Longitudinal genomic and epigenomic changes in glioblastoma brain tumours

Submitted in accordance with the requirements for the degree of
Doctor of Philosophy

Yousef Mohammed Alghamdi

University of Leeds

School of Medicine

Leeds Institute of Medical Research

August 2025

The candidate confirms that the work submitted is his own and that appropriate credit has been given where reference has been made to the work of others.

Acknowledgements

I would like to express my deepest gratitude to my supervisors, Dr. Lucy Stead, Dr. James Poulter, and Dr. Georgette Tanner. Lucy, thank you for your incredible support, enthusiasm throughout this journey. Your dedication to research and your group has been truly inspiring. James, I'm especially grateful for your insight, constructive feedback, and your calm, thoughtful guidance at every stage. Georgette, although you left academia during my PhD, I will never forget the impact you had on my project and on me personally. Your support and patience during the early stages of my research gave me confidence and direction when I needed it most.

To my wife, thank you from the bottom of my heart. Balancing a PhD with the responsibilities of family life is challenging for anyone, but raising a child with Autism Spectrum Disorder has added an extra layer of complexity and strength to our journey. Your resilience, patience, and unconditional support have been extraordinary. I know this has not been easy, and I am endlessly grateful for everything you've done to keep our family grounded and loved through it all.

To my children, thank you for your love, laughter, and for being the light in my life. I know that at times this PhD took me away from you more than I would have liked. I missed moments that I can never get back, but every step of this journey was for you, and I hope one day you'll understand and feel proud of it.

To my parents, thank you for your prayers, encouragement, and unwavering belief in me. Your love and support have been the foundation that helped carry me through.

I would also like to thank the Health Affairs at the Ministry of National Guard for sponsoring my PhD and making this opportunity possible.

Lastly, I want to thank the Glioma Genomics group. Working alongside such a generous and collaborative group of people made this experience so much richer. Whether it was troubleshooting code or analysing data, I always felt supported and never alone.

Finally, during this PhD, I experienced the heartbreaking loss of my sister. I wish she were here with me now to share this moment. If years could be shared, I would gladly give her some of mine just to keep her in our lives a little longer. Her memory has walked beside me through every stage of this journey and this thesis is, in part, for her.

Abstract

Glioblastoma is a highly aggressive brain tumour with a poor prognosis and inevitable recurrence following standard treatment. Understanding the molecular basis of treatment resistance and tumour progression is critical to improving therapeutic outcomes. This PhD thesis aimed to explore the genetic and epigenetic evolution of GBM through three phases: optimisation of sequencing pipelines, identification of altered biological pathways under therapeutic pressure, and DNA methylation profiling of recurrent disease.

In the first phase, whole-exome and whole-genome sequencing pipelines were optimised for use with challenging clinical material, including FFPE-derived samples. Custom adjustments, including the correction of overlapping read pairs and mitigation of FFPE artefacts, significantly improved variant calling accuracy and tumour mutational burden estimation.

The second phase focused on uncovering treatment-associated pathway alterations using paired primary and recurrent GBM samples from 2 cohorts. By tracking changes in variant allele frequency pre- and post-treatment, I identified variants either selected for or against by therapy. Pathway analysis using PathScore revealed several significant biological pathways under selection pressure, notably involving the ERBB signalling family. Disruption of ERBB4 signalling was associated with treatment sensitivity, suggesting that its inhibition may enhance therapeutic efficacy in a subset of patients.

The final phase applied genome-wide DNA methylation profiling using Illumina Infinium arrays. Although recurrence-associated changes were subtle at the cohort level, stratification by JARID2-related transcriptional response revealed subtype-specific epigenetic dynamics. A quadrant-based analysis highlighted greater methylation shifts in Down responders, potentially reflecting adaptive responses to treatment.

Altogether, this work provides insight into GBM evolution under therapy, demonstrating how both genetic and epigenetic shifts contribute to recurrence. The identification of ERBB4 signalling as potentially associated with treatment sensitivity highlights a candidate pathway that warrants further functional validation. Future work, including targeted experimental studies of ERBB4 function, alongside single-cell and spatial profiling, may reveal actionable therapeutic insights and refine strategies to overcome treatment resistance.

Table of Contents

List of Figures

List of Tables

Abbreviations

CCF	Cancer cell fraction
CDKN2A/B	Cyclin-dependent kinase inhibitor 2A/B
CN	Copy number
CNA	Copy number aberration - somatic
DMP	Differentially methylated probes
DMR	Differentially methylated regions
ecDNA	Extrachromosomal DNA
EGFR	Epidermal growth factor receptor
GATK	Genome analysis toolkit
GBM	Glioblastoma
GLASS	Glioma Longitudinal AnalySiS consortium
IDH(1/2)	Isocitrate dehydrogenase (1/2)
InDel	Insertion or deletion
ITH	Intratumour heterogeneity
MAPK	Mitogen-activated protein kinase
MGMT	O6-methylguanine-DNA methyl-transferase
NGS	Next generation sequencing
PI3K	Phosphoinositide 3-kinases
PTEN	Phosphatase and tensin homolog
RTK	Receptor tyrosine kinase
SNV	Single nucleotide variant - somatic
TERT	Telomerase reverse transcriptase
TMZ	Temozolomide
VAF	Variant allele fraction
WES	Whole-exome sequencing
WGS	Whole-genome sequencing
wt	Wild-type

CHAPTER 1

1.1 Glioblastoma

1.1.1 Overview

Glioblastoma (GBM) is the most common and aggressive malignant primary brain tumour in adults, accounting for the majority of glioma-related deaths and a significant portion of all primary brain tumours (Ramirez et al., 2013, Perry and Wesseling, 2016). It is classified by the World Health Organisation (WHO) as a grade IV astrocytoma, which reflects its high level of malignancy, characterised by rapid growth, necrosis, and microvascular proliferation (Philips et al., 2018, Hanif et al., 2017). GBM belongs to the family of diffuse gliomas, which infiltrate surrounding brain tissue, making them particularly difficult to remove surgically (Perry and Wesseling, 2016, Ostrom et al., 2014).

Although GBM is relatively rare compared to other cancers, it has one of the worst prognoses. It is the most common primary malignant brain tumour in adults, accounting for approximately 45% of all primary malignant brain and central nervous system (CNS) tumours. The age-adjusted incidence rate is about 3 per 100,000 population per year, and the incidence increases markedly with age from around 1 per 100,000 in individuals under 40 years to over 10–15 per 100,000 in those above 75 years (Philips et al., 2018, Tamimi AF, 2017, Brodbelt et al., 2015, Ostrom et al., 2019). Males are affected more frequently than females. GBM is most often diagnosed in older adults, with a median age at diagnosis of around 64–65 years as shown in Figure 1-1 (Wen et al., 2021, Grochans et al., 2022).

The clinical presentation varies but often includes headaches, seizures, cognitive decline, or focal neurological symptoms (Wen et al., 2020, Chang et al., 2005). Diagnosis typically involves imaging, especially contrast-enhanced MRI, followed by histopathological and molecular analysis of a biopsy or resected tissue (Gilard et al., 2021). Even with aggressive treatment—usually combining surgical resection, radiotherapy, and chemotherapy—median overall survival is just 12 to 15 months, with a five-year survival rate under 7% (Philips et al., 2018, Tamimi AF, 2017).

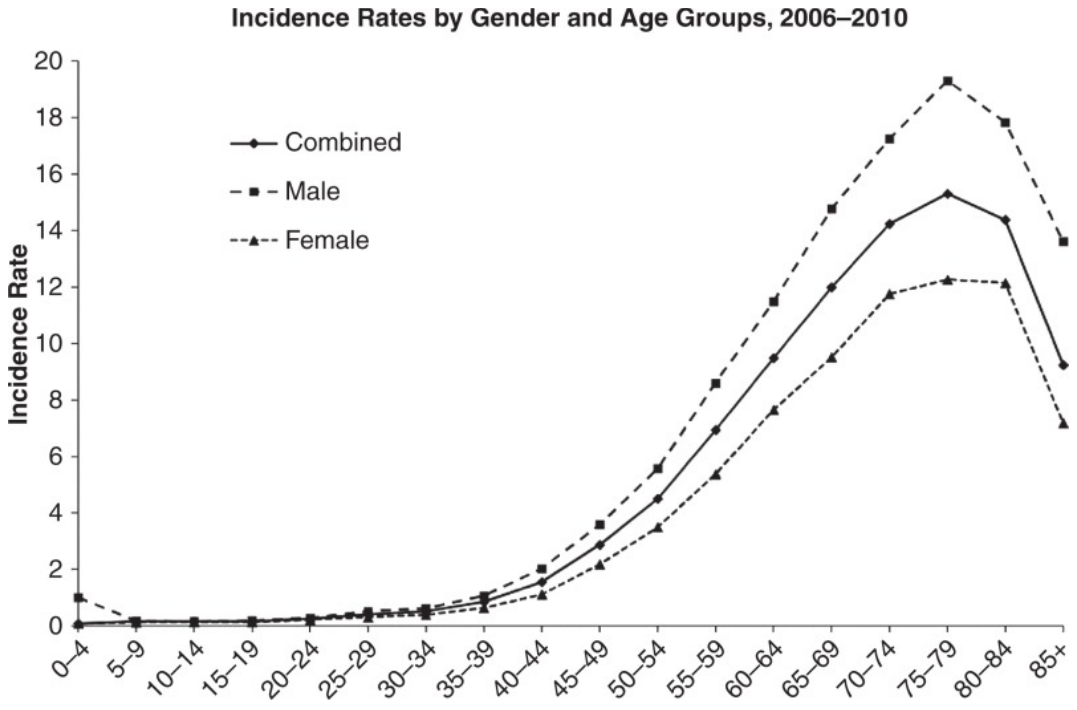


Figure 1-1: Age-adjusted and age-specific incidence rates for glioblastoma

*Adopted from (Tamimi AF, 2017), published under the Creative Commons Attribution 4.0 International (CC BY 4.0). https://creativecommons.org/licenses/by-nc/4.0/.

1.1.2 Types of GBM

GBMs can be divided into two main categories: primary GBMs, which arise de novo without any evidence of a precursor lesion, and secondary GBMs, which progress from lower-grade astrocytomas (grade II or III) (Alireza Mansouri, 2017, Ohgaki and Kleihues, 2013). Primary GBMs account for the majority of cases (~90%) and tend to occur in older adults (mean age ~62), whereas secondary GBMs are more often found in younger individuals (mean age ~45) (Alireza Mansouri, 2017, Ohgaki and Kleihues, 2013).

Although they appear histologically similar, the two types differ at the molecular level. Primary GBMs are usually IDH-wildtype and frequently show amplification of *EGFR*, *MDM2*, loss of heterozygosity on chromosome 10q, and loss of function mutations in *PTEN* (Ohgaki and Kleihues, 2007, Crespo et al., 2015). Secondary GBMs are typically IDH-mutant and are associated with mutations in *TP53* and *RB*, as well as LOH on chromosomes 17p and 19q (Ohgaki and Kleihues, 2007, Crespo et al., 2015).

This molecular distinction has become central to the classification system introduced by the WHO, which now emphasises IDH mutation status over clinical history (Dymova et al., 2021, Hanif et al.,

2017). According to this framework, only IDH-wildtype astrocytomas with grade IV features are now labelled as GBM.

Despite increasing knowledge about the molecular subtypes of GBM, treatment remains largely uniform across patients. All patients, regardless of subtype, are offered the standard of care: surgical resection followed by radiotherapy and temozolomide chemotherapy (Wen et al., 2020, Dymova et al., 2021). One molecular marker that does inform therapy is methylation of the O-6-methylguanine-DNA methyltransferase (*MGMT)* promoter. When methylated, the promoter silences *MGMT* expression, reducing tumour cells' ability to repair the DNA damage caused by temozolomide, thereby improving treatment response (Hegi et al., 2005).

1.1.3 Recurrence

Recurrence is a near-universal feature of GBM and remains one of the most difficult challenges in its treatment. Even with maximal standard therapy, tumours typically recur within 6–9 months (Stupp et al., 2005). In approximately 80% of patients, recurrence occurs close to the original resection site, suggesting that residual infiltrative cells are responsible (Birzu et al., 2020). Recurrent GBMs are often more aggressive, more treatment-resistant, and harder to manage. Less than half of patients are eligible for repeat surgery, and the survival benefit is modest—typically extending life by just 5 to 11 months (Ringel et al., 2016, Barbagallo et al., 2008, Suchorska et al., 2016, Woodroffe et al., 2020).

Over time, GBMs undergo significant molecular evolution. They become more genetically diverse and often acquire new mutations between diagnosis and recurrence. Multiple studies have shown that different regions of the same tumour can harbour distinct mutations and expression patterns, a concept known as intratumoural heterogeneity (ITH) (Sottoriva et al., 2013). This heterogeneity makes GBMs highly adaptable and contributes to therapy resistance.

Recent advances such as single-cell RNA sequencing have revealed that GBM cells exist in multiple transcriptional states and are capable of transitioning between them in response to treatment or environmental stress (Neftel et al., 2019). These findings provide insight into why GBM recurs and why existing treatments fail to provide long-term disease control.

Treatment options for recurrent GBM remain limited. While re-operation or re-irradiation may be considered in selected cases, these are not viable for all patients. Other therapies, such as

bevacizumab or lomustine, offer only modest benefit, and clinical trial enrolment is often the most realistic option (Wen et al., 2020, Gilbert, 2011). Unlike TMZ, which primarily methylates DNA, lomustine is a nitrosourea that also works as an alkylating agent, but its mechanism of action is distinct as it forms DNA interstrand cross-links, which are particularly difficult for tumour cells to repair (Weller and Le Rhun, 2020). In contrast, bevacizumab is a monoclonal antibody that targets vascular endothelial growth factor (VEGF), a key signalling protein in the formation of new blood vessels (angiogenesis). By inhibiting VEGF, bevacizumab effectively starves the tumour of its blood supply, thereby slowing its growth rather than directly causing cell death (Garcia et al., 2020).

Given the inevitability of recurrence and the lack of curative treatments, understanding how GBM evolves—and how heterogeneity shapes this evolution—is critical. This thesis focuses specifically on IDH-wildtype GBM and investigates the molecular and epigenetic changes that occur between primary and recurrent disease, with the aim of contributing to a better understanding of treatment resistance and tumour progression.

1.2 Treatment of Glioblastoma

1.2.1 Standard Therapy

The current standard of care for glioblastoma was established nearly two decades ago and remains largely unchanged since the introduction of the Stupp protocol in 2005 (Stupp et al., 2005, Wang et al., 2021). This approach combines maximal safe surgical resection, followed by radiotherapy and concomitant temozolomide (TMZ) chemotherapy, with additional adjuvant TMZ cycles (Wang et al., 2021).

Surgical resection is typically performed within two weeks of diagnosis and aims to remove as much of the tumour as possible without causing neurological damage (Muller et al., 2021). Techniques such as awake craniotomy and fluorescence-guided surgery using 5-aminolevulinic acid (5-ALA) are often employed to enhance tumour visibility and maximise resection margins (Zhang et al., 2020, Hadjipanayis and Stummer, 2019). Despite these efforts, infiltrative tumour cells inevitably remain in the brain tissue surrounding the resection cavity (Berens and Giese, 1999).

Following surgery, patients receive radiotherapy in daily fractions, delivered over a period of six weeks. Concurrently, TMZ is administered daily during the radiotherapy period (Wang et al., 2021). After a four-week break, patients begin adjuvant TMZ treatment (Stupp et al., 2005, Bjorland et al., 2021).

TMZ is an oral alkylating agent that readily crosses the blood-brain barrier due to its small, lipophilic structure (Arora and Somasundaram, 2019). At physiological pH, TMZ rapidly degrades into the active metabolite MTIC (5-(3-methyltriazen-1-yl)-imidazole-4-carboxamide), which then produces a methyl-diazonium ion. This highly reactive species methylates DNA at several sites, including N7 and O6 of guanine, and N3 of adenine (Arora and Somasundaram, 2019, Strobel et al., 2019). Among these, methylation at the O6 position of guanine (O6-MeG) is the most cytotoxic, as it results in replication errors and eventually triggers apoptosis through persistent mismatch repair cycles (Nagasaka et al., 2008, Thomas et al., 2017) Figure 1-2.

Despite its effectiveness in some patients, TMZ has limitations. Its short half-life (~1.9 hours) (Baker et al., 1999) and reliance on tumour sensitivity mean that not all patients benefit equally. The most consistent predictor of TMZ responsiveness is the methylation status of the MGMT promoter. When methylated, MGMT expression is suppressed, allowing O6-MeG lesions to persist and exert cytotoxic effects. Conversely, unmethylated MGMT allows the tumour to repair TMZ-induced damage, making the drug less effective (Sciuscio et al., 2011, Kitange et al., 2009).

Nonetheless, all patients are offered TMZ regardless of MGMT status, as responses are still occasionally seen in patients with unmethylated promoters (Weller et al., 2010). This may be due to variation in test sensitivity or tumour heterogeneity, which complicates accurate MGMT status determination (Choi et al., 2021).

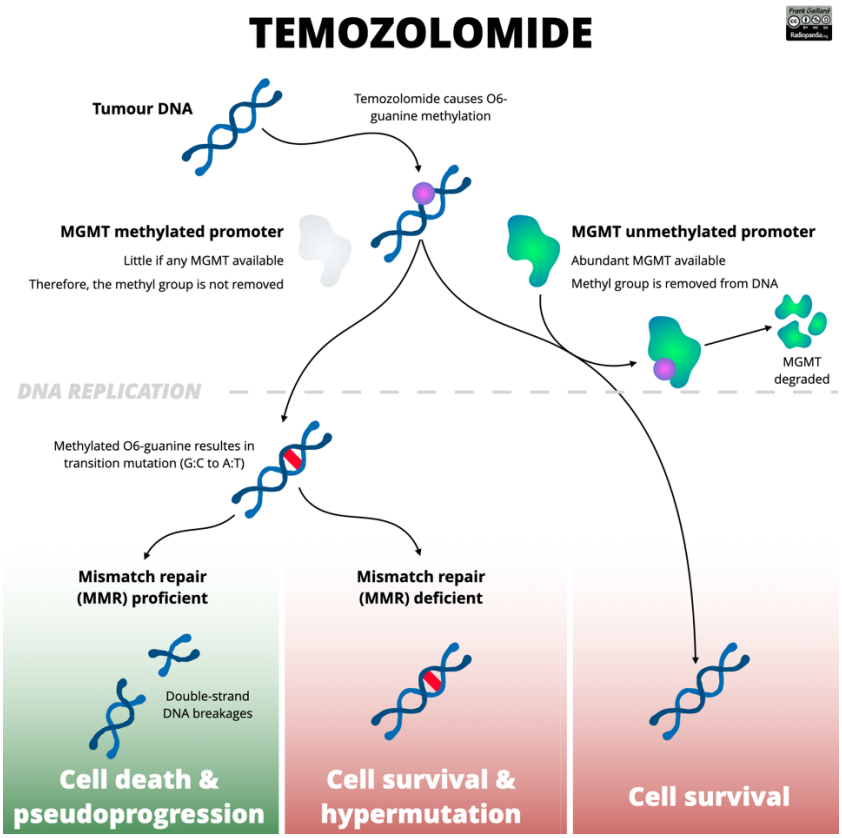


Figure 1-2: Temozolomide mechanism of action and resistance pathways in glioblastoma treatment.

TMZ is the standard chemotherapeutic agent used to treat glioblastoma. It induces DNA damage by adding a methyl group to the O6 position of guanine. In tumours with a methylated MGMT promoter, expression of the MGMT repair enzyme is reduced, allowing these methyl lesions to persist. During DNA replication, this leads to mismatches (G:C to A:T), which are detected by the mismatch repair (MMR) system. In MMR-proficient tumours, repeated attempts to correct the damage cause DNA double-strand breaks, leading to cell death and radiological condition known as pseudoprogression. However, MMR-deficient tumours tolerate these mismatches, enabling survival and promoting temozolomide-induced hypermutation. Tumours with an unmethylated MGMT promoter express active MGMT, which reverses the DNA methylation and confers resistance to TMZ.

*Adopted from (Gaillard, 2024)

1.2.2 Other Therapeutics

Despite numerous attempts to introduce new therapies for GBM, no alternative has demonstrated a clear clinical benefit in large-scale trials. Dozens of targeted agents, immunotherapies, and experimental drugs have progressed through early-phase clinical trials only to fail in phase III (Mandel et al., 2018). This lack of progress has been attributed to factors including poor translation from preclinical models, inter-patient variability, and the underlying heterogeneity of GBM (Bagley et al., 2022).

Some chemotherapeutic agents, such as etoposide and procarbazine, are still occasionally used in the recurrent setting, but there is no strong evidence that they improve survival outcomes in GBM (Wen et al., 2020). While both are chemotherapies, they have distinct mechanisms of action. Etoposide is a topoisomerase II inhibitor. It interferes with this enzyme, which is critical for DNA replication and repair, leading to DNA strand breaks and ultimately programmed cell death (Montecucco et al., 2015, Sevim et al., 2011). Procarbazine, in contrast, is a prodrug that is metabolised into an active alkylating agent. This agent primarily methylates DNA, similar to temozolomide, and also inhibits DNA, RNA, and protein synthesis, contributing to its cytotoxic effects (Kaina, 2023).

Similarly, targeted approaches such as Epidermal Growth Factor Receptor (EGFR) inhibitors, CAR-T cells, and vaccines targeting EGFRvIII (EGFR variant III, an isoform of EGFR with an in-frame deletion of exons 2-7) have shown initial promise, but ultimately failed due to mechanisms of adaptive resistance (O'Rourke et al., 2017, Nathanson et al., 2014). In many of these cases, tumours responded by eliminating or downregulating the targeted proteins, escaping the intended therapeutic effect (Brastianos et al., 2017).

1.2.3 Resistance

Resistance to standard treatment is one of the defining features of glioblastoma and the primary reason for its poor long-term prognosis. Although temozolomide increases survival by an average of 2–3 months, the majority of patients eventually relapse, and the disease becomes progressively harder to treat (Berens and Giese, 1999, Bjorland et al., 2021).

Mechanisms of resistance are multifactorial. At the molecular level, DNA repair pathways such as MGMT, mismatch repair (MMR), and base excision repair (BER) play important roles in diminishing TMZ efficacy (Lee, 2016). MGMT is the most prominent factor, but other contributors include epigenetic regulation, histone modifications, and miRNA-mediated silencing (Oldrini et al., 2020, Uno et al., 2011).

Interestingly, recent studies using matched primary and recurrent tumour samples have shown that genetic resistance mechanisms are relatively rare in GBM. Instead, many researchers now believe that transcriptional plasticity and cell state changes are the primary drivers of therapy resistance (Barthel et al., 2019, Neftel et al., 2019). Single-cell and lineage tracing studies have shown

that GBM cells can switch into slow-cycling, drug-tolerant states, often involving chromatin remodelling and developmental reprogramming (Eyler et al., 2020, Banelli et al., 2015).

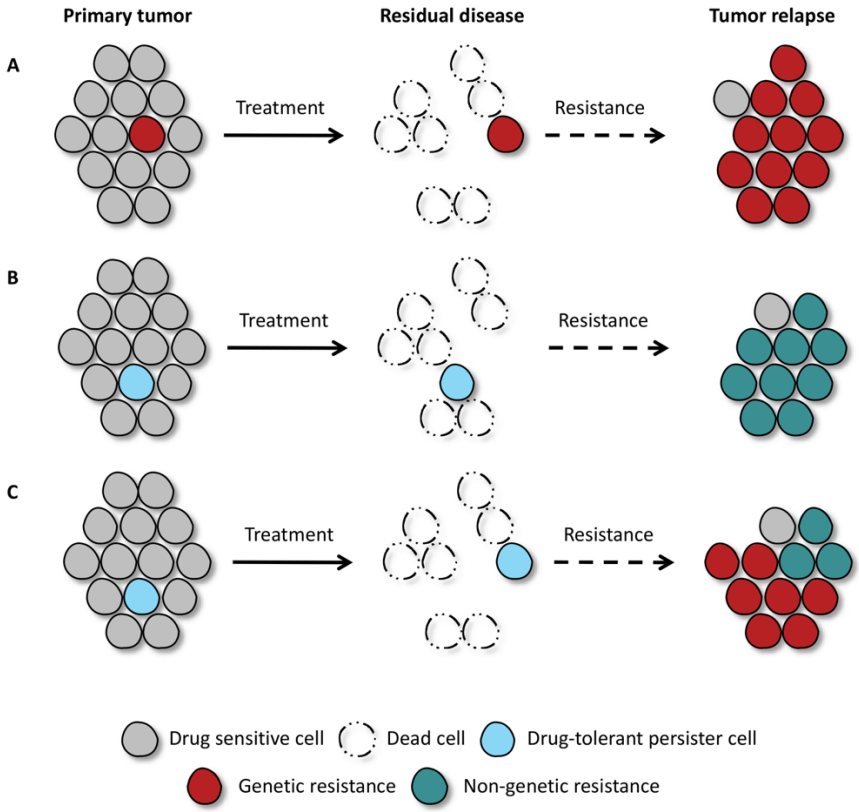


Figure 1-3: Models of cancer treatment resistance

Upon exposure to therapy, most cancer cells are eliminated, but a small subset can survive and eventually drive tumour relapse through distinct mechanisms.

(A) In the first scenario, treatment-resistant clones already exist before therapy begins. These genetically distinct subpopulations are not affected by the drug and expand over time, leading to recurrence.

(B) In the second scenario, no resistant mutations are initially present. However, a small number of drug-tolerant cells persist through a reversible, non-genetic state.

(C) These surviving cells can later acquire resistance through genetic mutations or non-genetic adaptations, enabling tumour regrowth.

*Adopted from (De Conti et al., 2021), published under the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).

In these cases, resistance arises not from a single mutation but from adaptive survival strategies, allowing a small subpopulation of tumour cells to persist during treatment and later repopulate the tumour. This transition appears to be reversible in some settings, while in others, persistent epigenetic changes lead to stable, fully resistant clones (Rabe et al., 2020).

Given this complexity, resistance in GBM is not well explained by simple mutational profiles. Instead, it's likely to involve a combination of mild genetic selection and strong transcriptional adaptation,

particularly in response to standard therapy. This highlights the need to study paired samples across disease progression, which is one of the major motivations for the work presented in this thesis.

1.3 Intratumoural Heterogeneity in Glioblastoma

Glioblastoma (GBM) is not a uniform entity. One of the major reasons it remains so difficult to treat lies in its remarkable intratumour heterogeneity (ITH) — the coexistence of genetically and phenotypically distinct cell populations within a single tumour (Eder and Kalman, 2014, Dymova et al., 2021). This diversity allows some subpopulations of tumour cells to evade treatment, survive, and ultimately drive recurrence, often in a more resistant and aggressive form.

ITH in GBM is observed at multiple levels. It includes differences in morphology (such as small vs. large anaplastic cells), molecular profiles (gene expression and mutations), and cellular identity (e.g. tumour cells, vascular cells, and immune cells, Figure 1-4) (Becker et al., 2021). These differences often result in distinct subclones within a tumour, each with its own set of genomic and transcriptomic alterations. Importantly, these subclones are not always evenly distributed; cells located in one region of the tumour can differ significantly from those in another, leading to sampling bias and challenges in accurately profiling the tumour as a whole (Eder and Kalman, 2014).

ITH is further complicated by the dynamic nature of tumour progression. Even if a particular clone is targeted by therapy, others may survive and expand. This clonal replacement is a key driver of treatment resistance and recurrence (Bergmann et al., 2020, Friedmann-Morvinski, 2014). Understanding how this heterogeneity emerges and evolves is essential to developing more effective, long-lasting treatment strategies.

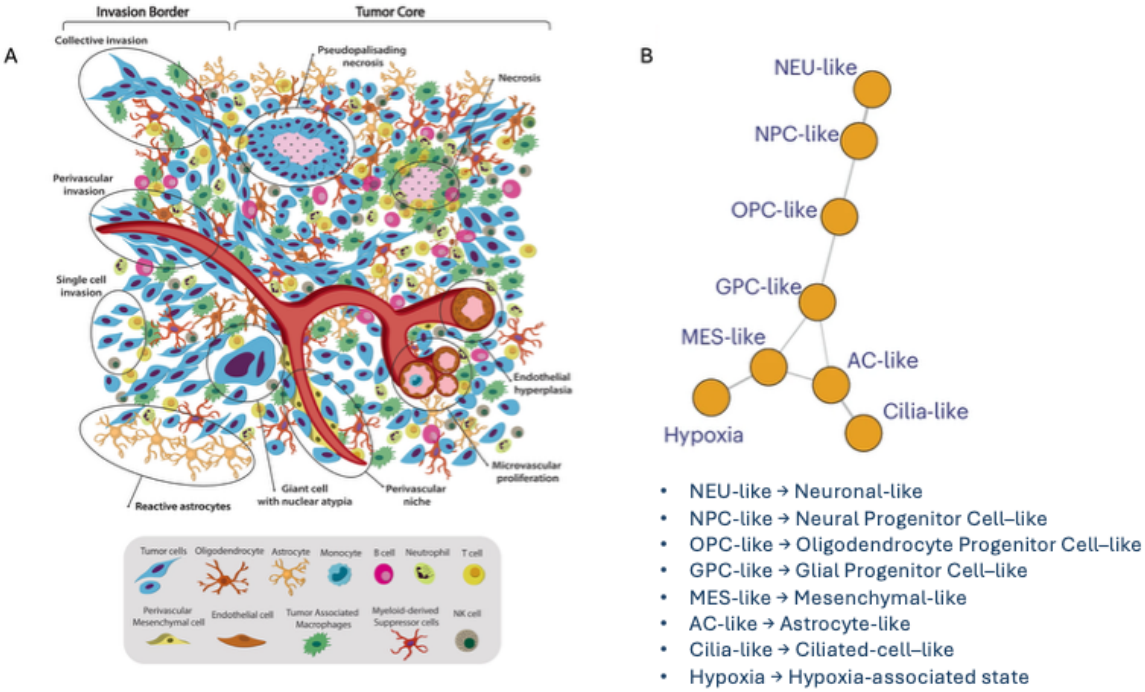


Figure 1-4: Heterogeneity in glioblastoma tumours.

Panel A: ITH IN GBM TUMOURS. This diagram provides a schematic overview of the complex spatiotemporal intratumoural heterogeneity of GBM. The main body of the figure highlights key histological and dynamic characteristics. Pathological hallmarks include areas of pseudopalisading necrosis, characterised by a garland-like arrangement of tumour cells at the edge of necrotic regions. The tumour microenvironment (TME) also features vascular abnormalities, such as endothelial hyperplasia and microvascular proliferation. Cellular diversity is further shown by the presence of large, pleomorphic glioma cells. The diagram illustrates how this heterogeneity extends to tumour cell movement, with different migratory patterns including collective invasion, single-cell invasion, and perivascular invasion at the tumour-brain interface. The bottom panel emphasises the striking cellular heterogeneity of the TME, which is comprised of both malignant and non-malignant cells. This diverse cellular ecosystem includes normal brain residents (e.g., astrocytes, microglia), endothelial cells from the vasculature, and infiltrating immune cells, with a prominent presence of tumour-associated macrophages (TAMs). This multifaceted composition at the histological, cellular, and dynamic levels is central to understanding GBM's complexity.

*Adopted from (Comba et al., 2021) under the Creative Commons Attribution (CC BY) license (http://creativecommons.org/licenses/by/4.0/).

Panel B: Heterogeneity amongst GBM cancer cells. This diagram presents a model of GBM cellular states, providing insight into the organisation and dynamics of intratumour heterogeneity within the malignant cell population. The model was constructed using single-nucleus RNA sequencing (snRNA-seq) to identify "high-frequency hybrid states," which are thought to represent cancer cells in transition between different identities. Each vertex (circle) in the diagram represents a distinct cellular state, and the connections between them represent these frequent hybrid states. This arrangement reveals a hierarchy of heterogeneity, where progenitor-like states (e.g., GPC-like) are located in the centre. This central position suggests they have the potential to differentiate into multiple specialised states (e.g., NEU-like, AC-like, MES-like) which are positioned at the edges of the model. This model provides a "multilayered transcriptional architecture of GBM"

*Adopted from (Nomura et al., 2025) under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License (http://creativecommons.org/licenses/by-nc-nd/4.0/).

1.3.1 Tumour Evolution and Models of Heterogeneity

Several models have been proposed to explain how ITH arises in GBM. The clonal evolution model suggests that normal cells acquire successive mutations over time, giving rise to increasingly aggressive subclones that adapt to environmental pressures like hypoxia, immune surveillance, and therapy (Noch et al., 2018). Selective pressures allow fitter clones to dominate while others die off, but many genetically distinct subclones can still coexist in the tumour at any one time (Dymova et al., 2021).

An alternative theory is the cancer stem cell (CSC) model, where a smaller subset of stem-like tumour cells drives tumour growth and gives rise to a hierarchy of cell types through asymmetric division (Minata et al., 2019). These CSCs can both self-renew and differentiate, contributing to both inter- and intra-tumoural diversity. While the CSC model explains phenotypic variation and plasticity, it does not fully account for the complex clonal architectures observed in GBM, particularly those emerging after treatment.

In practice, evidence suggests that both models may coexist in GBM, with clonal evolution explaining the accumulation of mutations and CSCs contributing to functional heterogeneity and adaptability (Ramón y Cajal et al., 2020).

1.3.2 Genomic Alterations Contributing to ITH

At the genetic level, several types of alterations underlie the subclonal architecture in GBM.

Somatic Mutations (SNVs, Indels):

Point mutations and small insertions/deletions (indels) accumulate as tumours evolve. Commonly mutated genes in GBM include *TP53*, *PTEN*, *NF1*, *EGFR*, and *TERT* promoter regions (Gan et al., 2013, Olympios et al., 2021). Many of these mutations arise early and are clonal, but additional mutations occur subclonally over time, leading to spatially distinct profiles. Whole-genome sequencing from primary and recurrent GBM samples has shown that most point mutations are retained over time, but a subset of subclonal variants can emerge or disappear between timepoints (Korber et al., 2019).

