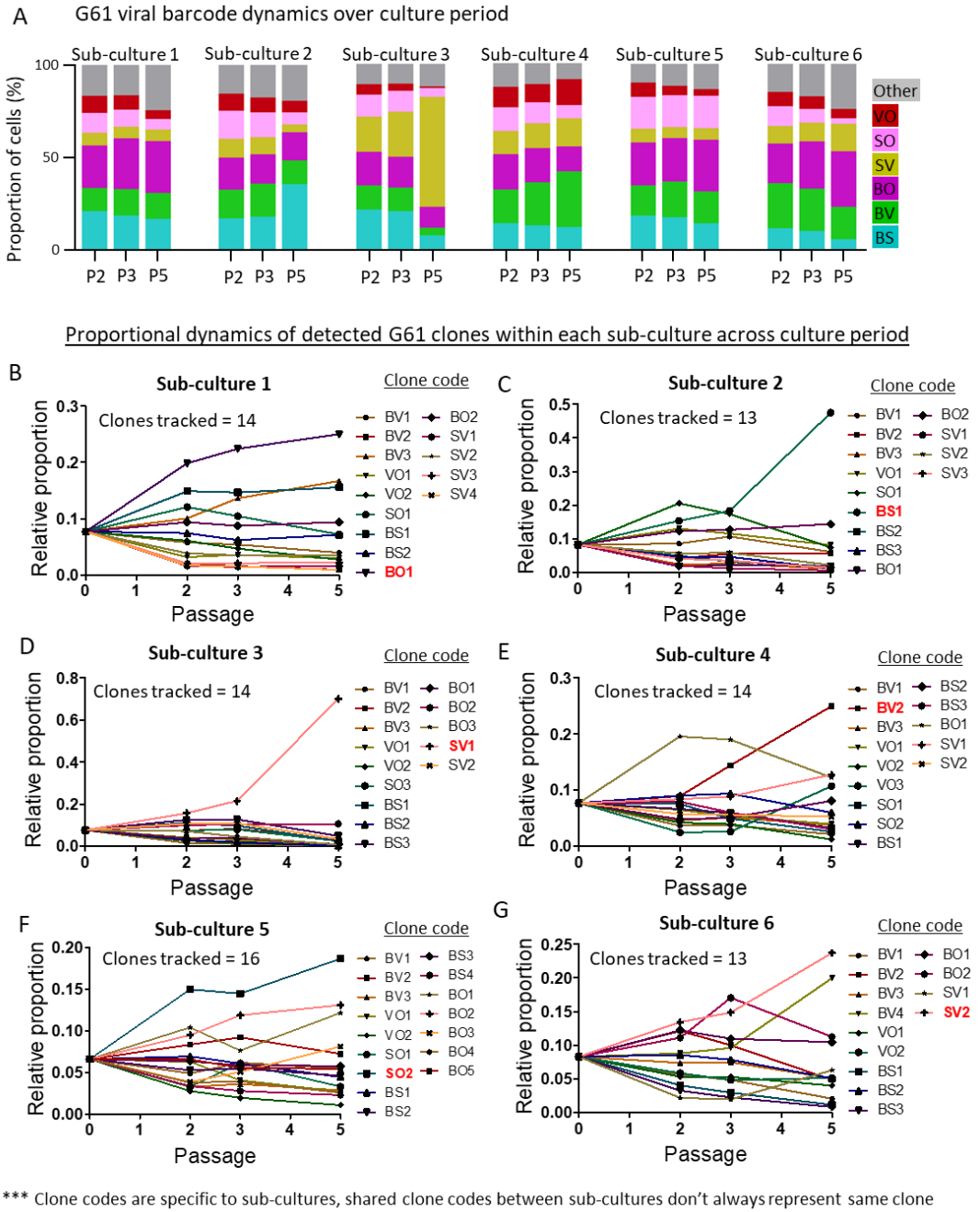


Figure 5.5: Clonal tracking in 500-cell derived cultures of barcoded G61 reveal clonal populations with different growth propensities. (A) Quantification of viral label groups (BS, BV, BO, SV, SO, VO and other labels) across passages 2, 3 and 5 (P2, P3 and P5). (B – G) Relative proportions of cell numbers in tracked clones across P2, P3 and P5 detected in different viral label groups. Clone codes indicate the viral label group a clone is from and the number deviates clones within the same label group. Most prominent clone at P5 in each culture is highlighted in red in list of clone codes. N = 1 reading of clonal proportion at each passage. At passage 0 (i.e the seeded population of 500-cells), clonal proportions were assumed to be equal although not quantified. Scales on y-axis of each graph adjusted to best display data from each culture Line colours are arbitrary and applied randomly to each plot. A variety of clonal growth dynamics can be seen across the subcultures with either a single clone (Sub-culture 2 and 3) or 2 – 3 clones out performing the majority (Sub-culture 1, 4, 5 and 6)

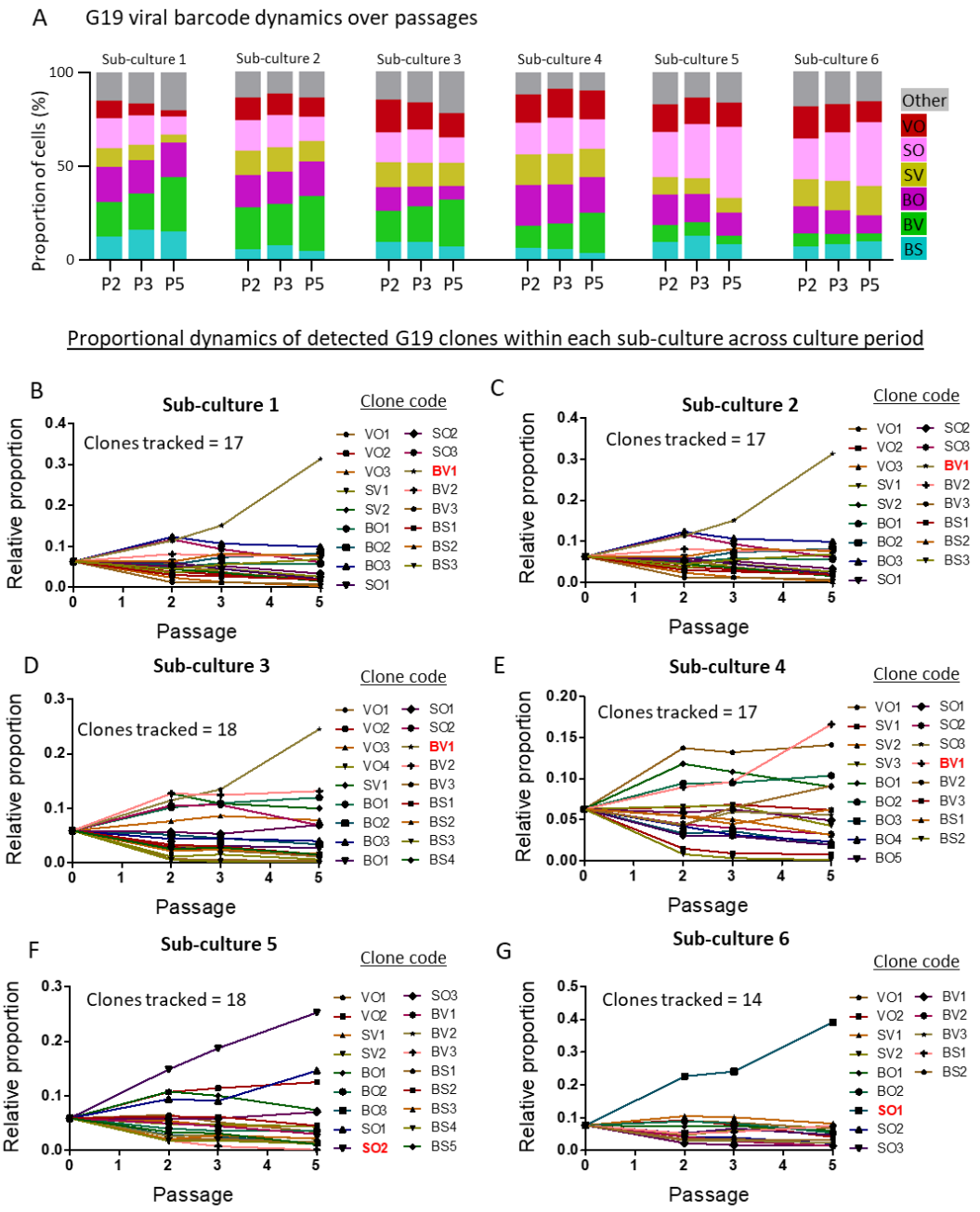


Figure 5.6 Clonal tracking in 500-cell derived cultures of barcoded G19 reveal clonal populations with different growth propensities. (A) Quantification of viral label groups (BS, BV, BO, SV, SO, VO and other labels) across passages 2, 3 and 5 (P2, P3 and P5). (B – G) Relative proportions of cell numbers in tracked clones across P2, P3 and P5 detected in different viral label groups. Clone codes indicate the viral label group a clone is from and the number deviates clones within the same label group. Most prominent clone at P5 in each culture is highlighted in red in list of clone codes. N = 1 reading of clonal proportion at each passage. At passage 0 (i.e the seeded population of 500-cells), clonal proportions were assumed to be equal although not quantified. Scales on y-axis of each graph adjusted to best display data from each culture Line colours are arbitrary and applied randomly to each plot. A seen by raw flow cytometry plots the same clone appears to be predominating in sub-cultures 1 – 4. With clones form a different viral barcode predominating in sub-cultures 5 and 6.

A variety of growth dynamics could be observed in the clones tending toward predominance in both cell lines. In some cultures the predominant clone at P5 was already detected as the predominant clone at P2. However, in certain cultures the predominant clone at P2 or P3 was superseded by a less prevalent clone by P5 (G61: Sub-cultures 2, 4 and 6 & G19: Sub-culture 3 & 4). Results of flow cytometry quantification were supported by imaging results of each culture at P2 and 5 displayed in Figure 5.7

In conclusion, our barcode label approach could detect and quantifying the proportions of up to 18 clones in a culture over 5 passages. Tracking suggested each culture was tending toward clonal predominance with variable growth patterns observed between clones in the same culture.

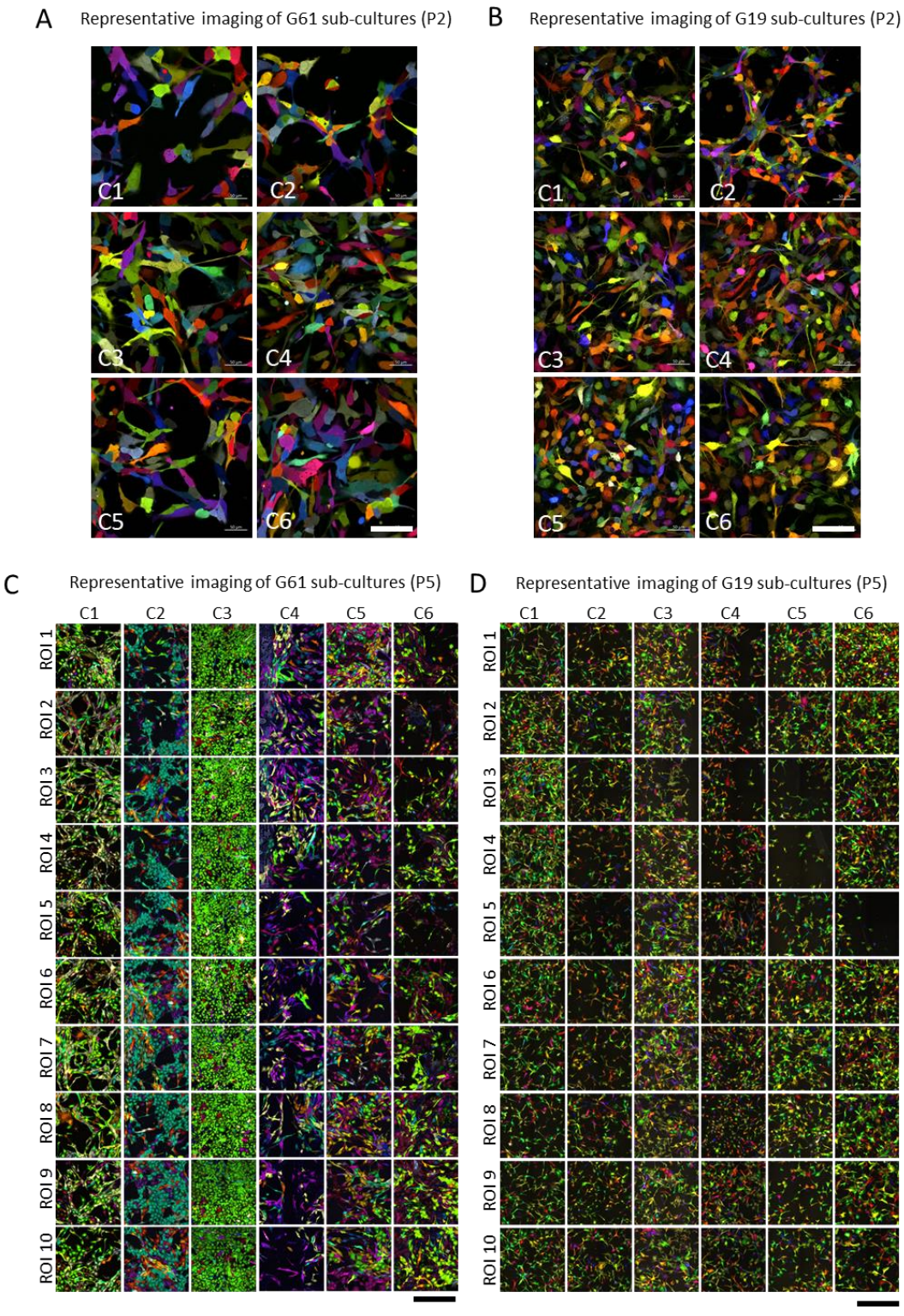


Figure 5.7 Fluorescent imaging of sub-cultures at P5 show different pattern of clonal outgrowth between glioma CSC lines. (A) representative images of culture regions from the (n=6) serially passaged G61 sub-cultures at passage 2 (P2). (B) Same as A but showing G19. (C) Panel displaying multiple regions of interest (ROI) of cells expanded from the G61 P5 population. The predominant colours observed across the ROIs corroborate with flow cytometry data. (D) Multiple ROIs of cells across the (n = 3) G19 subcultures (C1 – C6) after expansion of cells form passage 5. G19 cultures do not show the distinct colour differences seen across G61 sub-cultures.

5.2.3 Adjunct clonal tracking and surface marker profiling reveal clonal dynamics underlying marker plasticity in glioma CSC lines

In addition to reading of lentiviral labels, each sub-culture was also stained and assessed for its marker expression at P2, P3 and P5 (Figure 5.8A). Prior to seeding sub-cultures, the bulk population of barcoded cells was also profiled for cellular marker expression and this reading is displayed as P0 in Figure 5.8A. Our Monte-Carlo (95% confidence intervals) modelling suggested that sampling of 500 cells from the seeding population would lead to maximum cosine similarity deviation of 0.98. In both cell lines, we found that marker expression was not stable across passages and proportions of marker profiles fluctuated between sub-cultures. However, by P5, sub-cultures C1 and C3 for G61 reported a notably larger representation of the $CD44^{+}/A2B5^{+}$ marker expression profile compared to the other G61 sub-cultures (Figure 5.8A). G19 produced a much larger portion of marker negative cells than any of the G61 cultures and interestingly, by P5 all G19 sub-cultures shared a more consistent marker expression profile. We applied cosine similarity analysis to perform pair-wise comparisons of each sub-culture's marker profile at P2, P3 and P5 with the P0 value recorded for G61 and G19 (Figure 5.8B). Black dotted line denotes threshold cosine similarity that could be attributed to sampling error (0.98). These results further suggest the two cell lines show distinct behaviours in marker plasticity across the passages. It appears that G19 sub-cultures adapt their marker expression with variable degrees of similarity to the seeding P0 population. Interestingly, each sub-cultures similarity to P0 remains relatively stable at P3, but by P5, sub-cultures converge and share a more consistent degree of similarity in their marker expression compared with the seeding population.

This array of similarities is retained at P3 but by P5 all cultures cluster with a similar degree of similarity to the P0 population. Conversely for G61, sub-cultures show a large degree of similarity with the P0 population at P2 which then diverges for all populations at P3. By P5 all G61 sub-cultures other than C3 again increase their similarity with the starting population. Interestingly, G61 sub-culture 3 (C3) contained an SV clone that constituted over 60% of all cells in the culture that

diverges considerably in its marker expression compared to other cell clones (Figure 7.7). This suggests the divergence of C3 from the general pattern of marker

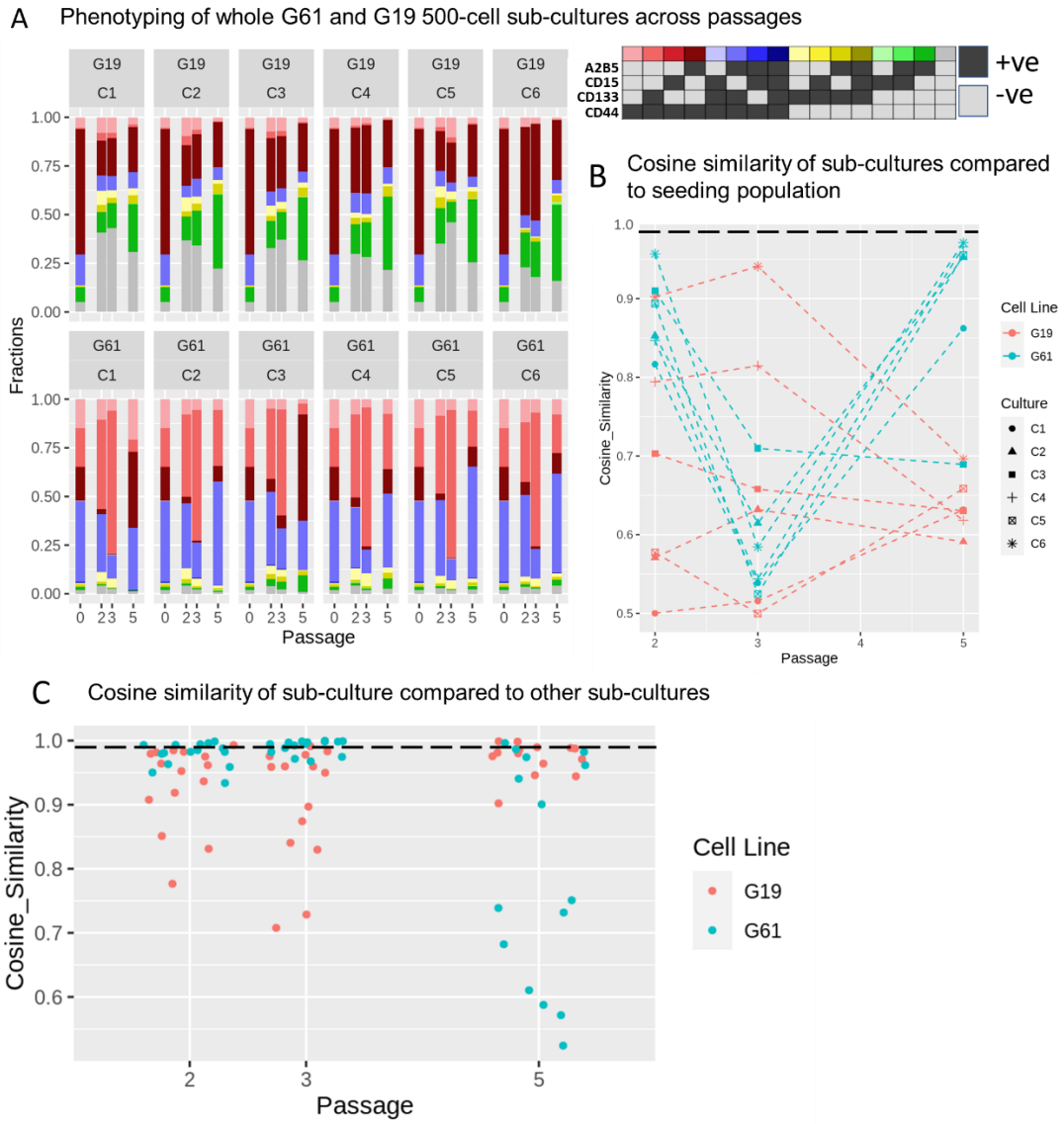


Figure 5.8: Whole culture surface marker analysis of serially passaged G61 and G19 sub-cultures reveals unique dynamics in marker plasticity. (A) Stacked bar plots showing cellular proportions of marker expression in glioma CSC (G61 or G19) bulk sample (Passage 0) and sub-cultures (C1 – C6) over passages 2, 3 and 5. Binary plot explains how colours denote marker expression profile. (B) Cosine similarity analysis of passaged (P2, P3 and P5) sub-culture (C1 – C6) marker profiles compared to bulk sample (passage 0) which seeded sub-cultures. (C) Pair-wise cosine similarity comparisons of sub-cultures (C1 – C6) for G61 and G19 across passages 2, 3 and 5.

expression shared across passages of the other 5 G61 subcultures is solely due to the expansion of a single unique clone. In a second analysis approach, pairwise cosine similarity analysis comparing each of the six sub-cultures with one another at P2, P3

and P5 (i.e 5 comparisons per sub-culture per passage and 15 comparisons per cell line per passage). This analysis further supported the general trend observed when comparing sub-cultures with the P0 population. Namely, that G19 sub-cultures started varied and became more similar while G61 sub-cultures started similar and became more variable by P5 (Figure 5.8B & C).

To further probe the clonal influences on population surface marker expression we set about gating and extracting surface marker profiles for each of the identified clonal populations across the sub-cultures for G19 and G61. In total, 104 and 84 clones were extracted for G19 and G61 respectively. Examples of gated clones and a summary of extracted clones surface marker expression can be found in supplemental Figure 7.7. To summarise the clonal surface marker profiles and their distributions compared to one another, we again subjected the data to pair-wise comparisons of all clonal marker profiles from the same sub-culture across passages 2, 3 and 5, calculating a cosine similarity value for each pair-wise comparison. This provided a picture of the variation in marker expression across clones in each of the six sub-cultures, revealing distinct patterns of clonal surface marker heterogeneity between the two cell lines (Figure 5.9). Again dotted line displays the 0.995 Monte-Carlo CoS score deviation that could be attributed to sampling error. For both cell lines at all passages, the highest density of pairwise cosine similarity comparisons are towards a score of 1. Suggesting there is a common distribution of surface marker profiles which is shared by many clones within each sub-culture. However, both cell lines also show outlier clones that report low cosine similarity values (CoS < 0.25) compared to other clones in their sub-culture. Monte-Carlo modelling would suggest deviation below cosine similarity of 0.995 could not be explained by effects of random sampling. While both cell lines share the presence of these outlier clones the distribution of these outliers is markedly different. G19 clonal surface marker profile variation is spread across a spectrum of dissimilarity from near identical (CoS = 1) to highly divergent expression profiles and many intermediate profiles. However, G61 shows a more bi-modal distribution with clonal comparisons from C3, C4 and C6 diverging considerably from the clones that cluster around a cosine similarity score

of 1. The clonal profiles causing this bimodal distribution are highlighted in supplemental Figure 7.7 where bar charts of all clonal marker profiles are displayed.

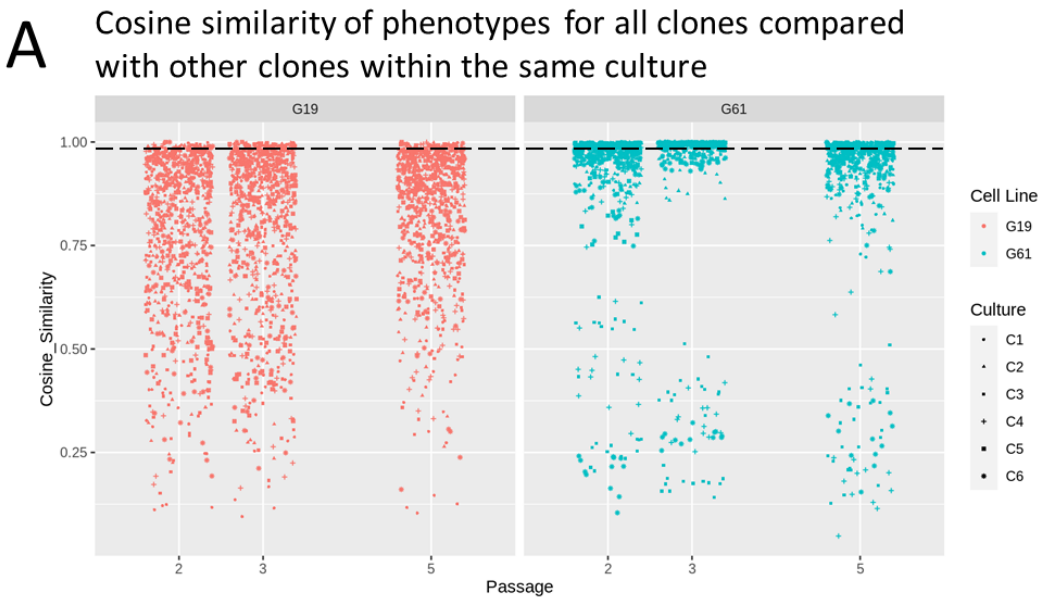


Figure 5.9: Cosine similarity reveals heterogneiety in clonal marker expression and apparent differences between G61 and G19. (A) Pairwise cosine similarity comparisons of clonal marker profiles between clones of the same sub-culture. High cosine similarity values suggest comparison of highly similar clones and low values suggest comparison of dissimilar clones. Results are displayed for each passage (2, 3 and 5) and comparisons of different sub-cultures (C1 – C6) are indicated by different shapes.

To further summarise the flow cytometry data and capture all inferable parameters of our assay, we calculated the mean of cosine similarity values for each clone at each passage (Figure 5.10). Providing a single summary measure of how similar a given clones marker expression is to other clones within the same sub-culture allowing us to calculate the mean fold change in cosine similarity for each clone between passages 2 and 5. A negative fold change indicates a clone has decreased in similarity to other clones in its sub-culture, whereas, a positive value indicates a given clone has increased in similarity to other clones in its sub-culture from P2 to P5. Clones which cluster around the value “0” would have shown no or very small change in their degree of similarity/dissimilarity to other clones in their culture. The second parameter calculated was the fold change in the fraction of cells a clone represented within its whole sub-culture between P2 and P5, providing a summary parameter of a clones behaviour in terms of outgrowth or attrition. A positive fold change indicates

this clone has expanded considerably while a negative fold change suggests either clonal loss or that a clone has remained the same size but become proportionally smaller owing to expansion of other clones (Figure 5.10).

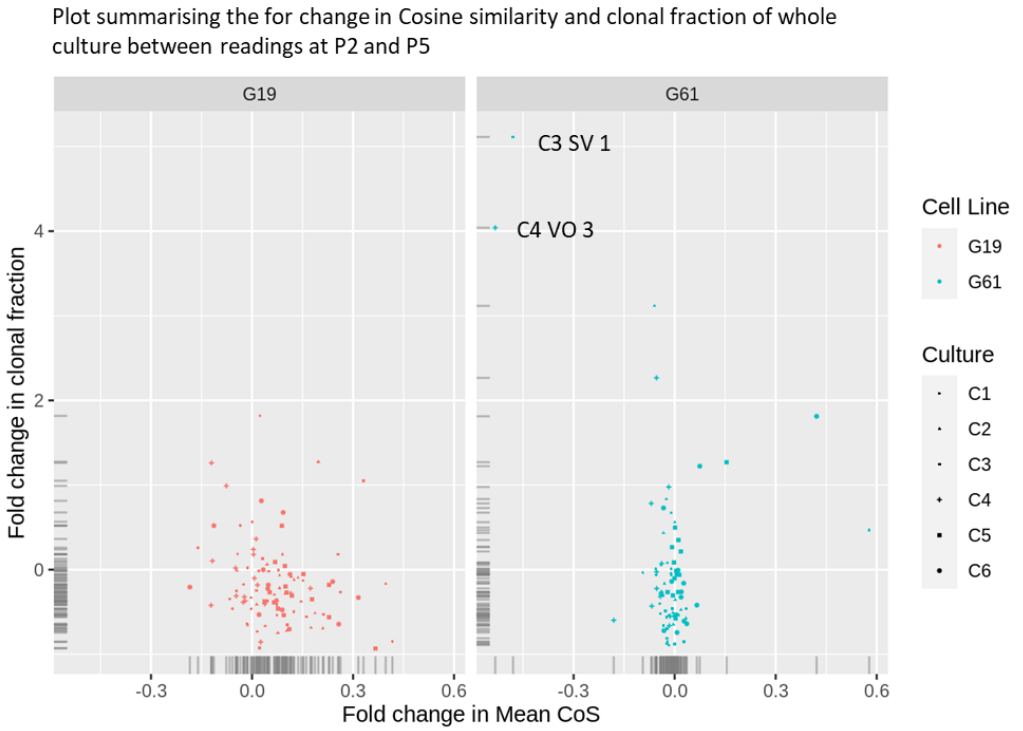


Figure 5.10: The relationship between clonal marker expression plasticity and growth. Fold change in clonal fraction; A given clones proportion of whole culture at P2/same clones proportion at P5). Fold change in mean CoS; Mean value of all pairwise cosine similarity calculations for a given clone at P2/same value at P5. Shapes dictate sub-culture corresponding to a given clone and are indicated in key on right handside.