Structural Variants and Copy Number Alterations:

Large-scale chromosomal alterations also play a major role in ITH. These include chromosome 7 gain and/or chromosome 10 loss, found in almost all IDH-wildtype GBMs, and focal amplifications of oncogenes like *EGFR* and *PDGFRA* (Wemmert et al., 2005, Louis et al., 2021, Korber et al., 2019). In

some cases, chromothripsis—a catastrophic chromosomal shattering and rearrangement event—can drive complex genomic rearrangements that produce sudden bursts of heterogeneity (Korber et al., 2019).

A significant structural feature contributing to GBM heterogeneity is the presence of extrachromosomal DNA (ecDNA). These circular DNA elements often carry amplified oncogenes such as *EGFR* and allow for highly variable copy number and uneven segregation during cell division, which fuels rapid subclonal diversification (Nathanson et al., 2014).

Epigenetic Alterations:

In addition to genetic mutations, epigenetic changes like DNA methylation contribute to ITH. Aberrant methylation patterns, such as promoter hypermethylation of *MGMT*, can influence response to treatment and may vary across tumour regions (Rippaus et al., 2019). More broadly, epigenetic alterations impact gene expression and chromatin state and may drive tumour cell identity and state transitions even in the absence of genetic changes. Importantly, studies have shown that epigenetic profiles (including methylation and histone modifications) are also heterogeneous across different tumour regions and persist through recurrence (Spitzer et al., 2025).

1.3.3 Evidence from Matched and Single-Cell Studies

Several recent studies have used multi-region sampling and single-cell approaches to map ITH in GBM more precisely. For example, Korber et al. (2019) performed whole-genome sequencing on matched primary and recurrent tumours and found that in most cases, no single subclone was selected through treatment — instead, clonal architecture was largely retained (Figure 1-5), suggesting that no strong genetic bottleneck occurs during recurrence (Korber et al., 2019).

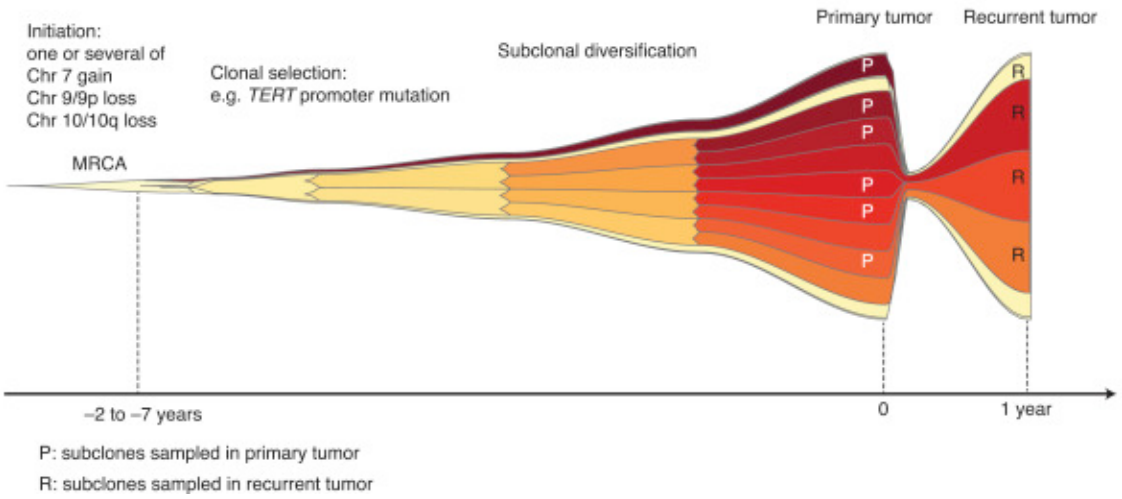


Figure 1-5: Clonal evolution and subclonal persistence in GBM.

This schematic illustrates the progression of glioblastoma from tumour initiation to recurrence. Individual subclones are represented by varying shades of yellow to red, originating from a most recent common ancestor (MRCA). Early driver events, such as chromosome 7 gain and loss of 9p or 10q, initiate tumour growth, followed by clonal selection events (e.g., TERT promoter mutations) and subclonal diversification. Profiling matched primary (P) and recurrent (R) tumours reveals that many subclones persist through therapy, supporting a model of minimal clonal selection during recurrence and limited treatment-induced bottlenecks.

*Adopted from (Korber et al., 2019) with permission from Elsevier (License number: 6074780895426).

A more recent study by (Spitzer et al., 2025) used single-cell longitudinal analysis to trace clonal trajectories in IDH-wildtype GBM (Figure 1-6). Although they found no universal genetic drivers of recurrence, they observed that certain low-frequency deletions and small alterations were associated with specific transcriptional programs related to treatment response. This indicates that even subtle genetic changes can shape cellular behaviour in a heterogeneous context.

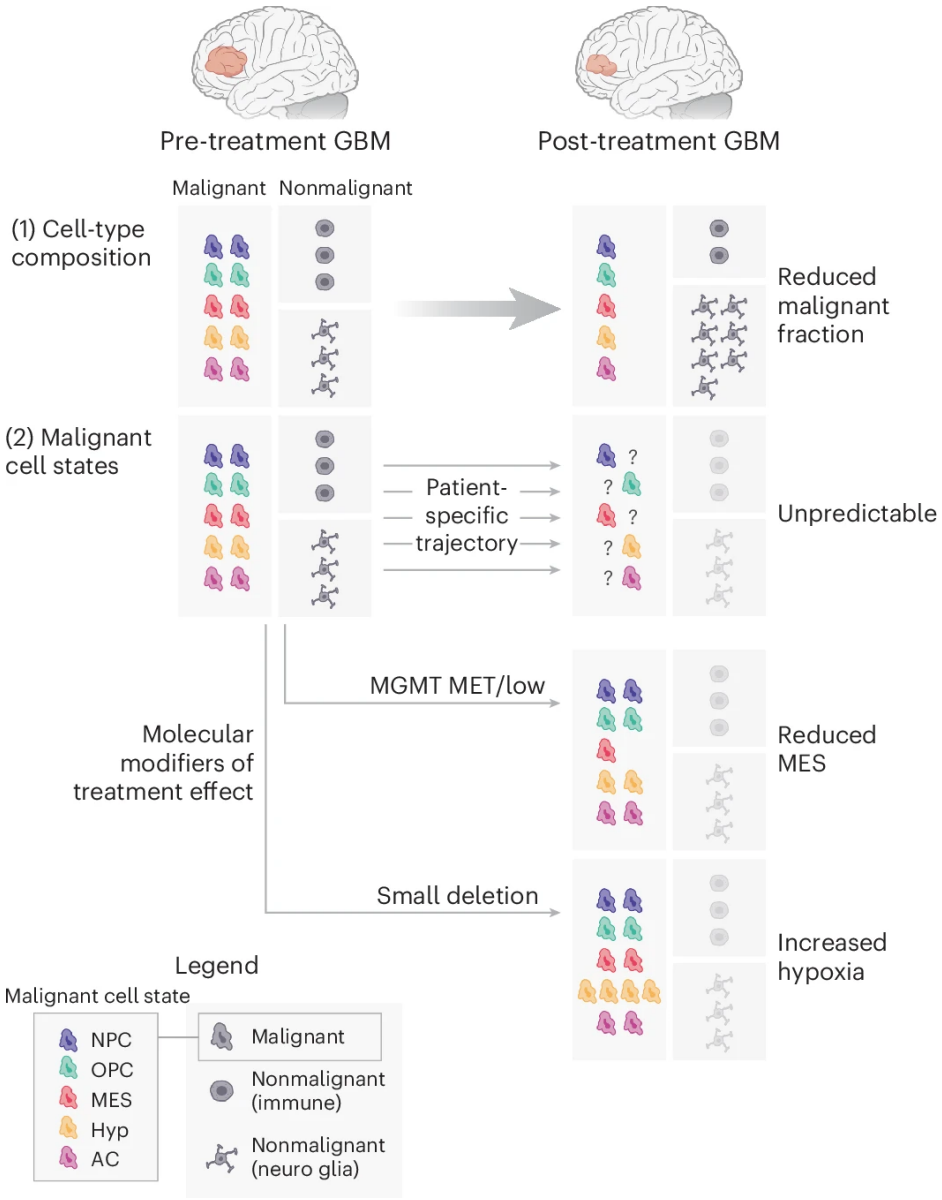


Figure 1-6: Single-cell longitudinal analysis reveals diverse transcriptional trajectories in glioblastoma following treatment.

This figure summarises key findings from (Spitzer et al., 2025), who profiled matched primary and recurrent IDH-wildtype glioblastoma samples using single-nucleus RNA sequencing (snRNA-seq) as part of the GBM CARE consortium. By quantifying changes in malignant and nonmalignant cell fractions and transcriptional states, they identified three major patterns in post-treatment tumour evolution:

(1) A consistent reduction in the malignant cell fraction at recurrence, accompanied by increased glial and neuronal (glio-neural) components within the tumour microenvironment.

(2) Highly variable, patient-specific shifts in malignant cell states, without a universally conserved recurrence trajectory.

(3) More predictable trajectories in subsets of patients, including a reduction in mesenchymal-like (MES-like) states in MGMT-methylated tumours and an increase in hypoxia-associated states in tumours with elevated small deletion burden following treatment.

Malignant states shown include neural progenitor cell-like (NPC), oligodendrocyte progenitor cell-like (OPC), MES-like (MES), hypoxia-associated (Hyp), and astrocyte-like (AC) populations.

*Adopted from (Spitzer et al., 2025), published under the Creative Commons Attribution-NonCommercial-NoDerivatives 4.0 International License (http://creativecommons.org/licenses/by-nc-nd/4.0/).

Other studies, including initiatives from the Pancancer Analysis of Whole Genomes (PCAWG) and TRAcking Cancer Evolution through therapy (TRACERx), have also shown that high ITH correlates with poor prognosis across many cancers. In GBM, however, the situation is more nuanced — genetic ITH is often preserved across recurrence, and resistance appears to be driven more by transcriptional plasticity than by clonal selection (Barthel et al., 2019, Wang et al., 2016b).

Previous studies have revealed that GBM recurrence often lacks a clear genetic bottleneck, with clonal architecture largely preserved over time. However, subtle genetic alterations, particularly in non-coding regions or subclonal populations, may still influence tumour evolution and therapy response. At the same time, transcriptional reprogramming and epigenetic adaptation appear to play a significant role in recurrence.

In this thesis, I examine paired primary and recurrent GBM samples using both whole genome/exome sequencing (WGS/WES) and DNA methylation arrays. This dual approach allows me to capture both clonal and subclonal genetic changes (e.g., SNVs, indels, CNVs) and epigenetic alterations (e.g. promoter methylation, pathway-level changes). By integrating these layers of information, the work aims to clarify how tumours evolve across treatment and identify converging biological programs that contribute to recurrence.

1.4 Tumour Cell Plasticity

While intratumour heterogeneity (ITH) describes the coexistence of distinct cell populations within a tumour, plasticity refers to the ability of individual tumour cells to transition between different phenotypic states. In glioblastoma (GBM), this plasticity adds a further layer of complexity, allowing cells to adapt dynamically in response to environmental cues, including therapy. This capacity for transcriptional reprogramming, even in the absence of new genetic mutations, has emerged as a key mechanism underlying tumour progression and treatment resistance.

The distinction between heterogeneity and plasticity is crucial. ITH can arise through either clonal evolution or the coexistence of functionally distinct cell states, but plasticity allows these states to interconvert, enabling tumour cells to evade selective pressures such as hypoxia, immune surveillance, or chemotherapy (Neftel et al., 2019). In other words, while ITH explains a snapshot of the tumour, plasticity helps explain how it changes.

1.4.1 Evidence of Plasticity in GBM

Compelling evidence for plasticity in GBM comes from single-cell transcriptomic studies. A landmark study by Neftel et al. (2019) used single-cell RNA sequencing of IDH-wildtype GBMs and identified four main transcriptional states: neural progenitor-like (NPC-like), oligodendrocyte progenitor-like (OPC-like), astrocyte-like (AC-like), and mesenchymal-like (MES-like). Cells within a single tumour could occupy any of these states, and transition between them, suggesting dynamic plasticity rather than fixed subclonal identity (Neftel et al., 2019).

Importantly, these cellular states correlate with tumour subtypes: for example, AC-like cells are enriched in classical GBMs, while MES-like cells dominate mesenchymal subtypes (Neftel et al., 2019). Yet, individual cells can shift between states depending on microenvironmental signals, such as hypoxia or inflammation. Lineage tracing experiments confirmed that each state has tumour-initiating capacity, and when isolated and transplanted, can regenerate the full spectrum of cellular diversity, reinforcing the concept of reversible plasticity (Neftel et al., 2019).

1.4.2 Adaptive Resistance and Therapy Response

Plasticity has been increasingly implicated in adaptive resistance to therapy. While many cancers show selection for resistant genetic subclones, recent work in GBM suggests that resistance often occurs without major changes in the mutational landscape (Korber et al., 2019, Barthel et al., 2019). Instead, tumour cells adapt by shifting into slow-cycling, drug-tolerant cell states.

In one study, slow-cycling persistent cells emerged after treatment with PDGFR inhibitors and exhibited a reversible transition into resistant states dependent on Notch signalling and chromatin remodelling (Eyler et al., 2020). Similar observations were made in GBM cells exposed to temozolomide (TMZ), where drug-tolerant cells showed widespread epigenetic reprogramming and eventually acquired stable resistance (Banelli et al., 2015, Rabe et al., 2020).

The study by (Spitzer et al., 2025) provided further support for this model. Using longitudinal single-cell transcriptomics in matched primary and recurrent GBMs, they found that although most genetic alterations remained stable, transcriptional state changes were extensive, with cells reprogramming toward mesenchymal or inflammatory phenotypes in response to treatment. These shifts were

associated with subtle deletions but were largely non-clonal, underscoring the importance of non-genetic adaptation in recurrence (Spitzer et al., 2025).

1.4.3 Clinical Implications

The ability of GBM cells to transition between functional states presents a major challenge for therapy. Unlike fixed subclonal mutations, plasticity cannot be targeted genetically, making it difficult to eliminate tumour cells based on a single vulnerability. It also contributes to temporal heterogeneity, where the tumour's molecular profile changes over time, limiting the effectiveness of therapies designed for the primary tumour (Wang et al., 2017, Singh et al., 2021).

Additionally, plasticity may explain why even patients with favourable biomarkers (e.g., MGMT promoter methylation) can eventually relapse. As treatment progresses, tumour cells may switch into resistant states or re-activate repair mechanisms through epigenetic pathways, bypassing the initial therapeutic advantage (Rippaus et al., 2019).

The ability of GBM cells to reversibly transition between transcriptional states, even in the absence of new mutations, has major implications for treatment resistance. However, the extent to which such transitions are driven by genetic alterations, transcriptional reprogramming, or a combination of both remains unclear.

This thesis takes a multi-faceted approach to characterise tumour adaptation following treatment by analysing both methylation and sequencing data from matched primary and recurrent GBM samples. I conduct independent analyses of both data types: methylation profiles are examined within stratified responder subtypes to identify epigenetic changes potentially associated with recurrence, while WGS and WES data are used to track changes in variant prevalence and infer clonal dynamics. By examining both the genetic and epigenetic layers of tumour evolution separately, this work provides a more comprehensive view of tumour adaptation than either method could offer alone.

1.5 Technical Approaches to Studying Tumour Heterogeneity and Evolution

Understanding the biological complexity of glioblastoma (GBM) — from intratumoural heterogeneity to treatment-induced plasticity — has been made possible through advances in high-throughput molecular technologies. These approaches have provided unprecedented resolution into the

genomic, transcriptomic, and epigenetic architecture of tumours. In this section, I briefly outline the key methodologies used in the field, focusing on the types of data they generate and how they contribute to studying tumour evolution. This technical overview is not GBM-specific but helps contextualise the analytical strategies used in this thesis and in the broader literature.

1.5.1 Bulk Genomic and Transcriptomic Profiling

DNA sequencing technologies, particularly whole-exome sequencing (WES) and whole-genome sequencing (WGS), have been foundational in mapping somatic mutations, copy number alterations, and structural variants across cancer genomes. These platforms are typically applied to bulk tumour samples and have enabled the identification of recurrently altered genes in GBM such as *TP53*, *PTEN*, *EGFR*, and *TERT* promoter mutations (Korber et al., 2019, Barthel et al., 2019).

RNA sequencing (RNA-seq) has similarly transformed how transcriptional programs are studied in cancer. In GBM, RNA-seq data was instrumental in defining the proneural, classical, and mesenchymal subtypes identified by The Cancer Genome Atlas (TCGA) (Cancer Genome Atlas Research, 2008). More recently, paired RNA-seq of primary and recurrent samples has revealed evidence of dynamic transcriptional reprogramming following treatment, even in the absence of clear genetic bottlenecks (Rippaus et al., 2019).

However, bulk approaches offer only a population-level average, potentially obscuring rare subclones or cell states present within the tumour. This limitation has led to increased use of more granular methods in recent years.

1.5.2 Bulk Epigenomic Technologies

Changes in DNA methylation represent a stable and functionally important layer of regulation, particularly relevant in GBM where promoter methylation of genes such as *MGMT* is predictive of treatment response (Rivera et al., 2010). Epigenomic profiling is typically performed using Illumina DNA methylation arrays, such as the 450K or EPIC platforms. These arrays quantify methylation at hundreds of thousands of CpG sites and are cost-effective and scalable, making them suitable for studies involving large patient cohorts or matched longitudinal samples.

Methylation arrays have also been used for tumour classification. The Heidelberg classifier, for instance, stratifies central nervous system tumours into biologically relevant subtypes based solely

on their methylation profiles — underscoring the potential of DNA methylation as both a diagnostic and analytical tool (Capper et al., 2018).

In this thesis, methylation arrays are used not only for classification but as a readout of epigenetic evolution between primary and recurrent GBMs, complementing genomic analysis by capturing non-mutational mechanisms of resistance and progression.

1.5.3 Single-Cell Technologies

More recently, single-cell RNA sequencing (scRNA-seq) has enabled the dissection of transcriptional heterogeneity at cellular resolution. In GBM, this approach revealed that individual tumour cells occupy a spectrum of distinct transcriptional states (e.g., NPC-like, OPC-like, AC-like, MES-like), and that these states are plastic, capable of interconversion under environmental stress or treatment (Neftel et al., 2019).

The study by (Spitzer et al., 2025) expanded on this by tracking single-cell trajectories across timepoints in matched GBM samples. While few new genetic drivers of recurrence were found, extensive shifts in transcriptional states were observed, highlighting the importance of plasticity in treatment response.

Despite their power, single-cell technologies remain costly, noisy, and technically demanding. As such, they are not yet widely adopted in large-scale clinical studies and were not used in this project. However, their findings help frame the interpretation of bulk methylation changes and support the view that transcriptional reprogramming may occur even without detectable genetic selection.

1.5.4 Computational Tools and Analytical Strategies

The complexity of analysing both genomic (WGS/WES) and epigenomic (DNA methylation array) data requires the use of specialised tools tailored to each data type. For variant calling from DNA sequencing, tools such as Mutect2 (Benjamin et al., 2019), and somatic copy number aberration callers were used to identify somatic mutations, copy number alterations across matched tumour samples. For methylation analysis, R packages like minfi (Aryee et al., 2014) and RnBeads (Muller et al., 2019) were employed to assess site-specific and regional methylation changes.

Many of these tools support paired-sample analyses, which made them suitable for comparing primary and recurrent tumour samples in this project. Throughout the thesis, tools were selected or adapted to enable:

- Detection of clonal and subclonal mutations over time using WES and WGS data
- Differential methylation analysis between primary and recurrent tumours, at both single CpG and regional levels

Optimisation of workflows was applied where necessary to better align with the study's design and biological questions.

1.6 Hypothesis

Recurrence in IDH-wildtype glioblastoma (GBM) occurs despite intensive treatment and is thought to arise through a combination of genetic selection and epigenetic reprogramming. Traditional approaches have focused on identifying individual driver mutations enriched at recurrence; however, this has yielded limited clinical insight due to the extreme heterogeneity and often subtle clonal dynamics of GBM.

This thesis hypothesises that:

A pathway-level approach, integrating genetic and epigenetic changes in paired primary and recurrent GBM samples, can reveal convergent biological programs that contribute to treatment resistance. Rather than focusing solely on recurrent single-gene alterations, this study investigates whether groups of variants and methylation changes acting within the same pathway exhibit coordinated patterns of selection, expansion, or suppression through therapy.

By analysing both somatic variants (from WGS/WES) and DNA methylation (from array data) in a longitudinal framework, the thesis aims to uncover pathway-level mechanisms underpinning resistance and to evaluate whether distinct patient subtypes exhibit different modes of tumour evolution and adaptation.

1.7 Aims and Objectives

The aims and objectives of this thesis are structured across three main chapters, each addressing a specific level of tumour biology and method development.

Chapter 2 – Development and optimisation of somatic variant calling pipelines for WGS and WES data

Aim:

To establish a robust, reproducible pipeline for calling high-confidence somatic variants from WGS and WES of paired GBM samples.

Objectives:

- Develop and standardise pipelines using benchmarked tools in the Glioma Genomics group, incorporating the best-practice filtering strategies.
- Optimise the pipeline to reduce false positives through post-mapping and post-calling refinement.
- Assess pipeline performance and variant quality using both internal QC metrics and visual inspection.
- Apply the workflow to paired primary and recurrent GBM samples to generate a validated mutation call set for downstream analyses.

Chapter 3 – Analysis of subclonal architecture and selection through therapy using variant allele frequency (VAF)

Aim:

To explore clonal dynamics in paired GBM samples using VAF-based metrics and identify variants under selective pressure during recurrence.

Objectives:

- Classify variants into clonal or subclonal categories based on VAF thresholds.
- Focus on shared variants across primary and recurrent pairs to track changes in subclonal prevalence.
- Use increases in VAF as a proxy for clonal expansion (resistance) and decreases in VAF for clonal depletion (sensitisation).
- Highlight pathways that are recurrently altered in expanding versus contracting subclones.

Chapter 4 – DNA methylation analysis of paired GBM samples stratified by transcriptional responder subtype

Aim:

To investigate epigenetic changes associated with GBM recurrence and determine whether distinct responder subtypes (Up and Down responders) exhibit different methylation dynamics.

Objectives:

- Process DNA methylation data using array-based platforms (450K/EPIC) and develop a custom analysis pipeline (e.g., minfi, RnBeads).
- Identify differentially methylated positions (DMPs) and regions (DMRs) between primary and recurrent samples within each transcriptionally stratified responder subtype group.
- Perform functional enrichment analysis to interpret the biological significance of subtype-specific epigenetic alterations.

1.8 REFERENCES

ALIREZA MANSOURI, J. K., SUNIT DAS 2017. *Molecular Genetics of Secondary Glioblastoma,* Brisbane (AU), Codon Publications; Chapter 2.

ARORA, A. & SOMASUNDARAM, K. 2019. Glioblastoma vs temozolomide: can the red queen race be won? *Cancer Biol Ther,* 20**,** 1083-1090.

ARYEE, M. J., JAFFE, A. E., CORRADA-BRAVO, H., LADD-ACOSTA, C., FEINBERG, A. P., HANSEN, K. D. & IRIZARRY, R. A. 2014. Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. *Bioinformatics,* 30**,** 1363-9.

BAGLEY, S. J., KOTHARI, S., RAHMAN, R., LEE, E. Q., DUNN, G. P., GALANIS, E., CHANG, S. M., NABORS, L. B., AHLUWALIA, M. S., STUPP, R., MEHTA, M. P., REARDON, D. A., GROSSMAN, S. A., SULMAN, E. P., SAMPSON, J. H., KHAGI, S., WELLER, M., CLOUGHESY, T. F., WEN, P. Y. & KHASRAW, M. 2022. Glioblastoma Clinical Trials: Current Landscape and Opportunities for Improvement. *Clin Cancer Res,* 28**,** 594-602.

BAKER, S. D., WIRTH, M., STATKEVICH, P., REIDENBERG, P., ALTON, K., SARTORIUS, S. E., DUGAN, M., CUTLER, D., BATRA, V., GROCHOW, L. B., DONEHOWER, R. C. & ROWINSKY, E. K. 1999. Absorption, metabolism, and excretion of 14C-temozolomide following oral administration to patients with advanced cancer. *Clin Cancer Res,* 5**,** 309-17.

BANELLI, B., CARRA, E., BARBIERI, F., WURTH, R., PARODI, F., PATTAROZZI, A., CAROSIO, R., FORLANI, A., ALLEMANNI, G., MARUBBI, D., FLORIO, T., DAGA, A. & ROMANI, M. 2015. The histone demethylase KDM5A is a key factor for the resistance to temozolomide in glioblastoma. *Cell Cycle,* 14**,** 3418-29.

BARBAGALLO, G. M., JENKINSON, M. D. & BRODBELT, A. R. 2008. 'Recurrent' glioblastoma multiforme, when should we reoperate? *Br J Neurosurg,* 22**,** 452-5.

BARTHEL, F. P., JOHNSON, K. C., VARN, F. S., MOSKALIK, A. D., TANNER, G., KOCAKAVUK, E., ANDERSON, K. J., ABIOLA, O., ALDAPE, K., ALFARO, K. D., ALPAR, D., AMIN, S. B., ASHLEY, D. M., BANDOPADHAYAY, P., BARNHOLTZ-SLOAN, J. S., BEROUKHIM, R., BOCK, C., BRASTIANOS, P. K., BRAT, D. J., BRODBELT, A. R., BRUNS, A. F., BULSARA, K. R., CHAKRABARTY, A., CHAKRAVARTI, A., CHUANG, J. H., CLAUS, E. B., COCHRAN, E. J., CONNELLY, J., COSTELLO, J. F., FINOCCHIARO, G., FLETCHER, M. N., FRENCH, P. J., GAN, H. K., GILBERT, M. R., GOULD, P. V., GRIMMER, M. R., IAVARONE, A., ISMAIL, A., JENKINSON, M. D., KHASRAW, M., KIM, H., KOUWENHOVEN, M. C. M., LAVIOLETTE, P.

S., LI, M., LICHTER, P., LIGON, K. L., LOWMAN, A. K., MALTA, T. M., MAZOR, T., MCDONALD, K. L., MOLINARO, A. M., NAM, D. H., NAYYAR, N., NG, H. K., NGAN, C. Y., NICLOU, S. P., NIERS, J. M., NOUSHMEHR, H., NOORBAKHSH, J., ORMOND, D. R., PARK, C. K., POISSON, L. M., RABADAN, R., RADLWIMMER, B., RAO, G., REIFENBERGER, G., SA, J. K., SCHUSTER, M., SHAW, B. L., SHORT, S. C., SMITT, P. A. S., SLOAN, A. E., SMITS, M., SUZUKI, H., TABATABAI, G., VAN MEIR, E. G., WATTS, C., WELLER, M., WESSELING, P., WESTERMAN, B. A., WIDHALM, G., WOEHRER, A., YUNG, W. K. A., ZADEH, G., HUSE, J. T., DE GROOT, J. F., STEAD, L. F., VERHAAK, R. G. W. & CONSORTIUM, G. 2019. Longitudinal molecular trajectories of diffuse glioma in adults. *Nature,* 576**,** 112-120.

BECKER, A. P., SELLS, B. E., HAQUE, S. J. & CHAKRAVARTI, A. 2021. Tumor Heterogeneity in Glioblastomas: From Light Microscopy to Molecular Pathology. *Cancers (Basel),* 13.

BENJAMIN, D., SATO, T., CIBULSKIS, K., GETZ, G., STEWART, C. & LICHTENSTEIN, L. 2019.

BERENS, M. E. & GIESE, A. 1999. "...those left behind." Biology and oncology of invasive glioma cells. *Neoplasia,* 1**,** 208-19.

BERGMANN, N., DELBRIDGE, C., GEMPT, J., FEUCHTINGER, A., WALCH, A., SCHIRMER, L., BUNK, W., ASCHENBRENNER, T., LIESCHE-STARNECKER, F. & SCHLEGEL, J. 2020. The Intratumoral Heterogeneity Reflects the Intertumoral Subtypes of Glioblastoma Multiforme: A Regional Immunohistochemistry Analysis. *Front Oncol,* 10**,** 494.

BIRZU, C., FRENCH, P., CACCESE, M., CERRETTI, G., IDBAIH, A., ZAGONEL, V. & LOMBARDI, G. 2020. Recurrent Glioblastoma: From Molecular Landscape to New Treatment Perspectives. *Cancers (Basel),* 13.

BJORLAND, L. S., FLUGE, O., GILJE, B., MAHESPARAN, R. & FARBU, E. 2021. Treatment approach and survival from glioblastoma: results from a population-based retrospective cohort study from Western Norway. *BMJ Open,* 11**,** e043208.

BRASTIANOS, P. K., NAYYAR, N., ROSEBROCK, D., LESHCHINER, I., GILL, C. M., LIVITZ, D., BERTALAN, M. S., D'ANDREA, M., HOANG, K., AQUILANTI, E., CHUKWUEKE, U. N., KANEB, A., CHI, A., PLOTKIN, S., GERSTNER, E. R., FROSCH, M. P., SUVA, M. L., CAHILL, D. P., GETZ, G. & BATCHELOR, T. T. 2017. Resolving the phylogenetic origin of glioblastoma via multifocal genomic analysis of pre-treatment and treatment-resistant autopsy specimens. *NPJ Precis Oncol,* 1**,** 33.

BRODBELT, A., GREENBERG, D., WINTERS, T., WILLIAMS, M., VERNON, S., COLLINS, V. P. & NATIONAL CANCER INFORMATION NETWORK BRAIN TUMOUR, G. 2015. Glioblastoma in England: 2007-2011. *Eur J Cancer,* 51**,** 533-542.

CANCER GENOME ATLAS RESEARCH, N. 2008. Comprehensive genomic characterization defines human glioblastoma genes and core pathways. *Nature,* 455**,** 1061-8.

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., HOLSKEN, 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., BRUCK, W., HOFER, S., BREHMER, S., SEIZ-ROSENHAGEN, M., HANGGI, 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., MUHLEISEN, H., MUELLER, W., POHL, U., PRINZ, M., BENNER, A., ZAPATKA, M., GOTTARDO, N. G., DRIEVER, P. H., KRAMM, C. M., MULLER, H. L., RUTKOWSKI, S., VON HOFF, K., FRUHWALD, 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.

CHANG, S. M., PARNEY, I. F., HUANG, W., ANDERSON, F. A., JR., ASHER, A. L., BERNSTEIN, M., LILLEHEI, K. O., BREM, H., BERGER, M. S., LAWS, E. R. & GLIOMA OUTCOMES PROJECT, I. 2005. Patterns of care for adults with newly diagnosed malignant glioma. *JAMA,* 293**,** 557-64.

CHOI, H. J., CHOI, S. H., YOU, S. H., YOO, R. E., KANG, K. M., YUN, T. J., KIM, J. H., SOHN, C. H., PARK, C. K. & PARK, S. H. 2021. MGMT Promoter Methylation Status in Initial and Recurrent Glioblastoma: Correlation Study with DWI and DSC PWI Features. *AJNR Am J Neuroradiol,* 42**,** 853-860.

COMBA, A., FAISAL, S. M., VARELA, M. L., HOLLON, T., AL-HOLOU, W. N., UMEMURA, Y., NUNEZ, F. J., MOTSCH, S., CASTRO, M. G. & LOWENSTEIN, P. R. 2021. Uncovering Spatiotemporal Heterogeneity of High-Grade Gliomas: From Disease Biology to Therapeutic Implications. *Front Oncol,* 11**,** 703764.

CRESPO, I., VITAL, A. L., GONZALEZ-TABLAS, M., PATINO MDEL, C., OTERO, A., LOPES, M. C., DE OLIVEIRA, C., DOMINGUES, P., ORFAO, A. & TABERNERO, M. D. 2015. Molecular and Genomic Alterations in Glioblastoma Multiforme. *Am J Pathol,* 185**,** 1820-33.

DE CONTI, G., DIAS, M. H. & BERNARDS, R. 2021. Fighting Drug Resistance through the Targeting of Drug-Tolerant Persister Cells. *Cancers (Basel),* 13.

DYMOVA, M. A., KULIGINA, E. V. & RICHTER, V. A. 2021. Molecular Mechanisms of Drug Resistance in Glioblastoma. *Int J Mol Sci,* 22.

EDER, K. & KALMAN, B. 2014. Molecular heterogeneity of glioblastoma and its clinical relevance. *Pathol Oncol Res,* 20**,** 777-87.

EYLER, C. E., MATSUNAGA, H., HOVESTADT, V., VANTINE, S. J., VAN GALEN, P. & BERNSTEIN, B. E. 2020. Single-cell lineage analysis reveals genetic and epigenetic interplay in glioblastoma drug resistance. *Genome Biol,* 21**,** 174.

FRIEDMANN-MORVINSKI, D. 2014. Glioblastoma heterogeneity and cancer cell plasticity. *Crit Rev Oncog,* 19**,** 327-36.

GAILLARD, F. 2024. Temozolomide mechanism of action and resistance. (Case study). *Radiopaedia.org*.

GAN, H. K., CVRLJEVIC, A. N. & JOHNS, T. G. 2013. The epidermal growth factor receptor variant III (EGFRvIII): where wild things are altered. *FEBS J,* 280**,** 5350-70.

GARCIA, J., HURWITZ, H. I., SANDLER, A. B., MILES, D., COLEMAN, R. L., DEURLOO, R. & CHINOT, O. L. 2020. Bevacizumab (Avastin(R)) in cancer treatment: A review of 15 years of clinical experience and future outlook. *Cancer Treat Rev,* 86**,** 102017.

GILARD, V., TEBANI, A., DABAJ, I., LAQUERRIERE, A., FONTANILLES, M., DERREY, S., MARRET, S. & BEKRI, S. 2021. Diagnosis and Management of Glioblastoma: A Comprehensive Perspective. *J Pers Med,* 11.

GILBERT, M. R. 2011. Recurrent glioblastoma: a fresh look at current therapies and emerging novel approaches. *Semin Oncol,* 38 Suppl 4**,** S21-33.

GROCHANS, S., CYBULSKA, A. M., SIMINSKA, D., KORBECKI, J., KOJDER, K., CHLUBEK, D. & BARANOWSKA-BOSIACKA, I. 2022. Epidemiology of Glioblastoma Multiforme-Literature Review. *Cancers (Basel),* 14.

HADJIPANAYIS, C. G. & STUMMER, W. 2019. 5-ALA and FDA approval for glioma surgery. *J Neurooncol,* 141**,** 479-486.

HANIF, F., MUZAFFAR, K., PERVEEN, K., MALHI, S. M. & SIMJEE SH, U. 2017. Glioblastoma Multiforme: A Review of its Epidemiology and Pathogenesis through Clinical Presentation and Treatment. *Asian Pac J Cancer Prev,* 18**,** 3-9.

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.

KAINA, B. 2023. Temozolomide, Procarbazine and Nitrosoureas in the Therapy of Malignant Gliomas: Update of Mechanisms, Drug Resistance and Therapeutic Implications. *J Clin Med,* 12.

KITANGE, G. J., CARLSON, B. L., SCHROEDER, M. A., GROGAN, P. T., LAMONT, J. D., DECKER, P. A., WU, W., JAMES, C. D. & SARKARIA, J. N. 2009. Induction of MGMT expression is associated with temozolomide resistance in glioblastoma xenografts. *Neuro Oncol,* 11**,** 281-91.

KORBER, V., YANG, J., BARAH, P., WU, Y., STICHEL, D., GU, Z., FLETCHER, M. N. C., JONES, D., HENTSCHEL, B., LAMSZUS, K., TONN, J. C., SCHACKERT, G., SABEL, M., FELSBERG, J., ZACHER, A., KAULICH, K., HUBSCHMANN, D., HEROLD-MENDE, C., VON DEIMLING, A., WELLER, M., RADLWIMMER, B., SCHLESNER, M., REIFENBERGER, G., HOFER, T. & LICHTER, P. 2019. Evolutionary Trajectories of IDH(WT) Glioblastomas Reveal a Common Path of Early Tumorigenesis Instigated Years ahead of Initial Diagnosis. *Cancer Cell,* 35**,** 692-704 e12.

LEE, S. Y. 2016. Temozolomide resistance in glioblastoma multiforme. *Genes Dis,* 3**,** 198-210.

LOUIS, D. N., PERRY, A., WESSELING, P., BRAT, D. J., CREE, I. A., FIGARELLA-BRANGER, D., HAWKINS, C., NG, H. K., PFISTER, S. M., REIFENBERGER, G., SOFFIETTI, R., VON DEIMLING, A. & ELLISON, D. W. 2021. The 2021 WHO Classification of Tumors of the Central Nervous System: a summary. *Neuro Oncol,* 23**,** 1231-1251.

MANDEL, J. J., YOUSSEF, M., LUDMIR, E., YUST-KATZ, S., PATEL, A. J. & DE GROOT, J. F. 2018. Highlighting the need for reliable clinical trials in glioblastoma. *Expert Rev Anticancer Ther,* 18**,** 1031-1040.

MINATA, M., AUDIA, A., SHI, J., LU, S., BERNSTOCK, J., PAVLYUKOV, M. S., DAS, A., KIM, S. H., SHIN, Y. J., LEE, Y., KOO, H., SNIGDHA, K., WAGHMARE, I., GUO, X., MOHYELDIN, A., GALLEGO-PEREZ, D., WANG, J., CHEN, D., CHENG, P., MUKHEEF, F., CONTRERAS, M., REYES, J. F., VAILLANT, B., SULMAN, E. P., CHENG, S. Y., MARKERT, J. M., TANNOUS, B. A., LU, X., KANGO-SINGH, M., LEE, L. J., NAM, D. H., NAKANO, I. & BHAT, K. P. 2019. Phenotypic Plasticity of Invasive Edge Glioma Stem-like Cells in Response to Ionizing Radiation. *Cell Rep,* 26**,** 1893-1905 e7.

MONTECUCCO, A., ZANETTA, F. & BIAMONTI, G. 2015. Molecular mechanisms of etoposide. *EXCLI J,* 14**,** 95-108.

MULLER, D. M. J., DE SWART, M. E., ARDON, H., BARKHOF, F., BELLO, L., BERGER, M. S., BOUWKNEGT, W., VAN DEN BRINK, W. A., CONTI NIBALI, M., EIJGELAAR, R. S., FURTNER, J., HAN, S. J., HERVEY-JUMPER, S., IDEMA, A. J. S., KIESEL, B., KLOET, A., MANDONNET, E., ROBE, P., ROSSI, M., SCIORTINO, T., VANDERTOP, W. P., VISSER, M., WAGEMAKERS, M., WIDHALM, G., WITTE, M. G. & DE WITT HAMER, P. C. 2021. Timing of glioblastoma surgery and patient outcomes: a multicenter cohort study. *Neurooncol Adv,* 3**,** vdab053.

MULLER, F., SCHERER, M., ASSENOV, Y., LUTSIK, P., WALTER, J., LENGAUER, T. & BOCK, C. 2019. RnBeads 2.0: comprehensive analysis of DNA methylation data. *Genome Biol,* 20**,** 55.

NAGASAKA, T., GOEL, A., NOTOHARA, K., TAKAHATA, T., SASAMOTO, H., UCHIDA, T., NISHIDA, N., TANAKA, N., BOLAND, C. R. & MATSUBARA, N. 2008. Methylation pattern of the O6-methylguanine-DNA methyltransferase gene in colon during progressive colorectal tumorigenesis. *Int J Cancer,* 122**,** 2429-36.

NATHANSON, D. A., GINI, B., MOTTAHEDEH, J., VISNYEI, K., KOGA, T., GOMEZ, G., ESKIN, A., HWANG, K., WANG, J., MASUI, K., PAUCAR, A., YANG, H., OHASHI, M., ZHU, S., WYKOSKY, J., REED, R., NELSON, S. F., CLOUGHESY, T. F., JAMES, C. D., RAO, P. N., KORNBLUM, H. I., HEATH, J. R., CAVENEE, W. K., FURNARI, F. B. & MISCHEL, P. S. 2014. Targeted therapy resistance mediated by dynamic regulation of extrachromosomal mutant EGFR DNA. *Science,* 343**,** 72-6.

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. & SUVA, M. L. 2019. An Integrative Model of Cellular States, Plasticity, and Genetics for Glioblastoma. *Cell,* 178**,** 835-849 e21.

NOCH, E. K., RAMAKRISHNA, R. & MAGGE, R. 2018. Challenges in the Treatment of Glioblastoma: Multisystem Mechanisms of Therapeutic Resistance. *World Neurosurg,* 116**,** 505-517.

NOMURA, M., SPITZER, A., JOHNSON, K. C., GAROFANO, L., NEHAR-BELAID, D., GALILI DARNELL, N., GREENWALD, A. C., BUSSEMA, L., OH, Y. T., VARN, F. S., D'ANGELO, F., GRITSCH, S., ANDERSON, K. J., MIGLIOZZI, S., GONZALEZ CASTRO, L. N., CHOWDHFURY, T., ROBINE, N., REEVES, C., PARK, J. B., LIPSA, A., HERTEL, F., GOLEBIEWSKA, A., NICLOU, S. P., NUSRAT, L., KELLET, S., DAS, S., MOON, H. E., PAEK, S. H., BIELLE, F., LAURENGE, A., DI STEFANO, A. L., MATHON, B., PICCA, A., SANSON, M., TANAKA, S., SAITO, N., ASHLEY, D. M., KEIR, S. T., LIGON, K. L., HUSE, J. T., YUNG, W. K. A., LASORELLA, A., VERHAAK, R. G. W., IAVARONE, A., SUVA, M. L. & TIROSH, I. 2025. The multilayered transcriptional architecture of glioblastoma ecosystems. *Nat Genet,* 57**,** 1155-1167.

O'ROURKE, D. M., NASRALLAH, M. P., DESAI, A., MELENHORST, J. J., MANSFIELD, K., MORRISSETTE, J. J. D., MARTINEZ-LAGE, M., BREM, S., MALONEY, E., SHEN, A., ISAACS, R., MOHAN, S., PLESA, G., LACEY, S. F., NAVENOT, J. M., ZHENG, Z., LEVINE, B. L., OKADA, H., JUNE, C. H., BROGDON, J. L. & MAUS, M. V. 2017. A single dose of peripherally infused EGFRvIII-directed CAR T cells mediates antigen loss and induces adaptive resistance in patients with recurrent glioblastoma. *Sci Transl Med,* 9.

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

OHGAKI, H. & KLEIHUES, P. 2013. The definition of primary and secondary glioblastoma. *Clin Cancer Res,* 19**,** 764-72.

OLDRINI, B., VAQUERO-SIGUERO, N., MU, Q., KROON, P., ZHANG, Y., GALAN-GANGA, M., BAO, Z., WANG, Z., LIU, H., SA, J. K., ZHAO, J., KIM, H., RODRIGUEZ-PERALES, S., NAM, D. H., VERHAAK, R. G. W., RABADAN, R., JIANG, T., WANG, J. & SQUATRITO, M. 2020. MGMT genomic rearrangements contribute to chemotherapy resistance in gliomas. *Nat Commun,* 11**,** 3883.

OLYMPIOS, N., GILARD, V., MARGUET, F., CLATOT, F., DI FIORE, F. & FONTANILLES, M. 2021. TERT Promoter Alterations in Glioblastoma: A Systematic Review. *Cancers (Basel),* 13.

OSTROM, Q. T., BAUCHET, L., DAVIS, F. G., DELTOUR, I., FISHER, J. L., LANGER, C. E., PEKMEZCI, M., SCHWARTZBAUM, J. A., TURNER, M. C., WALSH, K. M., WRENSCH, M. R. & BARNHOLTZ-SLOAN, J. S. 2014. The epidemiology of glioma in adults: a "state of the science" review. *Neuro Oncol,* 16**,** 896-913.

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.

PERRY, A. & WESSELING, P. 2016. Histologic classification of gliomas. *Handb Clin Neurol,* 134**,** 71-95.

PHILIPS, A., HENSHAW, D. L., LAMBURN, G. & O'CARROLL, M. J. 2018. Brain Tumours: Rise in Glioblastoma Multiforme Incidence in England 1995-2015 Suggests an Adverse Environmental or Lifestyle Factor. *J Environ Public Health,* 2018**,** 7910754.

RABE, M., DUMONT, S., ALVAREZ-ARENAS, A., JANATI, H., BELMONTE-BEITIA, J., CALVO, G. F., THIBAULT-CARPENTIER, C., SERY, Q., CHAUVIN, C., JOALLAND, N., BRIAND, F., BLANDIN, S., SCOTET, E., PECQUEUR, C., CLAIRAMBAULT, J., OLIVER, L., PEREZ-GARCIA, V., NADARADJANE, A., CARTRON, P. F., GRATAS, C. & VALLETTE, F. M. 2020. Identification of a transient state during the acquisition of temozolomide resistance in glioblastoma. *Cell Death Dis,* 11**,** 19.

RAMIREZ, Y. P., WEATHERBEE, J. L., WHEELHOUSE, R. T. & ROSS, A. H. 2013. Glioblastoma multiforme therapy and mechanisms of resistance. *Pharmaceuticals (Basel),* 6**,** 1475-506.

RAMÓN Y CAJAL, S., SESÉ, M., CAPDEVILA, C., AASEN, T., DE MATTOS-ARRUDA, L., DIAZ-CANO, S. J., HERNÁNDEZ-LOSA, J. & CASTELLVÍ, J. 2020. Clinical implications of intratumor heterogeneity: challenges and opportunities. *Journal of Molecular Medicine,* 98**,** 161-177.

RINGEL, F., PAPE, H., SABEL, M., KREX, D., BOCK, H. C., MISCH, M., WEYERBROCK, A., WESTERMAIER, T., SENFT, C., SCHUCHT, P., MEYER, B., SIMON, M. & GROUP, S. N. S. 2016. Clinical benefit from resection of recurrent glioblastomas: results

of a multicenter study including 503 patients with recurrent glioblastomas undergoing surgical resection. *Neuro Oncol,* 18**,** 96-104.

RIPPAUS, N., F-BRUNS, A., TANNER, G., TAYLOR, C., DROOP, A., BRÜNING-RICHARDSON, A., CARE, M. A., WILKINSON, J., JENKINSON, M. D., BRODBELT, A., CHAKRABARTY, A., ISMAIL, A., SHORT, S. & STEAD, L. F. 2019. JARID2 facilitates transcriptional reprogramming in glioblastoma in response to standard treatment. *biorxiv*.

RIVERA, A. L., PELLOSKI, C. E., GILBERT, M. R., COLMAN, H., DE LA CRUZ, C., SULMAN, E. P., BEKELE, B. N. & ALDAPE, K. D. 2010. MGMT promoter methylation is predictive of response to radiotherapy and prognostic in the absence of adjuvant alkylating chemotherapy for glioblastoma. *Neuro Oncol,* 12**,** 116-21.

SCIUSCIO, D., DISERENS, A. C., VAN DOMMELEN, K., MARTINET, D., JONES, G., JANZER, R. C., POLLO, C., HAMOU, M. F., KAINA, B., STUPP, R., LEVIVIER, M. & HEGI, M. E. 2011. Extent and patterns of MGMT promoter methylation in glioblastoma- and respective glioblastoma-derived spheres. *Clin Cancer Res,* 17**,** 255-66.

SEVIM, H., PARKINSON, J. F. & MCDONALD, K. L. 2011. Etoposide-mediated glioblastoma cell death: dependent or independent on the expression of its target, topoisomerase II alpha? *J Cancer Res Clin Oncol,* 137**,** 1705-12.

SINGH, N., MINER, A., HENNIS, L. & MITTAL, S. 2021. Mechanisms of temozolomide resistance in glioblastoma - a comprehensive review. *Cancer Drug Resist,* 4**,** 17-43.

SOTTORIVA, A., SPITERI, I., PICCIRILLO, S. G., TOULOUMIS, A., COLLINS, V. P., MARIONI, J. C., CURTIS, C., WATTS, C. & TAVARE, S. 2013. Intratumor heterogeneity in human glioblastoma reflects cancer evolutionary dynamics. *Proc Natl Acad Sci U S A,* 110**,** 4009-14.

SPITZER, A., JOHNSON, K. C., NOMURA, M., GAROFANO, L., NEHAR-BELAID, D., DARNELL, N. G., GREENWALD, A. C., BUSSEMA, L., OH, Y. T., VARN, F. S., D'ANGELO, F., GRITSCH, S., ANDERSON, K. J., MIGLIOZZI, S., GONZALEZ CASTRO, L. N., CHOWDHURY, T., ROBINE, N., REEVES, C., PARK, J. B., LIPSA, A., HERTEL, F., GOLEBIEWSKA, A., NICLOU, S. P., NUSRAT, L., KELLET, S., DAS, S., MOON, H. E., PAEK, S. H., BIELLE, F., LAURENGE, A., DI STEFANO, A. L., MATHON, B., PICCA, A., SANSON, M., TANAKA, S., SAITO, N., ASHLEY, D. M., KEIR, S. T., LIGON, K. L., HUSE, J. T., YUNG, W. K. A., LASORELLA, A., IAVARONE, A., VERHAAK, R. G. W., TIROSH, I. & SUVA, M. L. 2025. Deciphering the longitudinal trajectories of glioblastoma ecosystems by integrative single-cell genomics. *Nat Genet,* 57**,** 1168-1178.

STROBEL, H., BAISCH, T., FITZEL, R., SCHILBERG, K., SIEGELIN, M. D., KARPEL-MASSLER, G., DEBATIN, K. M. & WESTHOFF, M. A. 2019. Temozolomide and Other Alkylating Agents in Glioblastoma Therapy. *Biomedicines,* 7.

STUPP, R., MASON, W. P., VAN DEN BENT, M. J., WELLER, M., FISHER, B., TAPHOORN, M. J., BELANGER, K., BRANDES, A. A., MAROSI, C., BOGDAHN, U., CURSCHMANN, J., JANZER, R. C., LUDWIN, S. K., GORLIA, T., ALLGEIER, A., LACOMBE, D., CAIRNCROSS, J. G., EISENHAUER, E., MIRIMANOFF, R. O., EUROPEAN ORGANISATION FOR, R., TREATMENT OF CANCER BRAIN, T., RADIOTHERAPY, G. & NATIONAL CANCER INSTITUTE OF CANADA CLINICAL TRIALS, G. 2005. Radiotherapy plus concomitant and adjuvant temozolomide for glioblastoma. *N Engl J Med,* 352**,** 987-96.

SUCHORSKA, B., WELLER, M., TABATABAI, G., SENFT, C., HAU, P., SABEL, M. C., HERRLINGER, U., KETTER, R., SCHLEGEL, U., MAROSI, C., REIFENBERGER, G., WICK, W., TONN, J. C. & WIRSCHING, H. G. 2016. Complete resection of contrast-enhancing tumor volume is associated with improved survival in recurrent glioblastoma-results from the DIRECTOR trial. *Neuro Oncol,* 18**,** 549-56.

TAMIMI AF, J. M. 2017. Epidemiology and Outcome of Glioblastoma. *Codon Publications; 2017 Sep 27. Chapter 8*.

THOMAS, A., TANAKA, M., TREPEL, J., REINHOLD, W. C., RAJAPAKSE, V. N. & POMMIER, Y. 2017. Temozolomide in the Era of Precision Medicine. *Cancer Res,* 77**,** 823-826.

UNO, M., OBA-SHINJO, S. M., CAMARGO, A. A., MOURA, R. P., AGUIAR, P. H., CABRERA, H. N., BEGNAMI, M., ROSEMBERG, S., TEIXEIRA, M. J. & MARIE, S. K. 2011. Correlation of MGMT promoter methylation status with gene and protein expression levels in glioblastoma. *Clinics (Sao Paulo),* 66**,** 1747-55.

WANG, J., CAZZATO, E., LADEWIG, E., FRATTINI, V., ROSENBLOOM, D. I., 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. *Nat Genet,* 48**,** 768-76.

WANG, Q., HU, B., HU, X., KIM, H., SQUATRITO, M., SCARPACE, L., DECARVALHO, A. C., LYU, S., LI, P., LI, Y., BARTHEL, F., CHO, H. J., LIN, Y. H., SATANI, N., MARTINEZ-LEDESMA, E., ZHENG, S., CHANG, E., SAUVE, C. G., OLAR, A., LAN, Z. D., FINOCCHIARO, G., PHILLIPS, J. J., BERGER, M. S., GABRUSIEWICZ, K. R., WANG, G., ESKILSSON, E., HU, J., MIKKELSEN, T., DEPINHO, R. A., MULLER, F., HEIMBERGER, A. B., SULMAN, E. P., NAM, D. H. & VERHAAK, R. G. W. 2017. Tumor Evolution of Glioma-Intrinsic Gene Expression Subtypes Associates with Immunological Changes in the Microenvironment. *Cancer Cell,* 32**,** 42-56 e6.

WANG, Y., ZHANG, J., LI, W., JIANG, T., QI, S., CHEN, Z., KANG, J., HUO, L., WANG, Y., ZHUGE, Q., GAO, G., WU, Y., FENG, H., ZHAO, G., YANG, X., ZHAO, H., WANG, Y., YANG, H., KANG, D., SU, J., LI, L., JIANG, C., LI, G., QIU, Y., WANG, W., WANG, H., XU, Z., ZHANG, L. & WANG, R. 2021. Guideline conformity to the Stupp regimen in patients with newly diagnosed glioblastoma multiforme in China. *Future Oncol,* 17**,** 4571-4582.

WELLER, M. & LE RHUN, E. 2020. How did lomustine become standard of care in recurrent glioblastoma? *Cancer Treat Rev,* 87**,** 102029.

WELLER, M., STUPP, R., REIFENBERGER, G., BRANDES, A. A., VAN DEN BENT, M. J., WICK, W. & HEGI, M. E. 2010. MGMT promoter methylation in malignant gliomas: ready for personalized medicine? *Nat Rev Neurol,* 6**,** 39-51.

WEMMERT, S., KETTER, R., RAHNENFUHRER, J., BEERENWINKEL, N., STROWITZKI, M., FEIDEN, W., HARTMANN, C., LENGAUER, T., STOCKHAMMER, F., ZANG, K. D., MEESE, E., STEUDEL, W. I., VON DEIMLING, A. & URBSCHAT, S. 2005. Patients with high-grade gliomas harboring deletions of chromosomes 9p and 10q benefit from temozolomide treatment. *Neoplasia,* 7**,** 883-93.

WEN, J., CHEN, W., ZHU, Y. & ZHANG, P. 2021. Clinical features associated with the efficacy of chemotherapy in patients with glioblastoma (GBM): a surveillance, epidemiology, and end results (SEER) analysis. *BMC Cancer,* 21**,** 81.

WEN, P. Y., WELLER, M., LEE, E. Q., ALEXANDER, B. M., BARNHOLTZ-SLOAN, J. S., BARTHEL, F. P., BATCHELOR, T. T., BINDRA, R. S., CHANG, S. M., CHIOCCA, E. A., CLOUGHESY, T. F., DEGROOT, J. F., GALANIS, E., GILBERT, M. R., HEGI, M. E., HORBINSKI, C., HUANG, R. Y., LASSMAN, A. B., LE RHUN, E., LIM, M., MEHTA, M. P., MELLINGHOFF, I. K., MINNITI, G., NATHANSON, D., PLATTEN, M., PREUSSER, M., ROTH, P., SANSON, M., SCHIFF, D., SHORT, S. C., TAPHOORN, M. J. B., TONN, J. C., TSANG, J., VERHAAK, R. G. W., VON DEIMLING, A., WICK, W., ZADEH, G., REARDON, D. A., ALDAPE, K. D. & VAN DEN BENT, M. J. 2020. Glioblastoma in adults: a Society for Neuro-Oncology (SNO) and European Society of Neuro-Oncology (EANO) consensus review on current management and future directions. *Neuro Oncol,* 22**,** 1073-1113.

WOODROFFE, R. W., ZANATY, M., SONI, N., MOTT, S. L., HELLAND, L. C., PASHA, A., MALEY, J., DHUNGANA, N., JONES, K. A., MONGA, V. & GREENLEE, J. D. W. 2020. Survival after reoperation for recurrent glioblastoma. *J Clin Neurosci,* 73**,** 118-124.

ZHANG, J. J. Y., LEE, K. S., VOISIN, M. R., HERVEY-JUMPER, S. L., BERGER, M. S. & ZADEH, G. 2020. Awake craniotomy for resection of supratentorial glioblastoma: a systematic review and meta-analysis. *Neurooncol Adv,* 2**,** vdaa111.

CHAPTER 2

2.1 INTRODUCTION

2.1.1 DNA Sequencing Techniques

As GBM is a genetically complicated cancer, due to both inter- and intra- heterogeneity, it is important to understand the genetic changes that cause tumour growth, spread and resistance to treatment in order to improve testing results and patient health (Barthel et al., 2019, Korber et al., 2019, Brennan et al., 2013). Advances in next-generation sequencing (NGS) technologies have significantly enhanced our capacity to investigate these genomic alterations. WGS and WES are two of the most popular NGS methods. Each has its own advantages and disadvantages when it comes to finding somatic changes. To select the suitable technology, one should carefully think about the outcomes of each method to achieve the maximum benefit of it.

WGS captures the whole genome, which includes both coding (exonic) and non-coding regions. Because of that, WGS can detect many types of genetic changes, including single nucleotide variants, structural differences, copy number changes, and rare mutations. For complicated cancers like GBM, where mutations happen in many genomic areas, WGS is very helpful because it can find changes in regulatory elements like enhancers and promoters, as well as intergenic regions. Covering the non-coding regions helps us to identify broader genomic changes that drive GBM. Also, WGS is an ideal method to find large-scale genomic rearrangements and translocations, which are often very important for the growth and spread of the tumour.

Despite its extensive scope, WGS has some limitations. Because of its high cost and the amount of data it produces, high-performance processing power is needed to analyse it. WGS may become less commonly used as a result, especially in clinical settings. Furthermore, a sizable amount of WGS data originates from non-coding areas, many of which have ambiguous functional responsibilities. Prioritising and interpreting mutations becomes difficult, particularly when separating significant changes from noise. WGS generally reduced coverage is another disadvantage that may make it more difficult to find low-frequency somatic mutations. Formalin-fixed paraffin-embedded (FFPE) materials make this issue worse since DNA fragmentation and degradation further lower sequencing depth and introduce artefacts.

Unlike WGS, WES only looks at coding areas, which comprise 1% to 2% of the genome but are responsible for 85% of all currently known mutations linked to disease (Teer and Mullikin, 2010). WES

offers a more cost-effective method of finding mutations that are more likely to have functional effects on proteins by focusing on these areas. WES enables researchers to find protein-coding mutations that may be responsible for tumour growth and progression in cancer research, particularly GBM investigations. WES produces less data than WGS, which streamlines downstream analysis and facilitates interpretation of results. Because of its effectiveness, WES is especially helpful for clinical and translational studies that seek to find mutations that can be put into practice.

However, WES does have some problems. It can't find changes in non-coding regions, like promoters, because it only looks at coding regions. These areas may still be important in GBM biology though, like identifying *TERTp* mutations, which are associated with GBM progression and prognosis. Additionally, WES is not as good as WGS at finding structural changes at the chromosomal level. Also, when DNA quality is low in samples like FFPE-derived tissues, uneven coverage and biases caused by DNA fragmentation can make variant identification less accurate.

Both WGS and WES are crucial approaches, and each has unique advantages that could complement the other when both technologies utilised (Rotunno et al., 2020). While WES offers a targeted approach that focuses on protein-coding genes, WGS provides a complete picture of the genome, including non-coding regions and notable structural variations. To improve the breadth and precision of mutation identification in GBM samples, I used WGS and WES in this chapter. I aimed to get precise characterisation of somatic mutations in GBM by applying both approaches and implementing measures to mitigate their limitations. This dual methodology enabled me to overcome the intrinsic difficulties of studying extremely heterogeneous tumours such as GBM, yielding a more profound comprehension of its molecular landscape.

2.1.2 DNA Quality and Challenges with FFPE Samples

FFPE samples are extensively utilised in clinical and research environments because they effectively maintain tissue for prolonged durations. Nonetheless, the formalin fixation method presents numerous obstacles that hinder molecular analysis of FFPE samples, particularly in next-generation sequencing (NGS) investigations. Formaldehyde cross-links DNA to proteins, maintaining tissue architecture while compromising DNA integrity. This cross-linking inhibits effective DNA denaturation during sequencing, leading to shorter sequencing reads and diminishing the efficiency of subsequent processing workflows. Moreover, DNA in FFPE samples is frequently fragmented, an issue that exacerbates with extended storage in older specimens (Robbe et al., 2018, Einaga et al., 2017, Sah et al., 2013, Steiert et al., 2023). This fragmentation reduces overall read length and coverage quality, complicating the acquisition of the depth necessary for precise variant detection.

Environmental variables, such as suboptimal storage temperatures or elevated humidity, accelerate DNA degradation over time. Oxidative damage and enzymatic degradation additionally facilitate the generation of short, fragmented DNA molecules (Costello et al., 2013). These degraded fragments produce sequencing data of inferior quality, hindering alignment and variant detection operations. In WGS and WES, adequate read depth is essential for identifying low-frequency somatic mutations; nevertheless, the difficulties associated with FFPE-derived DNA might considerably compromise the reliability of the results. In cancer genomics, the accurate identification of somatic mutations necessitates addressing these challenges to yield significant results.

A common issue in FFPE samples is the elevated occurrence of cytosine-to-thymine (C>T) substitutions. The artefacts result from cytosine deamination, a chemical reaction that is expedited during formalin fixation, particularly at methylated cytosine sites. This alteration transforms cytosine into uracil, which is interpreted as thymine during sequencing, resulting in C>T transition inaccuracies (Kim et al., 2017). Such artefacts are particularly common in older FFPE samples or those preserved under suboptimal circumstances. The elevated frequency of these changes complicates the differentiation between genuine somatic mutations and sequencing artefacts. This poses significant challenges when evaluating aggressive malignancies such as glioblastoma, where precise identification of mutations is crucial for comprehending the tumour's genetic landscape.

Optimised bioinformatic procedures are necessary to address these problems (Steiert et al., 2023). Common approaches include applying filters to mitigate artefacts resulting from C>T substitutions. These filters relied on a variety of measures, including variable allele frequency (VAF) and strand bias assessments. Implementing these corrective procedures can enhance the precision of somatic variant calls, reducing the likelihood of sequencing artefacts. In addition to bioinformatic solutions, practical laboratory strategies such as using kits designed for FFPE-derived DNA can also be employed to improve the overall quality of the data.

In summary, FFPE processing substantially affects DNA quality by causing fragmentation, degradation, and the introduction of C>T mutations. These issues require meticulous consideration during both the experimental and analytical phases of NGS processes. The optimised pipeline reduced the influence of FFPE-induced artefacts on somatic mutation detection in GBM samples by employing rigorous quality control measures and customised bioinformatics approaches.

2.1.3 Double Counting of Variants in Overlapping Paired Reads

Although paired-end sequencing improves coverage and accuracy by sequencing DNA fragments from both ends, it may result in regions where forward and reverse reads overlap. These overlaps could result in "double counting" of variants, which would inflate their frequency if they were not properly treated (Pope et al., 2014). In cancer research, where accurate VAF measurements are essential for comprehending tumour growth, this problem is very important.

This issue is made worse in FFPE samples due to the shorter DNA fragments resulting from fragmentation, which raise the possibility of overlaps. Analysis of clonal and subclonal mutations, which are essential for assessing the progression of tumours in GBM, can be distorted by inflated VAFs, which can be deceptive for low-frequency variants.
Bioinformatic solutions can be used at various points in the analysis pipeline to resolve overlapping paired-end reads: pre-alignment corrections at the FASTQ level and post-alignment corrections at the BAM level. Every approach has unique advantages and disadvantages.

2.1.3.1 Correction at the FASTQ Level

FASTQ-level corrections address overlapping reads before alignment by merging overlapping paired-end reads into a consensus sequence. This approach eliminates the risk of double-counting mutations by ensuring each variant is counted only once per fragment. This method has many advantages:

- Merging overlapping reads prevents inflation of VAFs, ensuring reliable mutation frequency calculations.
- Consolidating overlapping regions into a single sequence simplifies downstream variant calling and data analysis.

However, it has a few limitations such as:

- Merging paired reads decreases the total read count, which can affect depth-sensitive analyses like copy number alteration (CNA) calling.
- Reduced depth may obscure subtle copy number changes and hinder detection of structural variations such as translocations and insertions.

BBMerge (Bushnell et al., 2017), a commonly used FASTQ-level tool, merges overlapping paired reads into a consensus sequence before alignment. While it effectively resolves overlaps, its impact on read

depth makes it unsuitable for depth-sensitive analyses, such as somatic copy number and structural variant detection.

2.1.3.2 Correction at the BAM Level

BAM-level corrections handle overlapping reads after alignment by modifying the aligned data without merging paired-end reads. This approach removes overlapping bases through hard or soft clipping, preventing double-counting of mutations while preserving read depth and the paired-end structure, ensuring accurate downstream analyses. This approach provides significant benefits:

- BAM-level clipping maintains read depth, rendering it suitable for depth-sensitive analysis like somatic copy number change (sCNA) detection.
- The paired-end read configuration is maintained, which is essential for detecting structural variants and other analyses dependent on read pairs.
- This method eliminates overlapping bases, hence preventing inflating in overlapping regions and providing precise variant allele frequency (VAF) estimates.

BAM-level processing tools, like fgbio-ClipBam, identify overlapping regions in paired reads and clip the overlapped bases. In this chapter I used fgbio ClipBam to repair overlapping paired-end reads from FFPE-derived GBM samples. This approach was selected due to its preservation of read depth, which is essential for copy number analysis and the detection of somatic mutations. As the sequencing coverage was maintained by masking overlapping areas rather than merging reads, accurate variant allele frequency estimates and detection of copy number changes are likely achievable.

Addressing double-counting artefacts via BAM-level clipping is a common approach to ensure accurate downstream analyses, including variant detection, and pathway analysis. This method improves the resolution of clonal versus subclonal mutations in GBM, providing significant insights into the tumour's mutational landscape and evolutionary dynamics. Such strategies provide a reliable framework for detecting somatic mutations in fragmented DNA, thereby advancing our understanding of GBM biology and progression.

2.1.4 Overview of Analytical Tools Available

To conduct a comprehensive and accurate analysis of both whole genome and whole exome sequencing data in glioblastoma samples, a carefully curated set of bioinformatic tools is required at each stage of the analysis pipeline. These tools are selected to address specific challenges associated

with NGS data, particularly in the context of somatic variant detection, data quality assessment, and copy number alteration analysis. Typically, the analytical process involves multiple steps, starting from raw sequencing data and progressing through quality control, alignment, variant and CNA calling, and downstream interpretation. A variety of tools are available for each of these steps, and their selection often depends on the characteristics of the data and the goals of the analysis. Some tools, including those used in this study, have been benchmarked in a published study by our group, which demonstrated superior sensitivity and specificity using in silico data for validation (Tanner et al., 2021). This section provides an overview of commonly used tools and their role in the analysis workflow.

FASTQC: Sequencing data is obtained in binary base call format (BCL), which is not a suitable input format to use in downstream applications. However, the raw data will be converted to FASTQ format to be used for downstream processing. FastQC (Andrews, 2010) is the most widely used tool to assess the quality of the fastq sequencing files. It is a critical step to check the validity of the raw sequencing data before proceeding to further analysis, such as variant calling. A wide range of quality metrics that FastQC can generate for individual fastq files includes per base sequence quality, GC content, sequence length distribution, sequence duplication proportions and adapter contaminations. These quality checks help remove low-base quality reads, trim the adapter remaining in the 5' or 3' ends of reads, and give an insight into the diversity of the data.