This analysis further highlights the differences in clonal and marker profile dynamics between the two cell lines and paints a similar picture to previous analysis. G19 shows an array of clonal readings in terms of proliferation and marker diversity while G61 is shows a cluster of clones with similar readings and a small number of outliers. Interestingly, there is a skew towards a positive fold change in mean CoS for all G19 clone suggesting clones within a culture are tending toward similarity with one another in terms of their marker expression (Figure 5.10). This is reflected in what was observed when assessing marker expression between whole sub-cultures for

G19 (section 5.2.2) where sub-culture marker expression was variable at P2 and became more similar by P5 (Figure 5.8C). Furthermore, in the G61 sub-cultures, the two clones (C3 SV1 and C4 VO3) showing the greatest fold change in their clonal fraction also showed the most negative fold change in their mean CoS (Figure 5.10). Indicating a divergence in marker expression of these two clones compared to the rest of their sub-culture. For both cell lines, the majority of clones report a negative fold change in clonal fraction suggesting most clones within the sub-cultures are reducing in their proportional representation (Figure 5.10).

In conclusion our sub-culturing, serial passaging, clonal tracking and surface marker assessment reveal extensive plasticity in marker expression and distinct behaviours of different patient derived lines. Our results also suggest that cellular surface marker expression is influenced by a cells heritage as we observed clones from the same cell line and indeed the same sub-culture with distinct distributions of marker profiles.

5.2.4 Transfer of sub-cultures to organoid culture results in more pronounced clonal outgrowth with predominant clones presenting distinct marker profiles

After 5 passages in adherent culture, selected sub-culture (C1 – C3) for both G61 and G19 were seeded into organoids. Each organoid was seeded with approximately 5000 cells and the number of repeat organoids sampled for each sub-culture are indicated on the figures. Once organoids had reached maturity (30-40 days), living organoids were embedded in agarose and sectioned on vibratome for imaging. The following day, organoid sections were dissociated, stained and assessed by flow cytometry. Once again cell lines showed very divergent behaviours in their clonal content and marker plasticity making direct comparison difficult. For this reason, results for the two cell lines are presented separately. Organoids were successfully derived from G61 sub-cultures 2 (n = 10, Figure 5.11) and 3 (n = 7, Figure 5.12), while organoids grown from sub-culture 1 developed very slowly and were not processed for analysis. For sub-culture 2 and 3, all mature organoids presented with two predominant clones with variable proportions observed between organoids (Figure 5.11 A&B and Figure 5.12 A&B). For both sub-cultures, the clone apparent through flow cytometry were also readily detectible through imaging (Subculture 2: Figure 5.11C & Sub-culture 13Figure 5.12C). Interestingly, clones detected in sub-culture 3 (O clone: Red and SV clone: Green) showed different phenotypes at 14 days culture with SV cell clones more elongated and spread throughout the Matrigel matrix and O cell clones growing more in clumps (Figure 5.12C). Further difference between the clones at 30 days was also noted with O clone growing externally and SV clone localised internally (Figure 5.12C). Flow cytometry quantification for relative clonal cell counts are shown in Figure 5.11D for sub-culture 2 and Figure 5.12D for sub-culture 3. The final metric shown is for the surface marker phenotypes reported for each detected clone after organoid culture. Clones retained in organoids grown from sub-culture 2 shared a similar distribution of marker profiles at P5 when seeded into organoids (Figure 5.11E). However, upon transplantation and growth in organoids, these two clones show divergent adaptation to their marker expression that remained consistent across organoids (Figure 5.11E).

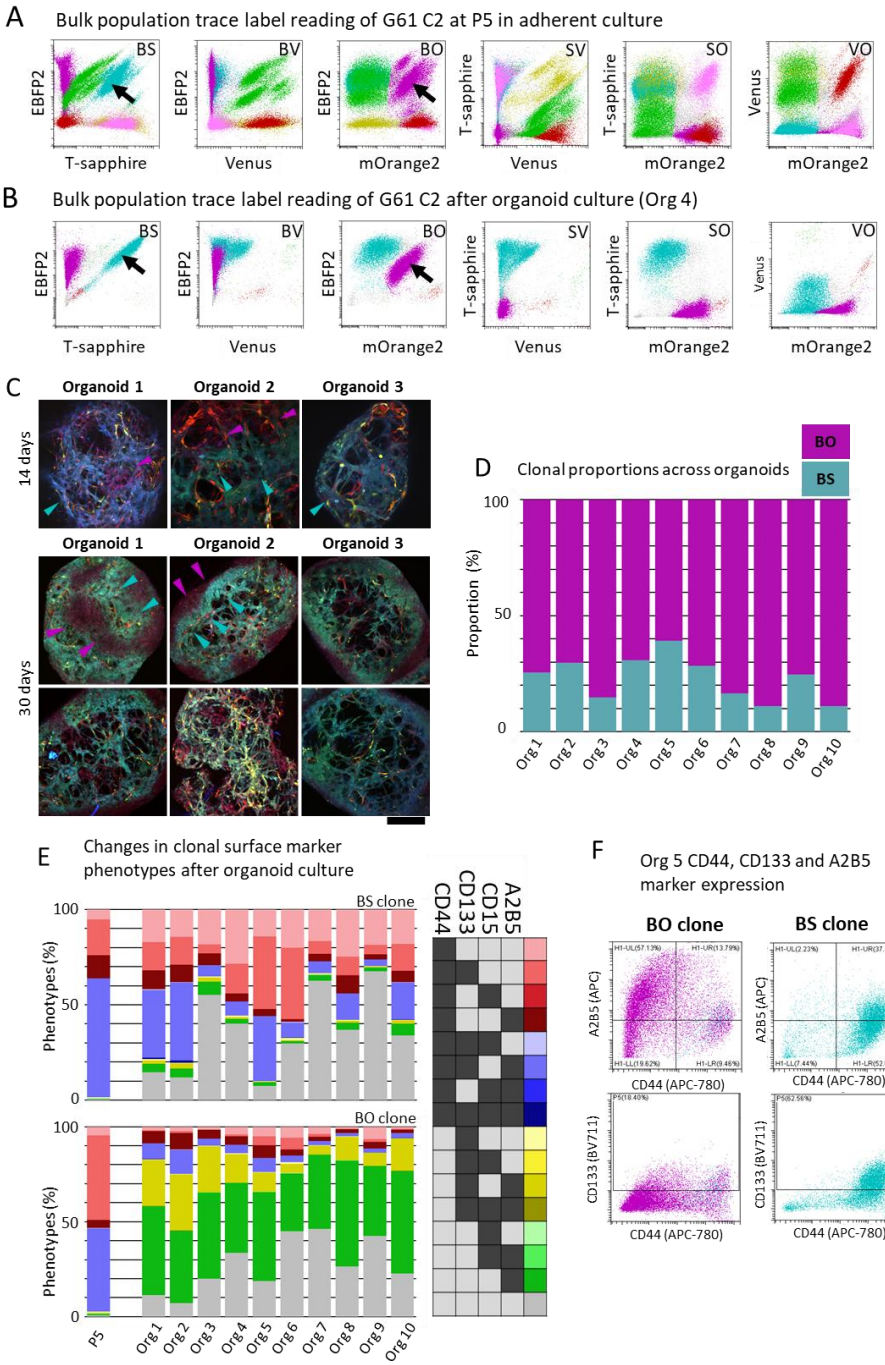


Figure 5.11: G61 (sub-culture 2) clones show substantial clonal outgrowth and divergent marker adaptation upon change of culture environment. (A) Flow cytometry dotplots of G61 sub-culture 2 at P5 before seeding cells into organoid culture. (B) Representative flow plots of G61 sub-culture 2 showing the two clones (BO and BS) indicated by arrows that predominated across all organoids (n = 10). (C) Representative imaging of organoids after 14 days of culture when expansion of both clones was already visible (purple arrows = BO clone, cyan arrows = (BS clone). And, imaging of organoids at maturity (30 days culture) suggesting BO (purple) clone is more readily found peripherally and BS clone (cyan)more readily found internally. (D) Quantification of clonal proportions in each organoids (n = 10) assessed through flow cytometry. (E) quantification of marker profiles for each clone (BS and BO) across all cultured organoids, colours used are consistent with previous displayes of marker expression. (F) Raw flowcytometry dot plots showing representative surface marker (CD44, CD133 and A2B5) expression of BO (purple) and BS (cyan) clones after being harvested from organoids.

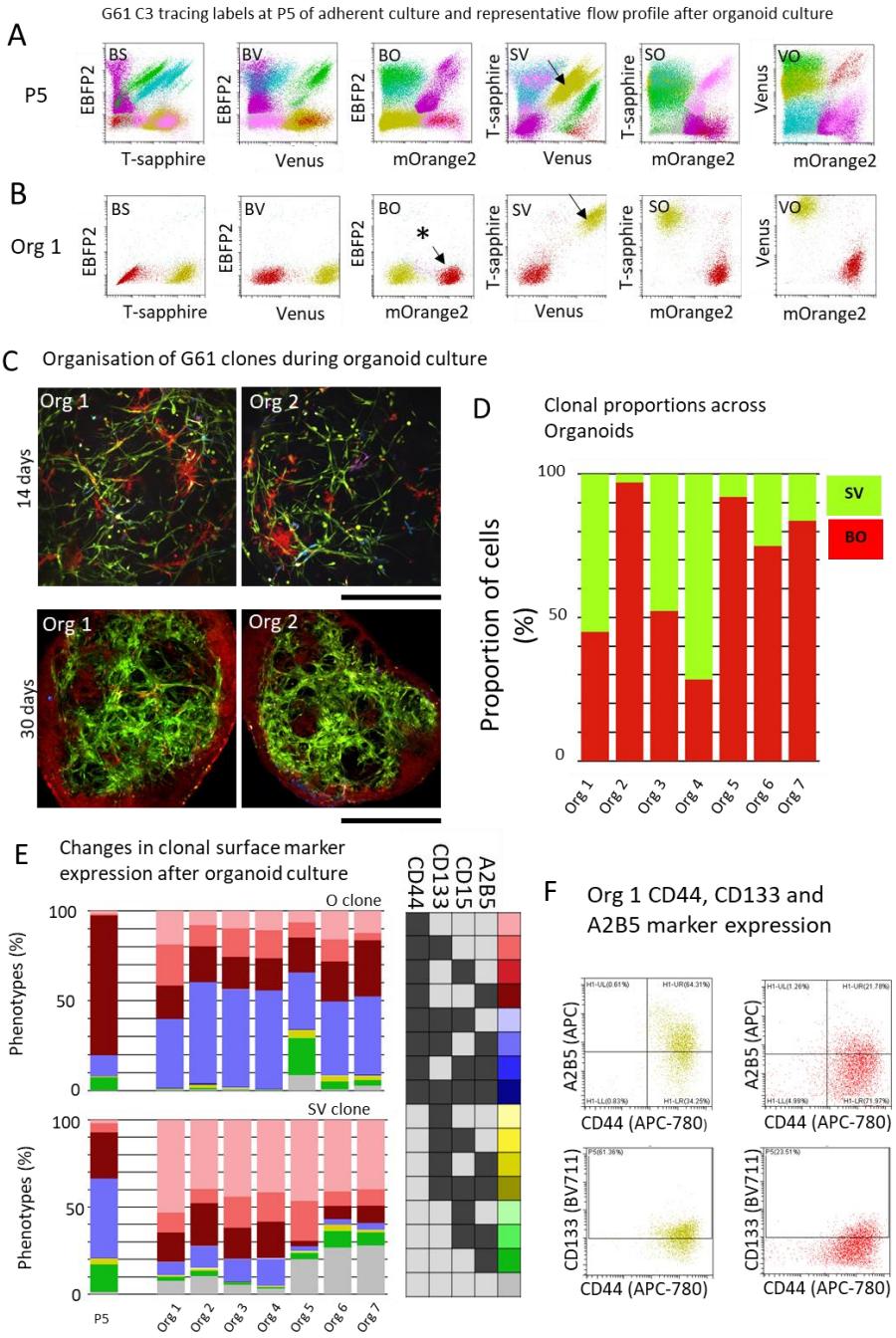


Figure 5.12: G61 (sub-culture 3) shows similar clonal outgrowth and differential adaptations to marker expression after change of culture enbironment: (A) Flow cytometry dotplots of G61 sub-culture 3 at P5 before seeding cells into organoid culture. (B) Representative flow plots of G61 sub-culture 3 showing the two clones (O and SV) that predominated across all organoids (n = 7). Arrows indicate clones predominating after organoid culture (SV and BO clones) and asterisk denotes a predominating clone that is seemingly expressing a single (mOrange2) rather than a dual lentiviral label. (C) Representative imaging of organoids at day 14 and at maturity (Day 30). Day 14 imaging suggests clone shave different growth and migratory phenotypes. Day 30 imaging shows O clone mainly localised to the periphery and SV clone localised internally. (D) Clonal proportions of G61 (sub-culture 3) adjudged by flow cytometry, showing O clone generally more prominant. (E) Quantification of marker profiles for each clone (O and SV) across all cultured organoids and at P5, marker combinations are indicated in key to right. (F) Raw flow cytometry dot-plots showing representative surface marker (CD44, CD133 and A2B5) expression of O (red) and SV (yellow) clones after being harvested from organoids.

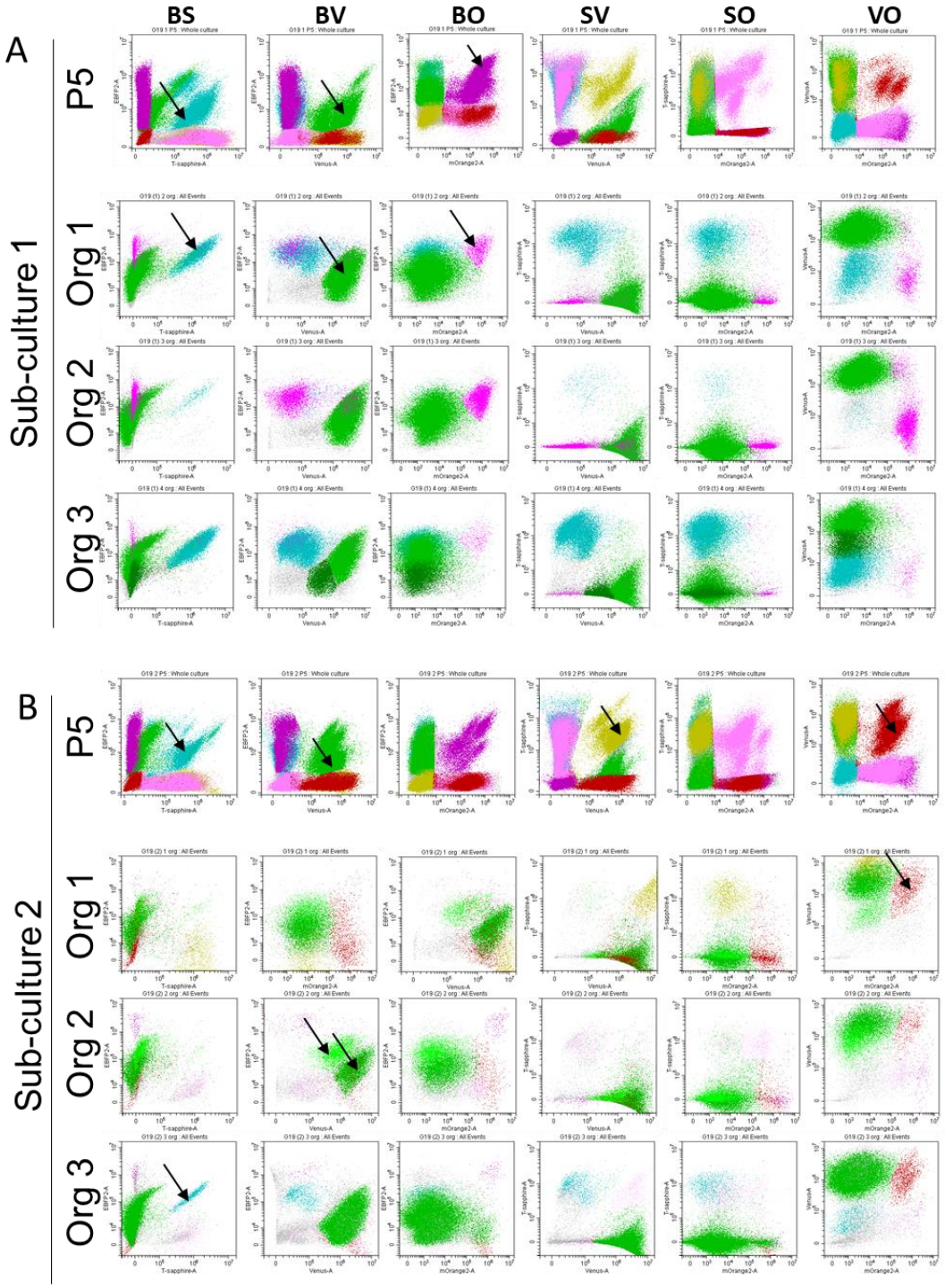
across organoids (Figure 5.11E). These differences are exemplified in representative flow plots shown in Figure 5.11F, where the clones differ in their expression of A2B5, CD44 and CD133. Similarly, clonal divergence in surface marker adaptation to culture in organoids is also observed for clones of G61 sub-culture 3 (Figure 5.12E). Clones from sub-culture 3 mainly differed in their expression of A2B5 (Figure 5.12F)

In contrast, G19 organoids grown from sub-cultured clonal mixtures did not report the same degree of clonal pruning or unique phenotypic behaviours. In general, a larger number of clones persisted in the organoids cultured from G19 which may just reflect the presence of more clones across G19 sub-cultures than in G61. Only a single organoid derived from sub-culture 2 contained as few as 2 clones, while all other organoids contained 3 or more. Again, unlike G61 which had apparently unique clones (judged by variable trace labelling) comprising their sub-cultures and organoids, G19 organoids all shared the same predominant BV clones indicated by arrows in Figure 5.13A, B and C; This observation in the flow cytometry data was corroborated by imaging which showed organoids derived from each sub-culture sharing a predominantly yellow signal. Owing to its position on the flow plot, this recurrent BV clone could be regarded as high expressor of Venus and low expressor of EBFP2 which would present as largely yellow (with minimal blue EBFP2 signal) by microscopy.

In G61, we observed apparent regional localisation of different clones centrally and at the surface (Figure 5.12E). This pattern of distribution was not observed for G19 where cells of individual clones were more widely spread through the organoid. A result mirroring the difference in spatial distribution of G61 and G19 clones when establishing the 500-cell derived sub-cultures (Figure 5.3A & B). In G61, we observed relatively consistent marker expression when comparing the same clone across multiple organoids. However, in G19 clonal marker expression after organoid culture did not show the same consistency. For example, the VO clone identified across all six organoids seeded from sub-culture 3 (Figure 5.14B) shows quite distinct surface marker expression in each organoid. In fact, the VO clones marker expression appears more similar to the clones it was cultured with in each organoid. These inferences are purely a qualitative assessment and would require the experiment to

be repeated on a larger scale to confirm. Overall, G19 and G61 behaviours in organoids echo the differences observed across passages in adherent culture.

In conclusion, changing the growth environment induces a selection pressure on the surface markers expressed by our two patient cell lines as previous work has indicated. However, our dual-barcoding also revealed considerable changes to the underlying clonal composition of the mixtures seeded into the organoids. Furthermore, as in adherent culture, we observed clones with divergent marker expression in both cell lines suggesting clones of the same tumour line can show variable responses in their marker expression after a change in culture conditions.



***Sub-culture 3 0rgnaoids shown on following page

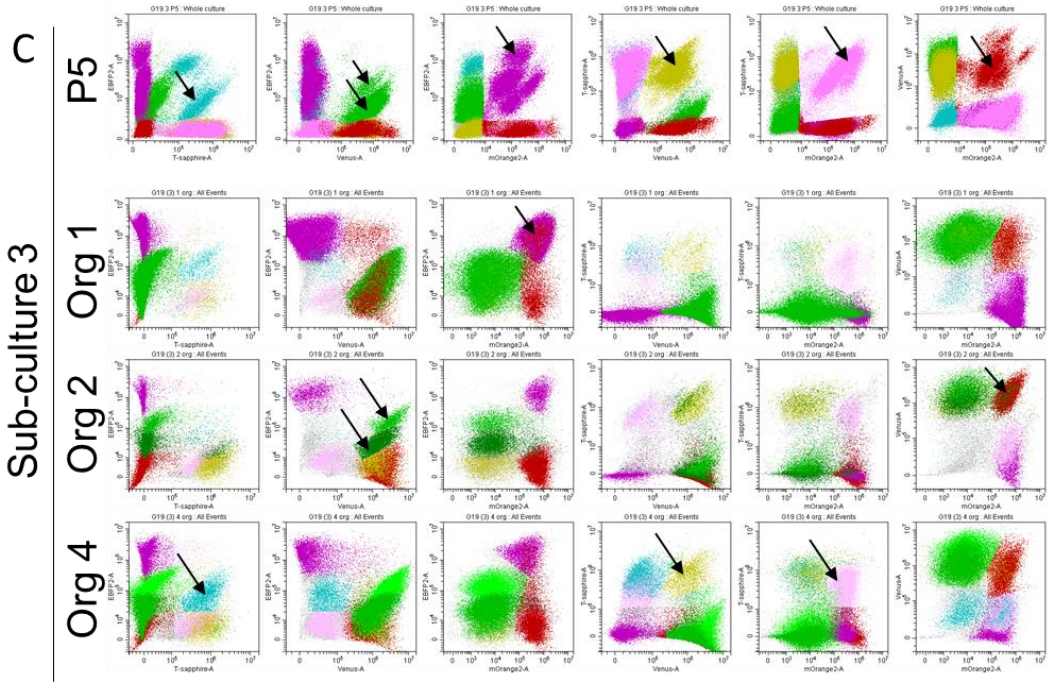


Figure 5.13: Organoids grown from G19 sub-cultures (1 – 3) all show outgrowth of the same clone. (A) Flow cytometry dot plots of (n = 3) organoids grown from G19 sub-culture 1 cells at P5. Columns are ordered as previously displayed (BS, BV, BO, SV, SO and VO). (B) Flow cytometry dot plots of (n = 3) organoids grown from G19 sub-culture 2 cells at P5. (C) Flow cytometry dot plots of (n = 3) organoids grown form G19 sub-culutre 3 cells at P5. All orgnaoids grown from all sub-cultures show a consistent outgrowth of a BV (Green) clone.

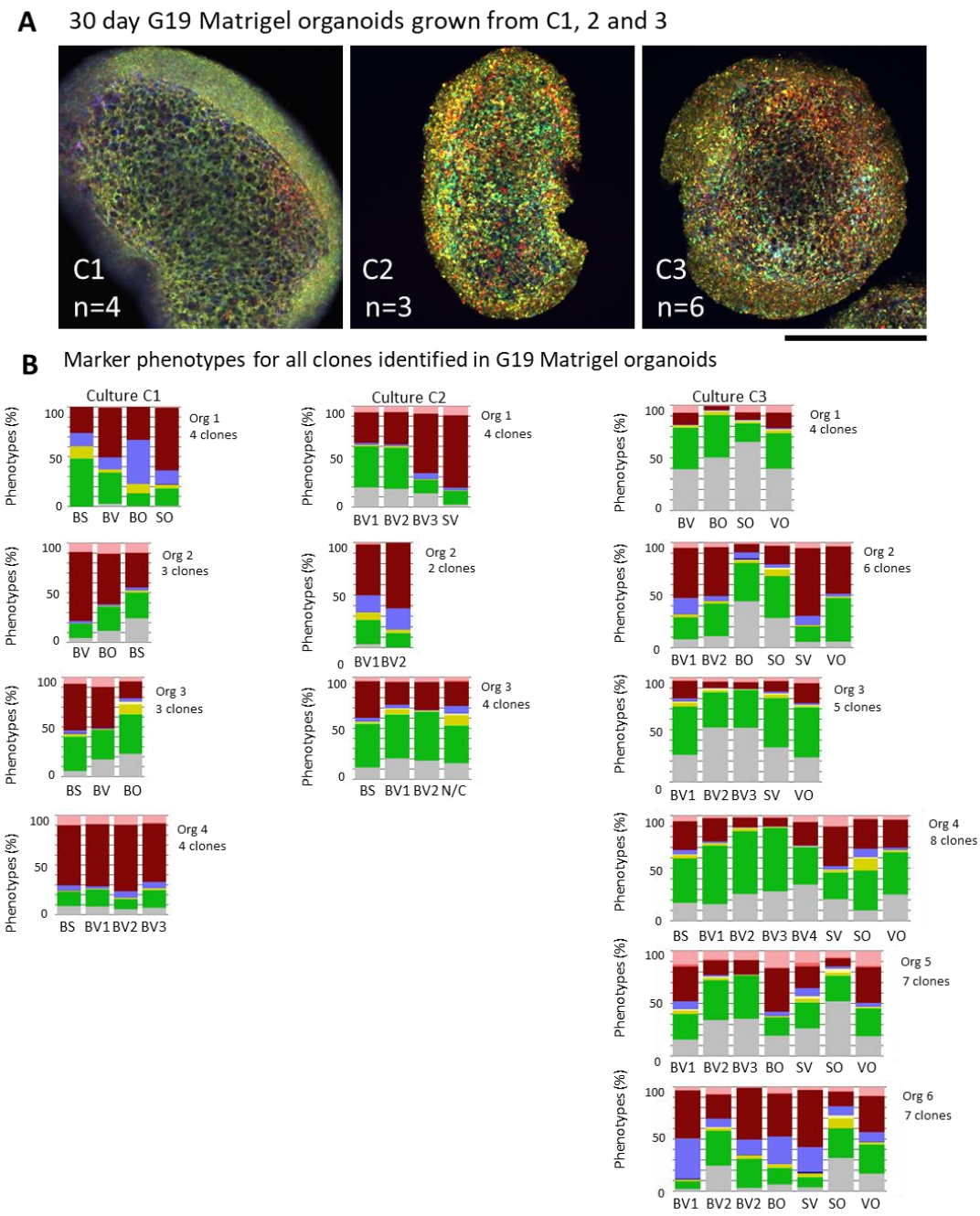


Figure 5.14: G19 Matrigel organoids retain more clones than G61 and also show divergent marker profile. (A) Representative confocal imaging of all the G19 sub-cultures after organoid culture (yellow appearance due to consistent BV clonal growth). (B) Summary of marker profiles for all identified clones in cultured G19 organoids, using same colour scheme as all previous dislays. A singel graph for each organoid is shown with a summary of the marker phenoypes of the clone spresent within that orgnaoids. Clone names are dictated by colour group in which they arrose (BS, BV, BO, SV, SO and VO).

5.2.5 Seeding of diversly labelled cells into 3D Matrigel organoids reports consistent observations as experiments using clonally enriched sub-clones

To further compare differences in clonal and marker expression phenotypes between G61 and G19 we seeded 5000 bulk diversely labelled cells of each cell line in a number of 3D Matrigel cultures. (G19 n = 21 & G61 n = 15). Organoids were cultured until maturity at around 30 – 40 days, vibratome sectioned and imaged without fixation then subsequently dissociated with Accumax® and stained for assessment of clonal content and surface marker expression. Imaging of clonal organisation within the organoids showed consistent observations in clonal organisation compared with sub-culture seeded organoids. G19 clones showed a greater degree of mixing throughout the culture matrix compared with G61 where clonal populations arose in spatially discrete clusters. Higher power images show that clones do mix in G61 (Figure 5.15B1, B2 & B3) but the degree of spatial clonal mixing in G19 is much greater (Figure 5.15A1, A2 & A3). Internally, both cell lines form an interconnected network of cells but with the clonal arrangement in this network strikingly different. G19 clones appear more diffusely spread through the Matrigel matrix while the G61 clones are more segregated in clusters throughout the matrix. Diffuse (G19) and clustered (G61) clonal organisation was also observed after culturing from 500 or 1000 cells in adherent culture. Flow cytometry analysis revealed further behaviours reflecting characteristics observed in adherent culture. For both lines, dual-barcode groups were seeded into organoids at near even proportions with a large variation in labels represented within each dual-barcode (Figure 5.16A & B). Quantification of colour groups across G61 organoids revealed variable shifts in label composition after organoid culture (Figure 5.16A). Furthermore, streaks present on flow plots suggest this expansion of certain label groups was due to predominance of a small number of clones in each organoid (Supplementary Figure 7.8 & Figure 7.9). Conversely, G19 organoids barcode label group composition was more conserved across different organoids with a lower level of proportional rearrangement. In most organoids, the BV label groups had expanded considerably with fluctuations observed across the other label groups (Figure 5.16B&D). Indeed, the number of clones of sufficient size for detection was much greater in G19 compared with G61 (Figure 5.16).

To investigate marker expression all clones were manually gated, and their surface marker expression extracted. A breakdown of marker expression for all organoids and all clones can be found in supplemental data (Figure 7.10). We subject this dataset to the same analysis performed on adherent sub-cultures looking in the first instance at similarity in marker expression between whole organoids and then between clones of the same organoids.

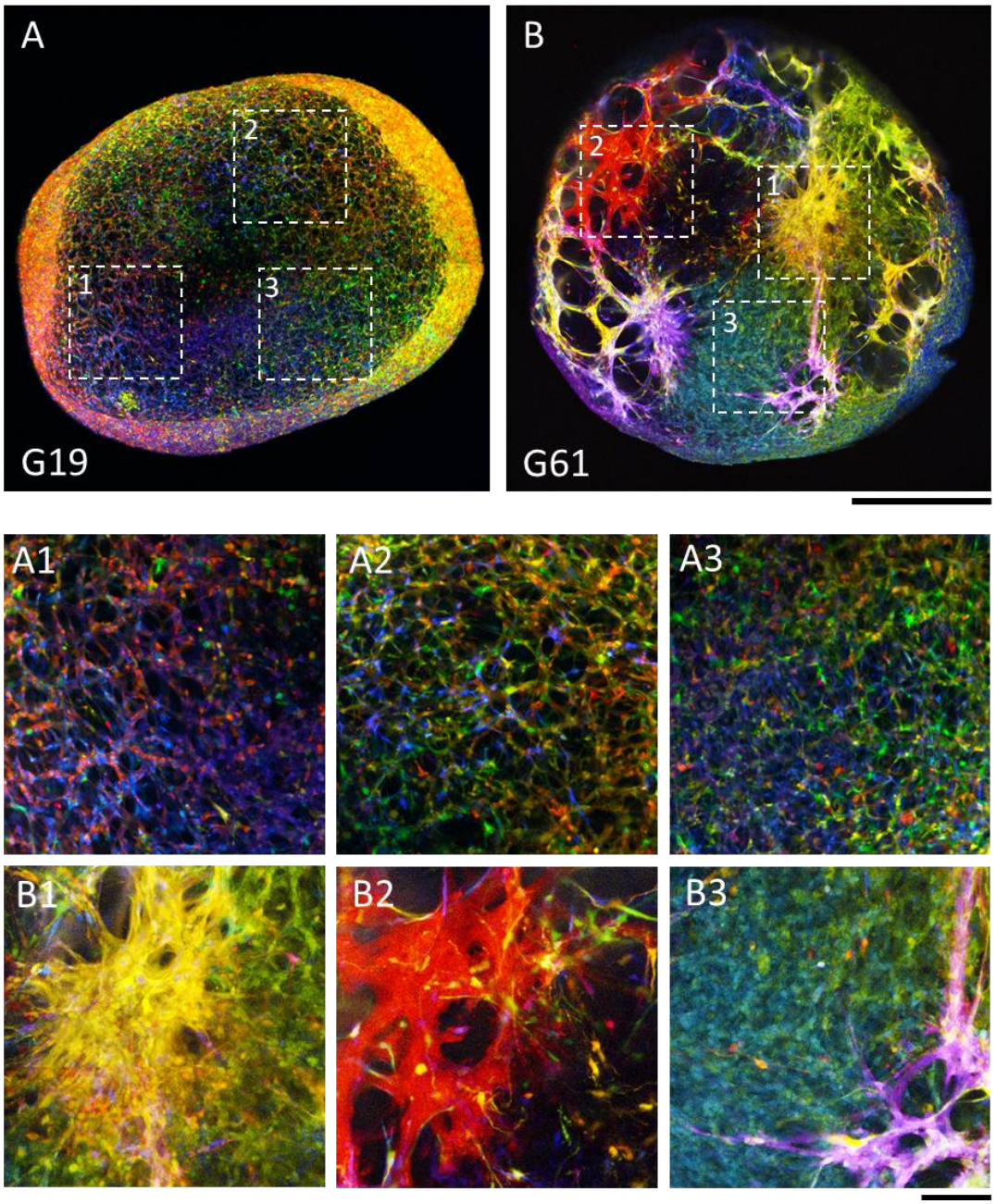


Figure 5.15: Diversely labelled cells seeded into orgnaoids deomnstrate phenotypic differences between G19 and G61. (A) Representative imaging of a living vibratome sectioned G19 orgnaoids seeded with diversely labelled (not sub-cultured) cells. Boxes indicate regions shown in higher power in A1 – A3. (B) Representative imaging of G61 organoid seeded with diversely labelled (not sub-cultured) cells. Boxes indicate regions shown in higher power in B1 – B3.

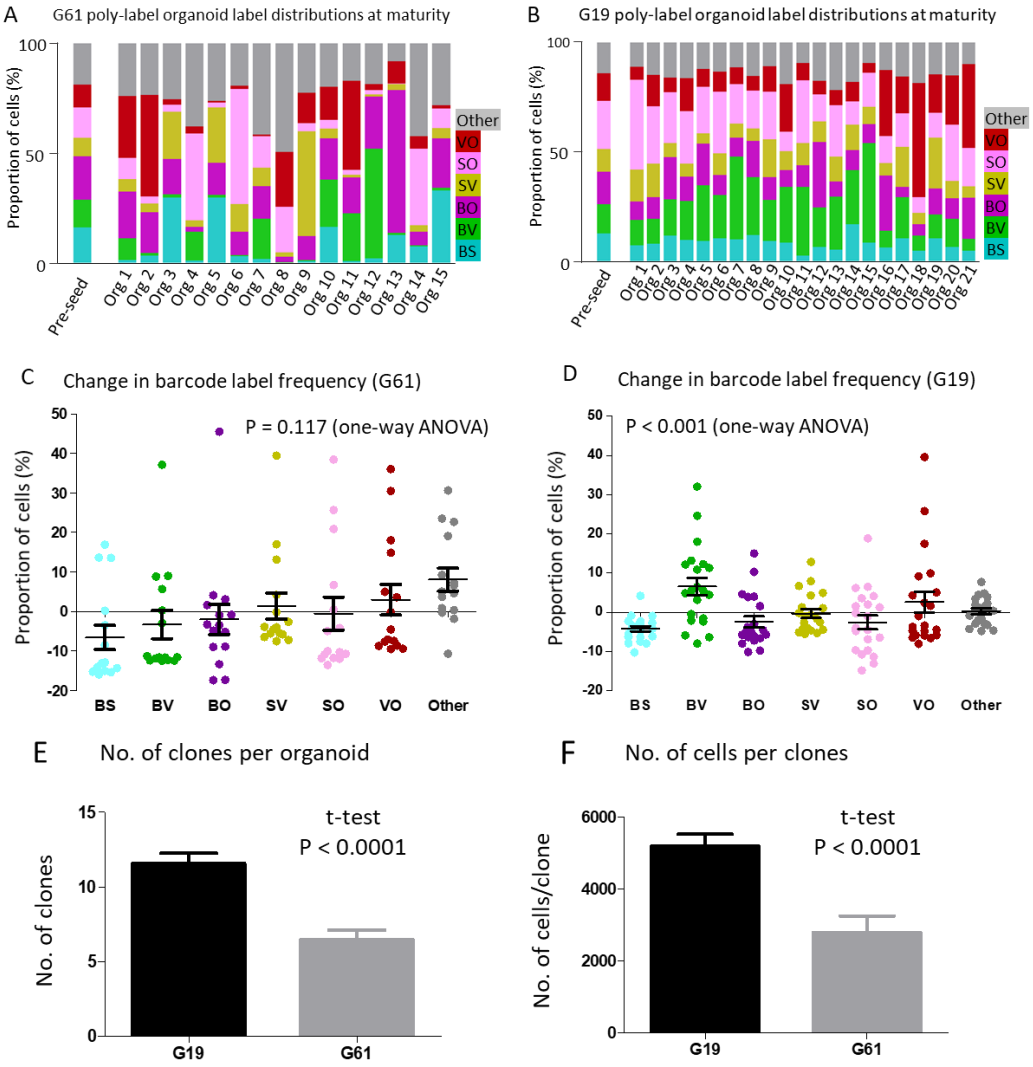


Figure 5.16: Trace-label quantification and identified clones support previous findings of differeing clonal dynamics between G61 and G19. (A) Raw flow cytometry dot plots of the 6-dual labelled G61 populations prior to seeding into organoid culture. (B) Quantification of colour group proportions across G61 or (n = 15) after 30 – 40 days of culture. (C)) Raw flow cytometry dot plots of the 6-dual labelled G19 populations prior to seeding into organoid culture. (D) Quantification of colour group proportions across G19 organoids (n = 21) after 30 – 40 days of culture. (E) Summary plot of the number of clones found per organoid between G61 and G19. (F) Summary plot displaying the number of cells per identified clone for G61 and G19.

Comparison of all cells in an organoid showed most organoids shared a similar proportion of surface marker profiles with the majority of pairwise comparisons producing a cosine similarity score of > 0.9 for both cell lines Figure 5.17A. Visual inspection of data in supplemental figures would further support this with organoids seeded from the same cell line showing a characteristic pattern of surface marker

expression (Figure 7.8 & Figure 7.9). Upon comparison of clonal surface marker expression, we find a larger degree of variability than was observed between whole organoids (Figure 5.17B). For both cell lines the majority of clones share a similar expression signature with pairwise comparisons yielding a similarity index of >0.8. However, for both cell lines there are a number of outlier clones which show divergent marker profiles (<0.6 cosine similarity), similar to what was observed in adherent culture (Figure 5.17. In adherent culture we found a different distribution between the cell lines with G61 reporting a bimodal distribution of clonal similarities and G19 reporting a more continuous distribution of highly similar to dissimilar expression profiles. In this analysis, both cell lines appear to show a continuous distribution of clonal marker expression similarities. However, G19 report a higher proportion of similar clones (>0.8 cosine similarity) and G61 reports a higher proportion of dissimilar clones (<0.6 cosine similarity).

In conclusion, our dual-label strategy is very well suited toward the identification and tracking of clones in a mixed population of human tumour cells. Moreover, this approach can be combined effectively with surface marker labelling to investigate marker expression plasticity and heterogeneity in a range culture settings. We find clonal populations of human tumour cells exist across a range of marker expression profiles, with the bulk of clones sharing a similar marker expression pattern and a number of clones reporting divergent, dissimilar marker expression distributions. These results further suggest, that in addition to environmental influences cellular heritage, plays a role in shaping cell identity and hierarchical tumour processes. Indeed, we find that bulk population plasticity in marker expression is underpinned by re-arrangements to the underlying clonal composition, supporting a role for clonal outgrowth and predominance in processes of tumour plasticity and adaptation.

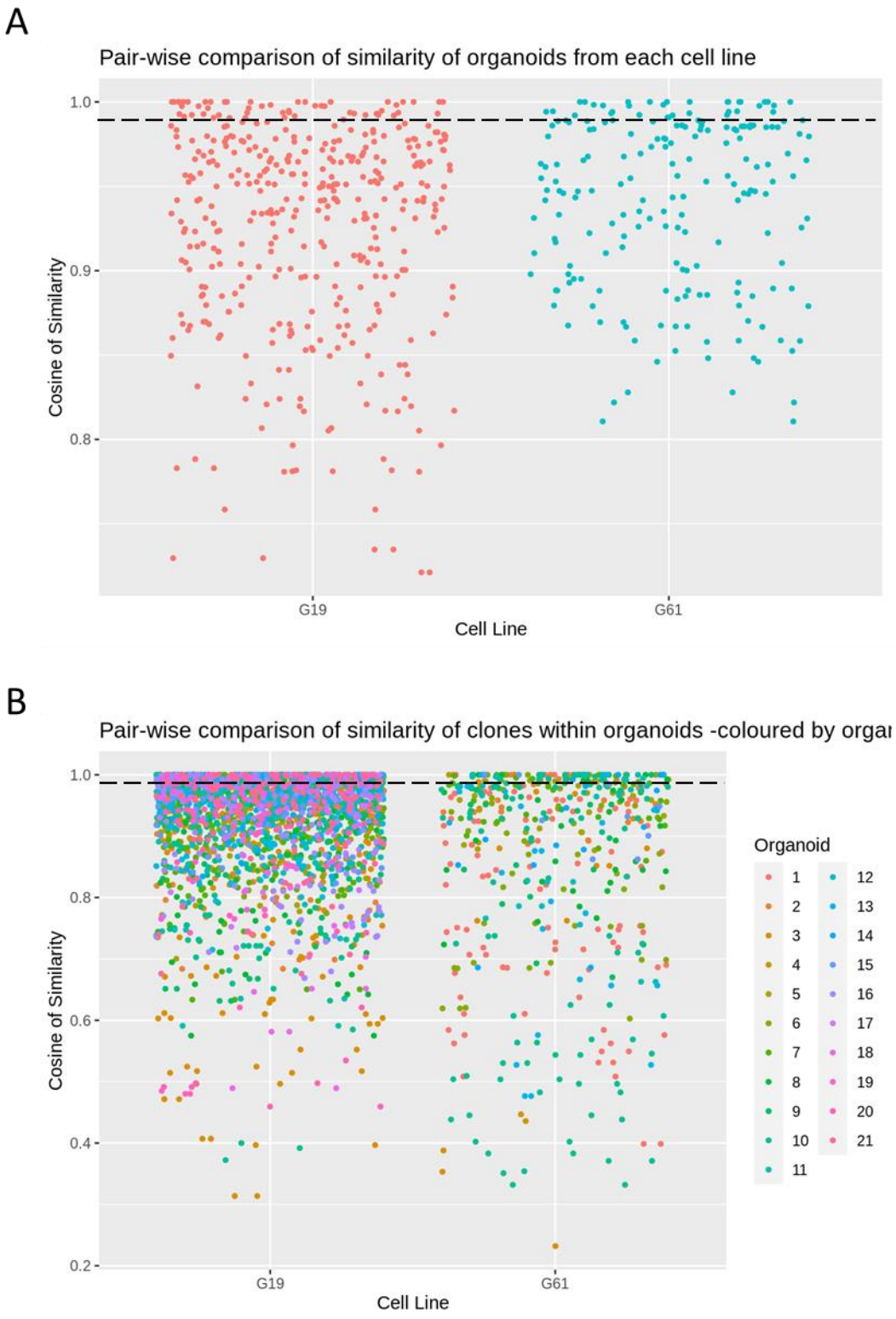


Figure 5.17: Inter orgnaoid and inter-clonal variation in suface marker expression profiles for diversely labelled G61 and G19 organoids. (A) Pair-wise cosine similarity comparisons of surface marker profiles for all G19 and G61 organoids, each point represents a cosine similarity calculation comparing two organoids. (B) Pair-wise cosine similarity comparisons of surface marker profiles between clones found in the same orgnaoid for both G19 and G61. Dots are coloured by orgnaoid with colour scheme shown on left of plot.

5.3 Discussion

Clonal evolution and plasticity of glioma CSCs giving rise to intra-tumour heterogeneity are key themes in GBM progression, treatment resistance and disease recurrence. Leveraging the formation of clonal populations, visualised as “clouds” or “streaks” on flow plots, we adapted the LeGO-vector based labelling method for rapid production of many traceable cell clones. In previous studies using cell labelling with LeGO-vectors, clones were produced by single-cell sorting and require parallel tissue culture maintenance (Mohme et al., 2017, Pericoli et al., 2020) or emerged after long-term serial passaging of RGB marked cells, for up to 36 passages (Brenière-Letuffe et al., 2018). In our approach, low density seeding of approximately 500 purified dual labelled cells consistently produced clonal populations after just two weeks of expansion with 13 – 16 (line G61) and 14 – 20 (Line G19) traceable clones in each sub-culture. Originally the barcoding procedure was described as single-cell sorting, expansion and maintenance of 21 separate clones from the murine glioma line GL261 (Mohme et al., 2017). This method provided a robust approach for investigating clonal behaviour but was limited in terms of its scalability. Single-cell sorting to derive clones can take more than two months to expand cells to sufficiently large numbers for downstream assays. Furthermore, repeating the process of sorting, expanding and maintaining 21 clones across multiple primary cell lines would have represented a technically demanding, protracted and financially challenging procedure. Instead, through parallel maintenance of just six 500-cell sub-cultures per cell line, we were able to identify 104 and 84 clones for G19 and G61 respectively. Thus, our modified approach represents a technically less demanding protocol which can produce more traceable clones and can be scaled more readily to simultaneously assess multiple primary cell lines. As such, our approach is better suited to capture patient specific differences in clonal behaviour.

Using our modified labelling protocol, we were able to simultaneously monitor clonal growth and glioma CSC surface marker expression. This is to our knowledge the first experimental approach to show how clonal dynamics underpin the recently characterised plasticity of human glioma cell surface marker expression (Brown et

al., 2017, Dirkse et al., 2019b). Not unexpectedly, and in keeping with previous studies (Dirkse et al., 2019a, Lan et al., 2017, Suva et al., 2014), there was considerable phenotypic diversity between the two cell lines (Figure 5.2D) but interestingly we also observed diversity between serially traced clonal populations of the same cell line (Figure 5.9 and Figure 5.10). This was observed for clones which shared the same culture well where we can speculate each clone was subject to broadly similar environmental pressures. This suggests that in addition to environmental factors and intrinsic plasticity, glioma stem cell surface marker expression may also be linked to clonal heritage. However, since clones were grown from cells expressing a range of combinations of surface marker phenotypes and not from purified phenotypic populations, it is also possible that the divergence between clones is a product of surface marker evolution when clones are derived from cells with different starting marker phenotypes. One approach to address this hypothesis would be to first establish a labelled clonal sub-culture and then sort this population into further sub-cultures of purified marker phenotypes. In this case, we would have multiple sub-cultures with same clonal composition, but each sub-culture is seeded with cells of a different marker phenotypes. Upon serial passaging, if divergence of clonal marker expression is again observed when all cells start with the same phenotype (irrespective of clonal identity), it would further suggest clonality and a cells heritage are factors shaping marker heterogeneity in glioma. This is an important consideration as it is possible the plasticity of marker expression found in previous studies (Brown et al., 2017, Dirkse et al., 2019b) was a product of underlying clonal outgrowth within the population rather than a plastic property of the population as a whole (Brown et al., 2017), or heightened plasticity assigned to certain marker phenotypes (Dirkse et al., 2019b).

Strikingly, the clonal surface marker diversity presented with patterns that appeared to be characteristic in each of the two cell lines, G19 and G61. G19 showed a continuum of divergent clonal phenotypic distributions while G61 showed a small number of highly divergent clones (Figure 5.9). This difference was further highlighted when clonal growth and marker plasticity were analysed (Figure 5.10). Again, G61 showed a small portion of outlier clones in terms of growth (fold change

in clonal fraction) and marker plasticity (fold change in mean CoS) while the majority of clones clustered quite closely together. The fold change in clonal fraction of the G19 line suggests a continuum of clonal growth rates while there was a skew of distribution toward a positive fold change in mean CoS for surface marker expression from P2 – P5. This suggests that marker phenotypes of G19 clones within sub-cultures were becoming more similar over passages. This finding appeared to be replicated also in whole sub-cultures (i.e. ES3 and ES5 in Figure 5.4) where sub-culture surface marker profiles were diverse at P2 and became more similar by P5. These trends highlighted the importance of using multiple patient lines to assess inter-tumour heterogeneity and indeed the potential of our technique to capture these differences. Nevertheless, refining and performing this experiment on a larger scale would be required to confirm and characterise these patient specific observations. One possibility would be to establish if patient lines harbouring similar copy number profiles and molecular profiles such as methylation class share any characteristics in clonal dynamics.

For both cell lines, transfer of sub-cultured clonal mixtures into organoid culture resulted in an attrition of clonal diversity as many clones present in the P5 population seeded into organoids (ES6 in Figure 5.4) were barely detectable after 30 days culture. This suggests that clones were subject to a more stringent selection pressure after a change of environment. However, whether organoid culture selected for clones with different properties or merely accentuated the clonal outgrowth already ongoing within each sub-culture was unclear. Flow results suggested that for G19 the most predominant clone from sub-cultures 1, 2 and 3 at P5 (Figure 5.6; Asterisk denoting BV clone), was also the most predominant population found after organoid culture (Figure 5.13), suggesting accentuation of already ongoing clonal predominance. Similarly, for G61, successfully cultivated organoids from sub-cultures 2 and 3 showed the persistence of clonal populations predominating at P5. However, in both instances, while the most predominant clone at P5 (sub-culture 2; BS clone & sub-culture 3; SV clone) persisted after organoid growth, it was no-longer the predominant population. This finding could be of relevance to multiple applications on the CSC field – it suggests that a certain clone has a growth advantage in adherent

culture and another clone grows better in three-dimensional Matrigel culture. However, as was put forward in a previous study of glioma CSC surface marker phenotyping (Dirkse et al., 2019b), propensity for plasticity maybe a possible driver of growth when cells are placed in a novel environment. Therefore, one possibility is that clones predominating after organoid culture were poised in states that facilitated a rapid transition into states effective for growth in Matrigel. Interestingly, the G61 clones predominating after organoid culture (Sub-culture 2; BO clone & sub-culture 3; O clone) show considerably distinct surface marker profiles (Figure 5.11). This observation is in contrast to a previous study (Dirkse et al., 2019), which suggest that environmental pressures drive cell lines to express a distinct pattern of surface markers. If this was the case, we would have expected dominant clones across organoids to share similar patterns of marker expression, which we did not find. However, it is possible that marker adaptation was still in progress and that clones may eventually have converged to the same marker phenotypes, when cultured for longer. Nevertheless, the contribution of clonality to marker plasticity was not addressed in this previous work by Dirkse at al. (Dirkse et al., 2019b), and given the considerable variability in marker expression observed between clones in our experiments, it is possible that clonal outgrowth was driving marker plasticity in previous studies rather than an intrinsic propensity for plasticity of all cells in the culture.

In our final experiment (ES7, Figure 5.4), we seeded the bulk labelled G61 or G19 cells into three-dimensional Matrigel culture for clonal tracking (flow cytometry and Imaging) and surface marker assessment. Imaging experiments reinforced the distinct properties of the cell lines with patient-specific organoids showing markedly different structural histology and clonal organisation. Assessment of trace label distributions (Figure 5.15B & D) and apparent streaks (Figure 7.8 & Figure 7.9) suggested clonal outgrowth was more prominent in G61 than in G19, where G61 showed fewer remaining clonal streaks when organoids were harvested, in keeping with observations in adherent sub-cultures. Imaging of G61 cultured organoids suggested there were many clones remaining that were not detected at sufficient quantities for analysis during flow cytometry. The majority of cells were localised to

the organoid surface and in G61 the surface regions were predominated by 2 or 3 different colours which broadly matched up with the flow cytometry streaks. It is possible these large clones masked detection of less prevalent clones within the organoids core. It was also further noted that for both lines (G61 and G19), clonal predominance appears to occur more rapidly at the organoid periphery with a more diverse repository of clones retained internally. This was also found during early experiments with RGB marked G61 cells (Figure 4.11). It can be hypothesised that the differing environments of the core of a sphere put different pressures on clonal development. For example, that cell division occurs at higher rates at the organoid surface, where cells have access to growth factors, oxygen, and nutrients, whilst cells at the core are deprived of these factors. After vibratome sectioning of live organoids it would be possible to dissect sections to separate the internal and external regions for further analysis. Whilst our study explores important aspects of relationships between clonal populations and stem cell phenotypes, there are limitations which will require further exploration.

Firstly, our study is limited to only two cell lines with variable diagnostic measures (i.e genetic profiles and methylation sub-class). It would be desirable to compare additional GBM cell lines with shared diagnostic criteria. Furthermore, performing this over a greater number of passages and in further defined micro-environments, would better elucidate the link between clonality and surface marker plasticity in primary GBM lines. We mitigated experimental variables by performing the entire series of experiments strictly in parallel using consistent culture conditions, but undoubtedly additional, massive parallel cultures will yield a more comprehensive dataset for computational analysis.

A further limitation to our study is the propagation of glioma CSC in adherent and organoid cultures that lack the complex environmental cues seen *in vivo.* Namely, stromal cell, endothelial cell, macrophages and other immune cells which could further modulate and influence surface marker expression. Whilst these extrinsic modifiers have been omitted, our model however provides in the first instance a picture how even comparatively well-controlled culture conditions can modulate marker expression of glioma CSCs. We also observed a potential limitation of our

labelling approach where single-label cells can dominate a culture (Figure 5.12) while triple labelled cells appeared to have a growth disadvantage. In this series of experiments, this imbalance was mitigated by eliminating single label cells and performing viral transduction such that no triple label cells are produced.

In conclusion, we present here an experimental approach to barcode glioma CSCs and create clonal populations, which can be phenotyped simultaneously. Through computational analysis we could pinpoint the fate of such populations, which can be used to interrogate the phenotypic plasticity in response to a change in culture environment. The simplification of the workflow of clonal selection, combined with reproducible and robust readouts to assess the functional properties of glioma CSCs renders this assay potentially very suitable for screening of newly established glioma CSCs for tumour-specific therapeutic vulnerabilities and to assess the impact of experimental genetic or epigenetic modifications on clonal development.

Chapter 6: General Discussion

6.1 Summary

This study examines lineage tracing, clonal dynamics, and evolution in gliomas, using several technically and conceptually distinct approaches. Initially, tumours expressing multiple, Cre-inducible fluorescent colours were successfully produced by combining a genetic model with constitutive expression of inducible fluorescent reporters (ROSA26-confetti labelling) and conditional inactivation of tumour suppressor genes (Lei et al., 2011, Snippert et al., 2010). In this model, tumours were either composed almost entirely of cells with the same label, suggesting a single clonal origin, or otherwise mixtures of labels suggesting multi-clonal origin. Furthermore, clonal composition showed regional variation, potentially mirroring findings in serial spatial biopsies of human tumours, with multiple distinct clonal populations (Sottoriva et al., 2013, Yu et al., 2020). However, the genetic model to induce tracer labels suggested that there was only limited value to provide for mechanistic insights in clonal behaviour of tumour cell populations. Specifically, this tumour model combined with confetti labelling used constitutive cre expression through retroviral integration, which resulted in permanent cre expression and thus continued flipping of the Lox cassettes flanking the fluorescent expression cassettes, resulting in repeated changes of expressed colours. Nevertheless, we successfully use these labelled murine tumours to develop three-dimensional in vitro models, flow cytometry approaches, live and three-dimensional imaging modalities for a readout of fluorescently traced tumours. These models also enabled us to modify our model systems to overcome the persistent expression cassette flipping and cell label changes.