Cutadapt: The presence of over-represented sequences indicated by FastQC means there could be sequencing adapters or primers in the data. This problem happens when the read becomes longer than the DNA fragment (insert) that is sequenced. As these sequences are artificial, they must be removed before using the data for further analysis. Amongst many tools used to trim these unwanted sequences is Cutadapt (Martin, 2011). It is fast, efficient and works with paired-end data. It also allows the user to drop reads that become low quality after trimming, making the data more reliable for sequence alignment.

BWA (Burrows-Wheeler Aligner): BWA (Li and Durbin, 2009) is the most widely used software for mapping the sequencing reads generated by Illumina against the human genome. It is fast and efficient, with a high mapping percentage and less alignment error rate. Briefly, BWA works on finding the maximal exact matches of alignment seeds and adapts the affine gap of the Smith-Waterman algorithm when extending the seeds. Users can specify a threshold for mapping quality to increase the accuracy of variant calling. BWA generates SAM files which

can be converted into BAM format as the latter is smaller in size and more reliable for downstream analysis.

Picard tools: Picard tools are a set of tools developed to process SAM, BAM and VCF files (https://broadinstitute.github.io/picard/). One of the commonly used tools is MarkDuplicates. In NGS, read depth may contain duplicated reads meaning that a variant might be supported by an artefact read. To ensure that alignment files do not contain duplicated reads, files must be processed before variant calling to eliminate any source of false positive variants. These duplicates can arise from library preparation or during cluster formation. At the end of tagging the duplicates process, MarkDuplicates software generates a metrics file to show the duplicate rate of the data.

ClipBam: ClipBam is a tool that eliminates overlap between the paired-end reads (http://fulcrumgenomics.github.io/fgbio/tools/latest/ClipBam.html). It works by clipping the sequencing reads from the same template. Clipping occurs at any end of read 1 and read 2 only if they are forward and reverse (FR) read pairs, with nearly half of overlapped bases being hard clipped as demonstrated in Figure 2-1. This is a crucial step for downstream processes, specifically variant calling, to avoid double-count evidence from the same fragment when both reads cover the variant locus in the same template. Hard clipping is the default setting; however, users can choose the soft-clipping parameter instead. Users must sort the BAM file by query name instead of coordinates before running the software to ensure correct clipping.

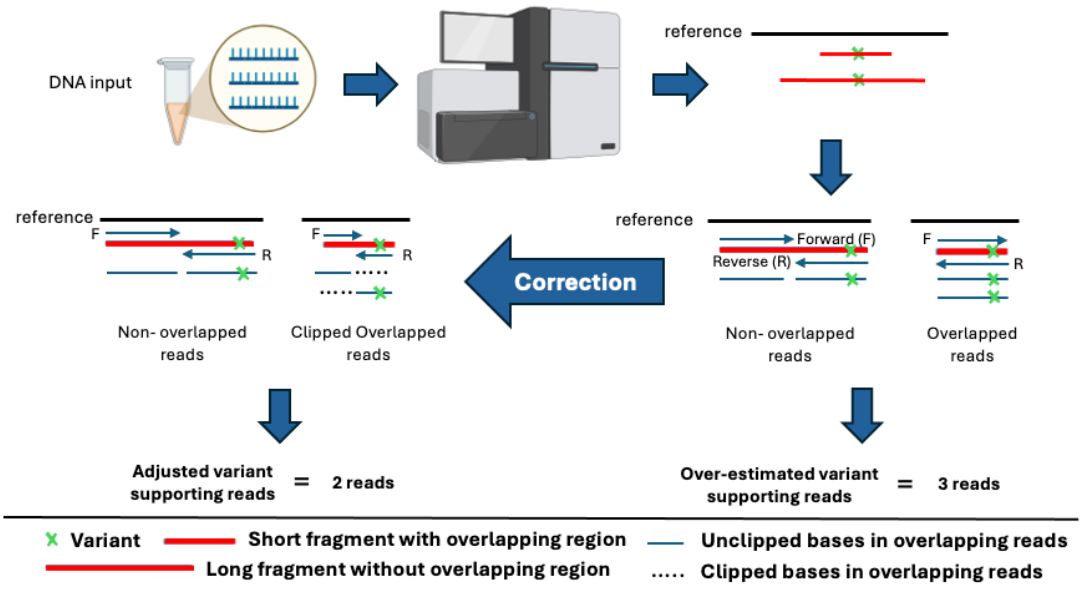


Figure 2-1: Clipping overlapping reads.

This illustration demonstrates how short DNA fragments can cause overlapping of paired-end reads, leading to overcounting of variants before overlap correction. The red lines represent two DNA fragments carrying the same SNV, one short and one long. In the before correction panel, the short fragment produces one paired-reads whose overlapping regions cause double counting (totalling two variant supporting reads), in addition to one variant supporting read from the non-overlapping reads of the long fragment. In the after correction panel, clipping removes the overlapping bases toward the 3' ends of short reads, reducing the overcounted two variant supporting reads back to the true one, while the non-overlapping reads remain unchanged. Dotted lines indicate the clipped bases.

Mutect2: Mutect2 is a somatic variant caller developed at the Broad Institute (Benjamin et al., 2019). It uses Bayesian classification to identify somatic alterations. Mutect2 doesn't process reads with low mapping quality reducing the chance of calling artefacts. It starts by defining regions of the genome called active regions, which have evidence of variation. The active regions will then comprise all the possible haplotypes in the data. Mutect2 then performs haplotype local realignment against the reference haplotype using Smith-Watermann algorithm to identify candidate variant sites. Using pair hidden Markov model, each read will then undergo pairwise alignment to compute the likelihood of alleles for every possible variant site. The likelihood of each genotype is calculated by using the likelihood of alleles calculated earlier by applying the Bayes rule.

FACETS: FACETS (Fraction and Allele-Specific Copy Number Estimates from Tumour Sequencing) infers allele-specific copy numbers, tumour purity, and ploidy from tumour-normal sequencing data (Shen and Seshan, 2016). It normalizes read depth, segments the

genome based on log-ratio signals, and uses allelic imbalance at heterozygous SNVs to distinguish clonal and subclonal alterations. This provides accurate copy number profiles and insights into tumour clonal architecture.

VEP (Variant Effect Predictor): VEP is a variant annotation tool developed at Ensembl (McLaren et al., 2016). It integrates multiple databases and allows using different plugins for other databases to predict the consequences of variants on the protein function. It annotates the variants using a set of sequence ontology consequence terms according to the variant's location, exonic, intronic or splice site. Variants are ranked based on the impact from high to low classifying nonsense or frameshift variants as highly impactful. By default, VEP uses SIFT and PolyPhen databases to predict the functional impact of missense variants on protein. SIFT reports variants with either deleterious or tolerated, while PolyPhen predicts variants to be either probably damaging, possibly damaging, benign or unknown. VEP also enables filtering variants based on pathogenicity predictions using the filter_vep plugin.

GISTIC: GISTIC is software utilised to identify gains and losses across the genome using somatic copy number aberrations data (Mermel et al., 2011). It analyses the segmentation data to identify genes targeted by these gains and losses. GISTIC uses a scoring scheme for each aberration, assigning the score based on the occurrences of the aberration across the samples and its amplitude. It considers the false discovery rate and calculates a q-value for every aberrant region. After analysis, GISTIC creates two plots, one for the amplified regions and one for the deleted regions. Each plot contains peaks which represents the altered locus and the regions within the boundaries of the peak are likely to have the candidate genes. GISTIC requires a file that contains the segmented data for all of the sample in order to generate the plots. Using the CNA calling data, the two prerequisite files were generated for primary and recurrent tumour profiles to identify the genetic changes through treatment.

Bam-readcount: Bam-readcount generates high-resolution depth information around specific genomic loci, providing detailed counts of reads supporting each variant (Khanna et al., 2022). This enables accurate variant allele frequency (VAF) estimation, making it particularly valuable for filtering out FFPE-induced artifacts and distinguishing true somatic mutations from sequencing noise.

Nextflow: Nextflow is one of the scientific workflow tools that have scalability and reproducibility (Di Tommaso et al., 2017). It is an example of a domain-specific language (DSL)

that simplifies the construction and integration of computational pipelines for genomic data analysis. It uses channels for input and output data and processes that can be written in many languages such as Python, Perl, Bash, etc. Nextflow enables the parallelisation of jobs and can be used in high-performance computing clusters or cloud systems.

Each of these tools contributed to a robust and accurate analysis pipeline, addressing the unique challenges presented by FFPE-derived DNA, overlapping read pairs, and somatic mutation detection in glioblastoma samples. Together, these tools enabled the comprehensive and reliable identification of somatic mutations, structural variations, and copy number alterations that form the basis of this study's insights into the genomic landscape of GBM.

2.2 METHODS

2.2.1 Workflow Automation for WGS and WES Analysis with Nextflow

To handle the high-throughput requirements of WGS and WES data processing efficiently, I developed a Nextflow pipeline to automate and streamline each step of the workflow on the high-performance computing (HPC) service, ARC3 and ARC4. Leveraging the HPC resources enabled efficient job submission across multiple samples, and Nextflow allowed me to design a modular and automated pipeline that covered essential tools for my sequencing data, including FASTQC, Cutadapt, BWA, BEDTools, Picard, GATK tools, Mutect2, VEP, Facets, and fgbio ClipBam.

Using Nextflow's built-in functionality for job scheduling, I was able to automate submission across the HPC's job queue, maximizing computational resources by parallelizing tasks such as quality control, alignment, and variant calling. Each tool was integrated into distinct Nextflow processes, which allowed for efficient execution while maintaining clear dependencies between steps. This setup not only reduced hands-on time but also minimized errors and variability between runs by enforcing a standardized process across all samples. Furthermore, Nextflow's robust error-handling features simplified troubleshooting, making it easier to monitor job statuses and rerun failed steps without restarting the entire pipeline, enhancing the pipeline's reproducibility and reliability. The Nextflow scripts developed for this pipeline are stored in a GitHub repository associated with this chapter for reference (https://github.com/umyma1/thesis_appendix/tree/main/chapter2).

2.2.2 Quality control and mapping of sequencing data

To ensure quality control for the WGS data, , which were generated using the Illumina platform with 150 bp paired-end reads at the MD Anderson Cancer Center (USA), I began by running FastQC on

each raw FASTQ file to assess the overall sequence quality, checking for common issues such as adapter content, GC content, and sequence duplication levels. Once the initial quality of the data was confirmed, I used Cutadapt with the following parameters: -m 20 --overlap 1 -q 20. These settings allowed for the removal of adapter sequences, ensuring that any reads shorter than 20 base pairs were discarded, requiring at least a one-base overlap for adapter trimming, and applying a minimum quality score of 20 across bases. Following adapter trimming, I used BWA to map the cleaned FASTQ files to the reference genome (GRCh38), aligning each read accurately to maximize downstream variant calling reliability. I then used Picard to mark duplicate reads, helping reduce potential biases in variant calling by identifying and flagging PCR and optical duplicates.

For the whole exome sequencing (WES) data, the samples were initially provided in BAM format aligned to an older version of the human genome (GRCh37). To reprocess these data in alignment with the WGS workflow, I first converted each BAM file back to FASTQ format using *bedtools bamtofastq*. Once in FASTQ format, I re-mapped the reads to the GRCh38 reference genome using BWA. After realignment, I applied the same processing pipeline as for the WGS data, using Picard to mark duplicates and GATK BQSR for recalibration, ensuring consistency across both WES and WGS datasets in preparation for downstream analyses.

2.2.3 Post alignment processing and optimisation

To prevent double-counting of variants in overlapping paired-end reads, I used fgbio ClipBam to clip overlapping regions in both WGS and WES BAM files. ClipBam processes BAM files alongside a reference genome, as required by this tool, and provides two clipping options: soft or hard. I selected hard clipping, which removes overlapping sequences from each read, reducing the risk of false-positive variant calls by counting each fragment only once. This method also reduces file size, easing storage demands on the HPC system. After clipping, I examined the resulting BAM files to confirm accurate clipping at high-depth positions, enhancing variant-calling precision. Subsequently, I performed GATK Base Quality Score Recalibration (BQSR) to correct for systematic sequencing errors based on known variant sites, further improving data quality for both WGS and WES data prior to variant calling.

To assess the impact of clipping on variant-supporting reads, I used bam-readcount, which provides detailed metrics on sequencing data at specified nucleotide positions, such as counts of observed bases, mapping and base quality summaries, and read position details. Running bam-readcount on both unclipped and clipped WGS and WES BAM files enabled me to compare variant-supporting read

counts pre- and post-clipping. This approach helped me confirm the effectiveness of clipping, especially in reducing overestimation of variant allele frequencies at overlapping regions.

2.2.4 Variant Calling and Filtering

2.2.4.1 Whole exome data

Since no normal samples were available for the exome datasets, I generated synthetic normal files. Given that the exome sequencing was performed using Agilent's WES kit, I used the Agilent exome enrichment BED file for consistency. This BED file was originally on genome build 37, so I converted it to genome build 38 using UCSC liftOver tool. I then used bcftools to generate synthetic normal exome files from whole genome data masking the off-target regions using the liftover bed file defined by Agilent.

After preprocessing, I used Mutect2 in GATK (version 4.2.0.0) for variant calling in multisample mode, analysing primary and recurrent samples alongside the synthetic normal files to capture somatic mutations specific to each tumour sample. Mutect2 generated an initial set of putative variants, which I refined with FilterMutectCalls to remove low-confidence variants based on GATK's somatic variant filtering thresholds. Additionally, to address potential orientation bias from sequencing artifacts, I applied LearnReadOrientationModel and included this model in FilterMutectCalls to remove false positives caused by read orientation artifacts. This comprehensive filtering process ensured a high confidence set of somatic variant calls across primary and recurrent samples.

2.2.4.2 Whole genome data

Due to computational constraints such as limited running time with a maximum of 48 hours per job on the ARC3 and ARC4 HPC systems, and mutect2 limitations such as unavailability of multithreading, I performed variant calling on a per-chromosome basis for the WGS data. For large chromosomes, I split the genome into regions with size ranges between 25-50 megabases to ensure jobs finished within 48 hours. Using Mutect2 in multisample mode, I applied the same protocol as for exome data, processing primary and recurrent tumour samples with their matched normal samples, but on a chromosome level. To further reduce false positives, I included a blacklist interval for Mutect2 to be excluded from mutation calling. The ENCODE blacklist file (Amemiya et al., 2019) contains the coordinates of many problematic genomic regions such as the low mappability islands, centromeric repeats, telomeric repeats and satellite repeats. Using this strategy, I aimed to speed up the process and avoid artefacts possibly being called from these regions. After filtering with FilterMutectCalls and

incorporating LearnReadOrientationModel to address orientation bias artifacts, I combined the results from each chromosome into a comprehensive, high-confidence variant dataset for the whole genome.

2.2.5 Functional Annotation of Variants with VEP

I used the Variant Effect Predictor (VEP) tool with the human genome reference GRCh38 and VEP release 104, which includes all necessary reference files for variant annotation. I ran VEP on a filtered VCF file of my variants in a Linux environment. To ensure reproducibility, I specified the --cache and --offline options, enabling VEP to access locally cached annotation data. I also included additional flags to enrich the output: --symbol for gene symbols, --canonical for canonical transcripts, --hgvs for standardized HGVS nomenclature, and --sift and --PolyPhen to predict protein impact.

VEP categorizes variants by impact level, including high, moderate, low, and modifier, providing a functional classification for prioritizing variants. High-impact variants are typically those with potentially severe consequences on gene function, such as stop-gain or frameshift mutations, which can result in truncated or nonfunctional proteins. Moderate-impact variants, like missense mutations, may alter protein function but with less certainty. Low-impact variants, such as synonymous mutations, generally have minimal effect on function, while modifier variants typically reside in non-coding regions. After running VEP, I examined the output VCF, which included new columns detailing each annotation field and impact classification. This comprehensive annotation file enabled a structured analysis of variant effects and prioritization for further study.

Following initial annotation with VEP, I applied a custom Python script to further classify variants as unique to the primary tumour, unique to the recurrent tumour, or common to both. This additional step enabled the distinction between mutations present at diagnosis and those that emerged or persisted upon recurrence, offering insights into tumour evolution and highlighting mutations that may contribute to tumour progression or serve as markers of recurrence and therapeutic resistance.

2.2.6 Somatic Copy Number Aberrations calling

In this chapter, I identified somatic copy number alterations (CNA) from whole genome sequencing (WGS) data using the FACETS tool, facilitated by the cnv_facets wrapper developed by Dario Beraldi (https://github.com/dariober/cnv_facets). FACETS is optimized for robust CNA calling, with cnv_facets providing an efficient pipeline for handling and parameter optimization. I began with --nbhd-snp auto to automatically determine neighbourhood SNP. The key parameter for segmentation

is cval (critical value), which controls the sensitivity of the segmentation algorithm; a lower cval yields finer, more sensitive segmentation, while a higher cval results in coarser segmentation. Following the authors' recommendation for a two-pass approach, I initially used cval values of 25 (for high sensitivity, fine segmentation) and 400 (for low sensitivity, coarse segmentation), with the high-cval run used to establish tumour purity and baseline Log R ratio.
However, these initial values resulted in over-segmentation. Following the authors' guidance to adjust cval based on dataset-specific factors like data quality and sequencing method, I systematically tested higher cval values: first 50 and 500, and later 100 and 1000. This approach enabled control over segmentation precision, reducing over-segmentation and improving the reliability of the CNA calls. By refining cval values in this manner, a well-balanced CNA profile was achieved, with segmentation sensitivity and continuity adapted to the unique characteristics of the data.

Since FACETS requires comparing tumour samples to matched normal samples, I processed the data by running each primary tumour sample against its matched normal sample and each recurrent tumour sample against the same normal.

2.2.7 Analysis of Somatic Copy Number Alterations Using GISTIC

After running cnv_facets, I obtained VCF files for each sample containing somatic CNA information. These VCF files included fields such as chromosome, position, structural variant type (e.g., duplication or deletion), and copy number data. These fields differ slightly from conventional copy number call formats. For GISTIC analysis, FACETS authors recommend mapping specific fields from the VCF file to match GISTIC's input requirements. Specifically, I used the CHROM field for "Chromosome," POS for "Starting Position," and parsed the INFO field to extract END for "Ending Position," NUM_MARK for "Number of Markers in Segment," and CNLR.MEDIAN for "Seg.CN. i.e. Segment Copy Number" I then reformatted these extracted fields into a segmentation file, as required by GISTIC, with columns for Sample ID, Chromosome, Start Position, End Position, Number of Markers, and Segment Copy Number in log2 ratio.

I ran GISTIC using the default parameters, which are optimized for identifying recurrent, significant copy number alterations across primary and recurrent GBM cohorts. This approach allowed me to identify regions with recurrent CNAs and to highlight key genomic areas involved in tumorigenesis across the cohort.

2.2.8 Data visualisation

Some of the figures in this chapter were generated using a tool named Maftools (Mayakonda et al., 2018), designed to facilitate the visualization and analysis of mutation data. I used Maftools extensively to represent and interpret variant data derived from whole exome sequencing (WES) and to incorporate copy number variation (CNV) data from whole genome sequencing (WGS) analyses.

Starting with the variant data, I converted my VEP-annotated VCFs to the MAF format using vcf2maf.pl, a tool developed by the Genome Data Science group at Memorial Sloan Kettering Cancer Center. While vcf2maf requires VEP and its annotation cache files to function, it is not included in the standard VEP Conda package. However, as a standalone Perl script, it can be easily obtained from the group's GitHub repository and used alongside VEP for VCF to MAF conversion. Once converted, I loaded the MAF files into R, creating MAF objects that allowed for efficient data handling and analysis in Maftools.

After creating the MAF objects, I applied several Maftools visualization functions to analyse these data, starting with *plotTiTv* to examine mutation patterns in terms of transitions and transversions, and *plotVaf* to explore the variant allele frequency distributions. These visualizations helped me understand the clonality and prevalence of mutations across the cohort.

Using *tcgaCompare,* I compared the tumour mutation burden (TMB) of Stead's cohort to the TMB observed in the TCGA cohort, providing a broader overview of how similar Stead's cohort to other cancers in TCGA beside GBM. Finally, I used *oncoplot* function to generate an oncoplot to highlight the key alterations and patterns across patients. This approach provided a detailed, comparative view of the genetic landscape in my study cohort.

Other figures for alignment optimization and distribution of variants and single base substitutions were generated using python custom scripts and can be accessed on this GitHub link for this chapter (https://github.com/umyma1/thesis_appendix/tree/main/chapter2).

2.3 RESULTS

2.3.1 Cohort description

The cohort consists of 34 patients with IDH wild-type GBM, analysed using WGS and WES of longitudinal samples. Each patient had a sample from the primary tumour site collected during initial surgical resection, followed by a sample from a local recurrence after the tumour recurred. Two

patients were excluded from the analysis: one lacked a recurrent sample, and another did not have a primary tumour sample, leaving 32 patients with paired primary-recurrent samples for analysis. The primary samples represent untreated tumours, while the recurrent samples were collected after standard treatment.

Of the 32 patients included, 26 received the standard GBM therapy, which involves surgical resection followed by radiotherapy (RT) and temozolomide (TMZ) (Stupp et al., 2005, Johnson and O'Neill, 2012, Riganti et al., 2014, Huang and Zhou, 2020). Seven patients received only RT after surgery, and one patient underwent treatment with radioactive iodine (RAI) following tumour removal. For 12 patients, immunohistochemistry (IHC) was performed to confirm IDH wild-type status, while no IHC records were available for the remaining patients. Figure 2-1 provides a summary of the cohort characteristics.

Table 2-1: Discovery cohort data

Patient ID	Originating centre	Days between primary and recurrent surgery	Age at diagnosis	Gender	Vital status	Days to death (from primary surgery)	RT after primary	Chemo after primary
A1	A	585	31	F	DECEASED	1795	Y	TMZ
A2	A	825	44	F	DECEASED	1473	Y	PCV
B1	B	276	60	F	DECEASED	351	Y	N
B2	B	338	41	M	DECEASED	440	Y	TMZ
B3	B	295	45	M	DECEASED	376	Y	PCV
B4	B	49	53	M	DECEASED	250	Y	N
B5	B	721	58	M	DECEASED	1106	Y	TMZ
B6	B	677	58	F	DECEASED	1069	Y	TMZ & PCV
B7	B	246	66	M	DECEASED	424	Y	TMZ
B8	B	870	42	F	DECEASED	1150	Y	TMZ
B9	B	49	40	M	UNKNOWN	UNKNOWN	Y	TMZ
B10	B	519	67	M	DECEASED	872	Y	TMZ
B11	B	178	68	M	DECEASED	269	Y	PCV
B12	B	476	47	M	DECEASED	570	Y	TMZ
B13	B	424	45	F	DECEASED	551	Y	TMZ
B14	B	657	33	M	ALIVE	N/A	Y	TMZ
B15	B	293	68	M	ALIVE	N/A	Y	TMZ
B16	B	294	55	M	ALIVE	N/A	Y	TMZ
B17	B	369	53	F	ALIVE	N/A	Y	TMZ
C1	C	652	36	F	DECEASED	1296	Y	TMZ
C2	C	1341	63	F	DECEASED	1464	Y	TMZ & CCNU
C3	C	1265	49	M	ALIVE	N/A	Y	TMZ
C4	C	695	53	M	DECEASED	973	Y	TMZ
C5	C	N/A	35	M	DECEASED	N/A	Y	TMZ
C6	C	799	60	F	DECEASED	1260	Y	TMZ
C7	C	523	55	M	DECEASED	721	Y	TMZ
C8	C	438	60	F	DECEASED	818	Y	TMZ & CCNU
C9	C	315	35	M	DECEASED	N/A	Y	N
C10	C	438	29	M	DECEASED	796	Y	N
C11	C	2141	26	M	DECEASED	3511	N	Radioactive iodine
D1	D	730	56	F	UNKNOWN	UNKNOWN	Y	TMZ & Carmustine
D2	D	365	61	M	UNKNOWN	UNKNOWN	Y	TMZ & Carmustine
D3	D	1095	41	F	UNKNOWN	UNKNOWN	Y	TMZ
D4	D	730	55	M	UNKNOWN	UNKNOWN	Y	TMZ

2.3.2 Assembling a robust and scalable bioinformatics pipeline (workflow)

Next-generation sequencing (NGS) data are massive and require multiple intensive computational steps to achieve reliable results from a high-quality analysis. One can manage to analyse a few

datasets using the local computational resource, but this is not the typical situation. The dataset I'm using in this study comprises 126 x 150bp paired-end genome fastq sequencing files and 79 BAM files generated from exome data for the same cohort. The genome files were primary-recurrent pairs with matched normal blood samples whereas exome files were primary-recurrent pairs without matched normals (Table 2-1). These files include samples collected from patients who underwent multiple recurrent surgeries, as well as re-sequenced samples or samples obtained from different regions of the same tumour. This requires processing the datasets robustly and efficiently to produce accurate results. The tumour samples were from FFPE tissues and the matched normal samples were from blood. Therefore, I opted to use a tool that helped automate executing the required analytical tasks: Nextflow. Nextflow (Di Tommaso et al., 2017) is an example of a domain-specific language (DSL), and it is specialised in the biological field, mainly the genomic data analysis domain. It facilitates the writing of bioinformatics pipelines and allows the integration of multiple programs. Among the advantages that it provides, is its compatibility with the high-performance computing service available for researchers at the University of Leeds. I successfully developed a pipeline starting from Fastqc and ending with annotating the VCF files. Figure 2-2 shows the software employed in the pipeline.

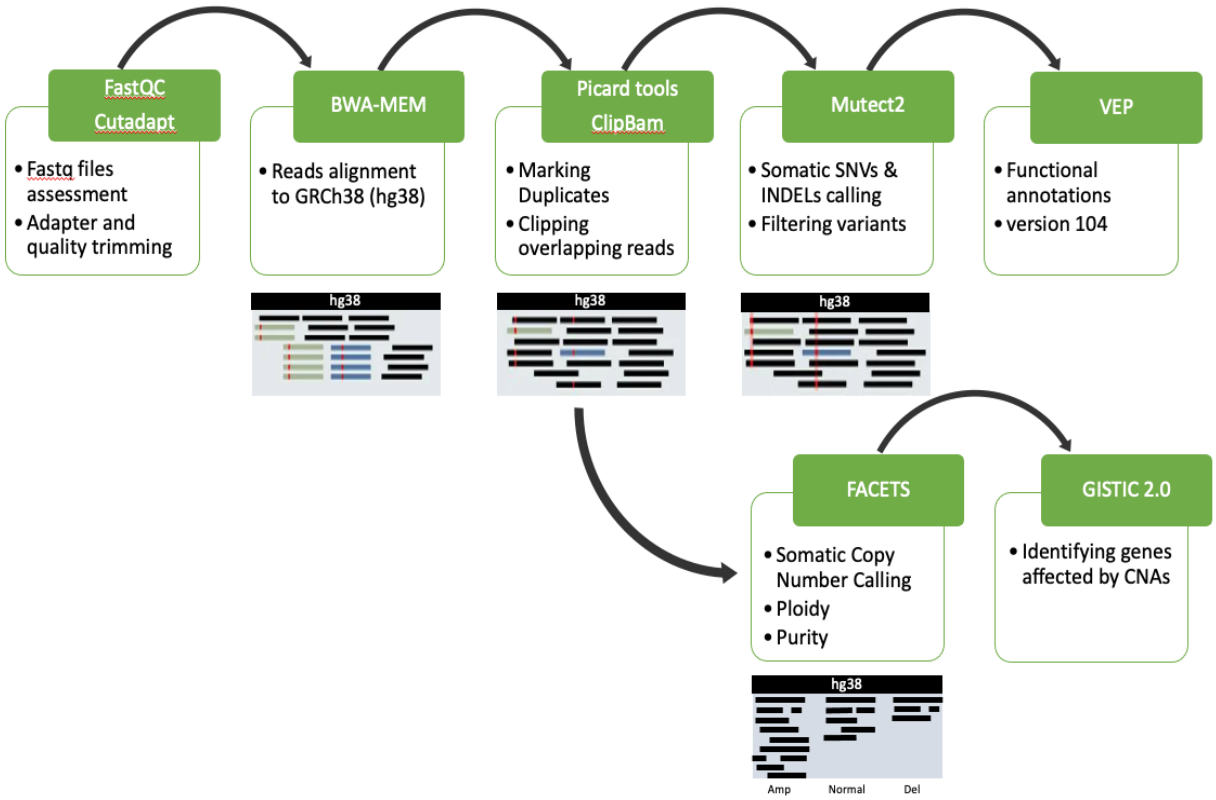


Figure 2-2: Bioinformatics workflow for WGS and WES analysis implemented in Nextflow.

A summary of the multiple processes with the associated tools applied to the raw data to call somatic SNVs and CNAs.

2.3.3 Data quality

2.3.3.1 Pre-alignment processing

Before mapping the reads to the reference genome, I followed the best practice of data preprocessing to ensure that the genome and exome datasets were free of bad-quality reads and sequencing adapters. The FastQC indicated that all files had a high base quality; however, adapter contamination was evident. As those adapters are artificial and not part of the genomic DNA, I had to remove them to minimise the sources of possible wrong base calls. Having known that the libraries were prepared using an Agilent kit, I initially used Agilent universal adapter sequences with the adapter removal tool, Cutadapt (Martin, 2011); however, rerunning FastQC indicated that the adapters have not been removed and suggested that Illumina universal adapter sequences had been used, which is similar to some adapters used in some Agilent library preparation kits . I trimmed these off with Cutadapt, along with bases with a call quality lower than a Phred score of 20 towards the end of the reads.

2.3.3.2 Post-alignment processing

The overall quality of the data was high, as indicated in the FastQC report of the fastq files. Removing the adapters, the overrepresented sequences and reads with quality scores below 20 resulted in higher quality data. However, sequence length distribution was affected due to the applied trimming parameters. The average read length after processing the genome fastq files with Cutadapt is 132 bp per fastq file. In addition, an average of 44% of GC content was reported per raw data file. The mapping rate of the data using BWA was also high. An average of 99% of reads were mapped to the genome. Of the mapped reads, an average of 97% were proper pairs. Finally, the average rate of duplicated sequences was 12%. On the other hand, exome data had a mapping rate of 99%, 51% GC content, 37% duplication. The GC content in the WES data was higher than in the WGS data due to the design of capture probes, which target exonic regions that are typically GC-rich. Also, the duplication rate was higher in the WES data because the capture probes often target short sequences. This results in the enrichment of similar DNA fragments, leading to higher duplication rates.

2.3.4 Diagnostic Checks for Accurate Variant Identification

Before I started the analysis of the WGS data, I searched the literature and found a study that was done on GBM by Korber et al., where they utilised primary and recurrent tumour samples, similar to

my analysis, and used it to for benchmarking (Korber et al., 2019). The whole genome samples were sequenced as paired end with an average depth of coverage 25x for the blood samples and 20x for the tumour samples. Initial variant calling results revealed that a median of 183,187 SNVs and INDELs were found in the set of primary tumours, and a median of 155,264 SNVs and INDELs were found in the recurrent tumours. These findings are far higher than what was reported by other GBM studies that utilised WGS data. Korber et al., 2019, have reported a median of 12,800 somatic variants per sample with read depth of 149x for tumour samples and 78x for matched blood samples. Another study by (Barthel et al., 2019) reported an average of 4,224 somatic variants per sample with >100x depth of coverage. To investigate the higher number of variants, I combined all samples and classified the variants into four categories: shared primary, shared recurrent, unique primary and unique recurrent. I then made diagnostic plots to check the distributions of variants using variant allele frequency (VAF), variant allele depth, and total depth (Figure 2-3). The hypothesis behind the four classifications is that shared primary and shared recurrent variants are likely true variants. This would help me to assign a cut-off to identify the artefacts, which are expected to share features such as low depth, few variants supporting reads and low VAF, and confidently eliminate them.

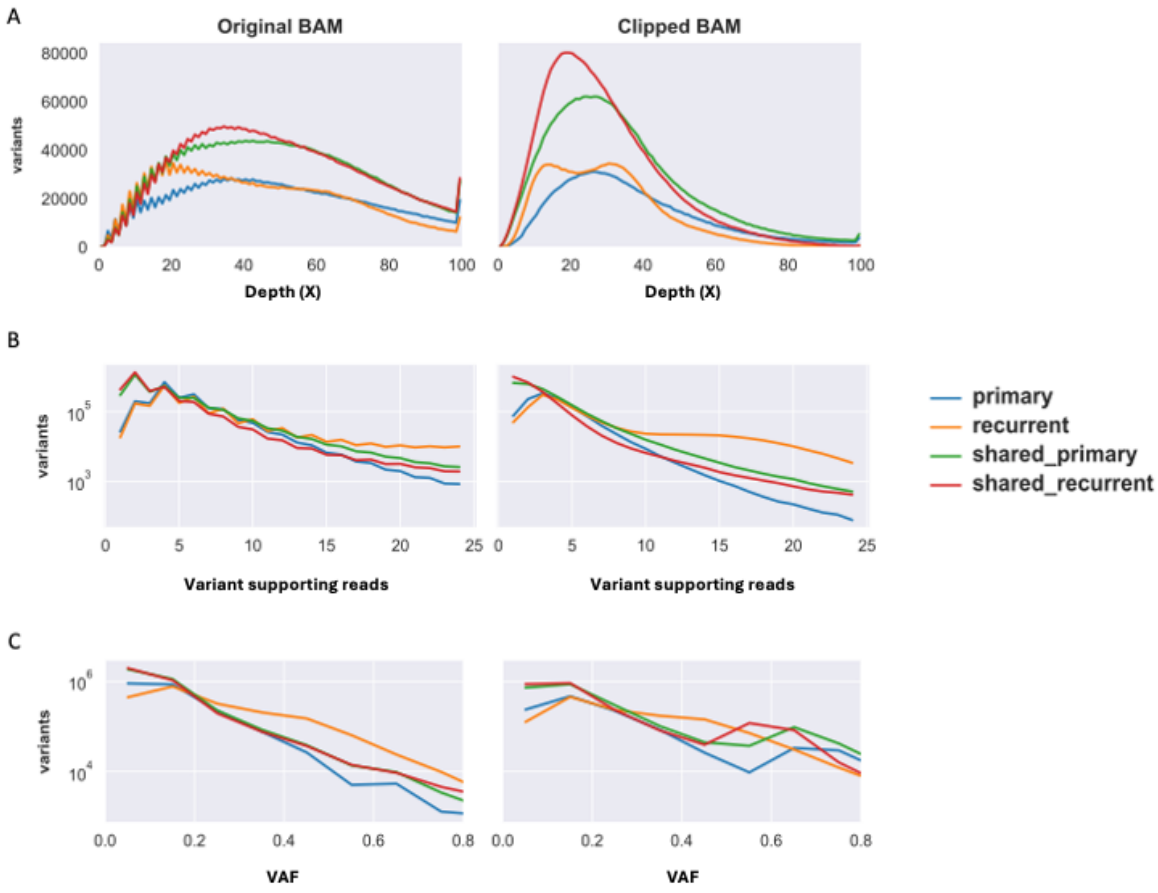


Figure 2-3: Diagnostic plots to investigate calling high number of variants in the original and clipped reads.

A: Distribution of total depth.

B: Distribution of variant allelic depth.

C: Distribution of variant allele frequency.

As a further diagnostic step, I examined the distribution of total and variant allelic depths, where a spiked pattern was observed: even read depth values occurred more frequently than odd read depth values (Figure 2-3A). This suggested that there could be double counting of evidence, possibly resulting from short DNA fragments where paired reads sequenced the variant in both forward and reverse directions. To investigate this possibility, I used fgbio-ClipBam (http://fulcrumgenomics.github.io/fgbio/tools/latest/ClipBam.html). ClipBam clips paired reads from the ends of read 1 and read 2. At the stage of having mapped reads ready for mutation calling, ClipBam was a reliable option because it can be applied directly to aligned read files, meaning that I wouldn't need to repeat the raw reads mapping process.

In addition, I compared the numbers of mutations for each substitution type in the WGS data, as FFPE samples are known to have a high proportion of C>T substitutions that are mostly false positives (Williams et al., 1999, Quach et al., 2004). C>T variants were predominant, with a higher proportion than other substitutions, followed by T>C variants (Figure 2-4), suggesting that many of the called variants are likely FFPE artefacts.

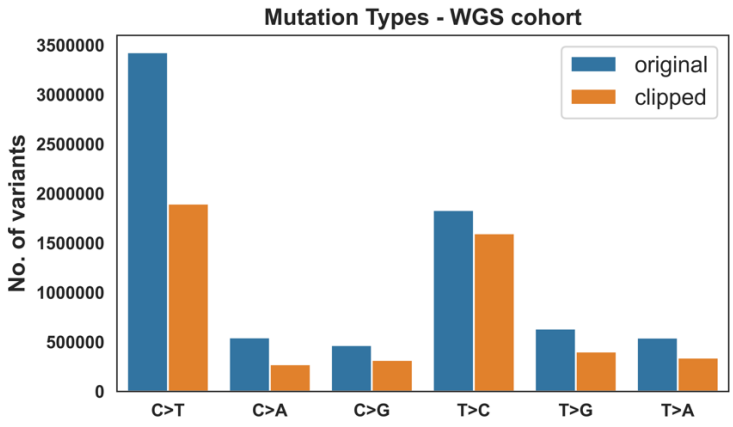


Figure 2-4:Proportions of single base substitution (SBS).

A diagnostic plot to show the prevalence of each mutation type. C>T known to be an FFPE artefact is the predominant substitution in the WGS cohort.

To see if clipping the overlapping reads improves the data and reduces the artefacts, I randomly tested a sample from the exome data as it is smaller than the genome, does not require high hardware specifications like the whole genome and is easy to optimise. I applied ClipBam on the exome bam file, and re-ran Mutect2 to call the somatic variants. I then compared the results before

and after using ClipBam. This showed that Mutect2 double counts the variants in the overlapped reads (Figure 2-5). Pearson correlation scores were significantly high between variant supporting reads before and after clipping the overlapped reads further confirming that variant supporting reads were overestimated because of double counting the evidence.

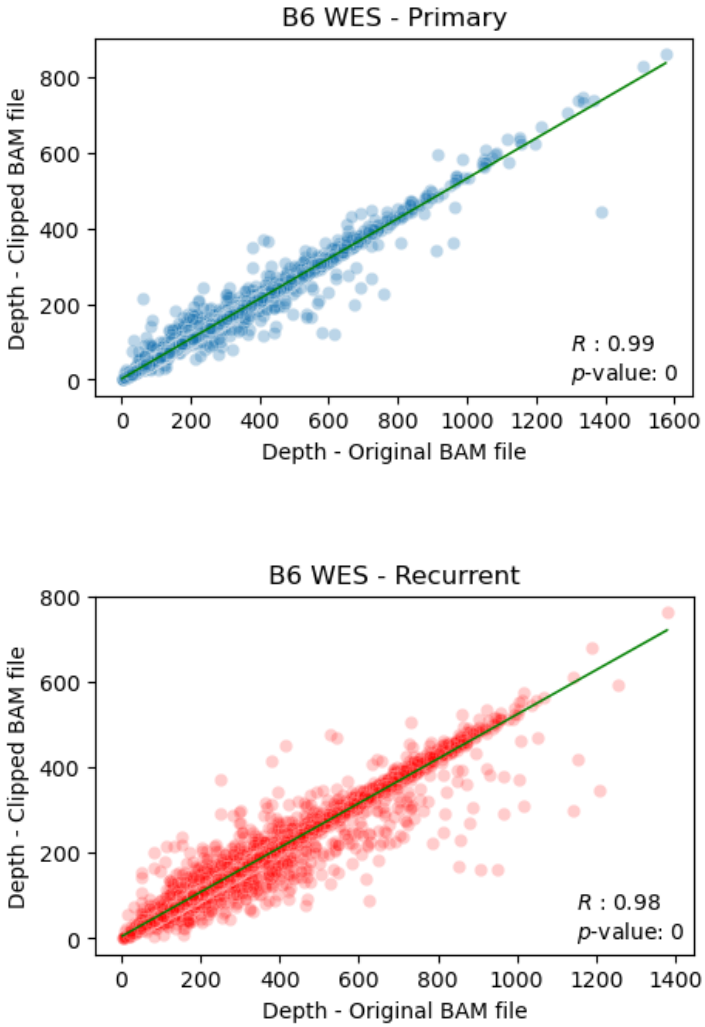


Figure 2-5: Total depth of variants in the pre and post clipping the overlapping reads.
Variants of the clipped exome sample were extracted and checked for coverage in the original bam using Bam-readcount (Khanna et al., 2022).

2.3.5 Clipping the overlapped reads

After confirming the double counting of evidence, I applied ClipBam to both WGS and WES cohorts. This step was performed after identifying the duplicate reads to maximise the accuracy of post-alignment processing. The initial bam files contained soft-clipped bases as a result of the alignment process. Soft clipping is a way to handle mismatches or low-quality bases without discarding the entire read. Nonetheless, ClipBam provides metrics afterwards to show how much of the data were clipped. This allowed me to compare the metrics before and after applying ClipBam on both WGS

and WES data. For the WGS data, the median percentage of bases clipped during alignment was 1.7% and the median percentage of clipped bases due to overlapping was 45%, whereas the WES data had 20% of bases clipped during alignment and 35% due to overlapping reads. Clipping the overlapping reads then allowed me to call variants from 53.3% of the bases for the WGS and 45% of the WES data. The extensive clipping was acceptable given that the median insert length was 126 bp using the 150 bp paired-end sequencing method. Figure 2-6 shows the percentage of usable and clipped bases for each cohort.

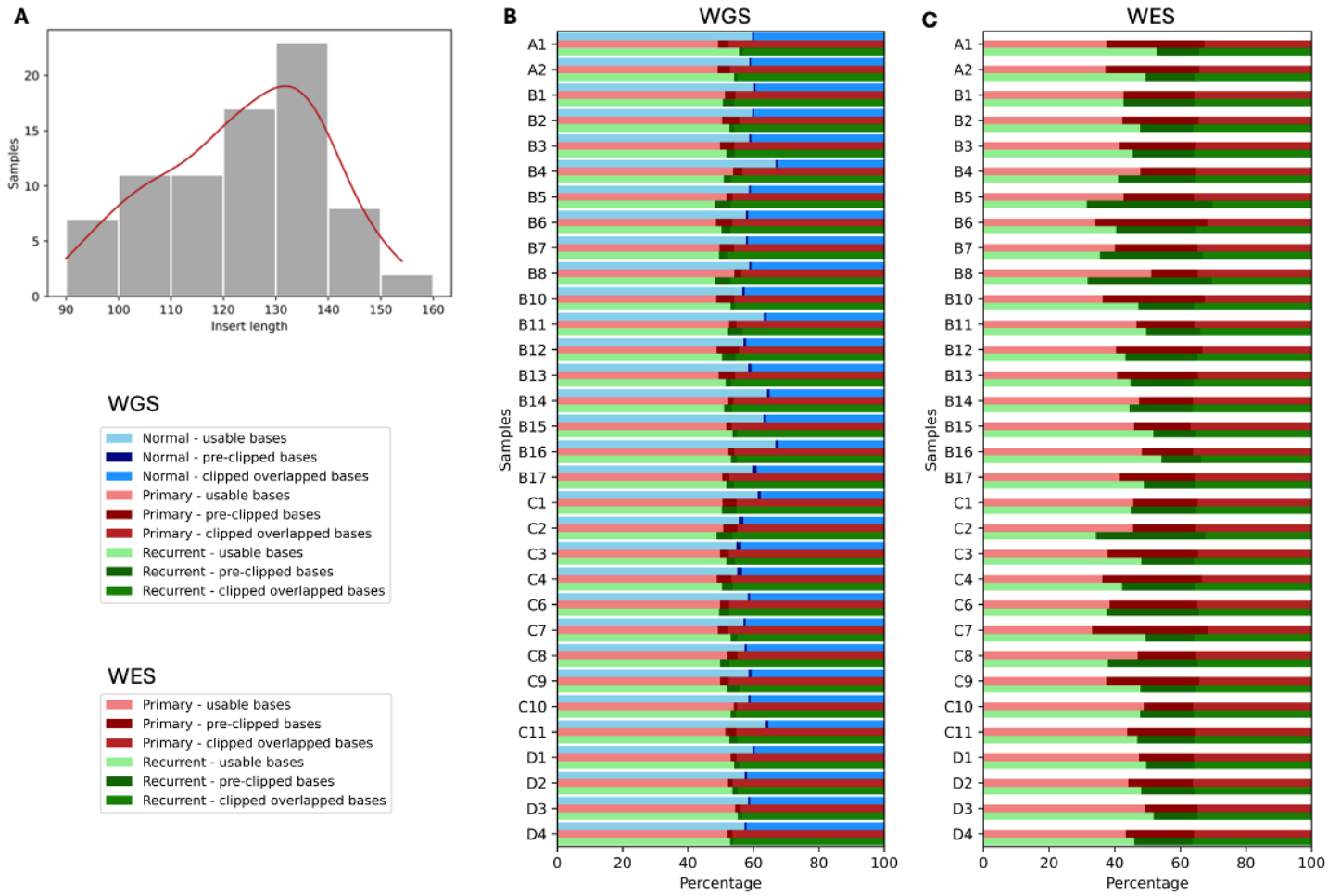


Figure 2-6: Investigation the overlapping reads per sample.

A: The distribution of the insert length of short reads.

B: The proportions of usable bases and clipped bases before and after clipping the overlapped reads in whole genome cohort. Each patient has a pair of primary-recurrent samples matched with a normal sample. Light colours for usable reads, medium colours for bases clipped due to overlapping, and the dark colours are for bases clipped during mapping to find best alignment.

C: Same as B but for the whole exome cohort. Exome data didn't include matched normal samples.

2.3.6 Whole Exome Cohort

After converting the BAM files to the raw read format and re-aligning them, the exome data had an average read depth of 99x for the tumour samples. Mutect2 detected a median of 6,509 somatic variants in the primary tumours and a median of 6,075 somatic variants per sample in the recurrent tumours.

2.3.6.1 Common mutated genes in GBM

To further ensure comparability of our data with that found in other large cohort studies, I inspected the most commonly mutated genes. GBMs are characterized by genetic alterations commonly in *EGFR*, *PTEN*, *TP53* and *RB1* genes (Korber et al., 2019, Brennan et al., 2013). The Stead cohort shows comparable findings with these studies. In comparison with a study done by Brennan et al., 2014, *EGFR* was mutated in 26% of their cohort while the Stead cohort had 30% of mutated *EGFR* cases. *PTEN* mutations were also found at a similar frequency, at 28% in both cohorts. *TP53* was mutated in 39% of Stead cohort, while Brennan et al's had 26%. Finally, *RB1* was mutated in 19% of the Stead cohort whereas Brennan et al's reported 8% of the cases with *RB1* mutations. In addition to the common genes, I observed high mutation rates in other top 20 genes. However, these genes are among the longest in the human genome, making them more prone to accumulating mutations due to their size.

The exome data analysis revealed that *TTN* was the most frequently mutated gene in the Stead cohort. This finding aligns with previous studies, such as the work by (Oh et al., 2020), which reported that *TTN* had the highest mutational load in GBM samples (Figure 2-7A). *MUC16* was also among the top 20 most frequently mutated genes, with variants identified in 25% of cases. This result is supported by a preprint study by (Ferrer, 2022), which highlighted *MUC16* as a recurrently mutated gene in GBM. Of the four common GBM mutated genes (Figure 2-8B), *EGFR* had variants with VAF less than 10% suggesting a presence of subclonal mutations while the VAF of *PTEN*, *TP53* and *RB1* variants ranged from ~20% to approximately 70% suggesting that these alterations are likely clonal and occurred earlier during tumour development.

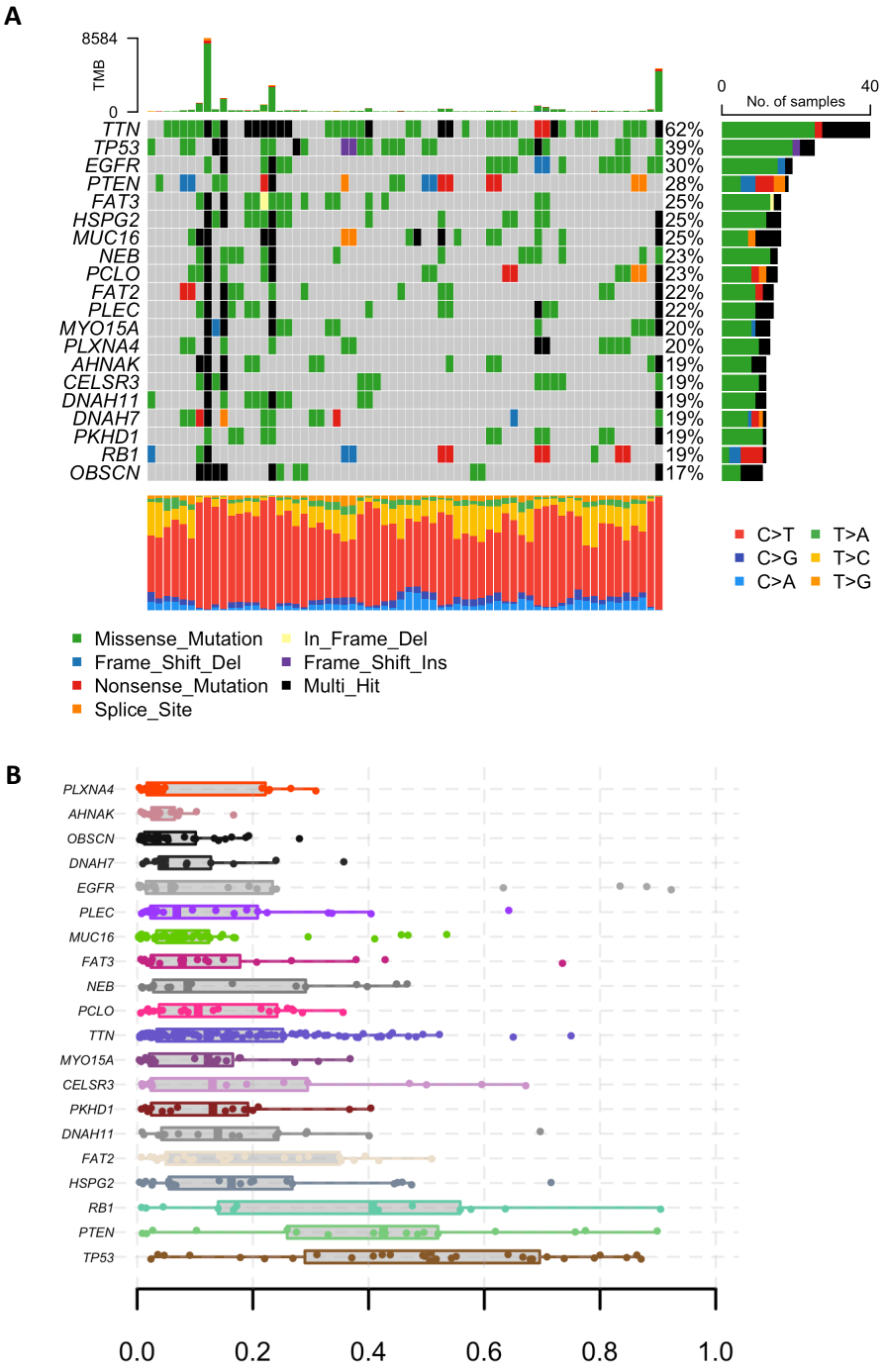


Figure 2-7: Mutational landscape and variant allele frequency (VAF) distribution in the WES cohort.

(A) Oncoplot showing the top 20 most frequently mutated genes across the cohort. Each column represents a tumour sample, and each row corresponds to a gene. Mutation types are colour-coded as indicated in the legend. The bar plot on the top indicates the total number of mutations per sample, while the bar plot on the right shows the mutation frequency (number of samples with at least one mutation) for each gene across the cohort. The stacked bar plot below depicts the proportion of single base substitution classes within each sample.

(B) Distribution of variant allele frequencies for mutations in the same top 20 genes. Each box plot represents the range of VAFs per gene across all samples, illustrating clonal versus subclonal patterns (higher vs. lower VAFs).

2.3.6.2 Tumour mutational burden

Tumour mutational burden (TMB) is another metric I used to compare my exome results to other studies. I compared the TMB of the Stead cohort (64 samples) with the Cancer Genome Atlas (TCGA), which has 398 samples, mostly from primary tumours (Figure 2-8A). The exome data shows a slightly higher median TMB rate than the TCGA, with 4.5 and 5.2 mutations per MB respectively, when considering only deleterious mutations. The primary tumours had three hypermutator samples where TMB exceeded 10 mutations/MB, while the recurrent tumours had 5 hypermutated samples as indicated in Figure 2-8 A and B. Kim et al., 2015 had similar results to ours and TMB ranges between 4-5.5 mutations/Mb. Furthermore, 134 IDHwt GBM samples analysed in the study of (Barthel et al., 2019) showed less TMB rate of 2.85 mutations/Mb.

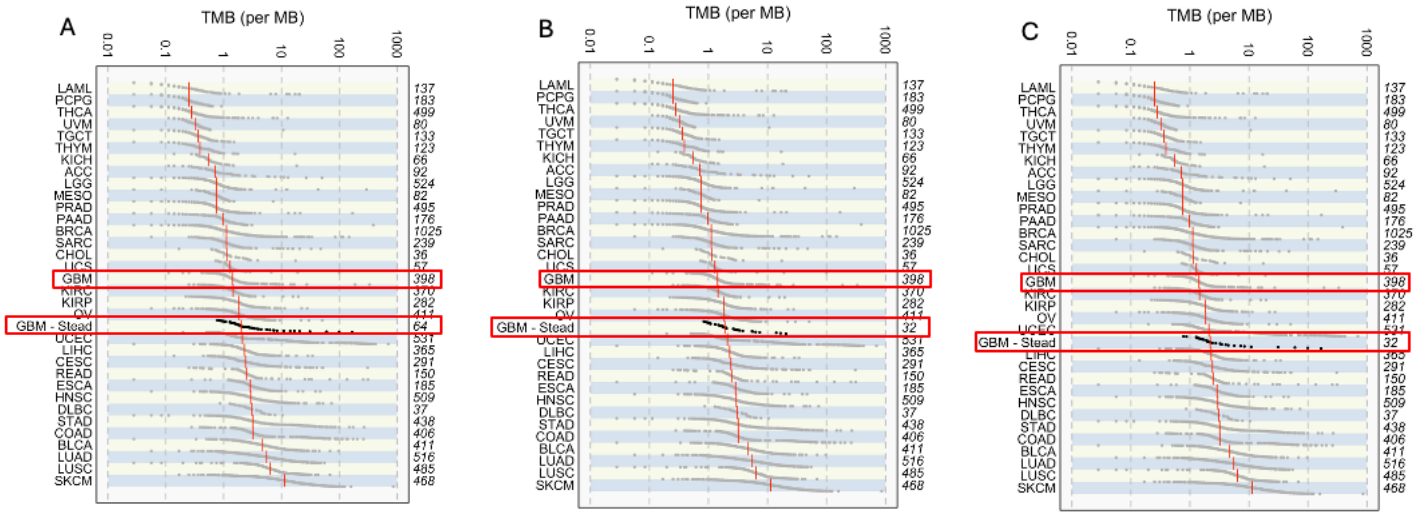


Figure 2-8: TMB rate across different cancers in TCGA database, and the Stead cohort.

A: TMB rates for primary and recurrent samples combined against TCGA datasets. Number of samples is indicated on the right Y axis.

B: TMB rates for primary samples only against TCGA datasets.

C: TMB rates for recurrent samples only against TCGA datasets.

2.3.7 Whole Genome Cohort

After applying ClipBam, I repeated the mutation calling by mutect2. The number of primary tumour variants reduced from a median of 183,187 variants per sample in the pre-clipped data to a median of 117,810 in the clipped data, while recurrent variants reduced from 155,264 to 89,257 variants per sample. The read depth remained the same even after clipping the overlapped reads with 25x for the normal samples and 20x for the tumour samples. Although clipping the overlapping mapped reads has successfully reduced the number of variants by 35-40%, the genome profiles still have 5-10 times the number of somatic variants reported by (Barthel et al., 2019, Korber et al., 2019) making the WGS

data not suitable for variant calling. The lower read depth, and the nature of DNA samples being from FFPE tissue, played roles in detecting higher rate of false positives. Also, it is not applicable to apply a hard filter for "limit of detection" in samples with low coverage and heterogenous tumours such as GBM (Steiert et al., 2023). It was therefore decided that WGS data will be used only to call somatic copy number aberrations (sCNA).

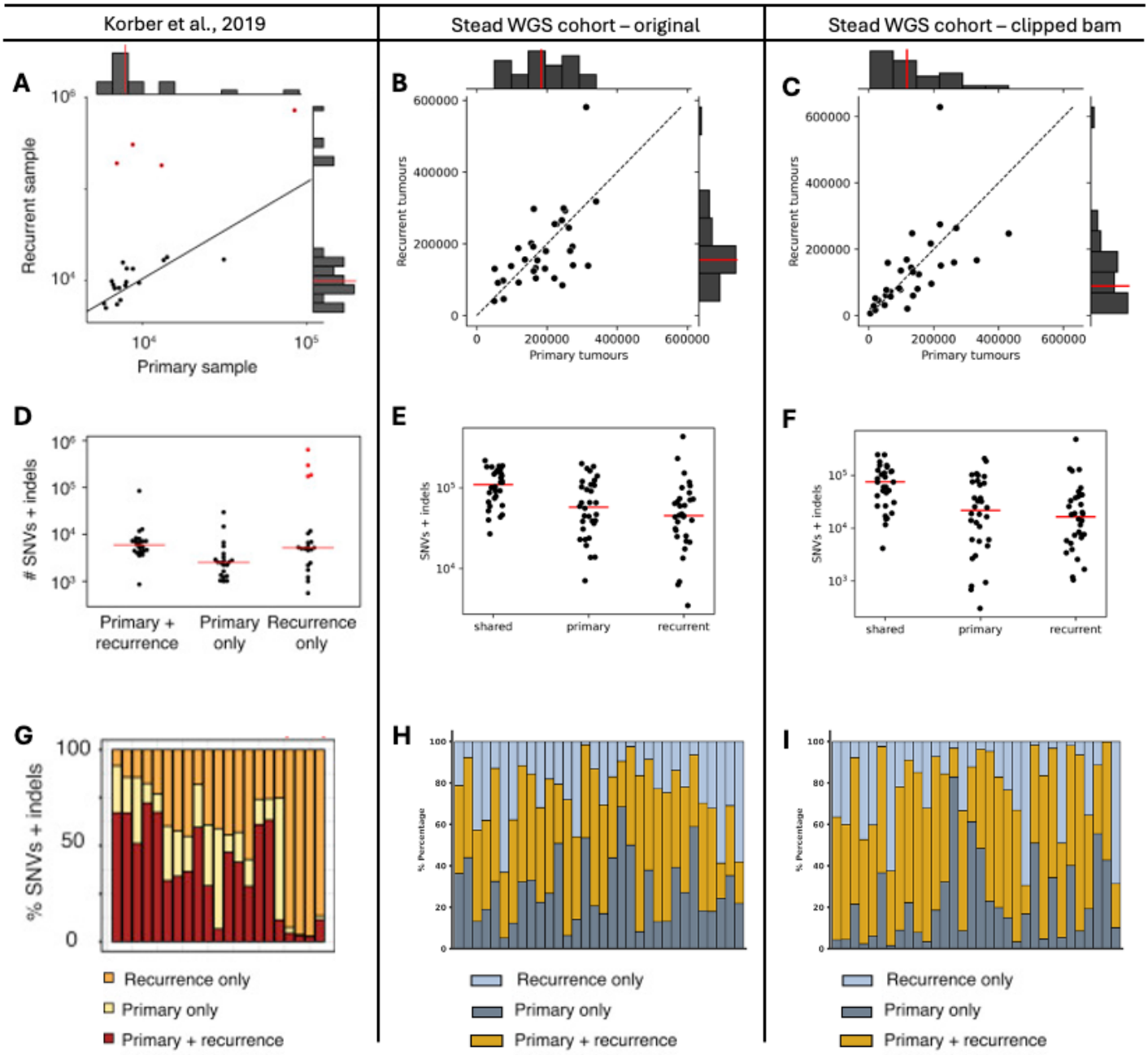


Figure 2-9: Variant metrics before and after clipping the overlapped reads.

A-C: Number of somatic SNVs and indels called in primary and recurrent tumours (red lines, median)

D-F: Numbers of shared and private SNVs and indels to each tumour (red lines, median)

G-I: Proportion of mutations per patient relative to D, E and F.

**Plots A, D and G are published in (Korber et al., 2019)

2.3.7.1 Regulatory region mutations

The Telomerase Reverse Transcriptase (*TERT*) promoter is commonly mutated in GBM patients. This promotor region was not included in the hybrid capture probes used in generating the Stead cohort WES data. Therefore, exomes lack information on *TERTp* mutation status. As an alternative, I used the genome data to check the fraction of samples that were mutated. Of 32 patients, 21 carried one of the common mutations in the *TERT* promoter region. After receiving the therapy, 19 patients retained the mutation in their recurrent tumours. C228T was predominantly mutated in 70% of the patients, while C250T was found in 20%. These mutations are commonly referred to using the nomenclature C228T and C250T, which correspond to C>T transitions located 124 bp and 146 bp upstream of the TERT transcription start site. The remaining 10% of the patients had no mutation in either sample. Similar results were reported by (Lombardi et al., 2021), showing that the *TERT* promoter region was mutated in approximately 70% of the recurrent GBM.

2.3.7.2 Chromosomal aberration findings

Calling copy number aberrations requires a well optimised pipeline, especially in the presence of short reads produced from FFPE samples. Using Facets SCNA caller, I went through multiple attempts at optimisation until I found the suitable set of parameters, such as the low and high critical values that yielded the best segmentation. These critical values set the statistical thresholds for merging adjacent segments. The quality of the segmentation determines whether further refinement is needed, as over-segmentation can introduce noise and lead to inaccurate profiles, while under-segmentation may miss important changes. Since there are no universal or standardised parameters for copy number calling, the process can vary depending on the tool used, the nature of the sequencing data, and even lab-specific protocols. In this context, parameter refinement was particularly important because BAM files were processed to clip overlapping bases in paired-end reads. Better segmentation was achieved progressively, as shown when comparing Figure 2-10A through 2-10C. The final parameters reduced noise and over-segmentation, producing more reliable copy number calls. This was crucial for downstream analyses such as estimating the cancer cell fraction (CCF), subclonal deconvolution, and pathway analysis (Tanner et al., 2021). Figure 2-10 illustrates the optimisation stages using different critical values tested to achieve the optimal segmentation for the Stead cohort.

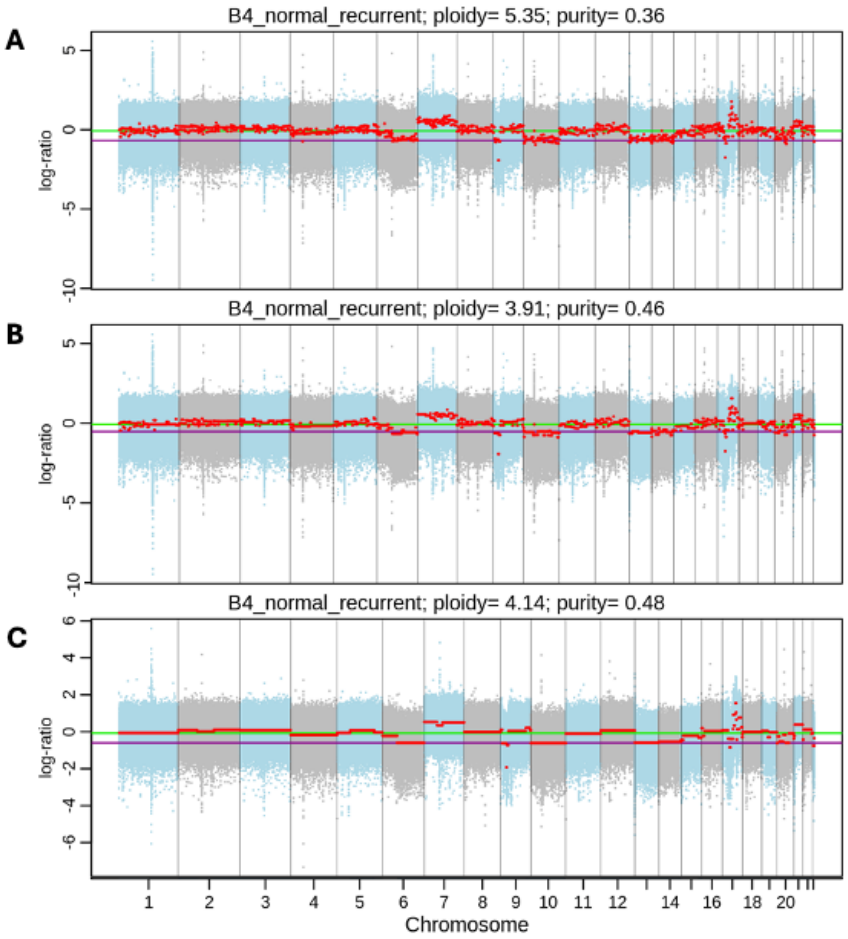


Figure 2-10: Optimising stages of copy number calling using FACETS.

A: Joint segmentation using the default critical values for preprocessing and final segmentation.

B: Joint segmentation using adjusted critical values to reduce the over segmentation.

C: Joint segmentation using critical values that were applied the WGS dataset.

Tumour purity and ploidy were inferred by FACETS, which jointly models log2 copy-number ratios and allelic imbalance (B-allele frequencies) from matched tumour–normal sequencing data. FACETS iteratively searches for the purity–ploidy combination that best fits the observed data, yielding an optimal estimate of the fraction of tumour DNA (purity) and the average DNA copy number per cell (ploidy).

The segmentation data was then further analysed using GISTIC to identify copy number changes across the cohort. The analysis conducted by GISTIC revealed significant losses and gains in profiles of both primary and recurrent tumours, including chr 7 gain, chr 9 loss and chr 10 loss (Figure 2-11A,C). This finding is comparable with the results reported by (Korber et al., 2019). Among the genes that were amplified in the chromosome 7 gain event is *EGFR*. This gene was mutated in 84% of the primary tumours and 59% of the recurrent tumours. Chromosome 10, which includes *PTEN*, had partial deletions of more than 80% of the primary tumours and 70% of the recurrences. Additionally, chromosome 9p acquired deletions in more than 50% of both primary and recurrent tumours. These

findings are also aligned with results reported previously (Brennan et al., 2013) where they analysed copy number profiles from 543 GBM samples, including amplifications in *EGFR* and *PDGFRA*, and deletions in *PTEN*, *CDKN2A* and *RB1* (Figure 2-11B). Finally, GISTIC revealed gains and losses in the recurrent tumours that were not initially detected in the primary tumours (Figure 2-11C); novel partial deletions were found in 10q and 13q arms and novel partial amplifications in 3q and 17q arms.