Subsequently, we delivered fluorophores with lentiviral vectors to achieve more diverse labelling in primary human GBM lines. Iterative adaptations of this labelling method had been reported previously, and we adopted this protocol into a simple, dual-label barcode approach to derive and detect multiple clonal lineages. These clonal populations were simultaneously characterised with a panel of CSCs surface markers, CD133, CD44, A2B5, and CD15. Recent single cell (sc)RNA-sequencing studies suggest that GBM cells have a considerable propensity for plasticity in state

transitions, a view supported by recent work describing environment induced adaptations to glioma CSC marker expression (Couturier et al., 2020, Yu et al., 2020, Neftel et al., 2019) (Dirkse et al., 2019b, Scott et al., 2019). However, the relationship between clonal populations and the associated expression of combinations of CSC surface markers, i.e. interrogating plasticity of marker expression, has not been thoroughly addressed. In keeping with previous findings, we observe patient specific patterns of glioma CSC surface marker expression but also considerable variation in marker expression between co-cultured clones of the same patient-line. Through Monte Carlo-modelling, we demonstrate this variability could not be observed through the effect of random sampling, suggesting the differences are truly due to the clonal origin of different populations. Furthermore, we find that population marker plasticity in response to culture in a changed environment (laminin adherent culture to Matrigel® three-dimensional culture) is underpinned by outgrowth and attrition of certain clones. this manifested as marker adaptation in a new environment appearing notably distinct between clones within line G61 while the relationship between clonality and plasticity in G19 appeared less well defined. This highlights the efficacy of different patient lines in capturing tumour dependent behaviours and the potential of our assay to rapidly characterise clonal and properties as well as phenotypes of patient lines. Beyond addressing basic questions associated with glioma CSC marker plasticity our labelling approach has potential to identify tumour specific vulnerabilities, discussed in the subsequent chapter.

6.2 Characterising patient lines and investigating disease specific gene disruptions

To build further on our studies combining clonality and marker expression, a set of future experiments would assess further cell lines, produce additional biological replicas and explore additional surface markers. Modern spectral flow cytometry cell analysers can decipher up to 21 fluorescent signals, highlighting the potential to increase the number of CSC markers investigated. Furthermore, advanced *in vitro* culture methods such as Glioma Cerebral Organoid (GLICO) could be employed to better model specific aspects of microenvironmental factors of GBM. The use of GLICO culture between patient matched iNSCs and glioma CSCs was explored here

but is not presented. We found it difficult to identify and extract sufficient numbers of glioma CSCs after GLICO culture to comprehensively study clonal marker expression dynamics (data not shown). Nevertheless, improving experimental conditions could make use of complex *in vitro* models more feasible. With such improvements, our approach could be used as a rapid assessment to characterise newly established patient lines in terms of their marker expression plasticity and clonal dynamics. Automation of the flow cytometry analysis pipeline which separates clonal populations and isolates their specific marker expression profile could be easily developed and allow assessment of many cell lines, repeat sub-cultures and clones. Through characterisation of cell lines with similar diagnostic profiles it could be possible to better understand how a tumour's underlying molecular profile influences their clonal dynamics and plasticity.

A further possibility, also explored here, examines the effects of gene disruption on marker plasticity and clonality. We developed a pipeline for knockdown of Nuclear Factor I X (NFIX), also known as CCAAT-Binding Transcription Factor, a gene found universally overexpressed across glioma CSC lines after comparison with skin fibroblast derived, patient-match induced neural stem cells (iNSCs) (Vinel et al., 2021). We successfully developed an experimental pipeline for achieving and confirming gene disruption without the need for a fluorescent reporter, as these interfered with downstream labelling and clonal tracking. Preliminary results of these experiments are shown in supplemental Figure 7.12, Figure 7.13 and Figure 7.14 but owing to time constraints we were not able to pursue these experiments further.

6.3 Methodological considerations for future studies

In our study, we found outgrowth of one clone across a number of sub-cultures for line G19 (Figure 5.6) . It is possible this represented an example of clonal evolution specific to G19 but may also be a result of clonal predominance that occurred within the bulk of diversely labelled cell populations during expansion phases before sub-culturing. In any future experiments establishing clonal sub-cultures from dual-barcoded cells, sub-cultures should be seeded immediately after labelling, instead of viral labelling, followed by expansion, FACS enrichment of dual barcoded cells, further expansion and then seeding of sub-cultures. Sub-cultures could be seeded

directly during the initial FACS enrichment step. With this approach, initial accurate titres and equal viral delivery will not be required as FACS could allow seeding of sub-cultures with even starting distributions of fluorescent barcodes. Such an adaptation should reduce the chance of clones sharing the same label growing out across repeat sub-cultures. This is desirable because as demonstrated here, clonal behaviours can be diverse and repeat assessment of the same predominant clone would limit any potential to capture this diversity.

6.4 Clonal heterogeneity in passaged tumour cells

In the form presented here, our clonal tracking strategy has limitations, which need to be considered when interpreting our results. Patient-derived lines were labelled at around 10 passages after derivation. During propagation it is likely that a proportion of clonal diversity has been lost (Brenière-Letuffe et al., 2018). In this case, our current experiments may be tracing newly arising or expanding clones which do not fully recapitulate the heterogeneity present in the original tumour. However, when extensively studied cancer cell lines such U87MG and HeLa, passaged over hundreds or thousands of cycles, have been genetically characterised in different research groups, variable characteristic genetic and epigenetic alterations are observed (Ben-David et al., 2019). This indicates that certain aspects of cell plasticity and evolution are ongoing in cultured cell lines and further stresses the importance of functionally assessing patient lines at early passages to more faithfully model biology of the original tumour. A recent study used fluorescent labelling to identify rare spontaneous cell fusion events where parasexual genetic recombination appeared to occur (Lu and Kang, 2009, Miroshnychenko et al., 2021) both *in vivo* and *in vitro*. Authors mixed cell individually labelled for EGFP or mCherry and detected cell fusion events when cells expressed both markers. Fused cells contained variable combinations of parent cell alleles and modelling predicted these events may amplify clonal heterogeneity and facilitate tumour plasticity. Cell fusions have also been observed in GBM models where extra-cellular vesicles were shown to play also a role in sharing genetic material (Gao et al., 2020). In our clonal sub-cultures, flow results suggested the presence of cells sharing labels derived from two separate clones

(Figure 7.11), however, evidence for this in imaging data was not found and these detections were potentially cell doublets of cells from each clone. Nevertheless, our labelling approach would be well suited for identifying cell fusion events and clones with a higher propensity for this behaviour which may identify the conditions under which fusion events occur. Therefore, even when aspects of tumour heterogeneity are lost upon passaging and use of early passage tissue is preferable for modelling tumour heterogeneity, previous evidence and our results here suggest certain processes of clonal evolution can be studied in early passaged cell lines.

6.5 Clonal outgrowth assay to identify mechanisms of treatment resistance

In this part of our study we explore the plasticity of different clones in response to culturing in a changing culture environment. This method could be used to investigate clonal outgrowth under exposure to treatments, for example chemotherapy or pre-clinical candidate small molecules, such as receptor tyosin kinase inhibitors (RTKIs) (Montor et al., 2018). Outgrowth of mixtures of clonal populations (derived through sub-culturing as described in Chapter 5:) could be assessed in parallel, with outgrowth in normal conditions compared to outgrowth whilst exposed to therapy. Clones which overcome and successfully outgrow while challenged with a therapeutic agent could be identified and isolated from both treatment exposed and normal culture conditions. Clones which are susceptible and "die out" during treatment could also be isolated from the normal culture conditions. Under such conditions, surface marker labelling would not be required, and additional fluorophore labels could be introduced to obtain more viral barcodes and potentially improved clonal detection. Characterisation and comparison of treatment resistant clones in therapeutic conditions or normal culture conditions could reveal how cells have adapted to continue proliferation. Furthermore, comparison of clones grown in normal conditions which show either resistance or susceptibility in the therapy-exposed culture, would further inform the molecular and genetic conditions which prime certain populations with the potential to adapt and overcome treatment. Establishing clonal mixture through sub-culturing as

described in our study would make it possible to assess many different mixtures and identify many examples of populations which achieve resistance. Through single cell sorting and expansion, achieving 3 mixtures composed of 20 different clones would require initial establishment of 60 isolated clonal cultures, whereas, with our approach the same conditions can be achieved through just three 500-cell sub-cultures. Finding mechanisms for adaptation which are consistent across resistant clones would add further confidence in identifying their disease resistance mechanisms. However, performed *in vitro,* these experiments would have a limited ability to assess resistance to promising immune based therapies which require an *in vivo* context. In this case, clonal labelling could be adapted to suit murine allograft brain tumour models where the same clonal mixtures could be allografted in the presence or absence of immunotherapy. Nevertheless, *in vitro* assays assessing patient derived cells could be performed at lower cost and at higher through-put, while capturing patient variability. A further consideration would include assessing the effectiveness of treatment combinations on eradicating clones of patient lines, an approach which could identify treatment regimens, tailored to individual patients.

6.6 Lineage-Tracing in CRISPR screens

In the current research climate, there is a need for non-omics based approaches for single-cell techniques as presented here. However, with technological advancements and reduced cost of sequencing approaches, fluorescent labelling for cell tracking may become obsolete in the future. Sequencing based approaches can achieve higher dimensional datasets and more accurate delineation of clonal populations which can be further linked to expression states and genomic alterations at single-cell level (Sun et al., 2021, Wagner and Klein, 2020). Critically, fluorescent approaches are suitable for *in situ* assessment of tumour clonality, but again, further development spatial transcriptomic approaches may supersede fluorescent labelling for cell tracking. A further desirable aspect of this is to mitigate the need for viral infection and any potentially cytotoxic effects conferred by cytoplasmic fluorophore expression (Ansari et al., 2016). Such factors may influence cell behaviour in our assays and spatial transcriptomics surpasses the need for invasive cell manipulations.

Extensive work has identified the genetic underpinnings of GBM, however, the epigenetic mechanisms which likely govern aspects of cell plasticity described in this work, remain relatively poorly understood. Indeed, in a sample of 135 sequenced GBM biopsies, 46% of samples contained mutations related to chromatin architecture (Brennan et al., 2014). Future experiments could utilise CRISPR screens of alterations affecting epi-genetic regulation, combined with single cell sequencing to better understand how GBM cells harness epi-genetic regulation to promote and achieve plasticity. It is possible that targeted epi-genetic manipulations can direct tumour cells toward distinct expressional states or limit their capacity for state-transitions. The effect of such a manipulation ability of a tumour cell population to overcome treatment are not well established but may represent an attractive novel therapeutic avenue where tumour cells are directed toward expressional states which have established therapeutic susceptibilities.

References

ALCANTARA LLAGUNO, S. R. & PARADA, L. F. 2016. Cell of origin of glioma: biological and clinical implications. *Br J Cancer,* 115**,** 1445-1450.

ALCANTARA LLAGUNO, S. R., WANG, Z., SUN, D., CHEN, J., XU, J., KIM, E., HATANPAA, K. J., RAISANEN, J. M., BURNS, D. K., JOHNSON, J. E. & PARADA, L. F. 2015. Adult Lineage-Restricted CNS Progenitors Specify Distinct Glioblastoma Subtypes. *Cancer Cell,* 28**,** 429-440.

ALDAPE, K., ZADEH, G., MANSOURI, S., REIFENBERGER, G. & VON DEIMLING, A. 2015. Glioblastoma: pathology, molecular mechanisms and markers. *Acta Neuropathologica,* 129**,** 829-848.

ANSARI, A. M., AHMED, A. K., MATSANGOS, A. E., LAY, F., BORN, L. J., MARTI, G., HARMON, J. W. & SUN, Z. 2016. Cellular GFP Toxicity and Immunogenicity: Potential Confounders in in Vivo Cell Tracking Experiments. *Stem Cell Reviews and Reports,* 12**,** 553-559.

ARTZI, M., BRESSLER, I. & BEN BASHAT, D. 2019. Differentiation between glioblastoma, brain metastasis and subtypes using radiomics analysis. *J Magn Reson Imaging,* 50**,** 519-528.

ASKARY, A., SANCHEZ-GUARDADO, L., LINTON, J. M., CHADLY, D. M., BUDDE, M. W., CAI, L., LOIS, C. & ELOWITZ, M. B. 2020. In situ readout of DNA barcodes and single base edits facilitated by in vitro transcription. *Nature Biotechnology,* 38**,** 66-75.

AZZARELLI, R., SIMONS, B. D. & PHILPOTT, A. 2018. The developmental origin of brain tumours: a cellular and molecular framework. *Development,* 145**,** dev162693.

BAEZA-KALLEE, N., BERGES, R., SOUBERAN, A., COLIN, C., DENICOLAI, E., APPAY, R., TCHOGHANDJIAN, A. & FIGARELLA-BRANGER, D. 2019. Glycolipids Recognized by A2B5 Antibody Promote Proliferation, Migration, and Clonogenicity in Glioblastoma Cells. *Cancers (Basel),* 11**,** 1267.

BANKHEAD, P., LOUGHREY, M. B., FERNÁNDEZ, J. A., DOMBROWSKI, Y., MCART, D. G., DUNNE, P. D., MCQUAID, S., GRAY, R. T., MURRAY, L. J., COLEMAN, H. G., JAMES, J. A., SALTO-TELLEZ, M. & HAMILTON, P. W. 2017. QuPath: Open source software for digital pathology image analysis. *Scientific Reports,* 7.

BARRANTES-FREER, A., RENOVANZ, M., EICH, M., BRAUKMANN, A., SPRANG, B., SPIRIN, P., PARDO, L. A., GIESE, A. & KIM, E. L. 2015. CD133 Expression Is Not Synonymous to Immunoreactivity for AC133 and Fluctuates throughout the Cell Cycle in Glioma Stem-Like Cells. *PLoS One,* 10**,** e0130519.

BARTKOVA, J., HAMERLIK, P., STOCKHAUSEN, M. T., EHRMANN, J., HLOBILKOVA, A., LAURSEN, H., KALITA, O., KOLAR, Z., POULSEN, H. S., BROHOLM, H., LUKAS, J. & BARTEK, J. 2010. Replication stress and oxidative damage contribute to aberrant constitutive activation of DNA damage signalling in human gliomas. *Oncogene,* 29**,** 5095-102.

BATLLE, E. & CLEVERS, H. 2017. Cancer stem cells revisited. *Nature Medicine,* 23**,** 1124-1134.

BEN-DAVID, U., BEROUKHIM, R. & GOLUB, T. R. 2019. Genomic evolution of cancer models: perils and opportunities. *Nature Reviews Cancer,* 19**,** 97-109.

BEN-PORATH, I., THOMSON, M. W., CAREY, V. J., GE, R., BELL, G. W., REGEV, A. & WEINBERG, R. A. 2008. An embryonic stem cell-like gene expression signature in poorly differentiated aggressive human tumors. *Nat Genet,* 40**,** 499-507.

BENEDYKCINSKA, A., FERREIRA, A., LAU, J., BRONI, J., RICHARD-LOENDT, A., HENRIQUEZ, N. V. & BRANDNER, S. 2016a. Generation of brain tumours in mice by Cre-mediated recombination of neural progenitors in situ with the tamoxifen metabolite endoxifen. .

BIAN, S., REPIC, M., GUO, Z., KAVIRAYANI, A., BURKARD, T., BAGLEY, J. A., KRAUDITSCH, C. & KNOBLICH, J. A. 2018. Genetically engineered cerebral organoids model brain tumor formation. *Nat Methods,* 15**,** 631-639.

BIEGING, K. T., MELLO, S. S. & ATTARDI, L. D. 2014. Unravelling mechanisms of p53-mediated tumour suppression. *Nat Rev Cancer,* 14**,** 359-70.

BOUGNAUD, S., GOLEBIEWSKA, A., OUDIN, A., KEUNEN, O., HARTER, P. N., MÄDER, L., AZUAJE, F., FRITAH, S., STIEBER, D., KAOMA, T., VALLAR, L., BRONS, N. H., DAUBON, T., MILETIC, H., SUNDSTRØM, T., HEROLD-MENDE, C., MITTELBRONN, M., BJERKVIG, R. & NICLOU, S. P. 2016. Molecular crosstalk between tumour and brain parenchyma instructs histopathological features in glioblastoma. *Oncotarget,* 7**,** 31955-71.

BRANDNER, S. & VON DEIMLING, A. 2015. Diagnostic, prognostic and predictive relevance of molecular markers in gliomas. *Neuropathology and Applied Neurobiology,* 41**,** 694-720.

BRAT, D. J., ALDAPE, K., COLMAN, H., HOLLAND, E. C., LOUIS, D. N., JENKINS, R. B., KLEINSCHMIDT-DEMASTERS, B. K., PERRY, A., REIFENBERGER, G., STUPP, R., VON DEIMLING, A. & WELLER, M. 2018. cIMPACT-NOW update 3: recommended diagnostic criteria for "Diffuse astrocytic glioma, IDH-wildtype, with molecular features of glioblastoma, WHO grade IV". *Acta Neuropathol,* 136**,** 805-810.

BRENIÈRE-LETUFFE, D., DOMKE-SHIBAMIYA, A., HANSEN, A., ESCHENHAGEN, T., FEHSE, B., RIECKEN, K. & STENZIG, J. 2018. Clonal dynamics studied in cultured induced pluripotent stem cells reveal major growth imbalances within a few weeks. *Stem Cell Res Ther,* 9**,** 165.

BROWN, D. V., FILIZ, G., DANIEL, P. M., HOLLANDE, F., DWORKIN, S., AMIRIDIS, S., KOUNTOURI, N., NG, W., MOROKOFF, A. P. & MANTAMADIOTIS, T. 2017. Expression of CD133 and CD44 in glioblastoma stem cells correlates with cell proliferation, phenotype stability and intra-tumor heterogeneity. *PLOS ONE,* 12**,** e0172791.

BURGER, P. C. & KLEIHUES, P. 1989. Cytologic composition of the untreated glioblastoma with implications for evaluation of needle biopsies. *Cancer,* 63**,** 2014-23.

CAI, D., COHEN, K. B., LUO, T., LICHTMAN, J. W. & SANES, J. R. 2013. Improved tools for the Brainbow toolbox. *Nature Methods,* 10**,** 540-+.

CANTANHEDE, I. G. & DE OLIVEIRA, J. R. M. 2017. PDGF Family Expression in Glioblastoma Multiforme: Data Compilation from Ivy Glioblastoma Atlas Project Database. *Scientific Reports,* 7.

CAPPER, D., JONES, D. T. W., SILL, M., HOVESTADT, V., SCHRIMPF, D., STURM, D., KOELSCHE, C., SAHM, F., CHAVEZ, L., REUSS, D. E., KRATZ, A., WEFERS, A. K., HUANG, K., PAJTLER, K. W., SCHWEIZER, L., STICHEL, D., OLAR, A., ENGEL, N. W., LINDENBERG, K., HARTER, P. N., BRACZYNSKI, A. K., PLATE, K. H., DOHMEN, H., GARVALOV, B. K., CORAS, R., HÖLSKEN, A., HEWER, E., BEWERUNGE-HUDLER, M., SCHICK, M., FISCHER, R., BESCHORNER, R., SCHITTENHELM, J., STASZEWSKI, O., WANI, K., VARLET, P., PAGES, M., TEMMING, P., LOHMANN, D., SELT, F., WITT, H., MILDE, T., WITT, O., ARONICA, E., GIANGASPERO, F., RUSHING, E., SCHEURLEN, W., GEISENBERGER, C., RODRIGUEZ, F. J., BECKER, A., PREUSSER, M., HABERLER, C., BJERKVIG, R., CRYAN, J., FARRELL, M., DECKERT, M., HENCH, J., FRANK, S., SERRANO, J., KANNAN, K., TSIRIGOS, A., BRÜCK, W., HOFER, S., BREHMER, S., SEIZ-ROSENHAGEN, M., HÄNGGI, D., HANS, V., ROZSNOKI, S., HANSFORD, J. R., KOHLHOF, P., KRISTENSEN, B. W., LECHNER, M., LOPES, B., MAWRIN, C., KETTER, R., KULOZIK, A., KHATIB, Z., HEPPNER, F., KOCH, A., JOUVET, A., KEOHANE, C., MÜHLEISEN, H., MUELLER, W., POHL, U., PRINZ, M., BENNER, A., ZAPATKA, M., GOTTARDO, N. G., DRIEVER, P. H., KRAMM, C. M., MÜLLER, H. L., RUTKOWSKI, S., VON HOFF, K., FRÜHWALD, M. C., GNEKOW, A., FLEISCHHACK, G., TIPPELT, S., CALAMINUS, G., MONORANU, C.-M., PERRY, A., JONES, C., et al. 2018. DNA methylation-based classification of central nervous system tumours. *Nature,* 555**,** 469-474.

CARON, H., VAN SCHAIK, B., VAN DER MEE, M., BAAS, F., RIGGINS, G., VAN SLUIS, P., HERMUS, M. C., VAN ASPEREN, R., BOON, K., VOÛTE, P. A., HEISTERKAMP, S., VAN KAMPEN, A. & VERSTEEG, R. 2001. The human transcriptome map: clustering of highly expressed genes in chromosomal domains. *Science,* 291**,** 1289-92.

CARUSO, F. P., GAROFANO, L., D'ANGELO, F., YU, K., TANG, F., YUAN, J., ZHANG, J., CERULO, L., PAGNOTTA, S. M., BEDOGNETTI, D., SIMS, P. A., SUVÀ, M., SU, X.-D., LASORELLA, A., IAVARONE, A. & CECCARELLI, M. 2020. A map of tumor–host interactions in glioma at single-cell resolution. *GigaScience,* 9.

CHA, H.-J. & YIM, H. 2013. The accumulation of DNA repair defects is the molecular origin of carcinogenesis. *Tumor Biology,* 34**,** 3293-3302.

CHALFIE, M., TU, Y., EUSKIRCHEN, G., WARD, W. W. & PRASHER, D. C. 1994. Green fluorescent protein as a marker for gene expression. *Science,* 263**,** 802-5.

CHALHOUB, N. & BAKER, S. J. 2009. PTEN and the PI3-Kinase Pathway in Cancer. *Annual Review of Pathology: Mechanisms of Disease,* 4**,** 127-150.

CHARLES, N. A., HOLLAND, E. C., GILBERTSON, R., GLASS, R. & KETTENMANN, H. 2012. The brain tumor microenvironment. *Glia,* 60**,** 502-14.

CHEN, C., ZHAO, S., KARNAD, A. & FREEMAN, J. W. 2018. The biology and role of CD44 in cancer progression: therapeutic implications. *J Hematol Oncol,* 11**,** 64.

CHOI, Y., NAM, Y., LEE, Y. S., KIM, J., AHN, K. J., JANG, J., SHIN, N. Y., KIM, B. S. & JEON, S. S. 2020. IDH1 mutation prediction using MR-based radiomics in glioblastoma:

comparison between manual and fully automated deep learning-based approach of tumor segmentation. *Eur J Radiol,* 128**,** 109031.

CHOW, L. M. L., ENDERSBY, R., ZHU, X. Y., RANKIN, S., QU, C. X., ZHANG, J. Y., BRONISCER, A., ELLISON, D. W. & BAKER, S. J. 2011. Cooperativity within and among Pten, p53, and Rb Pathways Induces High-Grade Astrocytoma in Adult Brain. *Cancer Cell,* 19**,** 305-316.

CLEVERS, H. 2011. The cancer stem cell: premises, promises and challenges. *Nature Medicine,* 17**,** 313-319.

COFFEY, S. E., GIEDT, R. J. & WEISSLEDER, R. 2013. Automated analysis of clonal cancer cells by intravital imaging. *IntraVital,* 2**,** e26138.

CORNILS, K., THIELECKE, L., HÜSER, S., FORGBER, M., THOMASCHEWSKI, M., KLEIST, N., HUSSEIN, K., RIECKEN, K., VOLZ, T., GERDES, S., GLAUCHE, I., DAHL, A., DANDRI, M., ROEDER, I. & FEHSE, B. 2014. Multiplexing clonality: combining RGB marking and genetic barcoding. *Nucleic Acids Res,* 42**,** e56.

COUTURIER, C. P., AYYADHURY, S., LE, P. U., NADAF, J., MONLONG, J., RIVA, G., ALLACHE, R., BAIG, S., YAN, X., BOURGEY, M., LEE, C., WANG, Y. C. D., WEE YONG, V., GUIOT, M.-C., NAJAFABADI, H., MISIC, B., ANTEL, J., BOURQUE, G., RAGOUSSIS, J. & PETRECCA, K. 2020. Single-cell RNA-seq reveals that glioblastoma recapitulates a normal neurodevelopmental hierarchy. *Nature Communications,* 11.

DA SILVA, B., MATHEW, R. K., POLSON, E. S., WILLIAMS, J. & WURDAK, H. 2018. Spontaneous Glioblastoma Spheroid Infiltration of Early-Stage Cerebral Organoids Models Brain Tumor Invasion. *SLAS Discov,* 23**,** 862-868.

DAVIS, A., GAO, R. & NAVIN, N. 2017. Tumor evolution: Linear, branching, neutral or punctuated? *Biochimica et Biophysica Acta (BBA) - Reviews on Cancer,* 1867**,** 151-161.

DIRKSE, A., GOLEBIEWSKA, A., BUDER, T., NAZAROV, P. V., MULLER, A., POOVATHINGAL, S., BRONS, N. H. C., LEITE, S., SAUVAGEOT, N., SARKISJAN, D., SEYFRID, M., FRITAH, S., STIEBER, D., MICHELUCCI, A., HERTEL, F., HEROLD-MENDE, C., AZUAJE, F., SKUPIN, A., BJERKVIG, R., DEUTSCH, A., VOSS-BOHME, A. & NICLOU, S. P. 2019a. Stem cell-associated heterogeneity in Glioblastoma results from intrinsic tumor plasticity shaped by the microenvironment. *Nat Commun,* 10**,** 1787.

DUNN, J., BABORIE, A., ALAM, F., JOYCE, K., MOXHAM, M., SIBSON, R., CROOKS, D., HUSBAND, D., SHENOY, A., BRODBELT, A., WONG, H., LILOGLOU, T., HAYLOCK, B. & WALKER, C. 2009. Extent of MGMT promoter methylation correlates with outcome in glioblastomas given temozolomide and radiotherapy. *British Journal of Cancer,* 101**,** 124-131.

ECK, M. J. & YUN, C.-H. 2010. Structural and mechanistic underpinnings of the differential drug sensitivity of EGFR mutations in non-small cell lung cancer. *Biochimica et Biophysica Acta (BBA) - Proteins and Proteomics,* 1804**,** 559-566.

ELINSON, R. P. & HOLOWACZ, T. 1995. Specifying the dorsoanterior axis in frogs: 70 years since Spemann and Mangold. *Curr Top Dev Biol,* 30**,** 253-85.

ENDERLING, H. 2015. Cancer stem cells: small subpopulation or evolving fraction? *Integrative Biology,* 7**,** 14-23.

ENGELMAN, J. A., ZEJNULLAHU, K., GALE, C. M., LIFSHITS, E., GONZALES, A. J., SHIMAMURA, T., ZHAO, F., VINCENT, P. W., NAUMOV, G. N., BRADNER, J. E., ALTHAUS, I. W., GANDHI, L., SHAPIRO, G. I., NELSON, J. M., HEYMACH, J. V., MEYERSON, M., WONG, K. K. & JANNE, P. A. 2007. PF00299804, an Irreversible Pan-ERBB Inhibitor, Is Effective in Lung Cancer Models with EGFR and ERBB2 Mutations that Are Resistant to Gefitinib. *Cancer Research,* 67**,** 11924-11932.

FEHSE, B., KUSTIKOVA, O. S., BUBENHEIM, M. & BAUM, C. 2004. Pois(s)on – It's a Question of Dose…. *Gene Therapy,* 11**,** 879-881.

FEIL, R., BROCARD, J., MASCREZ, B., LEMEUR, M., METZGER, D. & CHAMBON, P. 1996. Ligand-activated site-specific recombination in mice. *Proc Natl Acad Sci U S A,* 93**,** 10887-90.

FEIL, R., WAGNER, J., METZGER, D. & CHAMBON, P. 1997. Regulation of Cre recombinase activity by mutated estrogen receptor ligand-binding domains. *Biochem Biophys Res Commun,* 237**,** 752-7.

FILATOVA, A., ACKER, T. & GARVALOV, B. K. 2013. The cancer stem cell niche(s): the crosstalk between glioma stem cells and their microenvironment. *Biochim Biophys Acta,* 1830**,** 2496-508.

FRASER, H. & MCCONNELL, I. 1975. Experimental Brain Tumors. *Lancet,* 1**,** 44-44.

GALLI, R., BINDA, E., ORFANELLI, U., CIPELLETTI, B., GRITTI, A., DE VITIS, S., FIOCCO, R., FORONI, C., DIMECO, F. & VESCOVI, A. 2004. Isolation and characterization of tumorigenic, stem-like neural precursors from human glioblastoma. *Cancer Research,* 64**,** 7011-7021.

GALVAO, R. P., KASINA, A., MCNEILL, R. S., HARBIN, J. E., FOREMAN, O., VERHAAK, R. G., NISHIYAMA, A., MILLER, C. R. & ZONG, H. 2014. Transformation of quiescent adult oligodendrocyte precursor cells into malignant glioma through a multistep reactivation process. *Proc Natl Acad Sci U S A,* 111**,** E4214-23.

GAMBERA, S., ABARRATEGI, A., GONZÁLEZ-CAMACHO, F., MORALES-MOLINA, Á., ROMA, J., ALFRANCA, A. & GARCÍA-CASTRO, J. 2018. Clonal dynamics in osteosarcoma defined by RGB marking. *Nature Communications,* 9.

GAO, X., ZHANG, Z., MASHIMO, T., SHEN, B., NYAGILO, J., WANG, H., WANG, Y., LIU, Z., MULGAONKAR, A., HU, X.-L., PICCIRILLO, S. G. M., ESKIOCAK, U., DAVÉ, D. P., QIN, S., YANG, Y., SUN, X., FU, Y.-X., ZONG, H., SUN, W., BACHOO, R. M. & GE, W.-P. 2020. Gliomas Interact with Non-glioma Brain Cells via Extracellular Vesicles. *Cell Reports,* 30**,** 2489-2500.e5.

GHOSH, D., NANDI, S. & BHATTACHARJEE, S. 2018. Combination therapy to checkmate Glioblastoma: clinical challenges and advances. *Clinical and Translational Medicine,* 7**,** 33.

GIANGRECO, A., ARWERT, E. N., ROSEWELL, I. R., SNYDER, J., WATT, F. M. & STRIPP, B. R. 2009. Stem cells are dispensable for lung homeostasis but restore airways after injury. *Proceedings of the National Academy of Sciences,* 106**,** 9286-9291.

GIERMAN, H. J., INDEMANS, M. H. G., KOSTER, J., GOETZE, S., SEPPEN, J., GEERTS, D., VAN DRIEL, R. & VERSTEEG, R. 2007. Domain-wide regulation of gene expression in the human genome. *Genome Research,* 17**,** 1286-1295.

GOMEZ-NICOLA, D., RIECKEN, K., FEHSE, B. & PERRY, V. H. 2015. In-vivo RGB marking and multicolour single-cell tracking in the adult brain. *Scientific Reports,* 4**,** 7520.

HADJIECONOMOU, D., ROTKOPF, S., ALEXANDRE, C., BELL, D. M., DICKSON, B. J. & SALECKER, I. 2011. Flybow: genetic multicolor cell labeling for neural circuit analysis in Drosophila melanogaster. *Nature Methods,* 8**,** 260-266.

HEGI, M. E., DISERENS, A. C., GORLIA, T., HAMOU, M. F., DE TRIBOLET, N., WELLER, M., KROS, J. M., HAINFELLNER, J. A., MASON, W., MARIANI, L., BROMBERG, J. E., HAU, P., MIRIMANOFF, R. O., CAIRNCROSS, J. G., JANZER, R. C. & STUPP, R. 2005. MGMT gene silencing and benefit from temozolomide in glioblastoma. *N Engl J Med,* 352**,** 997-1003.

HEMMATI, H. D., NAKANO, I., LAZAREFF, J. A., MASTERMAN-SMITH, M., GESCHWIND, D. H., BRONNER-FRASER, M. & KORNBLUM, H. I. 2003. Cancerous stem cells can arise from pediatric brain tumors. *Proceedings of the National Academy of Sciences of the United States of America,* 100**,** 15178-15183.

HILSCHER, M. M., GYLLBORG, D., YOKOTA, C. & NILSSON, M. 2020. In Situ Sequencing: A High-Throughput, Multi-Targeted Gene Expression Profiling Technique for Cell Typing in Tissue Sections. *Methods Mol Biol,* 2148**,** 313-329.

HIRA, V. V. V., ADERETTI, D. A. & VAN NOORDEN, C. J. F. 2018. Glioma Stem Cell Niches in Human Glioblastoma Are Periarteriolar. *J Histochem Cytochem,* 66**,** 349-358.

HOCHHAUS, A., LARSON, R. A., GUILHOT, F., RADICH, J. P., BRANFORD, S., HUGHES, T. P., BACCARANI, M., DEININGER, M. W., CERVANTES, F., FUJIHARA, S., ORTMANN, C.-E., MENSSEN, H. D., KANTARJIAN, H., O'BRIEN, S. G. & DRUKER, B. J. 2017. Long-Term Outcomes of Imatinib Treatment for Chronic Myeloid Leukemia. *New England Journal of Medicine,* 376**,** 917-927.

HUBERT, C. G., RIVERA, M., SPANGLER, L. C., WU, Q., MACK, S. C., PRAGER, B. C., COUCE, M., MCLENDON, R. E., SLOAN, A. E. & RICH, J. N. 2016. A Three-Dimensional Organoid Culture System Derived from Human Glioblastomas Recapitulates the Hypoxic Gradients and Cancer Stem Cell Heterogeneity of Tumors Found In Vivo. *Cancer Research,* 76**,** 2465-2477.

JACOB, F., SALINAS, R. D., ZHANG, D. Y., NGUYEN, P. T. T., SCHNOLL, J. G., WONG, S. Z. H., THOKALA, R., SHEIKH, S., SAXENA, D., PROKOP, S., LIU, D. A., QIAN, X., PETROV, D., LUCAS, T., CHEN, H. I., DORSEY, J. F., CHRISTIAN, K. M., BINDER, Z. A., NASRALLAH, M., BREM, S., O'ROURKE, D. M., MING, G. L. & SONG, H. 2020. A Patient-Derived Glioblastoma Organoid Model and Biobank Recapitulates Inter- and Intra-tumoral Heterogeneity. *Cell,* 180**,** 188-204.e22.

JACQUES, T. S., SWALES, A., BRZOZOWSKI, M. J., HENRIQUEZ, N. V., LINEHAN, J. M., MIRZADEH, Z., MALLEY, C. O., NAUMANN, H., ALVAREZ-BUYLLA, A. & BRANDNER, S. 2010. Combinations of genetic mutations in the adult neural stem cell compartment determine brain tumour phenotypes. *Embo Journal,* 29**,** 222-235.

JENSEN, E. C. 2012. Use of Fluorescent Probes: Their Effect on Cell Biology and Limitations. *The Anatomical Record: Advances in Integrative Anatomy and Evolutionary Biology,* 295**,** 2031-2036.

JIN, X., KIM, L. J. Y., WU, Q., WALLACE, L. C., PRAGER, B. C., SANVORANART, T., GIMPLE, R. C., WANG, X., MACK, S. C., MILLER, T. E., HUANG, P., VALENTIM, C. L., ZHOU, Q.-G., BARNHOLTZ-SLOAN, J. S., BAO, S., SLOAN, A. E. & RICH, J. N. 2017. Targeting glioma stem cells through combined BMI1 and EZH2 inhibition. *Nature Medicine,* 23**,** 1352-1361.

JOHNSON, B. E., MAZOR, T., HONG, C. B., BARNES, M., AIHARA, K., MCLEAN, C. Y., FOUSE, S. D., YAMAMOTO, S., UEDA, H., TATSUNO, K., ASTHANA, S., JALBERT, L. E., NELSON, S. J., BOLLEN, A. W., GUSTAFSON, W. C., CHARRON, E., WEISS, W. A., SMIRNOV, I. V., SONG, J. S., OLSHEN, A. B., CHA, S., ZHAO, Y. J., MOORE, R. A., MUNGALL, A. J., JONES, S. J. M., HIRST, M., MARRA, M. A., SAITO, N., ABURATANI, H., MUKASA, A., BERGER, M. S., CHANG, S. M., TAYLOR, B. S. & COSTELLO, J. F. 2014. Mutational Analysis Reveals the Origin and Therapy-Driven Evolution of Recurrent Glioma. *Science,* 343**,** 189-193.

JURKOWSKI, M. P., BETTIO, L., WOO, E. K., PATTEN, A., YAU, S. Y. & GIL-MOHAPEL, J. 2020. Beyond the Hippocampus and the SVZ: Adult Neurogenesis Throughout the Brain. *Frontiers in Cellular Neuroscience,* 14.

KAZANIS, I., LATHIA, J., MOSS, L. & FFRENCH-CONSTANT, C. 2008. The neural stem cell microenvironment. *StemBook.* Cambridge (MA): Harvard Stem Cell Institute

KEGELMAN, T. P., HU, B., EMDAD, L., DAS, S. K., SARKAR, D. & FISHER, P. B. 2014. In vivo modeling of malignant glioma: the road to effective therapy. *Adv Cancer Res,* 121**,** 261-330.

KIM, C.-S., JUNG, S., JUNG, T.-Y., JANG, W.-Y., SUN, H.-S. & RYU, H.-H. 2011. Characterization of Invading Glioma Cells Using Molecular Analysis of Leading-Edge Tissue. *Journal of Korean Neurosurgical Society,* 50**,** 157.

KIM, H., KIM, M., IM, S. K. & FANG, S. 2018. Mouse Cre-LoxP system: general principles to determine tissue-specific roles of target genes. *Lab Anim Res,* 34**,** 147-159.

KIM, J., WOO, A. J., CHU, J., SNOW, J. W., FUJIWARA, Y., KIM, C. G., CANTOR, A. B. & ORKIN, S. H. 2010. A Myc Network Accounts for Similarities between Embryonic Stem and Cancer Cell Transcription Programs. *Cell,* 143**,** 313-324.

KOSCHMANN, C., LOWENSTEIN, P. R. & CASTRO, M. G. 2016. ATRX mutations and glioblastoma: Impaired DNA damage repair, alternative lengthening of telomeres, and genetic instability. *Molecular & Cellular Oncology,* 3**,** e1167158.

KRETZSCHMAR, K. & WATT, F. M. 2012. Lineage Tracing. *Cell,* 148**,** 33-45.

KRISHNA, BALASUBRAMANIYAN, V., VAILLANT, B., EZHILARASAN, R., HUMMELINK, K., HOLLINGSWORTH, F., WANI, K., HEATHCOCK, L., JOHANNA, LINDSEY, CONROY, S., LONG, L., LELIC, N., WANG, S., GUMIN, J., RAJ, D., KODAMA, Y., RAGHUNATHAN, A., OLAR, A., JOSHI, K., CHRISTOPHER, HEIMBERGER, A., SE, DANIEL, RAO, G., WILFRED,

HENDRIKUS, HEIDI, NAKANO, I., FREDERICK, COLMAN, H., ERIK & ALDAPE, K. 2013. Mesenchymal Differentiation Mediated by NF-κB Promotes Radiation Resistance in Glioblastoma. *Cancer Cell,* 24**,** 331-346.

KWON, C. H., ZHAO, D. W., CHEN, J., ALCANTARA, S., LI, Y. J., BURNS, D. K., MASON, R. P., LEE, E. Y. H. P., WU, H. & PARADA, L. F. 2008. Pten haploinsufficiency accelerates formation of high-grade astrocytomas. *Cancer Research,* 68**,** 3286-3294.

LAMBERT, G. G., DEPERNET, H., GOTTHARD, G., SCHULTZ, D. T., NAVIZET, I., LAMBERT, T., ADAMS, S. R., TORREBLANCA-ZANCA, A., CHU, M., BINDELS, D. S., LEVESQUE, V., NERO MOFFATT, J., SALIH, A., ROYANT, A. & SHANER, N. C. 2020. Aequorea's secrets revealed: New fluorescent proteins with unique properties for bioimaging and biosensing. *PLOS Biology,* 18**,** e3000936.

LAN, X., JORG, D. J., CAVALLI, F. M. G., RICHARDS, L. M., NGUYEN, L. V., VANNER, R. J., GUILHAMON, P., LEE, L., KUSHIDA, M. M., PELLACANI, D., PARK, N. I., COUTINHO, F. J., WHETSTONE, H., SELVADURAI, H. J., CHE, C., LUU, B., CARLES, A., MOKSA, M., RASTEGAR, N., HEAD, R., DOLMA, S., PRINOS, P., CUSIMANO, M. D., DAS, S., BERNSTEIN, M., ARROWSMITH, C. H., MUNGALL, A. J., MOORE, R. A., MA, Y., GALLO, M., LUPIEN, M., PUGH, T. J., TAYLOR, M. D., HIRST, M., EAVES, C. J., SIMONS, B. D. & DIRKS, P. B. 2017. Fate mapping of human glioblastoma reveals an invariant stem cell hierarchy. *Nature,* 549**,** 227-232.

LANCASTER, M. A., RENNER, M., MARTIN, C.-A., WENZEL, D., BICKNELL, L. S., HURLES, M. E., HOMFRAY, T., PENNINGER, J. M., JACKSON, A. P. & KNOBLICH, J. A. 2013. Cerebral organoids model human brain development and microcephaly. *Nature,* 501**,** 373-379.

LAPIDOT, T., SIRARD, C., VORMOOR, J., MURDOCH, B., HOANG, T., CACERES-CORTES, J., MINDEN, M., PATERSON, B., CALIGIURI, M. A. & DICK, J. E. 1994. A cell initiating human acute myeloid leukaemia after transplantation into SCID mice. *Nature,* 367**,** 645-648.

LATHIA, J. D. 2021. Fountain of chaos: cerebrospinal fluid enhancement of cancer stem cells in glioblastoma. *Neuro-Oncology,* 23**,** 530-532.

LATHIA, J. D., GALLAGHER, J., MYERS, J. T., LI, M. Z., VASANJI, A., MCLENDON, R. E., HJELMELAND, A. B., HUANG, A. Y. & RICH, J. N. 2011. Direct In Vivo Evidence for Tumor Propagation by Glioblastoma Cancer Stem Cells. *Plos One,* 6.

LATHIA, J. D., MACK, S. C., MULKEARNS-HUBERT, E. E., VALENTIM, C. L. L. & RICH, J. N. 2015. Cancer stem cells in glioblastoma. *Genes & Development,* 29**,** 1203-1217.

LEI, L., SONABEND, A. M., GUARNIERI, P., SODERQUIST, C., LUDWIG, T., ROSENFELD, S., BRUCE, J. N. & CANOLL, P. 2011. Glioblastoma Models Reveal the Connection between Adult Glial Progenitors and the Proneural Phenotype. *Plos One,* 6.

LI, S., CHEN, L. X., PENG, X. H., WANG, C., QIN, B. Y., TAN, D., HAN, C. X., YANG, H., REN, X. N., LIU, F., XU, C. H. & ZHOU, X. H. 2018. Overview of the reporter genes and reporter mouse models. *Animal Model Exp Med,* 1**,** 29-35.

LIGON, K. L., HUILLARD, E., MEHTA, S., KESARI, S., LIU, H., ALBERTA, J. A., BACHOO, R. M., KANE, M., LOUIS, D. N., DEPINHO, R. A., ANDERSON, D. J., STILES, C. D. & ROWITCH,

D. H. 2007. Olig2-Regulated Lineage-Restricted Pathway Controls Replication Competence in Neural Stem Cells and Malignant Glioma. *Neuron,* 53**,** 503-517.

LIM, D. A. & ALVAREZ-BUYLLA, A. 2016. The Adult Ventricular-Subventricular Zone (V-SVZ) and Olfactory Bulb (OB) Neurogenesis. *Cold Spring Harb Perspect Biol,* 8.

LINKOUS, A., BALAMATSIAS, D., SNUDERL, M., EDWARDS, L., MIYAGUCHI, K., MILNER, T., REICH, B., COHEN-GOULD, L., STORASKA, A., NAKAYAMA, Y., SCHENKEIN, E., SINGHANIA, R., CIRIGLIANO, S., MAGDELDIN, T., LIN, Y., NANJANGUD, G., CHADALAVADA, K., PISAPIA, D., LISTON, C. & FINE, H. A. 2019. Modeling Patient-Derived Glioblastoma with Cerebral Organoids. *Cell Reports,* 26**,** 3203-3211.e5.

LITTLEWOOD, T. D., HANCOCK, D. C., DANIELIAN, P. S., PARKER, M. G. & EVAN, G. I. 1995. A modified oestrogen receptor ligand-binding domain as an improved switch for the regulation of heterologous proteins. *Nucleic Acids Res,* 23**,** 1686-90.

LIU, C., SAGE, J. C., MILLER, M. R., VERHAAK, R. G. W., HIPPENMEYER, S., VOGEL, H., FOREMAN, O., BRONSON, R. T., NISHIYAMA, A., LUO, L. Q. & ZONG, H. 2011. Mosaic Analysis with Double Markers Reveals Tumor Cell of Origin in Glioma. *Cell,* 146**,** 209-221.

LIU, G., YUAN, X., ZENG, Z., TUNICI, P., NG, H., ABDULKADIR, I. R., LU, L., IRVIN, D., BLACK, K. L. & YU, J. S. 2006. Analysis of gene expression and chemoresistance of CD133+ cancer stem cells in glioblastoma. *Molecular Cancer,* 5**,** 67.

LIVET, J., WEISSMAN, T. A., KANG, H. N., DRAFT, R. W., LU, J., BENNIS, R. A., SANES, J. R. & LICHTMAN, J. W. 2007. Transgenic strategies for combinatorial expression of fluorescent proteins in the nervous system. *Nature,* 450**,** 56-+.

LLAGUNO, S. A., CHEN, J., KWON, C. H., JACKSON, E. L., LI, Y., BURNS, D. K., ALVAREZ-BUYLLA, A. & PARADA, L. F. 2009. Malignant Astrocytomas Originate from Neural Stem/Progenitor Cells in a Somatic Tumor Suppressor Mouse Model (vol 15, pg 45, 2009). *Cancer Cell,* 15**,** 240-240.

LOUIS, D. N., PERRY, A., REIFENBERGER, G., VON DEIMLING, A., FIGARELLA-BRANGER, D., CAVENEE, W. K., OHGAKI, H., WIESTLER, O. D., KLEIHUES, P. & ELLISON, D. W. .

LOUIS, D. N., WESSELING, P., ALDAPE, K., BRAT, D. J., CAPPER, D., CREE, I. A., EBERHART, C., FIGARELLA-BRANGER, D., FOULADI, M., FULLER, G. N., GIANNINI, C., HABERLER, C., HAWKINS, C., KOMORI, T., KROS, J. M., NG, H. K., ORR, B. A., PARK, S. H., PAULUS, W., PERRY, A., PIETSCH, T., REIFENBERGER, G., ROSENBLUM, M., ROUS, B., SAHM, F., SARKAR, C., SOLOMON, D. A., TABORI, U., VAN DEN BENT, M. J., VON DEIMLING, A., WELLER, M., WHITE, V. A. & ELLISON, D. W. 2020. cIMPACT-NOW update 6: new entity and diagnostic principle recommendations of the cIMPACT-Utrecht meeting on future CNS tumor classification and grading. *Brain Pathol,* 30**,** 844-856.

LOULIER, K., BARRY, R., MAHOU, P., LE FRANC, Y., SUPATTO, W., MATHO, K. S., IENG, S. H., FOUQUET, S., DUPIN, E., BENOSMAN, R., CHEDOTAL, A., BEAUREPAIRE, E., MORIN, X. & LIVET, J. 2014. Multiplex Cell and Lineage Tracking with Combinatorial Labels. *Neuron,* 81**,** 505-520.

LU, X. & KANG, Y. 2009. Cell Fusion as a Hidden Force in Tumor Progression. *Cancer Research,* 69**,** 8536-8539.

MACYSZYN, L., AKBARI, H., PISAPIA, J. M., DA, X., ATTIAH, M., PIGRISH, V., BI, Y., PAL, S., DAVULURI, R. V., ROCCOGRANDI, L., DAHMANE, N., MARTINEZ-LAGE, M., BIROS, G., WOLF, R. L., BILELLO, M., O'ROURKE, D. M. & DAVATZIKOS, C. 2016. Imaging patterns predict patient survival and molecular subtype in glioblastoma via machine learning techniques. *Neuro Oncol,* 18**,** 417-25.

MAIRE, C. L., MOHME, M., BOCKMAYR, M., FITA, K. D., RIECKEN, K., BÖRNIGEN, D., ALAWI, M., FAILLA, A., KOLBE, K., ZAPF, S., HOLZ, M., NEUMANN, K., DÜHRSEN, L., LANGE, T., FEHSE, B., WESTPHAL, M. & LAMSZUS, K. 2020. Glioma escape signature and clonal development under immune pressure. *Journal of Clinical Investigation,* 130**,** 5257-5271.

MAO, X., FUJIWARA, Y., CHAPDELAINE, A., YANG, H. & ORKIN, S. H. 2001. Activation of EGFP expression by Cre-mediated excision in a new ROSA26 reporter mouse strain. *Blood,* 97**,** 324-6.

MARINO, S., KRIMPENFORT, P., LEUNG, C., VAN DER KORPUT, H. A., TRAPMAN, J., CAMENISCH, I., BERNS, A. & BRANDNER, S. 2002. PTEN is essential for cell migration but not for fate determination and tumourigenesis in the cerebellum. *Development,* 129**,** 3513-22.

MARINO, S., VOOIJS, M., VAN DER GULDEN, H., JONKERS, J. & BERNS, A. 2000. Induction of medulloblastomas in p53-null mutant mice by somatic inactivation of Rb in the external granular layer cells of the cerebellum. *Genes Dev,* 14**,** 994-1004.

MARTÍNEZ-GONZÁLEZ, A., CALVO, G. F., PÉREZ ROMASANTA, L. A. & PÉREZ-GARCÍA, V. M. 2012. Hypoxic cell waves around necrotic cores in glioblastoma: a biomathematical model and its therapeutic implications. *Bull Math Biol,* 74**,** 2875-96.

MARTINS-MACEDO, J., LEPORE, A. C., DOMINGUES, H. S., SALGADO, A. J., GOMES, E. D. & PINTO, L. 2021. Glial restricted precursor cells in central nervous system disorders: Current applications and future perspectives. *Glia,* 69**,** 513-531.

MEINKE, G., BOHM, A., HAUBER, J., PISABARRO, M. T. & BUCHHOLZ, F. 2016. Cre Recombinase and Other Tyrosine Recombinases. *Chem Rev,* 116**,** 12785-12820.

MING, G. L. & SONG, H. 2011. Adult neurogenesis in the mammalian brain: significant answers and significant questions. *Neuron,* 70**,** 687-702.

MIROSHNYCHENKO, D., BARATCHART, E., FERRALL-FAIRBANKS, M. C., VELDE, R. V., LAURIE, M. A., BUI, M. M., TAN, A. C., ALTROCK, P. M., BASANTA, D. & MARUSYK, A. 2021. Spontaneous cell fusions as a mechanism of parasexual recombination in tumour cell populations. *Nature Ecology & Evolution,* 5**,** 379-391.

MOHME, M., MAIRE, C. L., RIECKEN, K., ZAPF, S., ARANYOSSY, T., WESTPHAL, M., LAMSZUS, K. & FEHSE, B. 2017. Optical Barcoding for Single-Clone Tracking to Study Tumor Heterogeneity. *Molecular Therapy,* 25**,** 621-633.

MONTOR, W. R., SALAS, A. R. O. S. E. & MELO, F. H. M. D. 2018. Receptor tyrosine kinases and downstream pathways as druggable targets for cancer treatment: the current arsenal of inhibitors. *Molecular Cancer,* 17.

MORGILLO, F., DELLA CORTE, C. M., FASANO, M. & CIARDIELLO, F. 2016. Mechanisms of resistance to EGFR-targeted drugs: lung cancer. *ESMO Open,* 1**,** e000060.

MORI, T., TANAKA, K., BUFFO, A., WURST, W., KÜHN, R. & GÖTZ, M. 2006. Inducible gene deletion in astroglia and radial glia-A valuable tool for functional and lineage analysis. *Glia,* 54**,** 21-34.

NANDU, H., WEN, P. Y. & HUANG, R. Y. 2018. Imaging in neuro-oncology. *Ther Adv Neurol Disord,* 11**,** 1756286418759865.

NAZARENKO, I., HEDE, S.-M., HE, X., HEDRÉN, A., THOMPSON, J., LINDSTRÖM, M. S. & NISTÉR, M. 2012. PDGF and PDGF receptors in glioma. *Upsala Journal of Medical Sciences,* 117**,** 99-112.

NEFTEL, C., LAFFY, J., FILBIN, M. G., HARA, T., SHORE, M. E., RAHME, G. J., RICHMAN, A. R., SILVERBUSH, D., SHAW, M. L., HEBERT, C. M., DEWITT, J., GRITSCH, S., PEREZ, E. M., GONZALEZ CASTRO, L. N., LAN, X., DRUCK, N., RODMAN, C., DIONNE, D., KAPLAN, A., BERTALAN, M. S., SMALL, J., PELTON, K., BECKER, S., BONAL, D., NGUYEN, Q.-D., SERVIS, R. L., FUNG, J. M., MYLVAGANAM, R., MAYR, L., GOJO, J., HABERLER, C., GEYEREGGER, R., CZECH, T., SLAVC, I., NAHED, B. V., CURRY, W. T., CARTER, B. S., WAKIMOTO, H., BRASTIANOS, P. K., BATCHELOR, T. T., STEMMER-RACHAMIMOV, A., MARTINEZ-LAGE, M., FROSCH, M. P., STAMENKOVIC, I., RIGGI, N., RHEINBAY, E., MONJE, M., ROZENBLATT-ROSEN, O., CAHILL, D. P., PATEL, A. P., HUNTER, T., VERMA, I. M., LIGON, K. L., LOUIS, D. N., REGEV, A., BERNSTEIN, B. E., TIROSH, I. & SUVÀ, M. L. 2019. An Integrative Model of Cellular States, Plasticity, and Genetics for Glioblastoma. *Cell,* 178**,** 835-849.e21.

NGUYEN, L. V., COX, C. L., EIREW, P., KNAPP, D. J. H. F., PELLACANI, D., KANNAN, N., CARLES, A., MOKSA, M., BALANI, S., SHAH, S., HIRST, M., APARICIO, S. & EAVES, C. J. 2014a. DNA barcoding reveals diverse growth kinetics of human breast tumour subclones in serially passaged xenografts. *Nature Communications,* 5**,** 5871.

NGUYEN, L. V., MAKAREM, M., CARLES, A., MOKSA, M., KANNAN, N., PANDOH, P., EIREW, P., OSAKO, T., KARDEL, M., CHEUNG, A. M., KENNEDY, W., TSE, K., ZENG, T., ZHAO, Y., HUMPHRIES, R. K., APARICIO, S., EAVES, C. J. & HIRST, M. 2014b. Clonal analysis via barcoding reveals diverse growth and differentiation of transplanted mouse and human mammary stem cells. *Cell Stem Cell,* 14**,** 253-63.

NGUYEN, L. V., PELLACANI, D., LEFORT, S., KANNAN, N., OSAKO, T., MAKAREM, M., COX, C. L., KENNEDY, W., BEER, P., CARLES, A., MOKSA, M., BILENKY, M., BALANI, S., BABOVIC, S., SUN, I., ROSIN, M., APARICIO, S., HIRST, M. & EAVES, C. J. 2015. Barcoding reveals complex clonal dynamics of de novo transformed human mammary cells. *Nature,* 528**,** 267-71.

NIGON, V. M. & FÉLIX, M. A. 2017. History of research on C. elegans and other free-living nematodes as model organisms. *WormBook,* 2017**,** 1-84.

NOWELL, P. C. 1988. Citation Classic - the Clonal Evolution of Tumor-Cell Populations. *Current Contents/Life Sciences***,** 19-19.

OCHOCKA, N., SEGIT, P., WALENTYNOWICZ, K. A., WOJNICKI, K., CYRANOWSKI, S., SWATLER, J., MIECZKOWSKI, J. & KAMINSKA, B. 2021. Single-cell RNA sequencing reveals

functional heterogeneity of glioma-associated brain macrophages. *Nature Communications,* 12.

OGAWA, J., PAO, G. M., SHOKHIREV, M. N. & VERMA, I. M. 2018. Glioblastoma Model Using Human Cerebral Organoids. *Cell Rep,* 23**,** 1220-1229.

OGAWA, K., KUROSE, A., KAMATAKI, A., ASANO, K., KATAYAMA, K. & KUROTAKI, H. 2020. Giant cell glioblastoma is a distinctive subtype of glioma characterized by vulnerability to DNA damage. *Brain Tumor Pathol,* 37**,** 5-13.