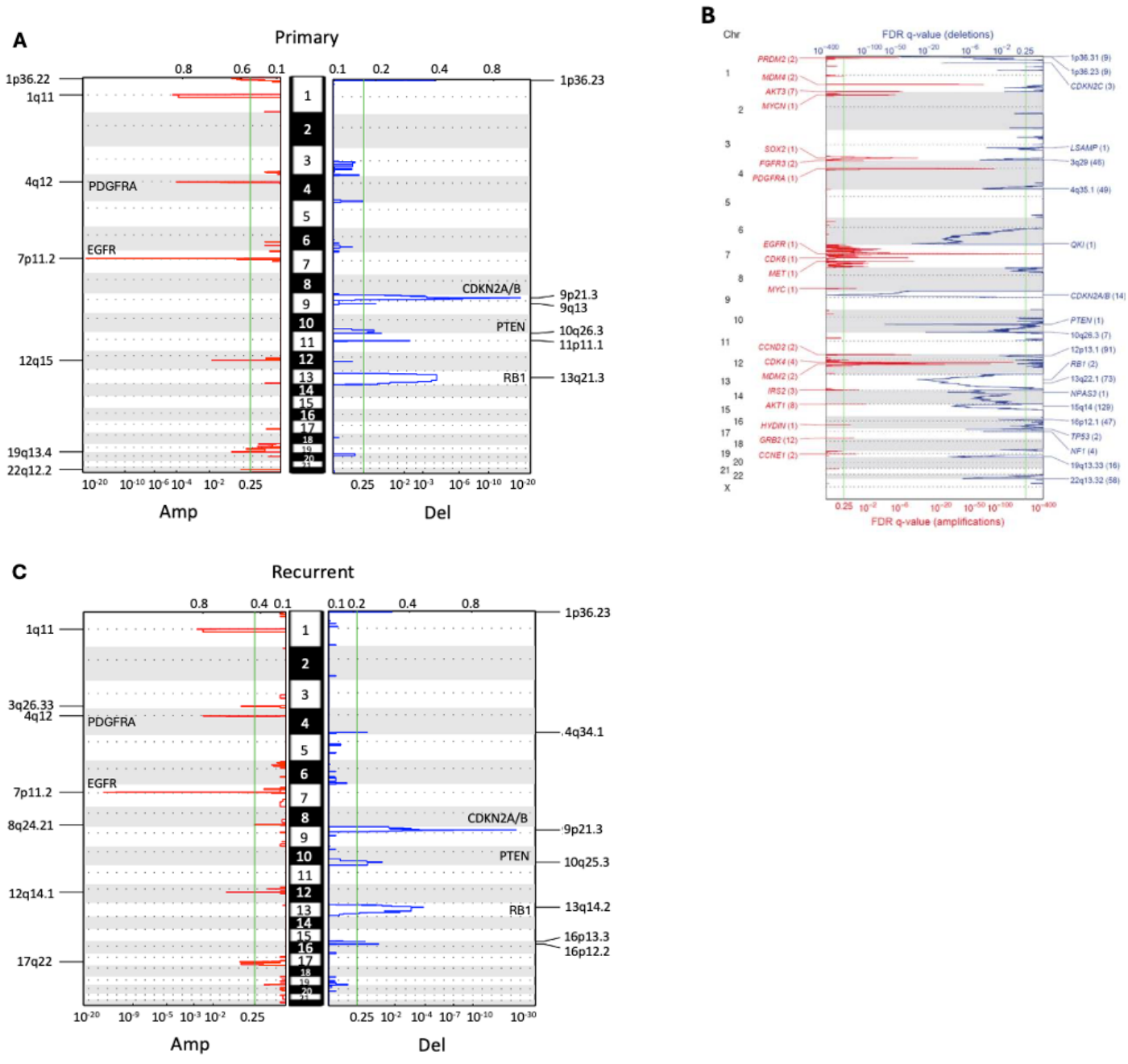


Figure 2-11: Somatic CNAs in Glioblastoma identified by GISTIC.

Genome-wide copy number alterations were identified using the GISTIC 2.0 algorithm, which detects regions of recurrent amplification (red peaks) and deletion (blue peaks) across the cohort. The x-axis represents the statistical confidence of each alteration, expressed as the false discovery rate (q-value). The vertical green line marks the significance threshold, with peaks

extending beyond this line indicating significantly recurrent events after multiple-testing correction. The height of each peak (G-score) reflects the combined amplitude and frequency of the aberration across the samples.

A: Primary tumours from the Stead cohort

B: TCGA primary glioblastoma dataset adapted from (Brennan et al., 2013) for comparison.

C: Recurrent tumours from the Stead cohort.

The recurrently altered genes highlighted in the TCGA reference plot (panel B) were also detected in the Stead cohort (panels A and C), demonstrating concordant amplification and deletion patterns across datasets.

2.4 DISCUSSION

Optimizing the whole-exome sequencing (WES) and whole-genome sequencing (WGS) analysis pipeline was necessary to address several technical issues to generate reliable results for downstream analyses. At first, the variant calling showed a very high number of somatic variants in both datasets. FFPE-induced alterations such an overabundance of C>T changes attributed to this extremely high number of variants. Overlapping paired-end readings that resulted in double counting of the evidence affected the VAF estimates. I implemented two processes in the pipeline: first, I applied the *LearnReadOrientationModel* function in Mutect2 to reduce the excess C>T counts. Second, I used *ClipBam* to clip the overlapping paired-end reads. After these two steps, I achieved a substantial improvement, with variant counts in WES data aligning closely with other GBM studies, indicating that the refined pipeline effectively minimised false positives while retaining true variants.

The WES data analysis confirmed that the optimised pipeline can produce reliable results. The variant counts were consistent with published studies, and I successfully detected mutations in key GBM genes, including *EGFR, PTEN, TP53*, and *RB1*. The estimates of the TMB are in line with other GBM cohorts, such as the TCGA. An efficient pipeline is particularly important for accurately computing TMB to avoid overestimations or underestimations. Results around the threshold of 10 mutations per megabase are prone to overestimation or underestimation of the number of true, problematic variants in the sample as FFPE-induced mutations could have skewed the results. After the pipeline adjustments, the hypermutated samples identified in my cohort were validated as true cases, not artifacts, which further proved the robustness of the pipeline.

The enhancements to the WGS data pipeline substantially decreased the variant counts, resulting in a reduction of 35–40% compared to before to the changes. Nonetheless, this still produced a greater number of variations than reported in other studies. The probable cause for this is our use of FFPE materials, which present significant challenges, coupled with a low sequencing depth of 20x for the tumour samples. The aforementioned factors rendered the WGS data inappropriate for identifying somatic SNVs. Conversely, it was appropriate for identifying copy number abnormalities. By executing

the optimised workflow I designed for the WGS data, I successfully identified prevalent CNAs linked to GBM, including gains on chromosome 7 and losses on chromosomes 10 and 9.

A significant enhancement I implemented was resolving the issue of overlapping paired-end reads, a prevalent contributor to the inflation of variation allelic ratios. I utilised ClipBam to mask the overlapping regions to prevent the variant caller from counting paired reads supporting variants twice. I conducted diagnostic assessments utilising VAF, read depth, and the count of supporting reads, and I verified that this modification significantly reduced the artefacts I had been observing. This was essential for VAF estimates, which are vital for monitoring variant prevalence over time in longitudinal datasets.

2.5 Future Directions

Despite resolving the identified issues, there is still opportunity for future improvements and optimizations. Higher depth sequencing would help in lowering the occurrence of low coverage artifacts in whole genome data. The troublesome overlaps will also be significantly reduced by using a read-length generation kits that is 75 bp paired-end, especially for FFPE-derived samples where short insert sizes are yielded by induced DNA fragmentation. Nearly 40% of the data had to be clipped because of overlaps when using 150 bp paired-end reads, as was the case with WGS and WES in this cohort. This decreased the amount of coverage that could be used. Some of these problems will be minimised with improved library preparation kits and large DNA input. These changes may result in increased sequencing depth and improved genome coverage.

Developing a panel of normals (PON) from non-malignant brain tissues is an additional step that might enhance the outcomes of subsequent research. The likelihood of false positives can be decreased by using a PON to assist filtering out recurrent technical artifacts unique to brain tissue sequencing. For instance, comparing tumour samples to a PON made from non-malignant brain tissue could help manage FFPE artifacts or sequencing-specific errors. A more precise baseline for determining true somatic variations would result from this. The pipeline may become more robust and dependable for research on brain cancer if this phase is added.

In the future, workflows for other data sources, including RNA sequencing (RNA-seq) or methylation analysis, could be developed utilizing the modular pipeline design methodology, like Nextflow. Even though this pipeline is designed for DNA sequencing data, the same automation, scalability, and adaptability may be used to create customized pipelines for analysing epigenetic alterations such as Bisulfite sequencing (BS-seq) or gene expression such as RNA-seq data.

2.6 Conclusion

This chapter concludes by showing how crucial thorough pipeline refining is to obtain accurate variant and copy number calling from next-generation sequencing data. By effectively addressing the main issues of overlapping reads, FFPE artefacts, and low sequencing depth, the approaches used here laid the groundwork for reliable downstream studies. The optimizations carried out guarantee that the data produced are of excellent quality and appropriate for further investigations, including the investigation of pathways and the long-term monitoring of genetic alterations.

I am confident that the optimised pipeline produced reliable results that can be built upon in the next chapter, which focuses on pathway analysis after resolving these issues. Critical investigations into treatment-driven tumour progression and clonal dynamics in GBM will be supported by accurate copy number aberration (CNA) profiles and variable allele frequency (VAF) estimates. Other studies using cancer sequencing data, especially those that depend on FFPE samples or low-coverage WGS datasets, will benefit significantly from the experience of this optimization approach. A fundamental step in producing accurate, repeatable, and therapeutically beneficial genetic data is optimizing the WGS and WES pipelines, which supports the larger objectives of advancing precision oncology and cancer research.

2.7 REFERENCES

BARTHEL, F. P., JOHNSON, K. C., VARN, F. S., MOSKALIK, A. D., TANNER, G., KOCAKAVUK, E., ANDERSON, K. J., ABIOLA, O., ALDAPE, K., ALFARO, K. D., ALPAR, D., AMIN, S. B., ASHLEY, D. M., BANDOPADHAYAY, P., BARNHOLTZ-SLOAN, J. S., BEROUKHIM, R., BOCK, C., BRASTIANOS, P. K., BRAT, D. J., BRODBELT, A. R., BRUNS, A. F., BULSARA, K. R., CHAKRABARTY, A., CHAKRAVARTI, A., CHUANG, J. H., CLAUS, E. B., COCHRAN, E. J., CONNELLY, J., COSTELLO, J. F., FINOCCHIARO, G., FLETCHER, M. N., FRENCH, P. J., GAN, H. K., GILBERT, M. R., GOULD, P. V., GRIMMER, M. R., IAVARONE, A., ISMAIL, A., JENKINSON, M. D., KHASRAW, M., KIM, H., KOUWENHOVEN, M. C. M., LAVIOLETTE, P. S., LI, M., LICHTER, P., LIGON, K. L., LOWMAN, A. K., MALTA, T. M., MAZOR, T., MCDONALD, K. L., MOLINARO, A. M., NAM, D. H., NAYYAR, N., NG, H. K., NGAN, C. Y., NICLOU, S. P., NIERS, J. M., NOUSHMEHR, H., NOORBAKHSH, J., ORMOND, D. R., PARK, C. K., POISSON, L. M., RABADAN, R., RADLWIMMER, B., RAO, G., REIFENBERGER, G., SA, J. K., SCHUSTER, M., SHAW, B. L., SHORT, S. C., SMITT, P. A. S., SLOAN, A. E., SMITS, M., SUZUKI, H., TABATABAI, G., VAN MEIR, E. G., WATTS, C., WELLER, M., WESSELING, P., WESTERMAN, B. A., WIDHALM, G., WOEHRER, A., YUNG, W. K. A., ZADEH, G., HUSE, J. T., DE GROOT, J. F., STEAD, L. F., VERHAAK, R. G. W. & CONSORTIUM, G. 2019. Longitudinal molecular trajectories of diffuse glioma in adults. *Nature,* 576**,** 112-120.

BENJAMIN, D., SATO, T., CIBULSKIS, K., GETZ, G., STEWART, C. & LICHTENSTEIN, L. 2019.

BRENNAN, C. W., VERHAAK, R. G., MCKENNA, A., CAMPOS, B., NOUSHMEHR, H., SALAMA, S. R., ZHENG, S., CHAKRAVARTY, D., SANBORN, J. Z., BERMAN, S. H., BEROUKHIM, R., BERNARD, B., WU, C. J., GENOVESE, G., SHMULEVICH, I., BARNHOLTZ-SLOAN, J., ZOU, L., VEGESNA, R., SHUKLA, S. A., CIRIELLO, G., YUNG, W. K., ZHANG, W., SOUGNEZ, C., MIKKELSEN, T., ALDAPE, K., BIGNER, D. D., VAN MEIR, E. G., PRADOS, M., SLOAN, A., BLACK, K. L., ESCHBACHER, J., FINOCCHIARO, G., FRIEDMAN, W., ANDREWS, D. W., GUHA, A., IACOCCA, M., O'NEILL, B. P., FOLTZ, G., MYERS, J., WEISENBERGER, D. J., PENNY, R., KUCHERLAPATI, R., PEROU, C. M., HAYES, D. N., GIBBS, R., MARRA, M., MILLS, G. B., LANDER, E., SPELLMAN, P., WILSON, R., SANDER, C., WEINSTEIN, J., MEYERSON, M., GABRIEL, S., LAIRD, P. W., HAUSSLER, D., GETZ, G., CHIN, L. & NETWORK, T. R. 2013. The somatic genomic landscape of glioblastoma. *Cell,* 155**,** 462-77.

BUSHNELL, B., ROOD, J. & SINGER, E. 2017. BBMerge - Accurate paired shotgun read merging via overlap. *PLoS One,* 12**,** e0185056.

COSTELLO, M., PUGH, T. J., FENNELL, T. J., STEWART, C., LICHTENSTEIN, L., MELDRIM, J. C., FOSTEL, J. L., FRIEDRICH, D. C., PERRIN, D., DIONNE, D., KIM, S., GABRIEL, S. B., LANDER, E. S., FISHER, S. & GETZ, G. 2013. Discovery and characterization of artifactual mutations in deep coverage targeted capture sequencing data due to oxidative DNA damage during sample preparation. *Nucleic Acids Res,* 41**,** e67.

DI TOMMASO, P., CHATZOU, M., FLODEN, E. W., BARJA, P. P., PALUMBO, E. & NOTREDAME, C. 2017. Nextflow enables reproducible computational workflows. *Nat Biotechnol,* 35**,** 316-319.

EINAGA, N., YOSHIDA, A., NODA, H., SUEMITSU, M., NAKAYAMA, Y., SAKURADA, A., KAWAJI, Y., YAMAGUCHI, H., SASAKI, Y., TOKINO, T. & ESUMI, M. 2017. Assessment of the quality of DNA from various formalin-fixed paraffin-embedded (FFPE) tissues and the use of this DNA for next-generation sequencing (NGS) with no artifactual mutation. *PLoS One,* 12**,** e0176280.

FERRER, V. P. 2022. MUC16 mutation is associated with tumor grade, clinical features, and prognosis in glioma patients.

HUANG, R. X. & ZHOU, P. K. 2020. DNA damage response signaling pathways and targets for radiotherapy sensitization in cancer. *Signal Transduct Target Ther,* 5**,** 60.

JOHNSON, D. R. & O'NEILL, B. P. 2012. Glioblastoma survival in the United States before and during the temozolomide era. *J Neurooncol,* 107**,** 359-64.

KHANNA, A., LARSON, D., SRIVATSAN, S., MOSIOR, M., ABBOTT, T., KIWALA, S., LEY, T., DUNCAVAGE, E., WALTER, M., WALKER, J., GRIFFITH, O., GRIFFITH, M. & MILLER, C. 2022. Bam-readcount - rapid generation of basepair-resolution sequence metrics. *Journal of Open Source Software,* 7.

KIM, S., PARK, C., JI, Y., KIM, D. G., BAE, H., VAN VRANCKEN, M., KIM, D. H. & KIM, K. M. 2017. Deamination Effects in Formalin-Fixed, Paraffin-Embedded Tissue Samples in the Era of Precision Medicine. *J Mol Diagn,* 19**,** 137-146.

KORBER, V., YANG, J., BARAH, P., WU, Y., STICHEL, D., GU, Z., FLETCHER, M. N. C., JONES, D., HENTSCHEL, B., LAMSZUS, K., TONN, J. C., SCHACKERT, G., SABEL, M., FELSBERG, J., ZACHER, A., KAULICH, K., HUBSCHMANN, D., HEROLD-MENDE, C., VON DEIMLING, A., WELLER, M., RADLWIMMER, B., SCHLESNER, M., REIFENBERGER, G., HOFER, T. & LICHTER, P. 2019. Evolutionary Trajectories of IDH(WT) Glioblastomas Reveal a Common Path of Early Tumorigenesis Instigated Years ahead of Initial Diagnosis. *Cancer Cell,* 35**,** 692-704 e12.

LI, H. & DURBIN, R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics,* 25**,** 1754-60.

LOMBARDI, G., GIUNCO, S., CAVALLIN, F., ANGELINI, C., CACCESE, M., CERRETTI, G., BONIS, P. D., ROSSI, A. D. & ZAGONEL, V. 2021. The clinical significance of telomerase reverse transcriptase (TERT) promoter mutations, telomere length and O6-methylguanine DNA methyltransferase (MGMT) promoter methylation status in newly diagnosed and recurrent IDH-wildtype glioblastoma (GBM) patients (PTS): A large mono-institutional study. *Journal of Clinical Oncology,* 39**,** 2053-2053.

MARTIN, M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal,* 17.

MAYAKONDA, A., LIN, D. C., ASSENOV, Y., PLASS, C. & KOEFFLER, H. P. 2018. Maftools: efficient and comprehensive analysis of somatic variants in cancer. *Genome Res,* 28**,** 1747-1756.

MCLAREN, W., GIL, L., HUNT, S. E., RIAT, H. S., RITCHIE, G. R., THORMANN, A., FLICEK, P. & CUNNINGHAM, F. 2016. The Ensembl Variant Effect Predictor. *Genome Biol,* 17**,** 122.

MERMEL, C. H., SCHUMACHER, S. E., HILL, B., MEYERSON, M. L., BEROUKHIM, R. & GETZ, G. 2011. GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. *Genome Biol,* 12**,** R41.

OH, J. H., JANG, S. J., KIM, J., SOHN, I., LEE, J. Y., CHO, E. J., CHUN, S. M. & SUNG, C. O. 2020. Spontaneous mutations in the single TTN gene represent high tumor mutation burden. *NPJ Genom Med,* 5**,** 33.

POPE, B. J., NGUYEN-DUMONT, T., HAMMET, F. & PARK, D. J. 2014. ROVER variant caller: read-pair overlap considerate variant-calling software applied to PCR-based massively parallel sequencing datasets. *Source Code Biol Med,* 9**,** 3.

QUACH, N., GOODMAN, M. F. & SHIBATA, D. 2004. In vitro mutation artifacts after formalin fixation and error prone translesion synthesis during PCR. *BMC Clin Pathol,* 4**,** 1.

RIGANTI, C., SALAROGLIO, I. C., PINZON-DAZA, M. L., CALDERA, V., CAMPIA, I., KOPECKA, J., MELLAI, M., ANNOVAZZI, L., COURAUD, P. O., BOSIA, A., GHIGO, D. & SCHIFFER, D. 2014. Temozolomide down-regulates P-glycoprotein in human blood-brain barrier cells by disrupting Wnt3 signaling. *Cell Mol Life Sci,* 71**,** 499-516.

ROBBE, P., POPITSCH, N., KNIGHT, S. J. L., ANTONIOU, P., BECQ, J., HE, M., KANAPIN, A., SAMSONOVA, A., VAVOULIS, D. V., ROSS, M. T., KINGSBURY, Z., CABES, M., RAMOS, S. D. C., PAGE, S., DREAU, H., RIDOUT, K., JONES, L. J., TUFF-LACEY, A., HENDERSON, S., MASON, J., BUFFA, F. M., VERRILL, C., MALDONADO-PEREZ, D., ROXANIS, I., COLLANTES, E., BROWNING, L., DHAR, S., DAMATO, S., DAVIES, S., CAULFIELD, M., BENTLEY, D. R., TAYLOR, J. C., TURNBULL, C., SCHUH, A. & PROJECT, G. 2018. Clinical whole-genome sequencing from routine formalin-fixed, paraffin-embedded specimens: pilot study for the 100,000 Genomes Project. *Genet Med,* 20**,** 1196-1205.

ROTUNNO, M., BARAJAS, R., CLYNE, M., HOOVER, E., SIMONDS, N. I., LAM, T. K., MECHANIC, L. E., GOLDSTEIN, A. M. & GILLANDERS, E. M. 2020. A Systematic Literature Review of Whole Exome and Genome Sequencing Population Studies of Genetic Susceptibility to Cancer. *Cancer Epidemiol Biomarkers Prev,* 29**,** 1519-1534.

SAH, S., CHEN, L., HOUGHTON, J., KEMPPAINEN, J., MARKO, A. C., ZEIGLER, R. & LATHAM, G. J. 2013. Functional DNA quantification guides accurate next-generation sequencing mutation detection in formalin-fixed, paraffin-embedded tumor biopsies. *Genome Med,* 5**,** 77.

SHEN, R. & SESHAN, V. E. 2016. FACETS: allele-specific copy number and clonal heterogeneity analysis tool for high-throughput DNA sequencing. *Nucleic Acids Res,* 44**,** e131.

STEIERT, T. A., PARRA, G., GUT, M., ARNOLD, N., TROTTA, J. R., TONDA, R., MOUSSY, A., GERBER, Z., ABUJA, P. M., ZATLOUKAL, K., ROCKEN, C., FOLSERAAS, T., GRIMSRUD, M. M., VOGEL, A., GOEPPERT, B., ROESSLER, S., HINZ, S., SCHAFMAYER, C., ROSENSTIEL, P., DELEUZE, J. F., GUT, I. G., FRANKE, A. & FORSTER, M. 2023. A critical spotlight on the paradigms of FFPE-DNA sequencing. *Nucleic Acids Res,* 51**,** 7143-7162.

STUPP, R., MASON, W. P., VAN DEN BENT, M. J., WELLER, M., FISHER, B., TAPHOORN, M. J., BELANGER, K., BRANDES, A. A., MAROSI, C., BOGDAHN, U., CURSCHMANN, J., JANZER, R. C., LUDWIN, S. K., GORLIA, T., ALLGEIER, A., LACOMBE, D., CAIRNCROSS, J. G., EISENHAUER, E., MIRIMANOFF, R. O., EUROPEAN ORGANISATION FOR, R., TREATMENT OF CANCER BRAIN, T., RADIOTHERAPY, G. & NATIONAL CANCER INSTITUTE OF CANADA CLINICAL TRIALS, G. 2005. Radiotherapy plus concomitant and adjuvant temozolomide for glioblastoma. *N Engl J Med,* 352**,** 987-96.

TANNER, G., WESTHEAD, D. R., DROOP, A. & STEAD, L. F. 2021. Benchmarking pipelines for subclonal deconvolution of bulk tumour sequencing data. *Nat Commun,* 12**,** 6396.

TEER, J. K. & MULLIKIN, J. C. 2010. Exome sequencing: the sweet spot before whole genomes. *Hum Mol Genet,* 19**,** R145-51.

WILLIAMS, C., PONTÉN, F., MOBERG, C., SÖDERKVIST, P., UHLÉN, M., PONTÉN, J., SITBON, G. & LUNDEBERG, J. 1999. A High Frequency of Sequence Alterations Is Due to Formalin Fixation of Archival Specimens. *The American Journal of Pathology,* 155**,** 1467-1471.

CHAPTER 3

3.1 INTRODUCTION

3.1.1 Overview

This chapter will focus on identifying biological mechanisms associated with treatment resistance in glioblastoma (GBM). In the previous chapter, I optimised my pipeline to call true somatic variants and accurately annotate their variant allele frequencies (VAFs) and coverage as further investigation will heavily rely on these two factors. The resulting mutation data, derived from whole exome sequencing, is comparable to what was reported in other GBM studies, instilling the confidence to proceed to apply different analytical approaches to gain further insights into the cellular processes involved in GBM that may influence resistance to therapy.

In this chapter, I analyse mutational data from two independent GBM cohorts. The first is the Discovery cohort; this includes all IDH wild-type (IDHwt) GBM cases that were collected and processed by Stead's group. For the second cohort, herein referred to as the Validation cohort, I was given access to data from the Glioma Longitudinal AnalySiS (GLASS) consortium as our group contributed GBM data to the consortium to allow high throughput analysis (Consortium, 2018). The Validation cohort mutation data is also from primary and recurrent tumours however I only had access to the variant call files, not the raw sequencing data, so not only is this an independent cohort, but it was processed using an independent analysis pipeline. My approach was to analyse the cohorts separately and look for findings that were shared under the rationale that such validated results are more likely to inform on the biology that underlies the progression of tumours through therapy, than result from cohort-specific artefacts.

3.1.2 GLASS data

The discovery cohort is thoroughly explained in Chapter 2 (section 2.3.1), hence herein I will explain the Validation cohort data. GLASS stands out as a collective effort that brings together crucial information from across the globe on all glioma subtypes. This initiative has gathered detailed genetic data from a large number of adults who've battled this disease, focusing on profiling of longitudinal paired samples to understand tumour progression. Within this collection is genetic data for a group of 94 patients diagnosed with IDHwt GBM who have undergone the usual treatment route, a combination of the drug temozolomide and radiotherapy. By looking at their cases, I can learn a lot about how the common treatment affects this aggressive tumour.

The GLASS repository doesn't provide the raw data from genome or exome sequencing. Instead, it provides high quality mutation calls which were called by Mutect2 (Benjamin et al., 2019). Eleven centres contributed 188 samples to the repository as indicated in the table 1. The exome data from the GLASS (validation cohort) and the discovery cohort were similarly processed by using Mutect2 in multi-sample mode to jointly call somatic point mutations including single nucleotide variants (SNVs) and short insertions and deletions (InDels). This is extremely useful for expanding my functional enrichment analysis as it makes the analysis more valuable by acquiring external data from patients who have been treated similarly allowing me to confidently study the effects of treatment on progression of response and highlight potential pathways that could be interesting for developing some targeted GBM therapies.

Table 3-1: Data sources for the validation cohort

Centre code	Centre Name	No. of patients
GLSS-19	Case Western	9
GLSS-AT	Medical University of Vienna – CeMM	7
GLSS-CU	Columbia University (USA)	16
GLSS-DF	Dana Farber Cancer Institute	3
GLSS-HF	Henry Ford Hospital	18
GLSS-LU	University of Leeds (UK)	8
GLSS-MD	MD Anderson Cancer Center	3
GLSS-MG	Massachusetts General Hospital	9
GLSS-SF	UC San Francisco	2
GLSS-SM	Samsung Medical Center	18
GLSS-14	Emory University	1

3.1.3 Variant distributions and treatment resistance

Using longitudinal i.e. primary and matched recurrent GBM tumours from the same patient is necessary to uncover how the molecular features of each tumour change over time. Genetic changes occur during the development of the tumour, and mutations keep accumulating as the tumour evolves. The expansion or eradication of subclones, containing specific mutations, under the selective pressure of treatment may imply that those mutations confer an advantage or disadvantage to the tumour cells, respectively (Figure 3-1).

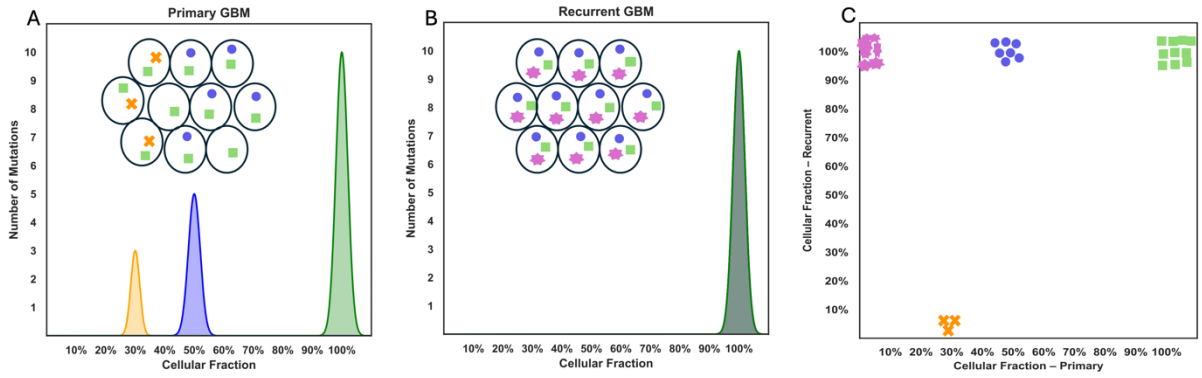


Figure 3-1: Changes in mutation prevalence between primary and recurrent GBM.

A: GBM cells in the primary tumour carry distinct mutations, represented by coloured symbols within each cell. The plot shows the cellular fraction on the x-axis (i.e., the percentage of tumour cells carrying each mutation) and the number of mutations in each group on the y-axis.

B: Same as **A** but it displays recurrent tumour.

C: This panel shows mutation frequency in the primary tumour (x-axis) compared to the recurrent tumour (y-axis). It allows identification of clonal mutations (high frequency in both tumours) versus subclonal mutations that were either lost or expanded between tumour stages.

In this chapter, I aim to investigate the change in mutational prevalence between primary and recurrent GBM, with a particular emphasis on distinguishing unique and common variants across tumour types. These variants may serve as crucial catalysts for tumorigenesis and therapeutic failure, and comprehending their functional enrichment may elucidate candidate mechanisms of treatment resistance, which could aid in developing novel therapeutic interventions and improve patient outcomes. Through genetic profiling of longitudinal pairs, variants can be categorized into three groups: primary-specific, recurrent-specific, and shared variants, with each group having a different potential biological interpretation.

3.1.4 Primary-specific

Variants unique to the primary tumour may arise from two biological scenarios. Firstly, the primary tumour sample may have included subclones that were wholly removed via surgery and thus were not present in the cells that gave rise to the recurrent tumour. Secondly, these variants could have been in cell subpopulation(s) that were lost during tumour evolution post-surgery, either by genetic drift or negative selection, particularly if they conferred treatment sensitivity to the tumour cells.

It is also important to consider tumour purity, as diagnostic biopsies often contain variable proportions of non-neoplastic cells. Variants detected at low allele frequencies or confined to poorly represented tumour regions may therefore appear unique to the primary sample simply due to

admixture with normal tissue or differences in sampling region. Additionally, non-biological factors could contribute to the identification of these unique variants. Sampling bias, where not all tumour heterogeneity is captured in the biopsy, may result in variants appearing unique to the diagnostic sample. Furthermore, technical artifacts, such as sequencing errors or DNA changes induced by formalin fixation of samples (Wong et al., 2014), can also lead to the identification of mutations that seem specific to the primary sample but may not reflect true biological differences.

3.1.5 Recurrence specific

Recurrent tumour-specific variants, which were not identified or were present at very low Variant Allele Frequencies (VAF; i.e., below the detection limit of standard variant calling) in the primary tumour, highlight the intricate dynamics of cancer evolution and the limitations of sampling and detection techniques. These variants might have existed in the primary tumour within a minor subclone, undetected due to their low abundance. Such subclonal populations, harbouring unique genetic alterations, might not have been effectively targeted or removed by initial treatments or surgery. Consequently, these subclones can persist, evolve, and eventually dominate in the recurrent tumour, especially under selective pressures such as therapy and the body's immune response. Furthermore, the initial sampling of the primary tumour might not have captured these specific subclonal populations, leading to an underrepresentation of their genetic diversity.

3.1.6 Shared variants

The presence of shared variants among primary and recurrent tumours may indicate the existence of clonal driver mutations, which may have happened early to facilitate the development and progression of this disease, and so are present in all tumour cells. The presence of shared variants shows that the recurrent tumours arise from the continued evolution of unresected cells from the primary tumour rather than from the emergence of a new tumour via de novo mechanisms. This subgroup of variants is of particular interest as it allows the inspection of subclonal dynamics, i.e., looking for non-clonal variants that define subclones and then seeing whether those subclones expanded or reduced over time.

Overall, the distribution of common and unique variants across primary and recurrent glioblastoma tumours can shed light on the processes underlying tumour growth and therapeutic response. Focusing on each classification of variants may reveal further insights. Shared variants, for instance, as well as those unique to either the primary or recurrent tumour, can be further studied by using Cancer Cell Fraction (CCF). CCF analysis enables the estimation of the proportion of tumour cells carrying a given mutation, thereby distinguishing between clonal and subclonal events. For shared variants, CCF helps confirm their clonal nature and persistence across time, whereas for unique variants, it can indicate the emergence or loss of subclones during tumour evolution. (Tanner et al., 2021), conducted a comprehensive assessment of subclonal deconvolution pipelines in cancer genomics using sophisticated tumour genome simulation tools. They created various datasets with different mutation rates, tumour complexities, and purities, and called variants at various depths. Their findings emphasize that higher sequencing depths, such as 250x, are more effective for accurately estimating CCF, especially for detecting rare mutations present in a small fraction of cancer cells.

However, in scenarios where high-depth sequencing is not feasible, the computation of CCF from sequencing data becomes challenging. Under these circumstances, a more simplistic proxy is the use of VAF, which can still provide approximate insights into subclonal composition and relative mutation abundance.

3.1.7 Variant Allele Frequency (VAF)

VAF measures the proportion of sequencing reads that support a particular variant within a tumour sample and is widely utilized in genomic studies such as whole-genome sequencing (WGS), whole-exome sequencing (WES), and targeted panels to estimate the frequency of somatic mutations in cancer genomes. In glioblastoma (GBM) and other cancers, VAF serves as a useful metric for assessing

clonal composition and tumour heterogeneity and may provide insights into therapeutic responses or disease progression (Nadeu et al., 2016). Under ideal conditions, VAF can act as a proxy for Cancer Cell Fraction (CCF), assuming that each cancer cell harbours one mutant allele and one wild-type allele. This assumption creates a direct relationship between the frequency of the mutant allele in the DNA sample (VAF) and the fraction of cancer cells carrying that mutation (CCF).

However, the use of VAF as a proxy for CCF can be problematic due to the complexity introduced by copy number variations (CNVs). CNVs result in multiple copies of genes or genomic regions within cancer cells, disrupting the straightforward correlation between VAF and CCF. For instance, CNVs can cause an overestimation or underestimation of CCF if VAF is used without appropriate adjustments (Figure 3-2). In cases where a gene is amplified, the VAF might appear higher, not because more cells carry the mutation, but because there are more copies of the gene in each cell. Conversely, deletions could lead to a reduction in VAF, underestimating the true CCF. These discrepancies highlight that while VAF can provide a rough estimate of CCF, it must be interpreted with caution, particularly in the presence of CNVs that can complicate the relationship.

Moreover, GBM tumours are highly heterogeneous, consisting of multiple subclones with distinct genetic profiles, and this heterogeneity, combined with CNVs, complicates the use of VAF as a proxy for CCF. Accurate estimation of CCF from VAF requires careful handling of CNAs, which can significantly distort VAF measurements. While some methods adjust for CNAs directly, others rely on prior correction or the masking of variants in regions with variable CNAs. Masking can be a safer approach, especially when dealing with low-coverage data, as it helps to minimize the impact of CNAs on CCF estimates. However, this strategy is not without its disadvantages; it can inadvertently exclude important mutations in regions know to be amplified or deleted in chromosome 7,9 and 10 which are common in GBM (Miura et al., 2018). Despite these challenges, masking remains a practical option when direct CNA adjustments are not feasible.

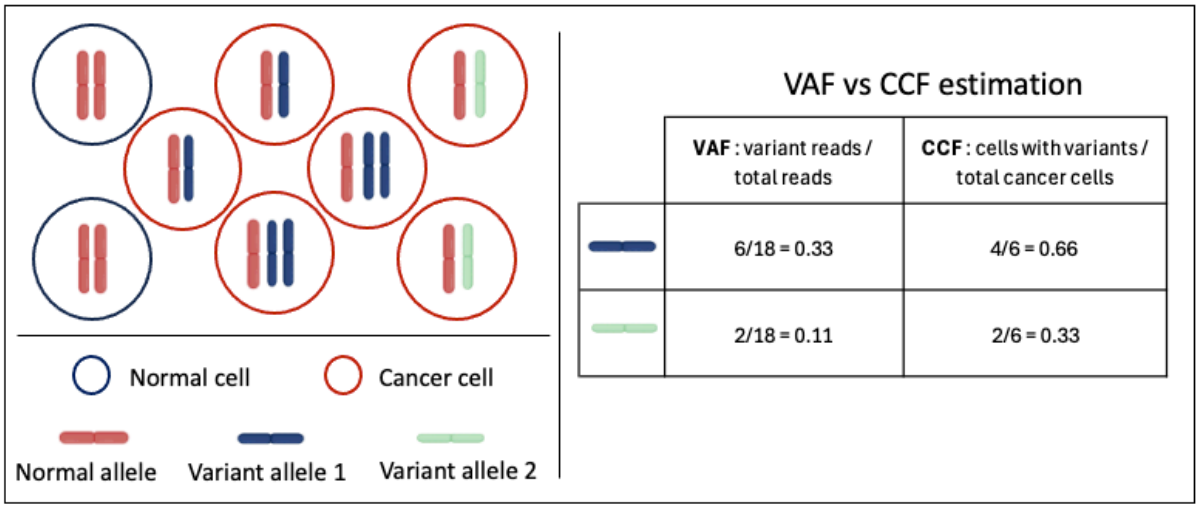


Figure 3-2: Demonstration of CCF estimation.

This diagram shows how VAF and CCF calculated. VAF is the fraction of variant supporting reads, whereas CCF accounts for ploidy, purity.

3.1.8 Using VAF to Investigate Clonal Evolution

In this chapter, I explore the subclassification of shared variants using VAF into four distinct groups, based on how the VAF of each mutation changes from primary to recurrence. I aim to investigate the characteristics of each group, starting with the first group which consists of mutations with low VAF in both tumours. This group contains mostly passenger mutations, which are characterised by having a low allele frequency. Passenger variants usually arise randomly during DNA replication and don't confer a selective advantage to the tumour cells. However, not all low allele frequency variants are necessarily passengers, some variants could be biased due to sequencing techniques, or they have not yet reached clonal dominance. However, using longitudinal samples suggests that this group of variants is likely to be passengers rather than drivers (Consortium, 2020, Aaltonen et al., 2020).

The second subgroup of shared variants are those with high allele frequency in primary tumours but low allele frequency in recurrent tumours. A decrease in VAF from primary to recurrent tumours could imply that treatment played a role and the tumour partially responded (Shomali and Gotlib, 2018). In GBM, understanding this dynamic is crucial for exploring potential treatment strategies. Specifically, a reduction in VAF may indicate a subset of tumour cells that were susceptible to the treatment, thereby offering a window into the tumour's heterogeneity and the effectiveness of the therapy employed.

The third group of shared variants is represented by the mutations that had an increase in VAF through treatment. The increase in VAF indicates that these variants, possibly subclonal in the primary tumour, gained a selective advantage, leading to clonal expansion and potentially contributing to treatment resistance. The presence of these variants in GBM is indicative of Minimal Residual Disease (MRD), a condition characterized by residual cancer cells that survive standard therapy and remain in the brain. These cells, often undetectable by radiological examinations, contribute to the shared genetic variants observed in recurrent GBM tumours (Qazi et al., 2022).) This phenomenon is a well-established fact in GBM pathology, reflecting the challenges in completely eradicating tumour cells with current treatment modalities.

The last group of shared variants represents mutations with high allelic fractions in both primary and recurrent tumours. The uniformly high allelic fractions of these mutations suggest their clonal nature, indicating that they were present early in the tumour's development and persist through to recurrence. This uniformity implies that these clonal mutations confer a survival and proliferative advantage to the tumour cells, playing a pivotal role in both the maintenance and recurrence of GBM (Korber et al., 2019).

Once classified into the different groups, it is possible to inspect the genes impacted by the different subsets of variants, to see whether any biological processes are implicated with regards the different interpretations for each grouping. One way to do this is by looking at functional enrichment analysis.

3.1.9 Functional enrichment

Functional enrichment analysis is a method used in the computational biology field to determine which biological functions, such as cellular processes, molecular functions, or biological pathways, are significantly associated with a gene list. These genes typically come from experimental data like genomic or expression datasets. One can provide a list of genes identified through proteomic analysis, which can reveal changes in protein expression; genes affected by epigenetic modifications from methylation studies, which can influence gene functions; and genes that are differentially expressed as determined by RNAseq, ranked according to a defined threshold. Alternatively, as I aimed to do here, the gene lists can be constructed from those that harbour significant variants. Over-representation analysis (ORA) describes a statistical method that determines whether a predefined set of biological functions or pathways is significantly overrepresented (or underrepresented) in a set of genes of interest. There are many tools that can be employed for

understanding the function of genes and gene sets. Functional enrichment analysis is increasingly being used to study cancer progression and recurrence and one of its applications widely used is Gene Ontology (GO)(Ashburner et al., 2000, Gene Ontology et al., 2023). GO categorizes genes into:

1- Biological Processes: This category encompasses terms describing molecular events within cells or organisms, such as cell division or signal transduction.
2- Cellular Components: Terms here describe the parts of a cell where gene products are located, such as the nucleus or cell membrane.
3- Molecular Functions: This involves terms describing the activities of gene products, like binding to other molecules or transporting them.

In ORA, the significance of gene set enrichment is determined using statistical test called hypergeometric testing. The hypergeometric test shows whether the overlap between the set of genes of interest and the set of genes with a particular function is greater than what would be expected randomly, indicating a potential functional significance. The hypergeometric test is formulated as follows:

$$P\ (X \geq x) = 1 - P(X \leq x - 1) = 1 - \sum_{i=0}^{x-1} \frac{\binom{M}{i}\binom{N-M}{n-i}}{\binom{N}{n}}$$

Here, *N* is the total number of background genes, n is the number of genes in the list of interest, *M* is the number of genes in a specific gene set, and x is the number of genes in the intersection of the list and the gene set. This equation calculates the probability that at least
x genes from the gene set are found in the gene list, contrasting with a random distribution. This test is pivotal in determining the statistical significance of the observed enrichment, providing a quantitative measure for the association between gene sets and biological functions.

However, functional enrichment primarily focuses on whether there are enough genes from the same gene set in the gene list (Fig 3-3A-B). This approach might overlook scenarios where a specific pathway or gene set is consistently affected across multiple patients (Fig 3-3C). In other words, the importance lies not only in the presence of multiple genes but also in their recurrence across different patients. This aspect requires careful inspection, highlighting the need for more comprehensive pathway analysis tools. These tools can delve deeper into understanding the commonalities and variations in gene expression or alterations across different patient samples

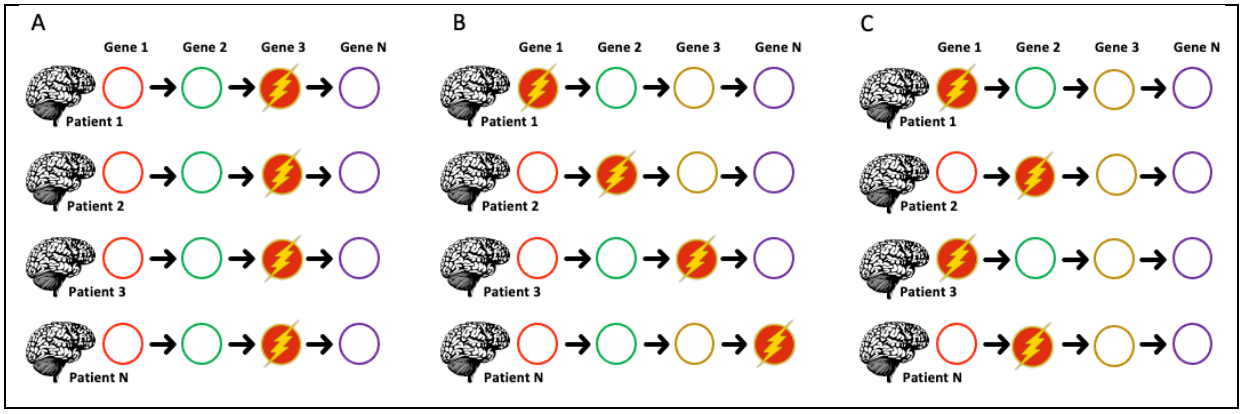


Figure 3-3: The difference between common approaches of pathway enrichment analysis and our novel approach.

A: This approach identifies significantly mutated genes. i.e. The same gene mutated in multiple patients.
B: This approach identifies significantly mutated pathways. It requires a significant number of pathway members to be mutated.
C: This novel approach identifies significantly mutated pathways based on their prevalence even in a subset number of patients.

3.1.10 Comparative pathways analysis of primary and recurrent GBM tumours

Previous studies have indicated that the GBM genetic profile maintains its heterogeneity during therapy (Korber et al., 2019, Barthel et al., 2019). Despite this, it is possible to identify driver pathways that may be exclusive to a small fraction of patients and can provide insight into the processes responsible for therapy resistance. This information can be used to develop drugs that can either slow or stop the tumour recurrence progression. In addition, identifying variants that disappear from primary to recurrent cells can provide information on the cellular processes that are involved in sensitising the cancer cells.

One aspect of understanding the cause of therapeutic resistance is to identify driver variants. However, these variants may present with allelic fractions similar to that of neutral passengers that have no significant effect on tumour progression (Consortium, 2020). Distinguishing driver from passenger variants is challenging and is a crucial aspect of cancer bioinformatics (Bailey et al., 2018). There are various methods for identifying driver genes, including looking for frequently mutated genes, identifying variants that cluster in specific regions of a gene sequence or protein structure, predicting the functional consequence of variants, and comparing the numbers of deleterious and benign variants within them. Pan-cancer studies have collectively used the aforementioned methods to identify driver genes across cancers. Combining these consensus methods has revealed nearly 600 driver genes across cancers, varying depending on the cancer type (Martinez-Jimenez et al., 2020).

The variant assessment strategies mentioned above focus on identifying driver genes at an individual level, which is not always possible, especially in smaller groups. Examining the frequency of variants

across multiple related genes in a specific pathway addresses this limitation. This approach increases statistical power and provides additional insights into the cellular mechanisms that are affected as a consequence of the variants. My aim is to apply this approach to two distinct cohorts, in order to identify variants that could potentially drive therapy resistance or sensitivity in glioblastoma (GBM). The objective is to utilize the VAF to differentiate between clonal and subclonal variants, particularly in those shared between primary and recurrent tumours. This is especially crucial as prior studies have demonstrated that there is a paucity of specific genetic alterations shared across multiple patients that re-emerge after treatment (Barthel et al., 2019, Korber et al., 2019, Wang et al., 2016b). This does not preclude there being pathways that are repeatedly mutated to confer treatment resistance though (Figure 3-3).

3.1.11 PathScore

In this chapter, I employ PathScore, a pathway analysis tool designed to compute pathway enrichment scores by considering the mutational load per patient, transcript length, and gene-specific background mutation rates (Gaffney and Townsend, 2016). PathScore calculates the actual and effective sizes of pathways in a given database using patient-gene pairs as input. The actual pathway size is based on the total number of DNA bases present in all genes of the pathway, while the effective pathway size incorporates gene-specific background mutation rates, gene transcript lengths, and per-patient mutation rates to estimate the maximum likelihood of the pathway size. PathScore also generates a P-value through a likelihood ratio test to assess the significance of the difference between the actual and effective pathway sizes, enabling direct comparisons of identical pathways across different inputs and illustrating the impact of alterations on pathways.

The selection of PathScore for this study is based on prior work by our research group, which compared various pathway analysis tools, including network analysis methods and *de novo* approaches. Network analysis methods, which utilize protein-protein interaction data and mutual exclusivity to identify driver subnetworks, are limited by their reliance on predefined network information that may not fully represent the altered physiology of tumour cells. *De novo* methods, which identify driver processes through patterns of altered genes without predefined pathways, offer greater flexibility but can suffer from reduced specificity, particularly when working with smaller subsets of altered genes, as in our study. This evaluation highlighted PathScore's distinct advantages, particularly its individualized approach that accounts for patient-specific mutation rates, gene lengths, and background mutation rates, making it well-suited to the specific needs of this analysis.

PathScore's per-patient evaluation allows for a more detailed analysis that is particularly effective for identifying pathways associated with therapy resistance or sensitivity in GBM. This individualized approach is crucial given the variability in mutational burden across patients and the need to distinguish between pathways implicated in therapy response versus those generally altered in GBM. Furthermore, PathScore enables direct comparisons of identical pathways across different subsets of altered genes, enhancing the ability to identify specific pathways relevant to therapy-induced changes. Overall, PathScore was chosen due to its robust handling of individualized data and its capacity for pathway comparisons across distinct patient subsets. These features align with the study's aim to elucidate cellular processes driving GBM progression through therapy, continuing and building upon the foundational work conducted within our research group.

3.2 METHODS

3.2.1 Data processing

In this chapter, I employed a comparative analysis approach to investigate treatment resistance and tumour recurrence, using data from two cohorts to allow cross validation of the findings. The discovery cohort represents the variants called from the exome data processed in chapter one as detailed in section 2.2. The validation cohort represents the variants retrieved from GLASS database. Variants of the discovery cohort were available in VCF file format while variants of the validation cohort were combined in one large, tabulated file. I used custom python scripts to process both cohorts' data, mainly the validation cohort where I filtered the data to include only the IDH-wt samples that were collected from patients that followed the standard multimodal therapy of GBM. I then converted samples into pairs to annotate shared variants between each pair of samples.

3.2.2 Variant classification

After processing the data and identifying shared variants between primary and recurrent tumours, my hypothesis was that these variants play a crucial role in tumour behaviour post-treatment. I classified these shared variants into four groups based on changes in their prevalence during treatment. This classification was essential in understanding the dynamics of these variants under therapeutic pressure. To achieve this classification, I first calculated the mean VAF for both primary and recurrent variants with a custom python script (https://github.com/umyma1/thesis_appendix/tree/main/chapter3) to establish a robust classification threshold. This step was essential in differentiating the variants according to their prevalence dynamics. The first group is for susceptible variants which had a VAF that is higher than the threshold of primary samples and below the threshold of recurrent samples. I then identified the second group, expanded variants, by looking at those that exceeded the VAF threshold of recurrent cohort and remained below the threshold of the primary samples. The third and fourth groups are those variants that exceeded the VAF threshold of both cohorts and the variants with VAF below the threshold of both cohorts.

3.2.3 Filtering variants

I began by investigating the raw variant data, where all mutation calls, without any filtering, were included to allow a comprehensive examination of the variants and to detect any fundamental patterns or groupings that could guide more detailed analyses. Following this, I investigated the data

after refining mutation calls through specific filtering criteria to understand the specific characteristics and implications of these variants in greater detail.

For the discovery cohort, I annotated variants using the Variant Effect Predictor (VEP) with the primary objective of filtering out variants with neutral effects on protein functions, i.e. variants likely to be benign or without significant functional consequences. VEP was used to rank variant impacts on proteins according to Gene Ontology (GO) terms, categorizing them as HIGH, MODERATE, LOW, or MODIFIER. To identify potentially deleterious variants that might be missed by VEP's ranking, I performed additional annotation on LOW impact variants using SIFT and PolyPhen-2. MODIFIER variants, typically non-coding and potentially introducing noise, were excluded from the analysis.

In the validation cohort, variants were annotated with Funcotator which has different annotation terms from VEP. This was because the GLASS consortium data were provided already annotated using Funcotator, and re-annotation from raw sequencing files was not feasible. To maintain consistency across the study, I applied the same Sequence Ontology (SO) terms used by VEP to re-classify variants in the validation cohort. This methodological consistency was critical for ensuring reliable comparative analysis between the cohorts. Finally, I filtered the variants following the same method of filtering applied to the discovery cohort. The SO terms used by VEP and the equivalent annotations by Funcotator are summarized in Table 3-2.

Table 3-2 VEP and Funcotator annotations

SO TERM	VEP	Funcotator	IMPACT (VEP)
splice_acceptor_variant	Splice acceptor variant	Splice_Site	HIGH
splice_donor_variant	Splice donor variant	Splice_Site	HIGH
stop_gained	Stop gained	Nonsense	HIGH
frameshift_variant	Frameshift variant	Frame_Shift_Del, Frame_Shift_Ins	HIGH
stop_lost	Stop lost	Nonstop	HIGH
start_lost	Start lost	Multiple Annotations	HIGH
inframe_insertion	Inframe insertion	In_Frame_Ins	MODERATE
inframe_deletion	Inframe deletion	In_Frame_Del	MODERATE
missense_variant	Missense variant	Missense	MODERATE
synonymous_variant	Synonymous variant	Silent	LOW

3.2.4 Gene lists construction and functional enrichment analysis

In my study, I conducted an enrichment analysis to explore significant biological terms associated with lists of genes. In preparing these gene lists for the enrichment analysis, I focused on including genes that harboured deleterious variants as described in section 3.3.

The primary objective of my pathway analysis approach was to pinpoint genes that play a key role in either resistance or response to treatment. To achieve this, I constructed gene lists based on two distinct groups of variants. The first group consisted of variants that showed an increase in frequency from primary to recurrent stages of cancer. The second group comprised variants that exhibited a decrease in allele frequency from primary to recurrent stages. The gene lists were then analysed using two different enrichment analysis tools, WebGestalt and PathScore.

3.2.5 GO enrichment analysis by WebGestalt:

For GO enrichment analysis, I created two gene lists for the two groups of interest and submitted them to WebGestalt (Liao et al., 2019). WebGestalt accepts gene lists with HGNC Symbol, and the analysis was performed using the reference genome GRCh38 and a background gene set of all protein-coding genes in the reference genome. Adjusted P-values were employed to identify pathways that were significantly enriched, considering a threshold below 0.05 as statistically significant. The results were visualized using bubble charts, created by a custom R script to identify the most significantly enriched biological processes (https://github.com/umyma1/thesis_appendix/tree/main/chapter3).

3.2.6 Pathway enrichment analysis by PathScore:

PathScore requires genes to have three annotations which are HGNC Symbol, Entrez ID, and patient ID, because it performs patient-specific pathway impact analysis. Unlike conventional enrichment tools such as WebGestalt (Liao et al., 2019), which operate on non-redundant gene lists, PathScore relies on variant-level data. This distinction is fundamental to its algorithm, which treats each mutation as an independent sampling event from the genome to estimate the probability that a pathway is affected in a particular patient. By modeling mutations rather than genes, PathScore accounts for the frequency, distribution, and background mutation rate (BMR) of variants, scaled by gene length. In other words, multiple variants occurring within the same gene or across different genes in the same pathway contribute cumulatively to the likelihood that the pathway is aberrated. Aggregating data to the gene level would obscure this information, as distinct mutations within one gene would be collapsed into a single observation, underestimating the mutation burden and

pathway impact. Therefore, the input file retains redundant gene entries, with each variant annotated by HGNC symbol, Entrez ID, and patient ID to maintain variant-specific resolution. PathScore was run using the Gene length & BMR-scaled algorithm with default background mutation rates. The enrichment score reflects the effect size of the aberrated pathway, and statistical significance was assessed using the Benjamini–Hochberg adjusted p-value.

3.2.7 Analysis of variants in diverse copy number regions

To assess the impact of Copy Number Alterations (CNAs) on pathway analysis, variants were categorized according to their location within regions of neutral copy number (CN), high CN, low CN, or stable abnormal CN in primary and recurrent tumours. In this approach, only variants within CN-neutral or CN-stable regions were retained, while those in highly variable CN regions were excluded to minimize the confounding influence of copy number–driven signal distortion. This filtering was particularly important given the absence of Cancer Cell Fraction (CCF) estimates, as CN variability can inflate variant allele frequencies (VAFs) in amplified regions or mask true allelic losses in deleted regions. Restricting the analysis to CN-stable regions therefore ensured that pathway enrichment more accurately reflected the burden of genuine point mutations rather than CN-related artefacts. For the discovery cohort (processing was detailed in Chapter 2.2), all VCF files were converted into BED files for analysis using BEDtools v2.30.0 (Quinlan and Hall, 2010) and annotated with CN status. Using BEDtools, I then identified the overlapped and unique regions in both samples and extracted the regions with stable CN status using a python script to perform pathway analysis. Conversely, samples in the validation cohort were processed using several CN callers, including TITAN, Sequenza, PyClone, and GATK. Copy number results were concatenated for the whole datasets resulting in multiple result files from each caller. I evaluated the annotations produced by each caller and decided to use the results called by TITAN as it has more annotations than the other callers. These additional annotations provided useful information not available from tools like Sequenza, PyClone, and GATK, which helped in filtering and processing the data more effectively using BEDtools. I then split the results by patient to create individual sample results and run BEDtools on each pair of samples following the same method applied to the discovery cohort. I wrote a python script to process the validation cohort and create BED files. BEDtools commands for processing both cohorts can be found in (https://github.com/umyma1/thesis_appendix/tree/main/chapter3).

3.2.8 Comparative analysis of primary versus recurrent profiles using GISTIC

The aim of this analysis was to track the evolution of copy number (CN) alterations from primary to recurrent tumours. The GISTIC tool was employed for this analysis due to its ability to identify regions

of the genome that are significantly altered in cancer. To assess the CN change between the pair of samples I utilised primary CN profiles as the baseline. This step is crucial for understanding the initial genomic state before any treatment or progression. First, I identified all regions exhibiting stable or neutral CN in both primary and recurrent tumours and I excluded them. This filtration was essential to focus solely on regions showing significant alterations, thus enhancing the specificity of our analysis. Second, I used BEDtools to identify the overlapping regions and calculated the logarithmic ratio of CN in recurrent tumours to CN in primary tumours using a python script . This step quantified the degree of CN alteration, offering a clear comparison between the two profiles. I then prepared the segmentation files which included the loci coordinates and the derived log ratios. These files are critical for GISTIC analysis as they represent the segmented CN data in a format that the tool can process. Finally, my method led to subsetting the genomes leaving out some regions without information. However, GISTIC requires that all chromosome segments should have coverage across all samples. To overcome this issue, I added CN neutral regions to each file to account for any missing regions to ensure that the GISTIC analysis was not biased by missing data. I prepared the required files for GISTIC using a python script and the segmentation files were assessed by IGV before and after adding CN neutral regions (https://github.com/umyma1/thesis_appendix/tree/main/chapter3)

3.3 RESULTS

3.3.1 Data processing

3.3.1.1 Datasets

In this chapter, I performed analysis on two independent cohorts. The first is labelled as the discovery cohort in this chapter; the cohort of samples processed locally, and which contains the variants that were called from the whole exome data that passed the stringent filtering criteria as thoroughly discussed in chapter 2.

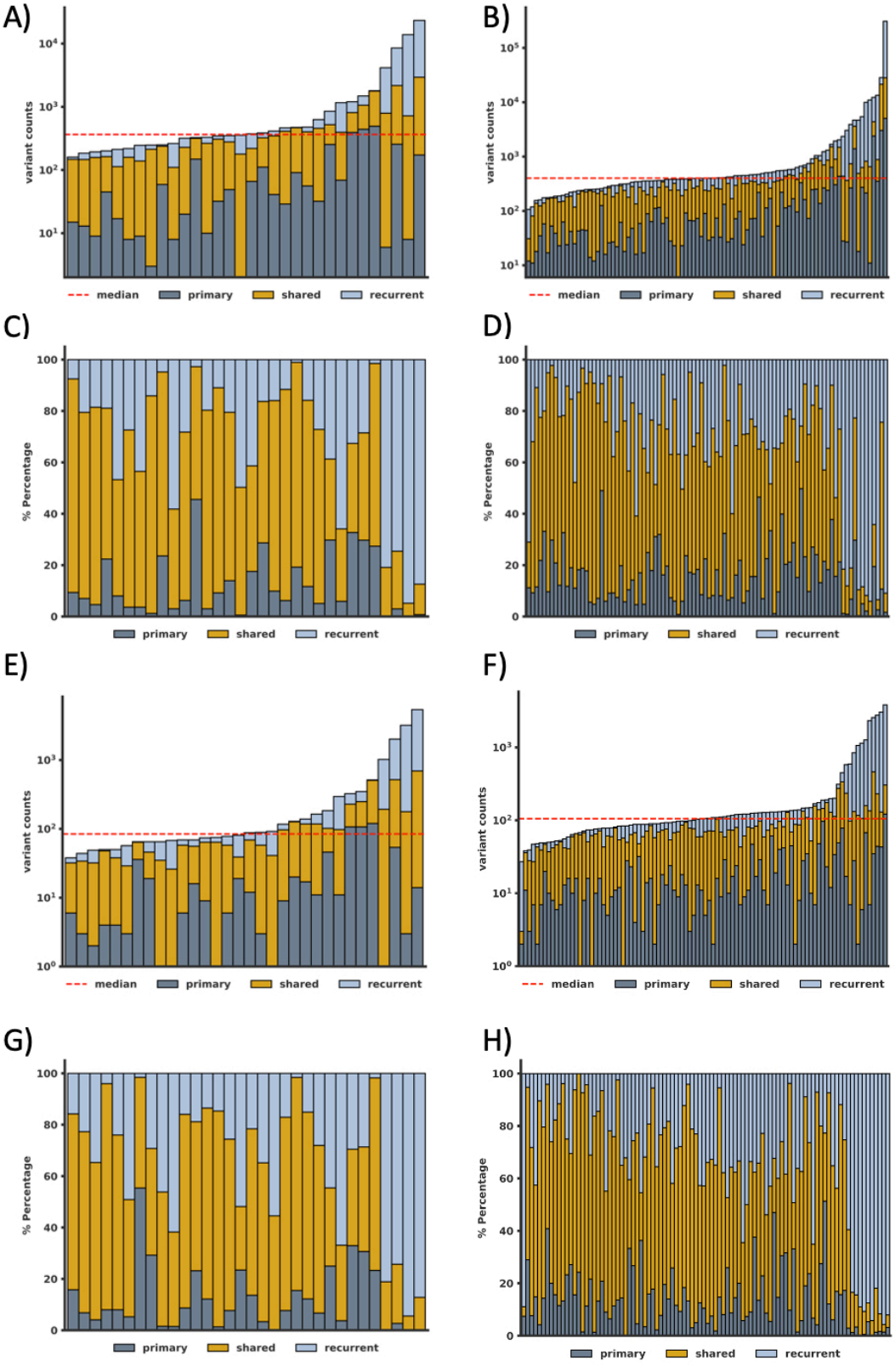
The second cohort, labelled as the validation cohort, consists of processed variant calls from exome data, provided by GLASS. This collection includes data from 392 patients, along with their clinical information. The samples are grouped into different tiers, such as the silver, gold, and platinum sets, based on data curation criteria, including whether they pass QC thresholds for DNA integrity and the availability of comprehensive molecular data like RNA expression or methylation profiles. The gold set of the GLASS dataset was produced from high-quality sequencing data for both primary and recurrent tumours. However, raw sequencing reads are not available for this set; instead, mutation call sets, identified using the Mutect2 tool, are provided.

Both the discovery and validation cohorts were processed using a similar approach. This was achieved by employing Mutect2 in a multi-sample mode, enabling the effective identification of somatic point mutations. The similarities in processing and findings between the two cohorts facilitated a robust comparative analysis, essential for drawing reliable conclusions in this chapter.

3.3.1.2 Comparative Analysis of Datasets

After conducting a thorough quality control process, I examined the distribution of somatic variants across tumour stages in both cohorts to assess dataset comparability (Fig. 3-4). Panels A–H illustrate per-patient variant counts and proportions before and after filtering for deleterious variants. The median total variant counts were 365 and 400 for the discovery and validation cohorts, respectively. After filtering for deleterious variants, these numbers reduced to 84 and 105.

To check for differences in variant composition among tumour categories (primary-specific, recurrent-specific, and shared) within and between cohorts, I applied a one-way ANOVA followed by Tukey's multiple-comparison test. The results are summarised in figure 3-4I-J. The analysis confirmed that the discovery and validation cohorts show no significant difference in overall variant distribution, validating their comparability, while in both cohorts the shared-variant group constitutes the largest and statistically most enriched category ($p < 0.01$ across most comparisons).


A)
B)
C)
D)
E)
F)
G)
H)
variant counts
% Percentage
median
primary
shared
recurrent

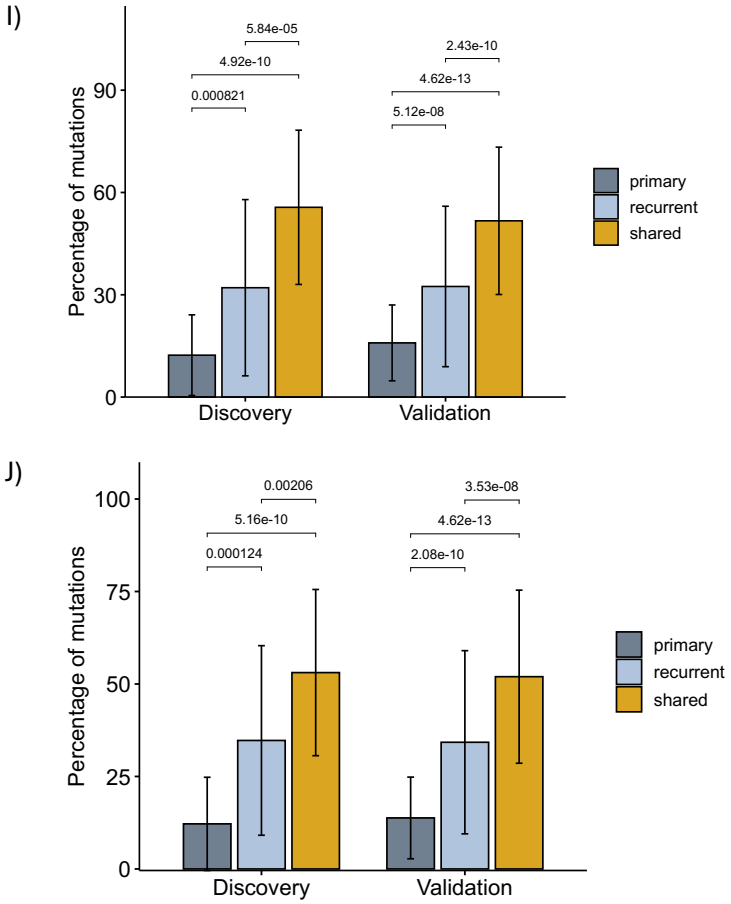


Figure 3-4: Distribution and classification of variants across cohorts and tumour stages.

Panels A-D show data before applying filtration criteria; panels E-H illustrate data after filtering benign variants.
A-B: Per-sample counts of SNVs + INDELs (log_{10} scale) in the discovery (n = 32) and validation (n = 94) cohorts. Bars are partitioned into variants private to the primary tumour, private to the recurrent tumour, or shared between both tumours. The red dashed line marks the cohort median.
C-D: Relative proportions of variant types in each patient for the discovery (C) and validation (D) cohorts.
E-F: Total counts of deleterious SNVs and INDELs per patient (after filtration) in the discovery (E) and validation (F) cohorts.
G-H: Relative proportions of variant types per patient in each cohort after filtration.
I-J: Results of one-way ANOVA comparing variant categories within each cohort before (I) and after (J) filtering, followed by Tukey post-hoc pairwise tests. Brackets and p-values indicate statistically significant pairwise differences. The shared-variant category is significantly higher than either primary- or recurrent-specific variants in both cohorts

3.3.1.3 Variant classification

As my investigation aimed to unravel the genetic underpinnings responsible for treatment resistance and tumour recurrence. I initiated this by analysing shared variants between primary and recurrent tumours in both discovery and validation cohorts. This analysis revealed that a significant proportion of the variants shared between primary and recurrent tumours were also shared across both the validation and discovery cohorts (Figure 3-4J), thereby directing our focus towards these shared variants for further insights into tumour recurrence post-treatment.

The shared variants were categorized into four groups based on their prevalence changes during treatment: susceptible variants (decreasing in frequency), expanded variants (indicative of potential clonal expansion), and stable variants further divided into high allele frequency (likely clonal) and low allele frequency groups.

Initial analysis of VAF was conducted to classify these variants. In the unfiltered data, the mean VAF of primary variants was 0.10 (SD = 0.14) in the discovery cohort and 0.20 (SD = 0.19) in the validation cohort. The mean VAF of unfiltered recurrent variants was 0.12 (SD = 0.15) in the discovery cohort and 0.19 (SD = 0.15) in the validation cohort as indicated in Figure 3-5A-B.