OGDEN, A. T., WAZIRI, A. E., LOCHHEAD, R. A., FUSCO, D., LOPEZ, K., ELLIS, J. A., KANG, J., ASSANAH, M., MCKHANN, G. M., SISTI, M. B., MCCORMICK, P. C., CANOLL, P. & BRUCE, J. N. 2008. IDENTIFICATION OF A2B5+CD133− TUMOR-INITIATING CELLS IN ADULT HUMAN GLIOMAS. *Neurosurgery,* 62**,** 505-515.

OHGAKI, H. & KLEIHUES, P. 2007. Genetic pathways to primary and secondary glioblastoma. *Am J Pathol,* 170**,** 1445-53.

OHLSTEIN, B., KAI, T., DECOTTO, E. & SPRADLING, A. 2004. The stem cell niche: theme and variations. *Current Opinion in Cell Biology,* 16**,** 693-699.

OPRITA, A., BALOI, S. C., STAICU, G. A., ALEXANDRU, O., TACHE, D. E., DANOIU, S., MICU, E. S. & SEVASTRE, A. S. 2021. Updated Insights on EGFR Signaling Pathways in Glioma. *International Journal of Molecular Sciences,* 22.

OSTROM, Q. T., CIOFFI, G., GITTLEMAN, H., PATIL, N., WAITE, K., KRUCHKO, C. & BARNHOLTZ-SLOAN, J. S. 2019. CBTRUS Statistical Report: Primary Brain and Other Central Nervous System Tumors Diagnosed in the United States in 2012-2016. *Neuro Oncol,* 21**,** v1-v100.

OSTROM, Q. T., GITTLEMAN, H., LIAO, P., ROUSE, C., CHEN, Y., DOWLING, J., WOLINSKY, Y., KRUCHKO, C. & BARNHOLTZ-SLOAN, J. 2014. CBTRUS statistical report: primary brain and central nervous system tumors diagnosed in the United States in 2007-2011. *Neuro Oncol,* 16 Suppl 4**,** iv1-63.

PATEL, A. P., TIROSH, I., TROMBETTA, J. J., SHALEK, A. K., GILLESPIE, S. M., WAKIMOTO, H., CAHILL, D. P., NAHED, B. V., CURRY, W. T., MARTUZA, R. L., LOUIS, D. N., ROZENBLATT-ROSEN, O., SUVA, M. L., REGEV, A. & BERNSTEIN, B. E. 2014. Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. *Science,* 344**,** 1396-1401.

PERICOLI, G., PETRINI, S., GIORDA, E., FERRETTI, R., AJMONE-CAT, M. A., COURT, W., CONTI, L. A., DE SIMONE, R., BENCIVENGA, P., PALMA, A., DI GIANNATALE, A., JONES, C., CARAI, A., MASTRONUZZI, A., DE BILLY, E., LOCATELLI, F. & VINCI, M. 2020. Integration of Multiple Platforms for the Analysis of Multifluorescent Marking Technology Applied to Pediatric GBM and DIPG. *International Journal of Molecular Sciences,* 21**,** 6763.

PLANCHARD, D., POPAT, S., KERR, K., NOVELLO, S., SMIT, E. F., FAIVRE-FINN, C., MOK, T. S., RECK, M., VAN SCHIL, P. E., HELLMANN, M. D. & PETERS, S. 2018. Metastatic non-small cell lung cancer: ESMO Clinical Practice Guidelines for diagnosis, treatment and follow-up. *Annals of Oncology,* 29**,** iv192-iv237.

PRAGER, B. C., BHARGAVA, S., MAHADEV, V., HUBERT, C. G. & RICH, J. N. 2020. Glioblastoma Stem Cells: Driving Resilience through Chaos. *Trends Cancer,* 6**,** 223-235.

PRESTON, A. N., CERVASIO, D. A. & LAUGHLIN, S. T. 2019. Visualizing the brain's astrocytes. Elsevier.

READ, T.-A., FOGARTY, M. P., MARKANT, S. L., MCLENDON, R. E., WEI, Z., ELLISON, D. W., FEBBO, P. G. & WECHSLER-REYA, R. J. 2009. Identification of CD15 as a Marker for Tumor-Propagating Cells in a Mouse Model of Medulloblastoma. *Cancer Cell,* 15**,** 135-147.

REEVES, M. Q., KANDYBA, E., HARRIS, S., DEL ROSARIO, R. & BALMAIN, A. 2018. Multicolour lineage tracing reveals clonal dynamics of squamous carcinoma evolution from initiation to metastasis. *Nature Cell Biology,* 20**,** 699-709.

REIFENBERGER, G., WIRSCHING, H. G., KNOBBE-THOMSEN, C. B. & WELLER, M. 2017. Advances in the molecular genetics of gliomas - implications for classification and therapy. *Nature Reviews Clinical Oncology,* 14**,** 434-452.

RICHARDSON, D. S. & LICHTMAN, J. W. 2015. Clarifying Tissue Clearing. *Cell,* 162**,** 246-257.

ROBERTSON, F. L., MARQUÉS-TORREJÓN, M. A., MORRISON, G. M. & POLLARD, S. M. 2019. Experimental models and tools to tackle glioblastoma. *Dis Model Mech,* 12.

RODRIGUEZ, E. A., CAMPBELL, R. E., LIN, J. Y., LIN, M. Z., MIYAWAKI, A., PALMER, A. E., SHU, X., ZHANG, J. & TSIEN, R. Y. 2017. The Growing and Glowing Toolbox of Fluorescent and Photoactive Proteins. *Trends Biochem Sci,* 42**,** 111-129.

SAKAGUCHI, R., LEIWE, M. N. & IMAI, T. 2018. Bright multicolor labeling of neuronal circuits with fluorescent proteins and chemical tags. *eLife,* 7.

SANAI, N., ALVAREZ-BUYLLA, A. & BERGER, M. S. 2005. Mechanisms of disease: Neural stem cells and the origin of gliomas. *New England Journal of Medicine,* 353**,** 811-822.

SASAKI, M., KNOBBE, C. B., ITSUMI, M., ELIA, A. J., HARRIS, I. S., CHIO, I. I. C., CAIRNS, R. A., MCCRACKEN, S., WAKEHAM, A., HAIGHT, J., TEN, A. Y., SNOW, B., UEDA, T., INOUE, S., YAMAMOTO, K., KO, M., RAO, A., YEN, K. E., SU, S. M. & MAK, T. W. 2012. D-2-hydroxyglutarate produced by mutant IDH1 perturbs collagen maturation and basement membrane function. *Genes & Development,* 26**,** 2038-2049.

SCHEPERS, A. G., SNIPPERT, H. J., STANGE, D. E., VAN DEN BORN, M., VAN ES, J. H., VAN DE WETERING, M. & CLEVERS, H. 2012. Lineage tracing reveals Lgr5+ stem cell activity in mouse intestinal adenomas. *Science,* 337**,** 730-735.

SCHIFFER, D., ANNOVAZZI, L., CASALONE, C., CORONA, C. & MELLAI, M. 2018. Glioblastoma: Microenvironment and Niche Concept. *Cancers,* 11**,** 5.

SCHINDELIN, J., ARGANDA-CARRERAS, I., FRISE, E., KAYNIG, V., LONGAIR, M., PIETZSCH, T., PREIBISCH, S., RUEDEN, C., SAALFELD, S., SCHMID, B., TINEVEZ, J. Y., WHITE, D. J., HARTENSTEIN, V., ELICEIRI, K., TOMANCAK, P. & CARDONA, A. 2012. Fiji: an open-source platform for biological-image analysis. *Nat Methods,* 9**,** 676-82.

SCHMIDT, U., WEIGERT, M., BROADDUS, C. & MYERS, G. 2018. Cell Detection with Star-Convex Polygons. Springer International Publishing.

SCOTT, J. G., DHAWAN, A., HJELMELAND, A., LATHIA, J., CHUMAKOVA, A., HITOMI, M., FLETCHER, A. G., MAINI, P. K. & ANDERSON, A. R. A. 2019. Recasting the Cancer Stem Cell Hypothesis: Unification Using a Continuum Model of Microenvironmental Forces. *Current Stem Cell Reports,* 5**,** 22-30.

SCOTT, J. G., HJELMELAND, A. B., CHINNAIYAN, P., ANDERSON, A. R. A. & BASANTA, D. 2014. Microenvironmental Variables Must Influence Intrinsic Phenotypic Parameters of Cancer Stem Cells to Affect Tumourigenicity. *PLoS Computational Biology,* 10**,** e1003433.

SELIGMAN, A. M., SHEAR, M. J. & ALEXANDER, L. 1939. Studies in Carcinogenesis: VIII. Experimental Production of Brain Tumors in Mice with Methylcholanthrene. *The American Journal of Cancer,* 37**,** 364.

SEQUERRA, E. B. 2014. Subventricular zone progenitors in time and space: generating neuronal diversity. *Front Cell Neurosci,* 8**,** 434.

SEQUIST, L. V., WALTMAN, B. A., DIAS-SANTAGATA, D., DIGUMARTHY, S., TURKE, A. B., FIDIAS, P., BERGETHON, K., SHAW, A. T., GETTINGER, S., COSPER, A. K., AKHAVANFARD, S., HEIST, R. S., TEMEL, J., CHRISTENSEN, J. G., WAIN, J. C., LYNCH, T. J., VERNOVSKY, K., MARK, E. J., LANUTI, M., IAFRATE, A. J., MINO-KENUDSON, M. & ENGELMAN, J. A. 2011. Genotypic and Histological Evolution of Lung Cancers Acquiring Resistance to EGFR Inhibitors. *Science Translational Medicine,* 3**,** 75ra26-75ra26.

SEQUIST, L. V., YANG, J. C.-H., YAMAMOTO, N., O'BYRNE, K., HIRSH, V., MOK, T., GEATER, S. L., ORLOV, S., TSAI, C.-M., BOYER, M., SU, W.-C., BENNOUNA, J., KATO, T., GORBUNOVA, V., LEE, K. H., SHAH, R., MASSEY, D., ZAZULINA, V., SHAHIDI, M. & SCHULER, M. 2013. Phase III Study of Afatinib or Cisplatin Plus Pemetrexed in Patients With Metastatic Lung Adenocarcinoma With EGFR Mutations. *Journal of Clinical Oncology,* 31**,** 3327-3334.

SEVER, R. & BRUGGE, J. S. 2015. Signal Transduction in Cancer. *Cold Spring Harbor Perspectives in Medicine,* 5**,** a006098-a006098.

SHCHERBO, D., SHEMIAKINA, I. I., RYABOVA, A. V., LUKER, K. E., SCHMIDT, B. T., SOUSLOVA, E. A., GORODNICHEVA, T. V., STRUKOVA, L., SHIDLOVSKIY, K. M., BRITANOVA, O. V., ZARAISKY, A. G., LUKYANOV, K. A., LOSCHENOV, V. B., LUKER, G. D. & CHUDAKOV, D. M. 2010. Near-infrared fluorescent proteins. *Nature Methods,* 7**,** 827-829.

SHIMOMURA, O., JOHNSON, F. H. & SAIGA, Y. 1962. Extraction, purification and properties of aequorin, a bioluminescent protein from the luminous hydromedusan, Aequorea. *J Cell Comp Physiol,* 59**,** 223-39.

SILVER, D. J. & LATHIA, J. D. 2018. Revealing the glioma cancer stem cell interactome, one niche at a time. *J Pathol,* 244**,** 260-264.

SINGH, S. K., CLARKE, I. D., TERASAKI, M., BONN, V. E., HAWKINS, C., SQUIRE, J. & DIRKS, P. B. 2003. Identification of a cancer stem cell in human brain tumors. *Cancer Research,* 63**,** 5821-5828.

SINGH, S. K., HAWKINS, C., CLARKE, I. D., SQUIRE, J. A., BAYANI, J., HIDE, T., HENKELMAN, R. M., CUSIMANO, M. D. & DIRKS, P. B. 2004. Identification of human brain tumour initiating cells. *Nature,* 432**,** 396-401.

SNIPPERT, H. J. & CLEVERS, H. 2011. Tracking adult stem cells. *Embo Reports,* 12**,** 113-122.

SNIPPERT, H. J., VAN DER FLIER, L. G., SATO, T., VAN ES, J. H., VAN DEN BORN, M., KROON-VEENBOER, C., BARKER, N., KLEIN, A. M., VAN RHEENEN, J., SIMONS, B. D. & CLEVERS, H. 2010. Intestinal Crypt Homeostasis Results from Neutral Competition between Symmetrically Dividing Lgr5 Stem Cells. *Cell,* 143**,** 134-144.

SNUDERL, M., FAZLOLLAHI, L., LE, L. P., NITTA, M., ZHELYAZKOVA, B. H., DAVIDSON, C. J., AKHAVANFARD, S., CAHIL, D. P., ALDAPE, K. D., BETENSKY, R. A., LOUIS, D. N. & IAFRATE, A. J. 2011. Mosaic Amplification of Multiple Receptor Tyrosine Kinase Genes in Glioblastoma. *Cancer Cell,* 20**,** 810-817.

SON, M. J., WOOLARD, K., NAM, D. H., LEE, J. & FINE, H. A. 2009. SSEA-1 is an enrichment marker for tumor-initiating cells in human glioblastoma. *Cell Stem Cell,* 4**,** 440-52.

SONABEND, A. M., BANSAL, M., GUARNIERI, P., LEI, L., AMENDOLARA, B., SODERQUIST, C., LEUNG, R., YUN, J., KENNEDY, B., SISTI, J., BRUCE, S., BRUCE, R., SHAKYA, R., LUDWIG, T., ROSENFELD, S., SIMS, P. A., BRUCE, J. N., CALIFANO, A. & CANOLL, P. 2014. The Transcriptional Regulatory Network of Proneural Glioma Determines the Genetic Alterations Selected during Tumor Progression. *Cancer Research,* 74**,** 1440-1451.

SONABEND, A. M., YUN, J., LEI, L., LEUNG, R., SODERQUIST, C., CRISMAN, C., GILL, B. J., CARMINUCCI, A., SISTI, J., CASTELLI, M., SIMS, P. A., BRUCE, J. N. & CANOLL, P. 2013. Murine cell line model of proneural glioma for evaluation of anti-tumor therapies. *J Neurooncol,* 112**,** 375-82.

SORIANO, P. 1999. Generalized lacZ expression with the ROSA26 Cre reporter strain. *Nat Genet,* 21**,** 70-1.

SOTTORIVA, A., SPITERI, I., PICCIRILLO, S. G. M., TOULOUMIS, A., COLLINS, V. P., MARIONI, J. C., CURTIS, C., WATTS, C. & TAVARE, S. 2013. Intratumor heterogeneity in human glioblastoma reflects cancer evolutionary dynamics. *Proceedings of the National Academy of Sciences of the United States of America,* 110**,** 4009-4014.

SOUSA, J. F. D., SERAFIM, R. B., FREITAS, L. M. D., FONTANA, C. R. & VALENTE, V. 2020. DNA repair genes in astrocytoma tumorigenesis, progression and therapy resistance. *Genetics and Molecular Biology,* 43.

STICHEL, D., EBRAHIMI, A., REUSS, D., SCHRIMPF, D., ONO, T., SHIRAHATA, M., REIFENBERGER, G., WELLER, M., HÄNGGI, D., WICK, W., HEROLD-MENDE, C., WESTPHAL, M., BRANDNER, S., PFISTER, S. M., CAPPER, D., SAHM, F. & VON DEIMLING, A. 2018. Distribution of EGFR amplification, combined chromosome 7 gain and chromosome 10 loss, and TERT promoter mutation in brain tumors and their potential for the reclassification of IDHwt astrocytoma to glioblastoma. *Acta Neuropathol,* 136**,** 793-803.

STUPP, R., HEGI, M. E., MASON, W. P., VAN DEN BENT, M. J., TAPHOORN, M. J., JANZER, R. C., LUDWIN, S. K., ALLGEIER, A., FISHER, B., BELANGER, K., HAU, P., BRANDES, A. A., GIJTENBEEK, J., MAROSI, C., VECHT, C. J., MOKHTARI, K., WESSELING, P., VILLA, S.,

EISENHAUER, E., GORLIA, T., WELLER, M., LACOMBE, D., CAIRNCROSS, J. G. & MIRIMANOFF, R. O. 2009. Effects of radiotherapy with concomitant and adjuvant temozolomide versus radiotherapy alone on survival in glioblastoma in a randomised phase III study: 5-year analysis of the EORTC-NCIC trial. *Lancet Oncol,* 10**,** 459-66.

SUN, G., LI, Z., RONG, D., ZHANG, H., SHI, X., YANG, W., ZHENG, W., SUN, G., WU, F., CAO, H., TANG, W. & SUN, Y. 2021. Single-cell RNA sequencing in cancer: Applications, advances, and emerging challenges. *Molecular Therapy - Oncolytics,* 21**,** 183-206.

SUVA, M. L., RHEINBAY, E., GILLESPIE, S. M., PATEL, A. P., WAKIMOTO, H., RABKIN, S. D., RIGGI, N., CHI, A. S., CAHILL, D. P., NAHED, B. V., CURRY, W. T., MARTUZA, R. L., RIVERA, M. N., ROSSETTI, N., KASIF, S., BEIK, S., KADRI, S., TIROSH, I., WORTMAN, I., SHALEK, A. K., ROZENBLATT-ROSEN, O., REGEV, A., LOUIS, D. N. & BERNSTEIN, B. E. 2014. Reconstructing and reprogramming the tumor-propagating potential of glioblastoma stem-like cells. *Cell,* 157**,** 580-94.

SWAIDAN, M. Y., HUSSAINI, M., SULTAN, I. & MANSOUR, A. 2012. Radiological findings in gliosarcoma. A single institution experience. *Neuroradiol J,* 25**,** 173-80.

SZERLIP, N. J., PEDRAZA, A., CHAKRAVARTY, D., AZIM, M., MCGUIRE, J., FANG, Y. Q., OZAWA, T., HOLLAND, E. C., HUSE, J. T., JHANWAR, S., LEVERSHA, M. A., MIKKELSEN, T. & BRENNAN, C. W. 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**,** 3041-3046.

TAN, C. S., CHO, B. C. & SOO, R. A. 2016. Next-generation epidermal growth factor receptor tyrosine kinase inhibitors in epidermal growth factor receptor -mutant non-small cell lung cancer. *Lung Cancer,* 93**,** 59-68.

TCHOGHANDJIAN, A. L., BAEZA, N., COLIN, C., CAYRE, M., METELLUS, P., BECLIN, C., OUAFIK, L. H. & FIGARELLA-BRANGER, D. 2010. A2B5 Cells from Human Glioblastoma have Cancer Stem Cell Properties. *Brain Pathology,* 20**,** 211-221.

THAKKAR, J. P., DOLECEK, T. A., HORBINSKI, C., OSTROM, Q. T., LIGHTNER, D. D., BARNHOLTZ-SLOAN, J. S. & VILLANO, J. L. 2014. Epidemiologic and molecular prognostic review of glioblastoma. *Cancer Epidemiol Biomarkers Prev,* 23**,** 1985-96.

TOMER, R., YE, L., HSUEH, B. & DEISSEROTH, K. 2014. Advanced CLARITY for rapid and high-resolution imaging of intact tissues. *Nature Protocols,* 9**,** 1682-1697.

TUNICI, P., BISSOLA, L., LUALDI, E., POLLO, B., CAJOLA, L., BROGGI, G., SOZZI, G. & FINOCCHIARO, G. 2004. *Molecular Cancer,* 3**,** 25.

UCHIDA, N., BUCK, D. W., HE, D., REITSMA, M. J., MASEK, M., PHAN, T. V., TSUKAMOTO, A. S., GAGE, F. H. & WEISSMAN, I. L. 2000. Direct isolation of human central nervous system stem cells. *Proceedings of the National Academy of Sciences,* 97**,** 14720-14725.

URRUTICOECHEA, A., ALEMANY, R., BALART, J., VILLANUEVA, A., VINALS, F. & CAPELLA, G. 2010. Recent advances in cancer therapy: an overview. *Curr Pharm Des,* 16**,** 3-10.

VALOR, L. M. & HERVÁS-CORPIÓN, I. 2020. The Epigenetics of Glioma Stem Cells: A Brief Overview. *Frontiers in Oncology,* 10.

VASAN, N., BASELGA, J. & HYMAN, D. M. 2019. A view on drug resistance in cancer. *Nature,* 575**,** 299-309.

VENDRAMIN, R., LITCHFIELD, K. & SWANTON, C. 2021. Cancer evolution: Darwin and beyond. *The EMBO Journal,* 40.

VERHAAK, R. G. W., HOADLEY, K. A., PURDOM, E., WANG, V., QI, Y., WILKERSON, M. D., MILLER, C. R., DING, L., GOLUB, T., MESIROV, J. P., ALEXE, G., LAWRENCE, M., O'KELLY, M., TAMAYO, P., WEIR, B. A., GABRIEL, S., WINCKLER, W., GUPTA, S., JAKKULA, L., FEILER, H. S., HODGSON, J. G., JAMES, C. D., SARKARIA, J. N., BRENNAN, C., KAHN, A., SPELLMAN, P. T., WILSON, R. K., SPEED, T. P., GRAY, J. W., MEYERSON, M., GETZ, G., PEROU, C. M., HAYES, D. N. & NETWORK, C. G. A. R. 2010. Integrated Genomic Analysis Identifies Clinically Relevant Subtypes of Glioblastoma Characterized by Abnormalities in PDGFRA, IDH1, EGFR, and NF1. *Cancer Cell,* 17**,** 98-110.

VESCOVI, A. L., GALLI, R. & REYNOLDS, B. A. 2006. Brain tumour stem cells. *Nature Reviews Cancer,* 6**,** 425-436.

VINEL, C., ROSSER, G., GUGLIELMI, L., CONSTANTINOU, M., POMELLA, N., ZHANG, X., BOOT, J. R., JONES, T. A., MILLNER, T. O., DUMAS, A. A., RAKYAN, V., REES, J., THOMPSON, J. L., VUONONVIRTA, J., NADKARNI, S., EL ASSAN, T., ALEY, N., LIN, Y. Y., LIU, P., NELANDER, S., SHEER, D., MERRY, C. L. R., MARELLI-BERG, F., BRANDNER, S. & MARINO, S. 2021. Comparative epigenetic analysis of tumour initiating cells and syngeneic EPSC-derived neural stem cells in glioblastoma. *Nat Commun,* 12**,** 6130.

WADDINGTON, C. H. 2014. *The strategy of the genes*, Routledge.

WAGNER, D. E. & KLEIN, A. M. 2020. Lineage tracing meets single-cell omics: opportunities and challenges. *Nature Reviews Genetics,* 21**,** 410-427.

WAGNER, D. E., WEINREB, C., COLLINS, Z. M., BRIGGS, J. A., MEGASON, S. G. & KLEIN, A. M. 2018. Single-cell mapping of gene expression landscapes and lineage in the zebrafish embryo. *Science,* 360**,** 981-987.

WALKER, M., PATEL, K. & STAPPENBECK, T. 2009. The stem cell niche. *The Journal of Pathology,* 217**,** 169-180.

WANG, J., CAZZATO, E., LADEWIG, E., FRATTINI, V., ROSENBLOOM, D. I. S., ZAIRIS, S., ABATE, F., LIU, Z., ELLIOTT, O., SHIN, Y.-J., LEE, J.-K., LEE, I.-H., PARK, W.-Y., EOLI, M., BLUMBERG, A. J., LASORELLA, A., NAM, D.-H., FINOCCHIARO, G., IAVARONE, A. & RABADAN, R. 2016. Clonal evolution of glioblastoma under therapy. *Nature Genetics,* 48**,** 768-776.

WANG, J., SAKARIASSEN, P. Ø., TSINKALOVSKY, O., IMMERVOLL, H., BØE, S. O., SVENDSEN, A., PRESTEGARDEN, L., RØSLAND, G., THORSEN, F., STUHR, L., MOLVEN, A., BJERKVIG, R. & ENGER, P. Ø. 2008. CD133 negative glioma cells form tumors in nude rats and give rise to CD133 positive cells. *International Journal of Cancer,* 122**,** 761-768.

WANG, J., ZHANG, K., XU, L. & WANG, E. 2011. Quantifying the Waddington landscape and biological paths for development and differentiation. *Proceedings of the National Academy of Sciences of the United States of America,* 108**,** 8257-8262.

WANG, L., BABIKIR, H., MÜLLER, S., YAGNIK, G., SHAMARDANI, K., CATALAN, F., KOHANBASH, G., ALVARADO, B., DI LULLO, E., KRIEGSTEIN, A., SHAH, S., WADHWA, H., CHANG, S. M., PHILLIPS, J. J., AGHI, M. K. & DIAZ, A. A. 2019. The Phenotypes of Proliferating Glioblastoma Cells Reside on a Single Axis of Variation. *Cancer Discovery,* 9**,** 1708-1719.

WEBER, K., BARTSCH, U., STOCKING, C. & FEHSE, B. 2008. A Multicolor Panel of Novel Lentiviral "Gene Ontology" (LeGO) Vectors for Functional Gene Analysis. *Molecular Therapy,* 16**,** 698-706.

WEBER, K., THOMASCHEWSKI, M., BENTEN, D. & FEHSE, B. 2012. RGB marking with lentiviral vectors for multicolor clonal cell tracking. *Nat Protoc,* 7**,** 839-49.

WEBER, K., THOMASCHEWSKI, M., WARLICH, M., VOLZ, T., CORNILS, K., NIEBUHR, B., TÄGER, M., LÜTGEHETMANN, M., POLLOK, J. M., STOCKING, C., DANDRI, M., BENTEN, D. & FEHSE, B. 2011. RGB marking facilitates multicolor clonal cell tracking. *Nat Med,* 17**,** 504-9.

WEISSMAN, T. A. & PAN, Y. A. 2015. Brainbow: New Resources and Emerging Biological Applications for Multicolor Genetic Labeling and Analysis. *Genetics,* 199**,** 293-306.

WICK, W., WELLER, M., VAN DEN BENT, M., SANSON, M., WEILER, M., VON DEIMLING, A., PLASS, C., HEGI, M., PLATTEN, M. & REIFENBERGER, G. 2014. MGMT testing--the challenges for biomarker-based glioma treatment. *Nat Rev Neurol,* 10**,** 372-85.

WINDREM, M. S., SCHANZ, S. J., ZOU, L., CHANDLER-MILITELLO, D., KUYPERS, N. J., NEDERGAARD, M., LU, Y., MARIANI, J. N. & GOLDMAN, S. A. 2020. Human Glial Progenitor Cells Effectively Remyelinate the Demyelinated Adult Brain. *Cell Rep,* 31**,** 107658.

WORBY, C. A. & DIXON, J. E. 2014. Pten. *Annu Rev Biochem,* 83**,** 641-69.

WU, J. W., TURCOTTE, R., ALT, C., RUNNELS, J. M., TSAO, H. & LIN, C. P. 2016. Defining Clonal Color in Fluorescent Multi-Clonal Tracking. *Scientific Reports,* 6**,** 24303.

XIONG, F., OBHOLZER, N. D., NOCHE, R. R. & MEGASON, S. G. 2015. Multibow: Digital Spectral Barcodes for Cell Tracing. *PLOS ONE,* 10**,** e0127822.

XU, Y., STAMENKOVIC, I. & YU, Q. 2010. CD44 Attenuates Activation of the Hippo Signaling Pathway and Is a Prime Therapeutic Target for Glioblastoma. *Cancer Research,* 70**,** 2455-2464.

YANG, B., TREWEEK, J. B., KULKARNI, R. P., DEVERMAN, B. E., CHEN, C. K., LUBECK, E., SHAH, S., CAI, L. & GRADINARU, V. 2014. Single-Cell Phenotyping within Transparent Intact Tissue through Whole-Body Clearing. *Cell,* 158**,** 945-958.

YU, K., HU, Y., WU, F., GUO, Q., QIAN, Z., HU, W., CHEN, J., WANG, K., FAN, X., WU, X., RASKO, J. E., FAN, X., IAVARONE, A., JIANG, T., TANG, F. & SU, X.-D. 2020. Surveying brain

tumor heterogeneity by single-cell RNA-sequencing of multi-sector biopsies. *National Science Review,* 7**,** 1306-1318.

YU, Z., PESTELL, T. G., LISANTI, M. P. & PESTELL, R. G. 2012. Cancer stem cells. *Int J Biochem Cell Biol,* 44**,** 2144-51.

ZAFAR, H., LIN, C. & BAR-JOSEPH, Z. 2020. Single-cell lineage tracing by integrating CRISPR-Cas9 mutations with transcriptomic data. *Nature Communications,* 11.

ZHANG, Y., DUBE, C., GIBERT, M., CRUICKSHANKS, N., WANG, B., COUGHLAN, M., YANG, Y., SETIADY, I., DEVEAU, C., SAOUD, K., GRELLO, C., OXFORD, M., YUAN, F. & ABOUNADER, R. 2018. The p53 Pathway in Glioblastoma. *Cancers,* 10**,** 297.

ZHANG, Y., PUSCH, S., INNES, J., SIDLAUSKAS, K., ELLIS, M., LAU, J., EL-HASSAN, T., ALEY, N., LAUNCHBURY, F., RICHARD-LOENDT, A., DEBOER, J., CHEN, S., WANG, L., VON DEIMLING, A., LI, N. & BRANDNER, S. 2019. Mutant IDH Sensitizes Gliomas to Endoplasmic Reticulum Stress and Triggers Apoptosis via miR-183-Mediated Inhibition of Semaphorin 3E. *Cancer Res,* 79**,** 4994-5007.

ZHANG, Y., ZENG, F., HAN, X., WENG, J. & GAO, Y. 2020. Lineage tracing: technology tool for exploring the development, regeneration, and disease of the digestive system. *Stem Cell Research & Therapy,* 11.

ZHENG, H. W., YING, H. Q., YAN, H. Y., KIMMELMAN, A. C., HILLER, D. J., CHEN, A. J., PERRY, S. R., TONON, G., CHU, G. C., DING, Z. H., STOMMEL, J. M., DUNN, K. L., WIEDEMEYER, R., YOU, M. J. J., BRENNAN, C., WANG, Y. A., LIGON, K. L., WONG, W. H., CHIN, L. & DEPINHO, R. A. 2008. p53 and Pten control neural and glioma stem/progenitor cell renewal and differentiation. *Nature,* 455**,** 1129-U13.

ZHU, Y., GUIGNARD, F., ZHAO, D. W., LIU, L., BURNS, D. K., MASON, R. P., MESSING, A. & PARADA, L. F. 2005. Early inactivation of p53 tumor suppressor gene cooperating with NF1 loss induces malignant astrocytoma. *Cancer Cell,* 8**,** 119-130.

ZIMMERMAN, H. M. & ARNOLD, H. 1943. Experimental Brain Tumors: II. Tumors Produced with Benzpyrene. *Am J Pathol,* 19**,** 939-55.

Chapter 7: Supplemental Figures

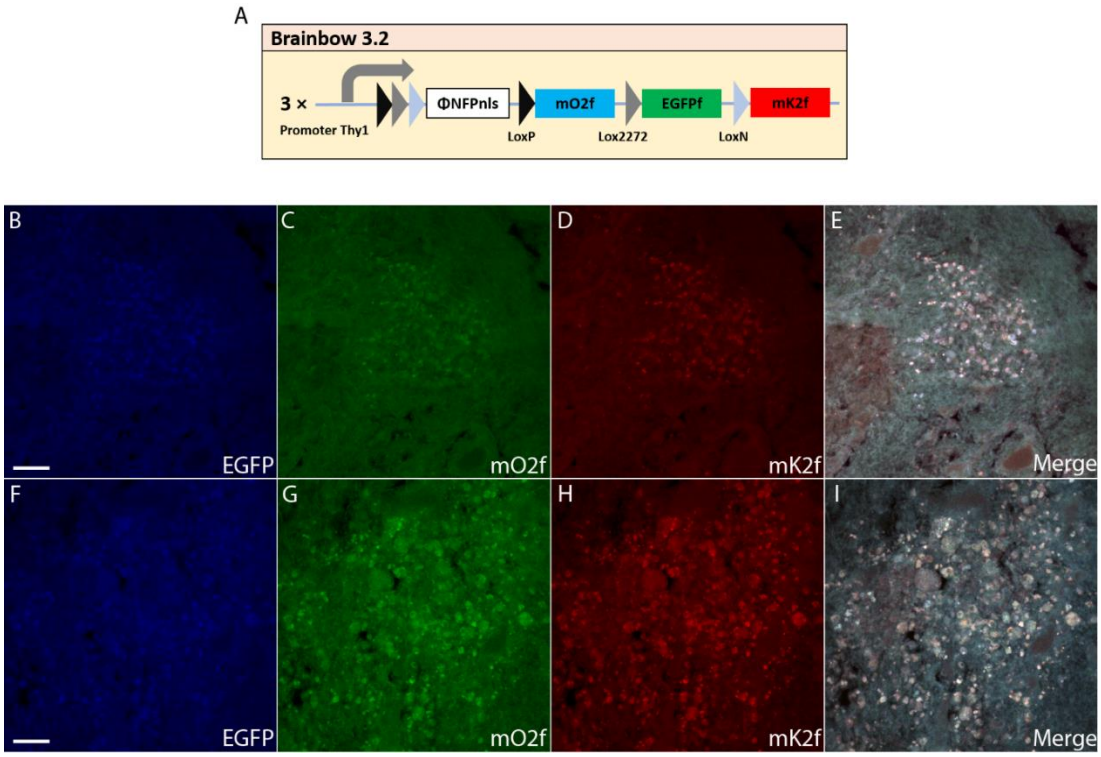


Figure 7.1: Alternative multicolour lineage tracing approach using Brianbow3.2 locus. (A) Brainbow makes use of incompatible LoxP sites to ensure expression of 1 out of 3 possible fluorophores. Multiple integrations of the cassette leads to overlapping expression and diversification of potential labels. (B-I) Example of imaging endogenous fluorescence from a Brainbow3.2/Pten/P53 animal injected with PIC. It was not possible to distinguish differently labelled cells

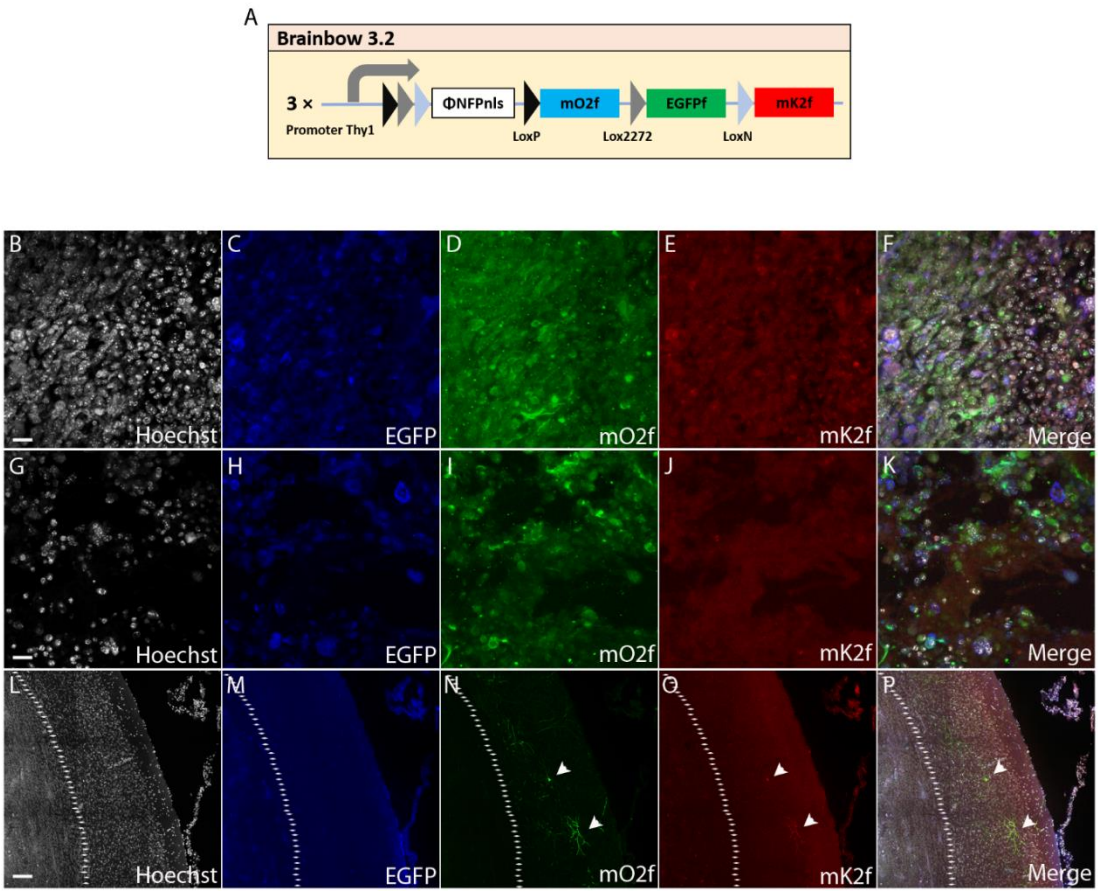


Figure 7.2: Antibody amplified fluorophore signal from PIC induced tumours in Brianbow3.2 animals. (A) Schematic of Brainbow3.2 locus. (B – K) Examples of staining for EGFP, mO2f and mK2f, again it was not possible to distinguish individual fluorophores. (L – F) Image showing well defined mO2f and mK2f signal from cortical neurons outside the tumour, white dotted line designates tumour border.

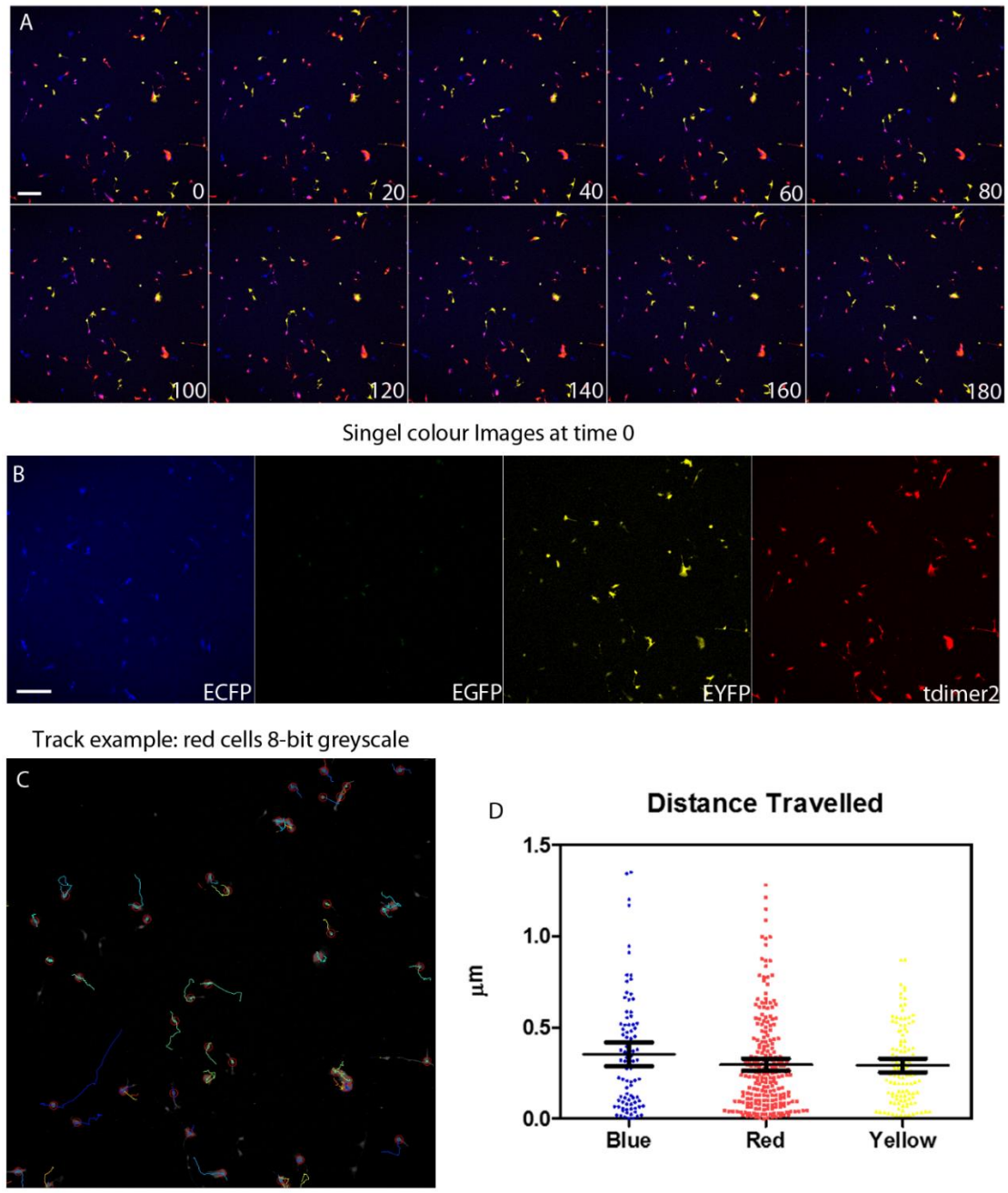


Figure 7.3: In development cell tracking approach for investigating motility of tumour cells homozygous for the r26-confetti locus. (A) Images in 20-minute increments of live r26-confetti labelled tumours, images were acquired every 5 minutes. (B) Split channels from Image at timepoint 0 in A. (C) Example of cell traces for yellow cells acquired using Fiji Trackmate. (D) Quantification of distance travelled from 6 sampled regions for ECFP, tdimer2 and EYFP expressing cells. Homozygous confetti cells were used therefore, many of the readings overlap between colours, we would like to apply this approach to cells after being cultured in tumours.

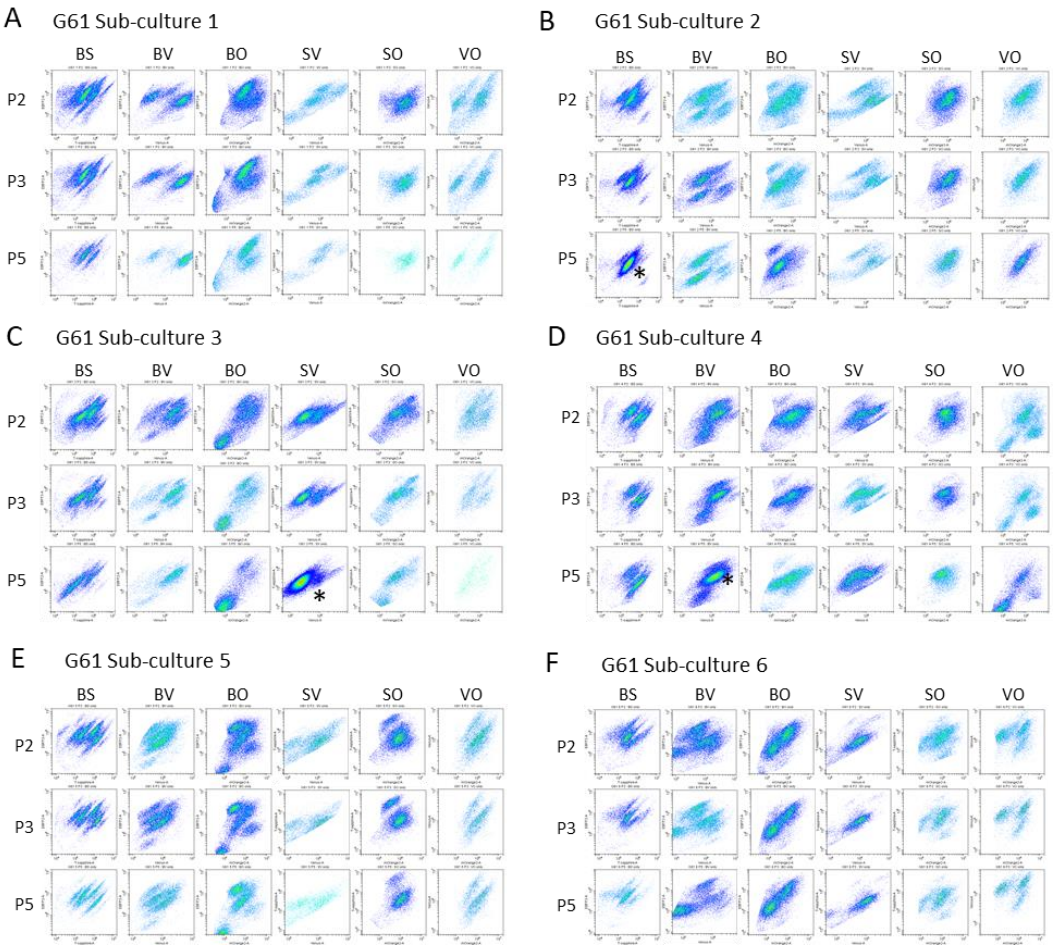


Figure 7.4: (A-F) Raw flow-cytometry data of all sub-cultures across passages showing clonal streaks. * denote clones that showed particularly strong outgrowth.

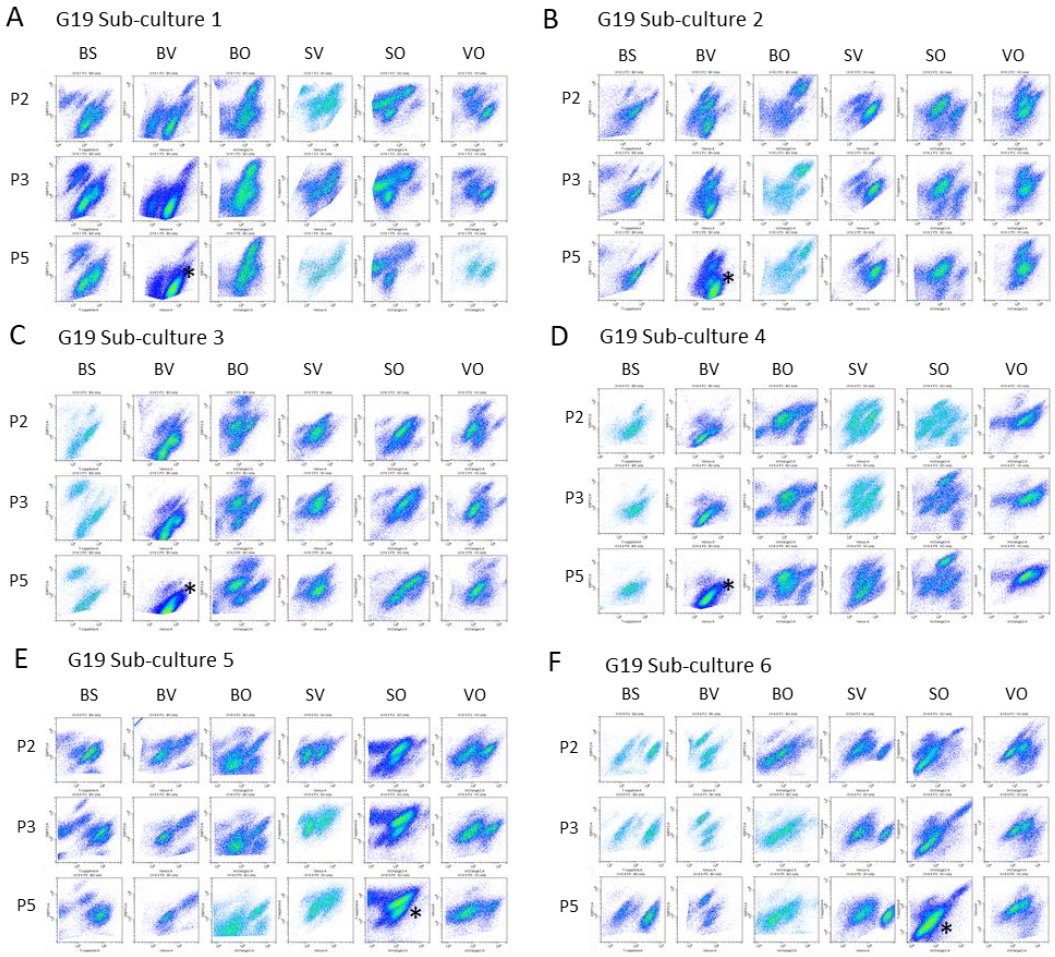


Figure 7.5: Dot-plots of all G19 clonal sub-cultures across P2, P3, and P5. (A-F) Raw flow-cytometry data of all sub-cultures across passages showing clonal streaks. * denote clones that showed particularly strong outgrowth.

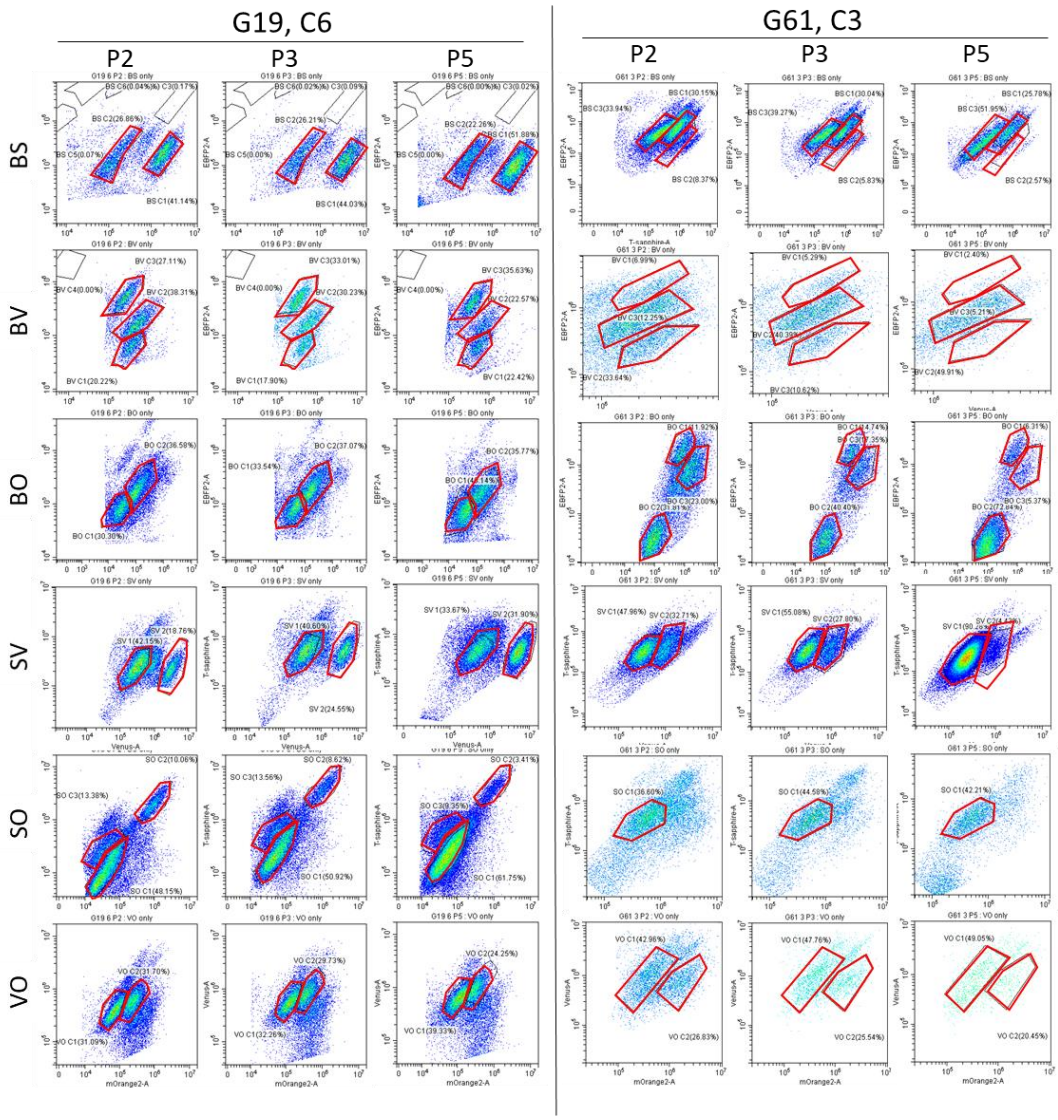


Figure 7.6: flow cytometry of barcode labelled cells to identify clonal population for selection. Each barcoded doublet (BS, BV, PO, SV, SO, VO) was identified and dominant subpopulations were "gated" to highlight clonal populations. Passaging of cells (P2, P3, and P5) show changing populations with emergence, growth and also reduction, represented by marked clusters in these flow diagrams.

Figure 7.7: plot of the surface marker profiles (phenotypes) of barcode-label populations. Data represent all 6 subcultures C1-C6, of cell lines G19 (A) and G61 (B). The selection process of barcode doublets is shown in Figure 7. The figure shows a selection of up to 5 barcode label doublets. Selection of fewer label doublets results in void areas. Boxed areas in G61-C3 and G61-C4 correspond to divergent and predominant populations in G61 sub-cultures 3 and 4.

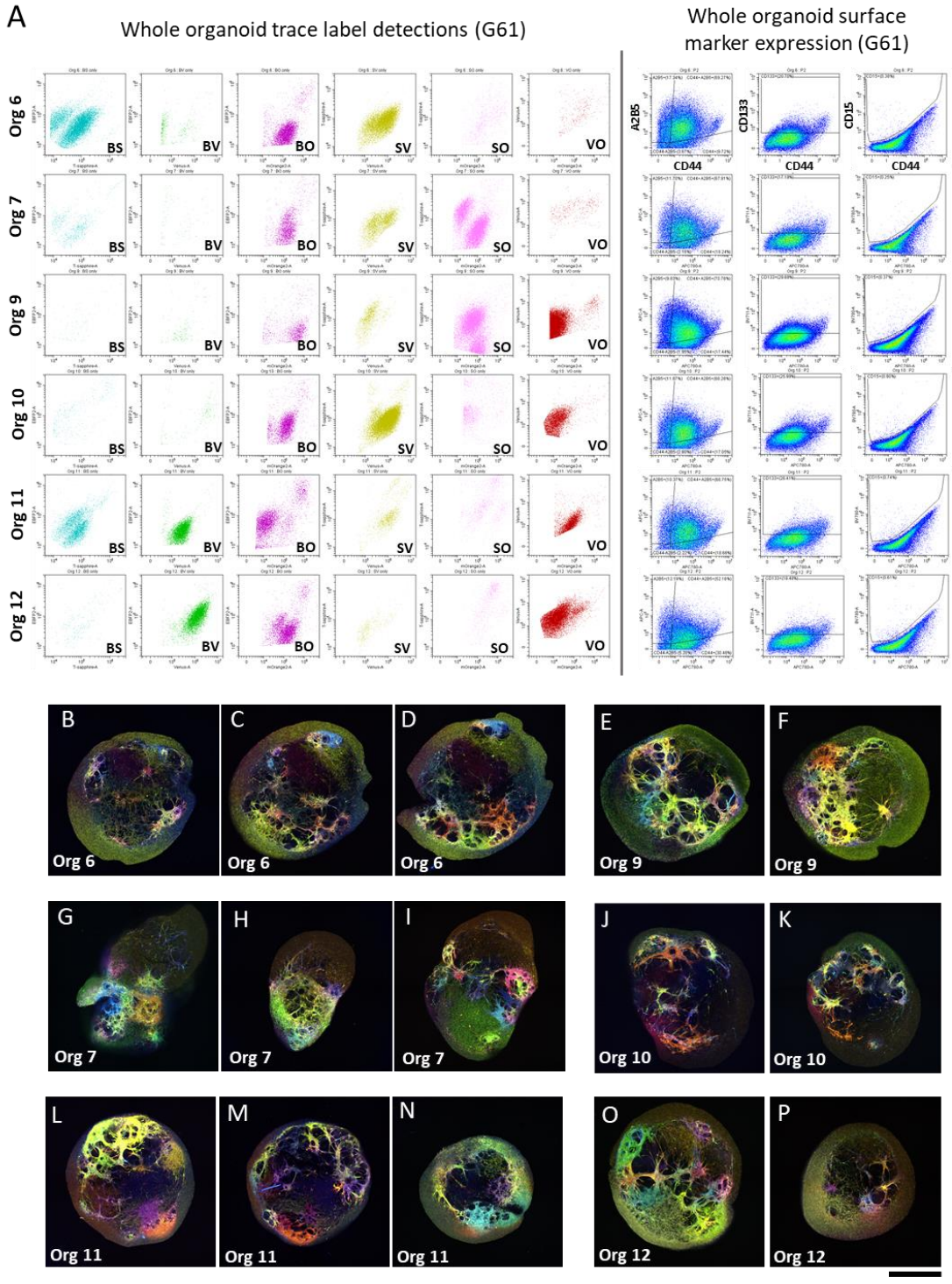


Figure 7.8: Summary of dataset collected for each of the G61 organoids grown from bulk diversely labelled cells. (A) Representative examples of flow cytometry streaks of barcode populations of different organoids on left and distribution of cell surface marker data on right. (B – P) show examples of serial vibratome sections and fluorescent imaging of living organoids corresponding to data displayed in part A.

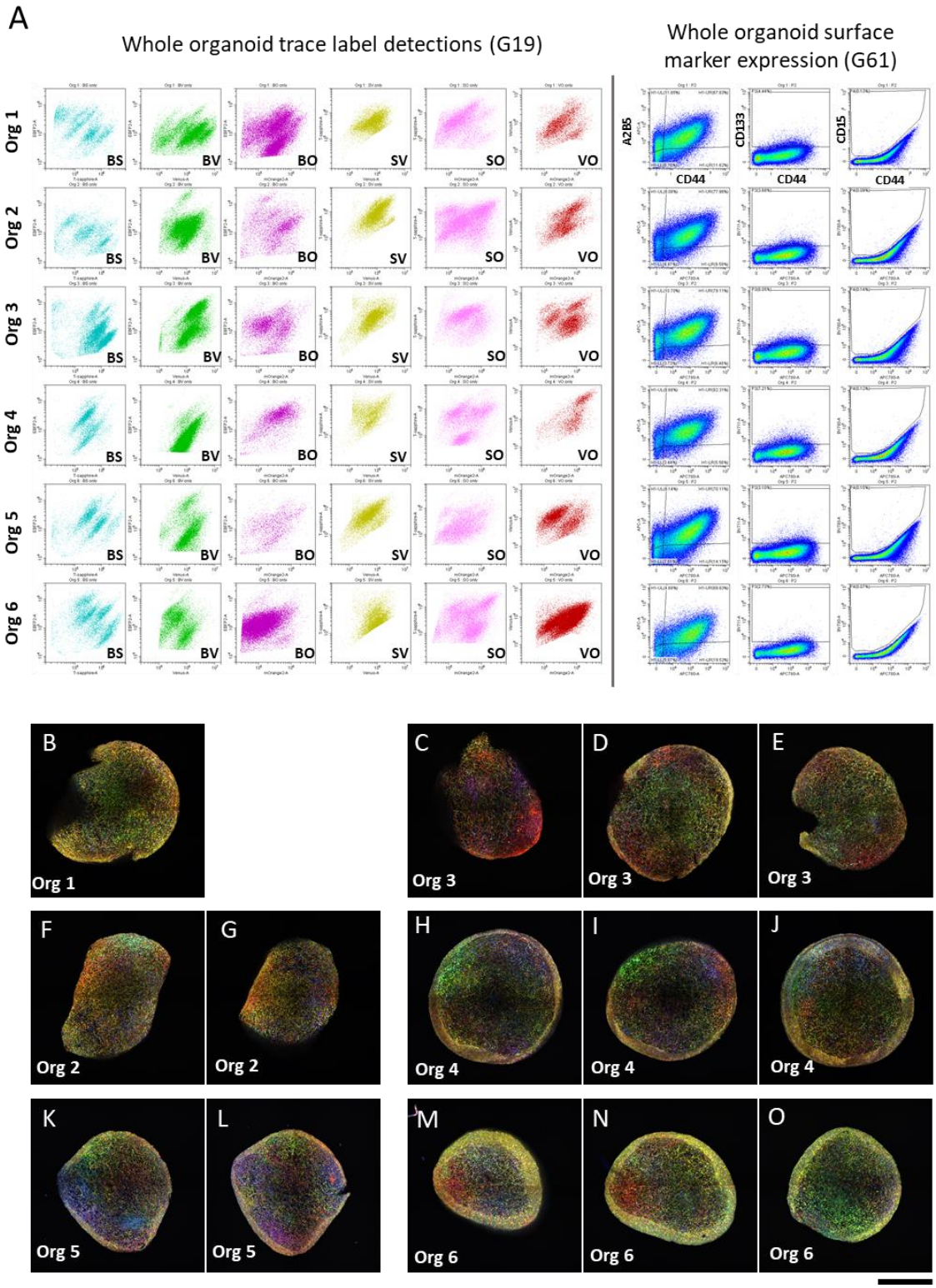


Figure 7.9: Summary of dataset collected for each of the G19 organoids grown from bulk diversely labelled cells. (A) Representative examples of flow cytometry streaks of barcode populations of different organoids on left and distribution of cell surface marker data on right. (B – P) show examples of serial vibratome sections and fluorescent imaging of living organoids corresponding to data displayed in part A.

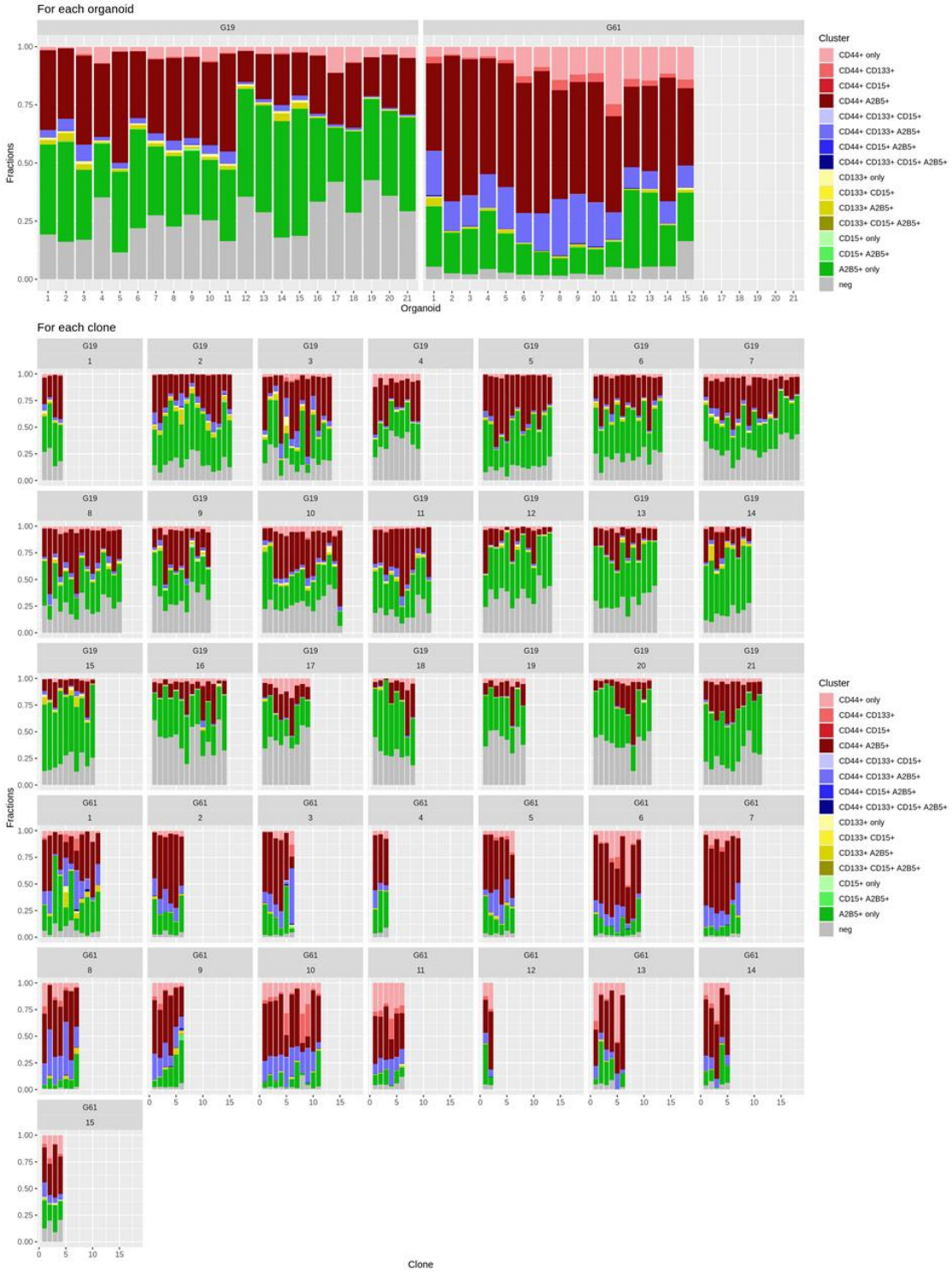


Figure 7.10 Summary of surface marker data from diversely labelled organoids (G61 and G19): **(**Top panel) distributions of marker phenotypes for all cells of different organoids. (Bottom Panel) Distirbution of marker phenotypes of different clones from eac organoid for G19 (top) and G61 (below).

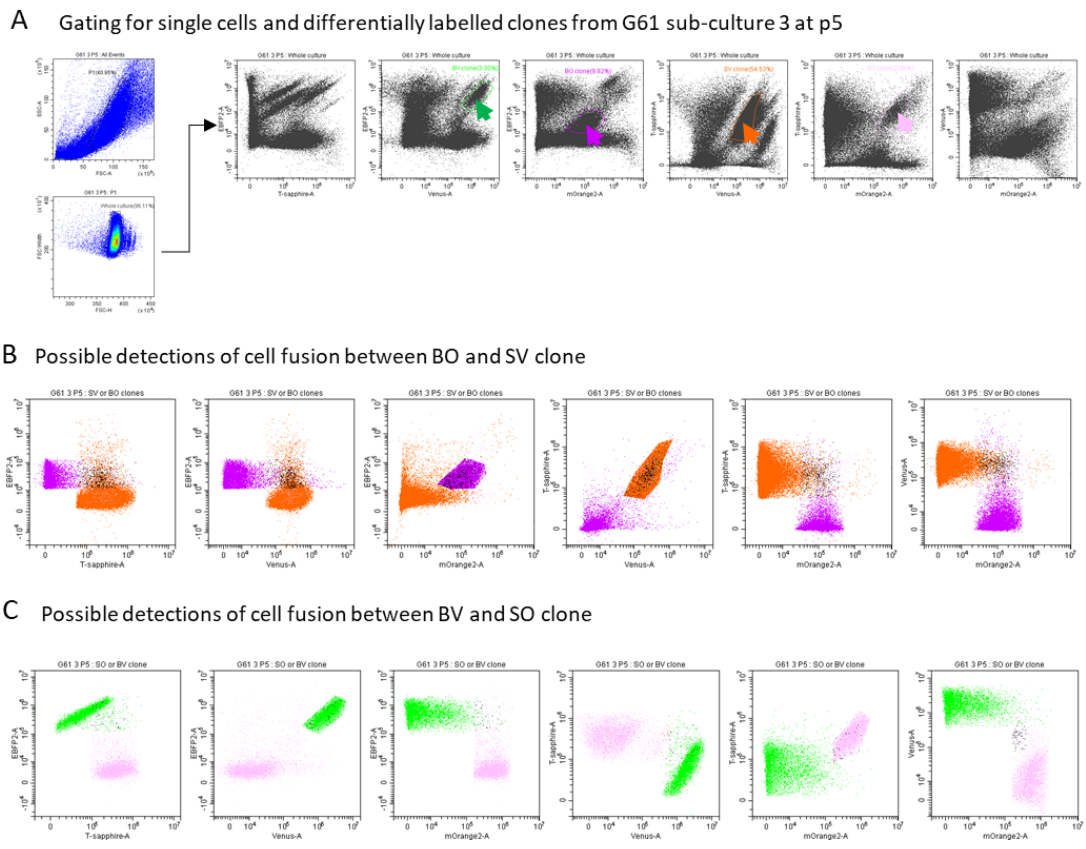


Figure 7.11: Potential flow cytometry evidence of cell fusion events between two clones. (A) Gating for single cells and different clone in G61 sub-culture 3 at P5. Green arrow and gate (BV clone), Purple arrow and gate (BO clone), Orange arrow and gate (SV clone), Pink arrow and gate (SO clone). (B) Combined population of either BO gated cells (Purple) or SV gated cells (Orange), a number of events (coloured in black) occur in the fluorescence space of both clones. While cells also bleed into other colour channels perhaps indicating sharing of extra-cellular vesicles between differently coloured clones. (C) Demonstrating the observation outlined in B, occurring between BV clone (green) and SO clone (pink), with small number of events occurring in gates for both clones (Black). It is also possible these are a product of reading doublet cells not fully detached but any inference would require further inspection.

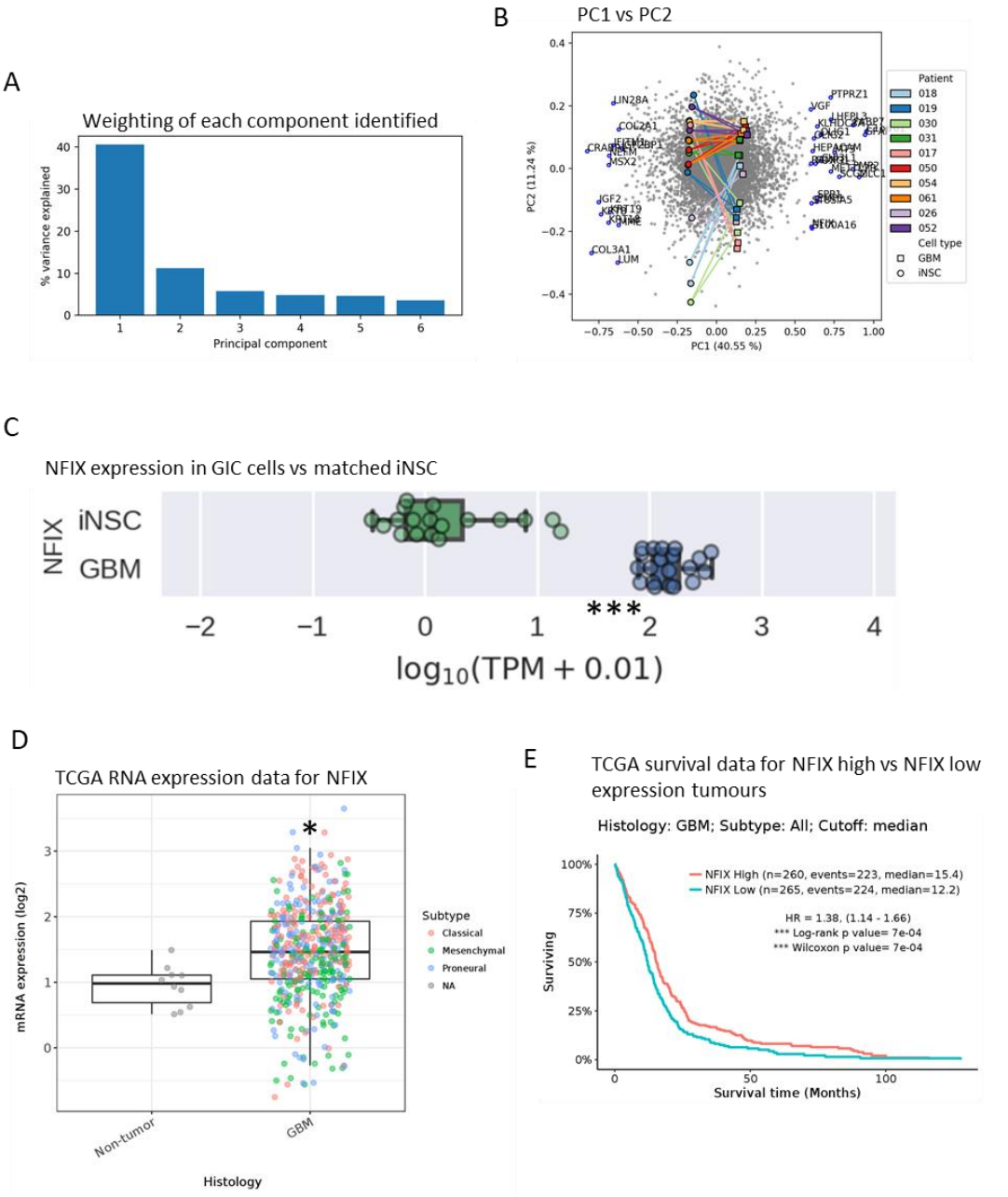


Figure 7.12: Patient matched glioma CSCs/iNSC expression comparisons reveals NFIX as a commonly aberrantly expressed gene in glioma CSCs. (A) Summary of differential gene expression principle component analysis showing the level of variance explained by each component. (B) Graphical plot of PC1 vs PC2 highlighting certain outlier genes and level of variance difference between patient matched iNSC and glioma CSC. (C) Summary of NFIX expression data for (n = 18) pairwise comparisons of glioma CSC and iNSC show NFIX consistently overexpressed in glioma CSC. (D) TCGA mRNA expression levels of NFIX also report significant overexpression in GBM tissue compared to non-tumour tissue (Pairwise t-test with Tukeys honest significant difference, p = 0.026 (CI = 95%). (E) Survival plot from TCGA data comparing high NFIX and low NFIX expressing tumour cases. High NFIX expression has statistically significant effect of increasing survival time.

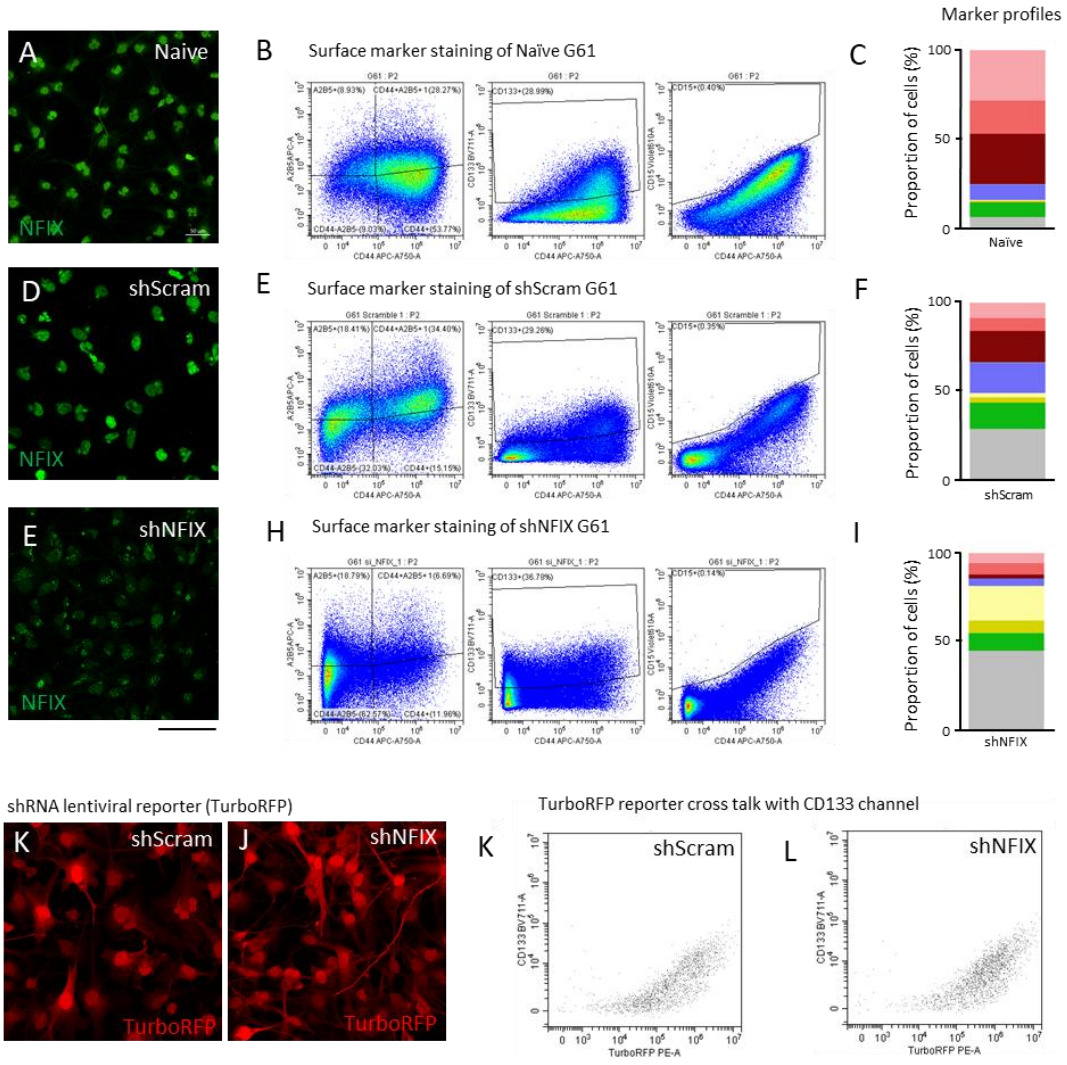
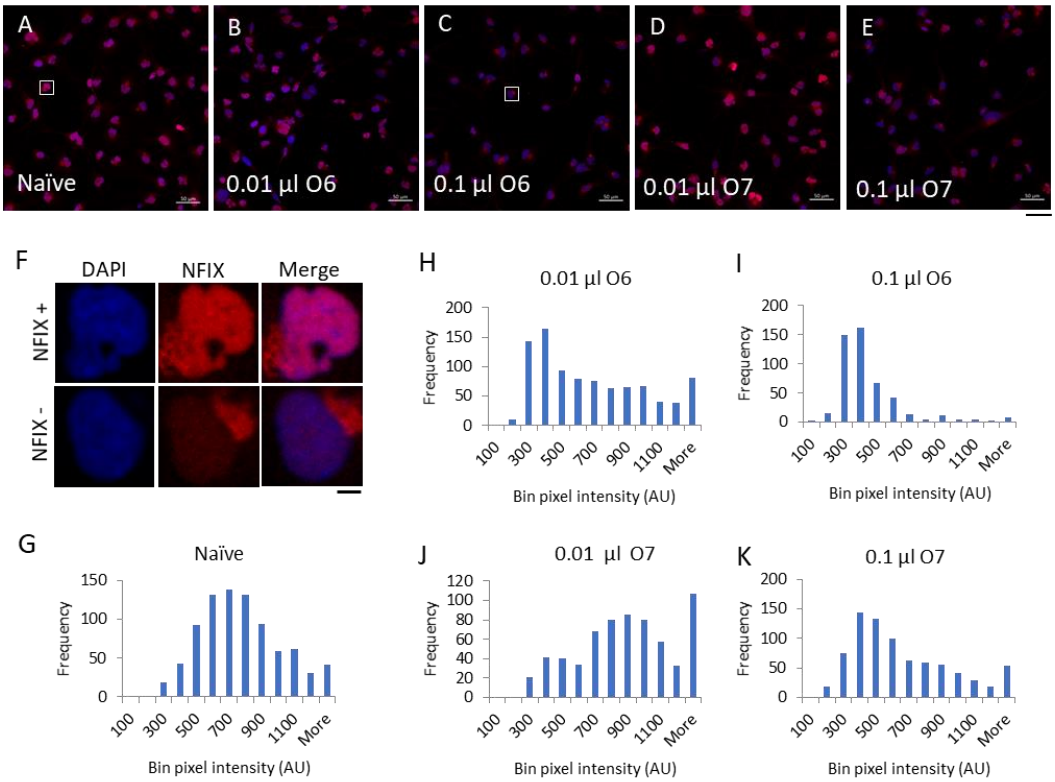


Figure 7.13: KD of NFIX in line G61 reduces CD44+ expression and increases overall marker negative fraction. (A, D & E) immunofluorescent staining for NFIX detects KD after exposure to combined shRNA's, scale bars = 50 µm. (B, E & H) Flow cytometry plots of Naïve, shScram and shNFIX G61 cells respectively. (C, F & I) Quantification of marker profiles displayed in B, E and H show loss of CD44+ marker profiles and expansion of negative cellular fraction (n = 3 per line). (K & J) Confocal imaging of fluorescent reporter for shRNA viral vectors. (K & L) Demonstration of how turboRFP bleeds through into the CD133 detection channel.



L

Condition	Total cell detections	Cells < 450 AU	Proportion infected	Titre (PFU)***	Infection MOI***
Naive	843	NA	NA	NA	NA
0.01 µl O6	923	373	0.40	2.02×10^9	0.24
0.1 µl O6	490	365	0.74	3.72×10^8	2.4
0.01 µl O7	647	81	0.13	6.26×10^8	0.082
0.1 µl O7	790	305	0.38	1.93×10^8	0.82

Titre = No. of cells(volume of virus/proportion infected)
*MOI = (titre x volume)/No. of cells

Figure 7.14: Immunofluorescence based assay confirms NFIX KD and and provides relative titre determination. (A – E) Merged immunofluorescence images of Hoechst and NFIX (Alexa594) staining in cells exposed to different concentrations of shRNA viruses O6 and O7 which each carry a different shRNA specific to NFIX mRNA. (F) Examples of NFIX staining in cells positive and negative for NFIX expression. (H – I) Distributions of cellular intensities for NFIX stain in naïve cells (G) and populations exposed to shRNAs targeting NFIX. Note the skew in distribution toward more negative cell intensities. Intensity distributions inform threshold intensity for cells positive or negative for NFIX expression, manually determined as 450 AU. (L) Table outlining quantification of NFIX KD in cells exposed to different virus' and different amounts of virus. These values used to determine relative titre of different virus' preparations.

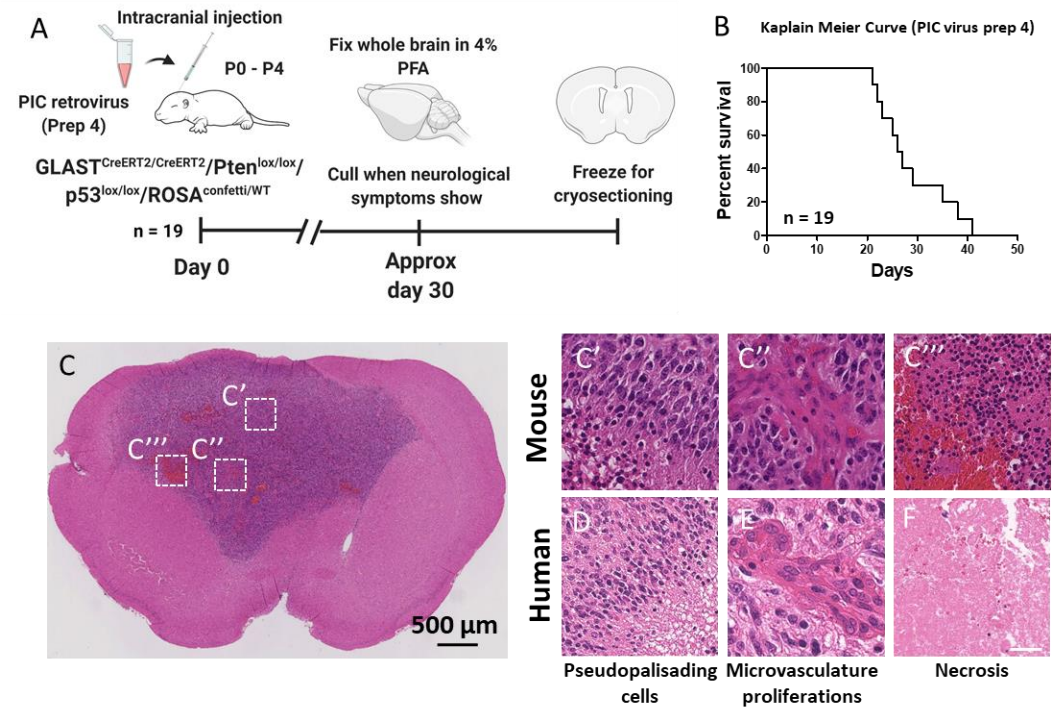


Figure 7.15: Intracranial injections of titred PIC virus' achieve rapid tumour formation with histological structures resembling GBM. (A) Schematic showing experimental time course using PIC virus (Prep 4) and animal genotypes. (B) Kaplain Meier curve showing GLAST$^{CreERT2/CreERT2}$/PTEN$^{lox/lox}$/P53$^{lox/lox}$/ROSA$^{confetti/wt}$ animal survival, median survival rate of 27.5 days. (C) H&E histology of PIC induced tumour showing larger tumour originating from the left ventricle. Boxes indicate images shown in C'; Pseudopalisades, C''; Microvasculature proliferations and C'''; Necrotic areas which are histological hallmarks of GBM. (D – E) show corresponding structures in human GBM acquired from pathology specimen