After filtering out non-deleterious variants, we observed a consistency in the mean VAF of the discovery cohort (0.10, SD = 0.15 for primary variants; 0.12, SD = 0.16 for recurrent variants). However, there was a notable reduction in the mean VAF of primary variants in the validation cohort (0.07, SD = 0.13), while the mean VAF of recurrent variants remained relatively unchanged (0.20, SD = 0.16) as shown in Figure 3-5C-D.

Despite these changes in mean VAF, especially in the validation cohort, the proportions of each variant group remained broadly similar before and after filtering. Specifically, variants decreasing in frequency accounted for 11%-12% in the discovery cohort and 13% in the validation cohort, both before and after filtering. Variants indicative of clonal expansion represented 16%-15% in the discovery cohort and 10%-13% in the validation cohort, before and after filtering, respectively. Variants with high allele frequency comprised 13%-14% in the discovery cohort and 25%-31% in the validation cohort. Finally, variants with low VAF represented 59% in the discovery cohort and 53%-44% in the validation cohort, both before and after filtering.

The overall proportions of variant groups post-filtering support the robustness of the classification strategy. While a reduction in the mean VAF of primary variants was observed in the validation cohort, the general distribution of variant groups remained relatively stable, with the exception of high VAF variants, which showed a more pronounced increase. Nonetheless, this level of consistency observed both before and after filtering non-deleterious variants reinforces the rationale for retaining only deleterious variants for downstream pathway analysis, helping to reduce background noise introduced by less functionally relevant mutations.

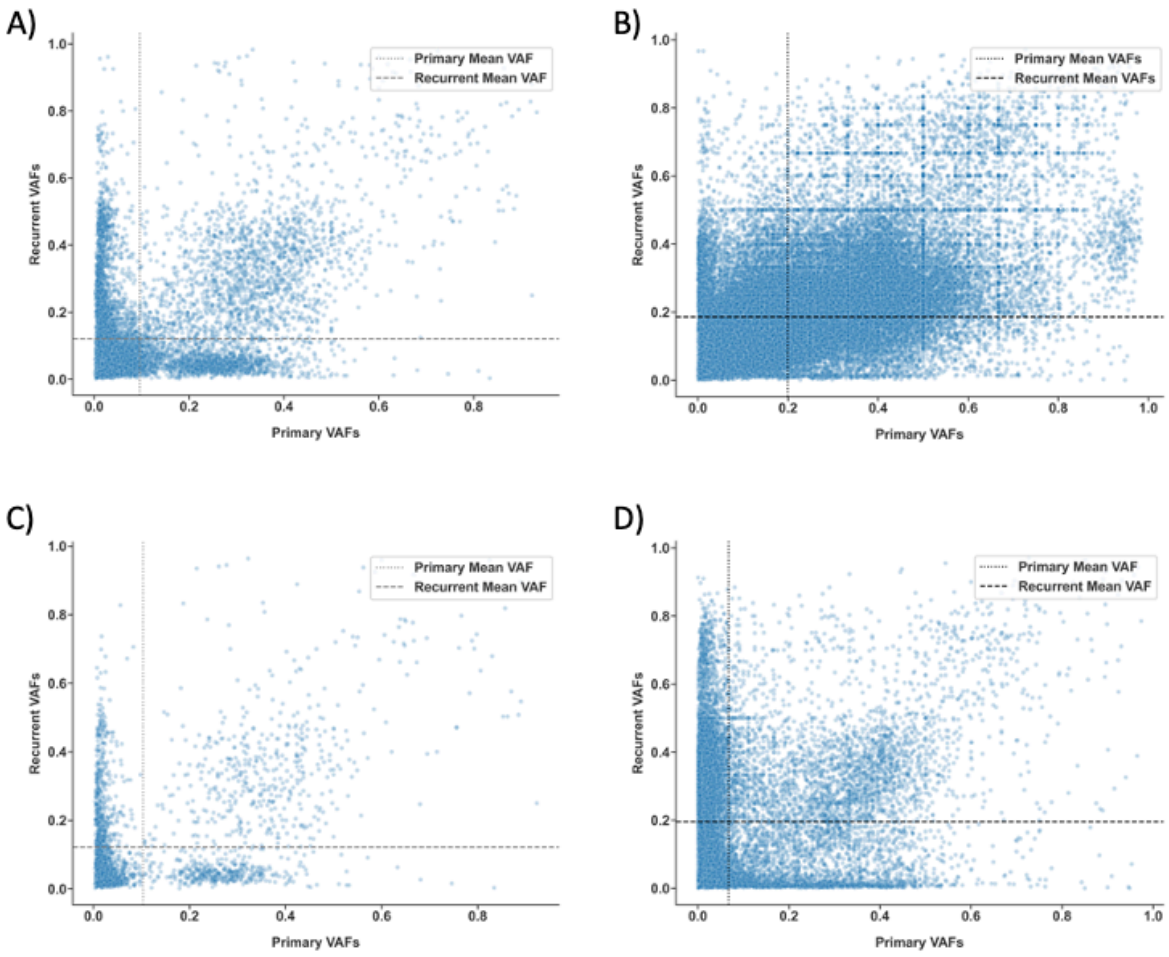


Figure 3-5: Variant classification using allele frequency.

A-B: scatter plots of shared variant VAFs in the discovery *(A)* and validation *(B)* cohorts. Variants are categorized into four groups based on their Variant Allele Frequencies (VAFs), relative to the mean VAF of primary and recurrent tumour variants. 1) Bottom left quadrant shows variants with VAF below the mean, 2) bottom right quadrant shows variants with VAF decreased from primary to recurrent tumour, 3) top left contains variants with VAF increased from primary to recurrent tumours, and 4) top right quadrant contains variants with VAF remained high through treatment.
C-D: Quadrant plots of discovery *(C)* and validation *(D)* cohorts, showing only the deleterious variant VAFs which are selected for the downstream analysis.

3.3.2 Gene Ontology (GO) enrichment analysis

In this section, I aimed to uncover the biological consequences of variant selection during therapeutic treatments by performing an in-depth Gene Ontology (GO) enrichment analysis. From the four identified groups, I selected two unique gene sets: the first comprised of variants likely to be selected for during therapy, which might suggest potential resistance mechanisms, and the second consisted of variants likely to be selected against, indicating possible susceptibilities or potential therapeutic responses. These two groups had variable VAF values indicating response or resistance. Therefore, it was hypothesised that analysis of these two groups would provide biologically meaningful outputs from the pathway analysis. The two dismissed groups represent variants with stable VAF that when

analysing variants with low VAF it is likely to include false positive or artefactual variants. For each gene set, I assessed the enrichment of GO terms across three fundamental categories: Biological Process, Cellular Component, and Molecular Function. The discovery dataset served as the foundation for pinpointing significant GO terms, which I then sought to corroborate using the validation dataset to confirm the consistency of the observed patterns. This methodical approach provided a comprehensive view of the selective pressures therapy may impose on the genome, offering insightful revelations into the gene processes in response to treatment. The comparative analysis revealed multiple enriched pathways, which I ranked based on the enrichment score and then selected the top ten pathways for further analysis and generated bubble charts to show the GO enrichment analysis results.

I focused on the terms which are shared, for each subgroup of expanded or reduced VAF variants, across the validation and discovery cohort and expand upon these results below.

3.3.2.1 Clonally expanded variants

3.3.2.1.1 Biological processes

In the discovery cohort (Figure 3-6A), the most significantly enriched pathways were related to the sinoatrial (SA) node cell function, with “SA node cell action potential” and “SA node cell to atrial cardiac muscle cell signalling” both showing an enrichment score of 11.5 (FDR=0.0054). Similarly, pathways involving communication between SA node cells and atrial cardiac muscle cells were highlighted “SA node cell to atrial cardiac muscle cell communication” with an enrichment score of 10. The importance of synaptic functions was also suggested by the enrichment in “regulation of synaptic vesicle clustering” (score=9) and “synaptic vesicle clustering” (score=7). Pathways associated with cardiac structure and function, such as “coronary vasculature morphogenesis” (score=6.5) and “cardiac muscle cell action potential involved in contraction” (score=4.5), were also notably enriched. While these pathways are annotated with cardiac or neuronal terminology, they likely reflect broader ion channel and membrane potential regulation processes that may be active in glial tumour cells, rather than representing true action potential generation.

The validation cohort presented a somewhat broader range of biological processes, though still with relevance to cardiac function. The "cellular response to heparin" and "membrane depolarization during SA node cell action potential" both showed an enrichment score of 6.9, aligning with the cardiac conduction system relevance seen in the discovery cohort. Interestingly, the "membrane depolarization during AV node cell action potential" was also highlighted with the same enrichment

value, suggesting a consistency in the relevance of cardiac action potential pathways across both cohorts. Developmental processes were also evident, including "cell migration involved in kidney development" and "trigeminal nerve morphogenesis," each with an enrichment score of 6.9, which may indicate a broader scope of the validation cohort's biological processes being examined. Neurological development and response were further exemplified by pathways like "positive regulation of axon extension involved in axon guidance" (score=6) and "cellular response to histamine" (score=5.4).

A key overlap between both cohorts is the enrichment of SA and AV node-related pathways, which play a crucial role in cardiac conduction. The SA node initiates the heart's electrical activity, sending signals to the atrial muscle cells for coordinated contraction, which involves membrane depolarization during the SA node cell action potential. The AV node then delays the signal, allowing the atria to contract before the ventricles, ensuring proper heart rhythm (Jalife, 1984, Lakatta et al., 2010, MacDonald et al., 2020). The presence of common SA and AV terms in both cohorts suggests that these processes may not be exclusive to the heart. It is possible that they play similar electrical signalling roles in brain functions, and when altered, may contribute to GBM progression.

3.3.2.1.2 Cellular components

In the discovery cohort (Figure 3-6A), the most enriched cellular component was the "laminin complex" (enrichment score=9.6), suggesting its prominent role in the cellular architecture under investigation. This was followed closely by components associated with platelet function, namely the "platelet dense tubular network" and its membrane (enrichment scores 9.4 and 8.6, respectively). The axonal structure was also significantly represented with "axonemal dynein complex" and "axoneme part" showing high enrichment scores (8.6 and 6.3, respectively). Looking closely at these two pathways, a significant overlap of six genes were identified suggesting a functional relationship (Figure 3-6C). The “myosin filament” and “muscle myosin complex” were highlighted as well (enrichment scores 5.8 and 5.7, respectively). While these components are typically associated with muscle function, their enrichment here likely reflects an overlap of actin- and myosin-related genes involved in cytoskeletal organisation and motility in GBM. The heatmap (Figure 3-6C) showed that myosin filament and muscle myosin complex pathways share four genes (*MYH13, MYH6, MYH7, MYOM1*), implying functional or structural relationships. The heatmap also showed that laminin complex and extracellular matrix component have overlapping genes in the discovery cohort, despite extracellular matrix component pathway not being identified in the validation cohort. This may link with the actin/myosin enrichment as these components are involved in cell movement and migration.

The validation cohort showed a shift towards extracellular matrix components, with “fibrillar collagen trimer” and “banded collagen fibril” both presenting the highest enrichment score of 8.2 (Figure 3-6B). This finding indicates a strong correlation with structural proteins involved in extracellular matrix organization. Eight genes were overlapping between the fibrillar collagen trimer, banded collagen fibril and complex of collagen trimers indicating similar functions by these pathways (Figure 3-6D). Cellular complexes related to muscle function, such as the "junctional membrane complex" and "junctional sarcoplasmic reticulum membrane" (enrichment scores 8.0 and 7.5), were also enriched in the validation cohort. Interestingly, the "laminin complex" was again observed but with a lower enrichment score (7.0) than in the discovery cohort and this suggests a potential conservation of its role.

Comparing both cohorts, the discovery cohort showed a particular emphasis on cellular components related to the axonal structure and platelet function, whereas the validation cohort highlighted the extracellular matrix, particularly collagen-related structures, and muscle-associated complexes. The "laminin complex" was the only cellular component to be significantly enriched in both, although with varying degrees of enrichment, which highlights its likely key role in the relevant cellular processes. Laminins constitute a group of glycoproteins essential for the foundational framework of basement membranes present in nearly all animal tissues. Composed of α, β, and γ chain subunits, each laminin forms a heterotrimer (Colognato and Yurchenco, 2000). These are secreted and become part of the extracellular matrices associated with cells. Serving multiple functions, laminins are involved in processes such as development, differentiation, and the movement of cells, owing to their ability to interact with a variety of cell surface proteins. In the context of gliomas, laminins are predominantly found in the microenvironment, particularly around the basal lamina of blood vessels and are notably present at the edge between the brain and tumour (Marino et al., 2023).

The distinction between the cohorts can be attributed to differential expression or involvement of these cellular components in GBM, or possibly due to the inherent biological variability between the discovery and validation populations. The validation cohort's significant emphasis on the extracellular matrix, especially collagen structures which are closely related to laminins in function, could suggest a pathophysiological process that involves tissue remodelling or fibrosis.

3.3.2.1.3 Molecular function

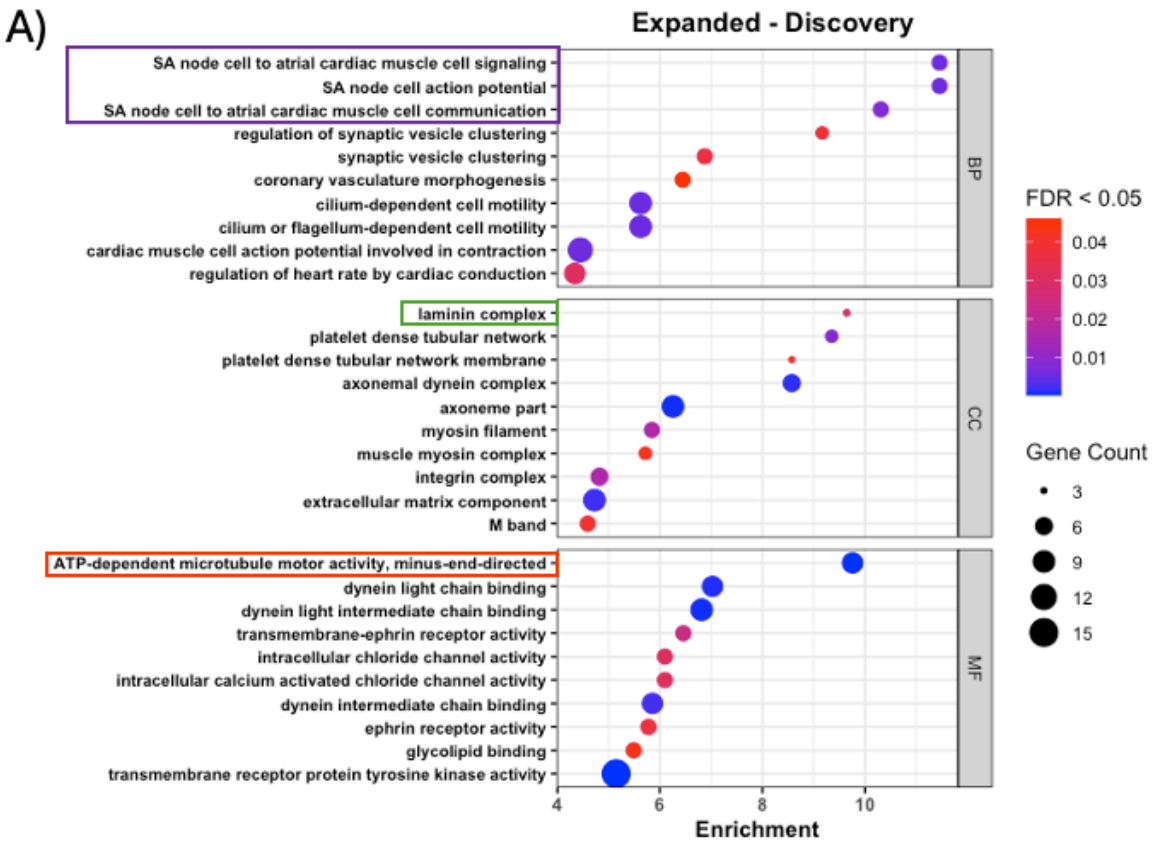
In the enrichment analysis of molecular functions, we observed significant functional categories in both discovery and validation cohorts. The discovery cohort displayed a notable enrichment for activities related to microtubule motor function and dynein binding (figure 3-6A). Specifically, 'ATP-dependent microtubule motor activity, minus-end-directed' showed the highest enrichment score (9.76) and was represented by multiple genes of *DNAH* family. In close relation, functions associated with 'dynein light chain binding' and 'dynein intermediate chain binding' were also significantly enriched, suggesting a concerted involvement of these molecular activities in the biological processes identified in the discovery cohort.

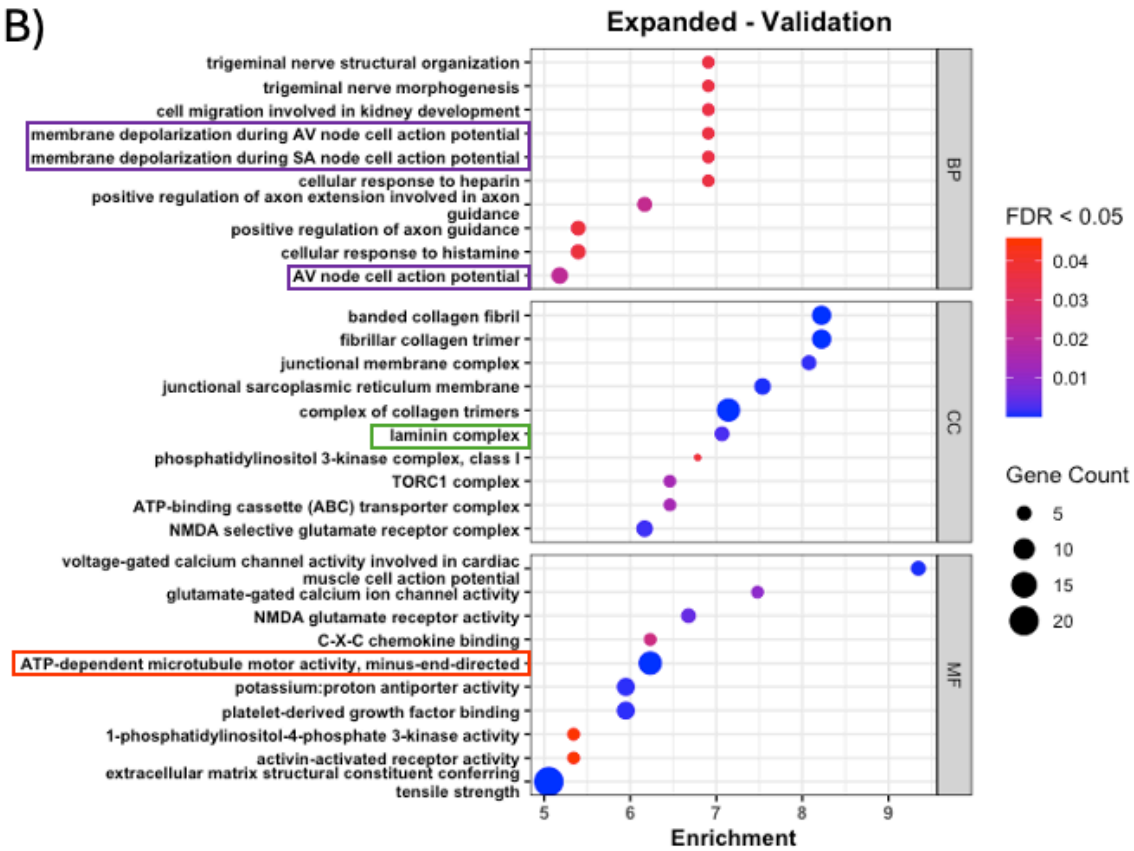
In contrast, the validation cohort showed a distinct set of molecular functions (figure 7B) with the highest enrichment observed in 'voltage-gated calcium channel activity involved in cardiac muscle cell action potential' (enrichment score=9.3). Additionally, 'glutamate-gated calcium ion channel activity' and 'NMDA glutamate receptor activity' were among the top enriched functions, emphasizing the importance of calcium ion dynamics in the validation cohort's biological context.

While there was some overlap in molecular functions such as 'ATP-dependent microtubule motor activity, minus-end-directed' between the two cohorts, the discovery cohort was characterized by a more diverse group of dynein-related activities, whereas the validation cohort was more focused on calcium ion channel activities and related functions.

Furthermore, both cohorts shared a general theme in the importance of ion channel activities, but with different specificities. The discovery cohort included 'transmembrane-ephrin receptor activity' and 'intracellular calcium activated chloride channel activity', while the validation cohort emphasized the 'glutamate-gated calcium ion channel activity' and 'NMDA glutamate receptor activity'.

It is also noteworthy that the validation cohort highlighted 'extracellular matrix structural constituent conferring tensile strength' with the highest number of associated genes (20), including multiple collagen genes, indicating a significant role of structural extracellular matrix components in GBM progression.


A)
Expanded - Discovery
SA node cell to atrial cardiac muscle cell signaling
SA node cell action potential
SA node cell to atrial cardiac muscle cell communication
regulation of synaptic vesicle clustering
synaptic vesicle clustering
coronary vasculature morphogenesis
cilium-dependent cell motility
cilium or flagellum-dependent cell motility
cardiac muscle cell action potential involved in contraction
regulation of heart rate by cardiac conduction
laminin complex
platelet dense tubular network
platelet dense tubular network membrane
axonemal dynein complex
axoneme part
myosin filament
muscle myosin complex
integrin complex
extracellular matrix component
M band
ATP-dependent microtubule motor activity, minus-end-directed
dynein light chain binding
dynein light intermediate chain binding
transmembrane-ephrin receptor activity
intracellular chloride channel activity
intracellular calcium activated chloride channel activity
dynein intermediate chain binding
ephrin receptor activity
glycolipid binding
transmembrane receptor protein tyrosine kinase activity
BP
CC
MF
4
6
8
10
Enrichment
FDR < 0.05
0.04
0.03
0.02
0.01
Gene Count
3
6
9
12
15


B)
Expanded - Validation
trigeminal nerve structural organization
trigeminal nerve morphogenesis
cell migration involved in kidney development
membrane depolarization during AV node cell action potential
membrane depolarization during SA node cell action potential
cellular response to heparin
positive regulation of axon extension involved in axon guidance
positive regulation of axon guidance
cellular response to histamine
AV node cell action potential
banded collagen fibril
fibrillar collagen trimer
junctional membrane complex
junctional sarcoplasmic reticulum membrane
complex of collagen trimers
laminin complex
phosphatidylinositol 3-kinase complex, class I
TORC1 complex
ATP-binding cassette (ABC) transporter complex
NMDA selective glutamate receptor complex
voltage-gated calcium channel activity involved in cardiac muscle cell action potential
glutamate-gated calcium ion channel activity
NMDA glutamate receptor activity
C-X-C chemokine binding
ATP-dependent microtubule motor activity, minus-end-directed
potassium:proton antiporter activity
platelet-derived growth factor binding
1-phosphatidylinositol-4-phosphate 3-kinase activity
activin-activated receptor activity
extracellular matrix structural constituent conferring tensile strength
BP
CC
MF
5
6
7
8
9
Enrichment
FDR < 0.05
0.04
0.03
0.02
0.01
Gene Count
5
10
15
20

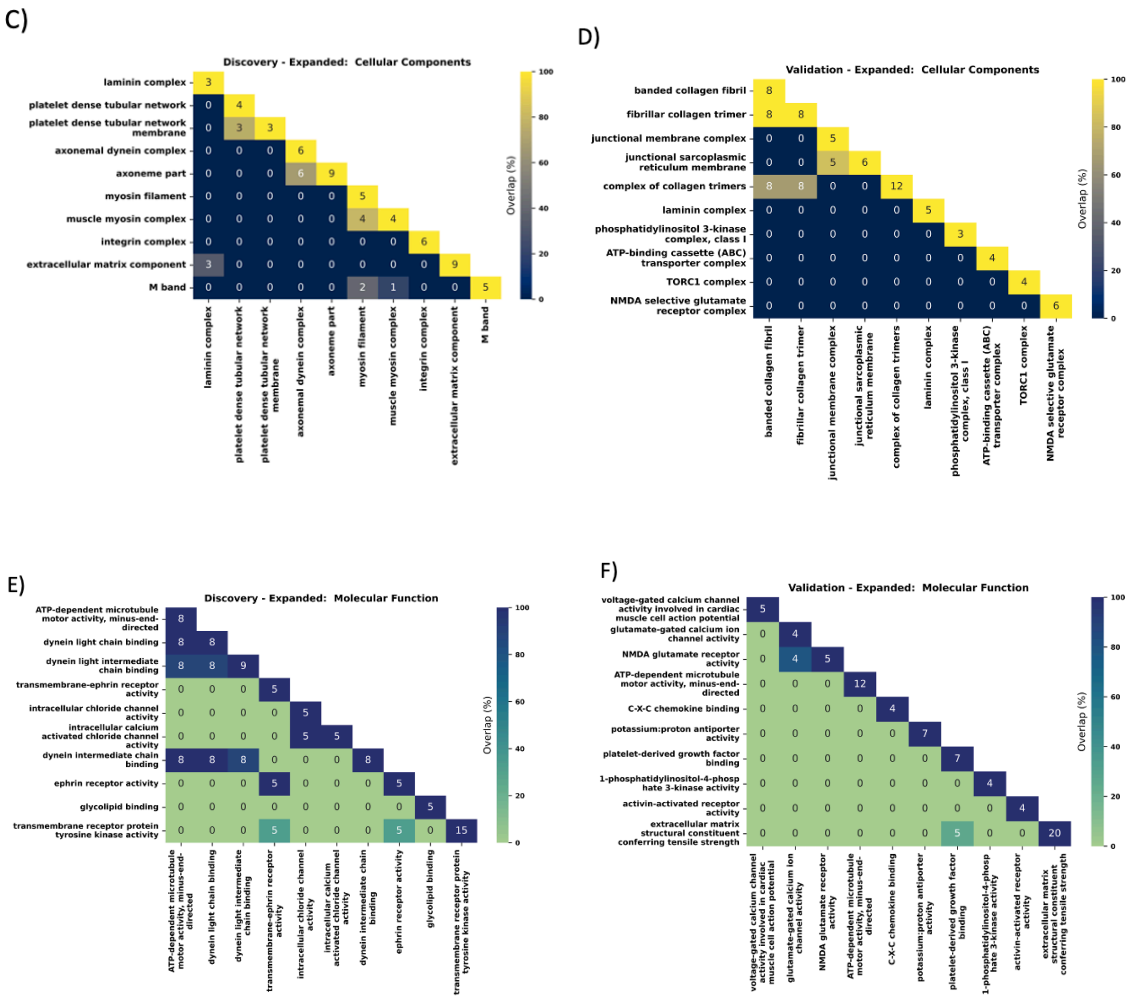


Figure 3-6: Functional enrichment analysis.

Combined Bubble plots illustrating the results of Gene Ontology (GO) function enrichment analysis. The three main categories of GO terms are represented: Biological processes (BP); Cellular components (CC); Molecular function (MF). The purple boxes show common biological processes, the green boxes show the common cellular component pathways and the red boxes shows common molecular function pathways between the cohorts. The y-axis shows pathway terms, while the x-axis denotes degree of enrichment. The size of the bubble determined by the number of genes associated with the term whereas the colour of the bubbles indicates the significance (adj *p* values). Top 10 altered pathways are visualized only for:

A: Expanded group of variants in the discovery cohort.

B: Expanded group of variants in the validation cohort.

C-D: Heatmap to show the relationship and the overlapping genes between the pathways of cellular component in the discovery (C) and the validation (D) cohorts.

E-F: Heatmap to show the relationship and the overlapping genes between the pathways of molecular function in the discovery (E) and the validation (F) cohorts.

3.3.2.2 Declining Variants

3.3.2.2.1 Biological processes

The analysis of the discovery cohort revealed several pathways significantly enriched in genes related to the regulation of neuronal structure and development (Figure 3-7A). The most prominent biological process identified was the "regulation of extent of cell growth," with an enrichment score

of 4.4. Similarly, the "positive regulation of cell morphogenesis involved in differentiation" and "regulation of axonogenesis" pathways, both integral to neuronal differentiation and axonal growth, displayed enrichment scores of 3.8 and 3.6, respectively. Notably, *BDNF* was a recurrent gene in both biological GO terms.

Processes associated with cell morphogenesis were repeatedly observed with high enrichment scores, indicating a concerted influence on cellular shape changes during differentiation, especially in the context of neuron development. "Axon development" was another significantly enriched process. Interestingly, cell morphogenesis and related processes such as "cell morphogenesis involved in neuron differentiation" were the most represented, implicating these biological processes as potential targets of selective pressure during therapy in the discovery cohort.

In contrast, the validation cohort presented a distinct set of biological processes, mostly related to cardiac function and calcium signalling (Figure 3-7B). The most significant pathway was "regulation of cardiac conduction," with an enrichment score of 4.5. Calcium-related pathways, such as "calcium ion transmembrane import into cytosol" and "calcium ion transport into cytosol," were also enriched, underscoring the essential role of calcium in signalling pathways. Moreover, pathways associated with "extracellular matrix organization" and "extracellular structure organization" had significant enrichment, suggesting that extracellular components play a crucial role in the cellular response to therapeutic pressure in the validation cohort.

Comparing the biological processes between the discovery and validation cohorts revealed distinct patterns of enriched pathways. The discovery cohort showed significant enrichment in pathways involved in neuronal development and morphogenesis, whereas the validation cohort pathways were predominantly related to calcium handling. This divergence may reflect the different selective pressures or cellular contexts between the two cohorts. The presence of genes such as *BDNF* in the discovery cohort and *ANK2* in the validation cohort highlights potential key players in the response to therapy within different biological processes.

3.3.2.2.2 Cellular components

Investigations into the discovery cohort revealed significant enrichment in various cellular components (Figure 3-7A). The mRNA cap binding complex exhibited the highest level of enrichment (13.67-fold). Another cellular feature, the "cell cortex region", showed a 7.40-fold enrichment. The “midbody” cellular component term showed 3.46-fold enrichment.

Additional cellular components such as the “cation channel complex, nuclear speck, and the cell cortex” were also significantly enriched, with scores 3.2, 3.0, 3.0, respectively. Membrane-associated structures like the "basolateral plasma membrane" and various transporter complexes showed enrichment values ranging from 2.85 to 3.02.

The validation cohort corroborated some of the discovery findings and also provided distinct insights (Figure 3-7B). Remarkably, the phosphatidylinositol 3-kinase complex, class I, presented a substantial 20.90-fold enrichment, comprising *PIK3CA, PIK3R1, and PIK3R6* genes. Collagen-associated structures such as the "fibrillar collagen trimer" and the "banded collagen fibril" both revealed an identical enrichment of 12.67-fold. Components of the sarcoplasmic reticulum membrane and collagen complexes demonstrated significant presence, with enrichment values of 11.61 and 11.00, respectively. Noteworthy was the voltage-gated sodium channel complex showing a 9.95-fold enrichment. The “extracellular matrix component” displayed a 7.82-fold enrichment.

3.3.2.2.3 Molecular function

The discovery cohort analysis uncovered a significant enrichment in molecular functions associated with various binding activities and ion transport (Figure 3-7A). Notably, "RNA 7-methylguanosine cap binding" showed the highest fold enrichment at 22.45. The "structural molecule activity conferring elasticity", presented a significant 15.15-fold enrichment, suggesting an important role in cellular structural integrity. The “RNA cap binding” activity was also notably enriched (11.22-fold), again implicating the mRNA processing machinery. Ion channel and transporter activities, including chloride, sodium, and various voltage-gated channels, displayed fold enrichments ranging from 3.08 to 4.08. These findings suggest an intricate network of ion homeostasis and signalling within the discovery cohort. Furthermore, "gated channel activity" was enriched 2.94, indicating a comprehensive role of these channels in cellular functions. The "Carbohydrate binding" activity was also observed to be enriched with a fold change of 2.87.

In the validation cohort (Figure 3-7B), "glutamate-gated calcium ion channel activity" was prominently enriched (16.50-fold), indicating a crucial role in neurotransmission. The "Structural molecules conferring elasticity" showed a 13.75-fold enrichment, highlighting its importance in cellular and extracellular structural stability.

Furthermore, the validation cohort confirmed significant enrichments in molecular functions related to sensory perception, with "melanocortin receptor activity" (13.75-fold) and "opioid receptor activity" (12.22-fold) being prominent. This suggests a potential role for these receptors in cellular signalling pathways related to these functions. Transporter activities for "aromatic amino acids transmembrane" and the "NMDA glutamate receptor activity" were also significantly enriched, pointing to essential functions in nutrient uptake and synaptic plasticity, respectively. Notably, the "extracellular matrix structural constituent conferring tensile strength" showed an 8.18-fold enrichment, involving a range of collagen genes, critical for maintaining the extracellular matrix's integrity.

The GO analysis of variants that became resistant to treatment across primary to recurrent tumours in the discovery and validation cohorts has provided insightful findings, particularly in identifying enriched biological processes, cellular components, and molecular functions specific to each cohort. For instance, the discovery cohort highlighted the significance of pathways related to cardiac and neuronal functions, such as SA node cell action potential and synaptic functions, which are crucial in understanding the cellular mechanisms underlying treatment resistance whereas the group of variants likely responded to treatment showed the regulation of cell growth and axonogenesis as a significant mechanism, with notable genes like *BDNF* recurrently implicated in these pathways.

However, while the GO analysis was informative, it had certain limitations. The inherent specificity and redundancy of GO terms can sometimes obscure broader biological insights, as many genes are categorized under multiple, overlapping terms. This can dilute the impact of distinct biological processes and complicate the extraction of clear, actionable insights. Moreover, GO analysis tends to focus on individual genes and their associated functions without providing a comprehensive view of how these functions interact within larger biological networks.

Given these limitations, we decided to move on to pathway analysis. Pathway analysis offers a more holistic view of cellular responses, allowing us to see how groups of genes interact within wider biological pathways and how these interactions change in response to treatment.

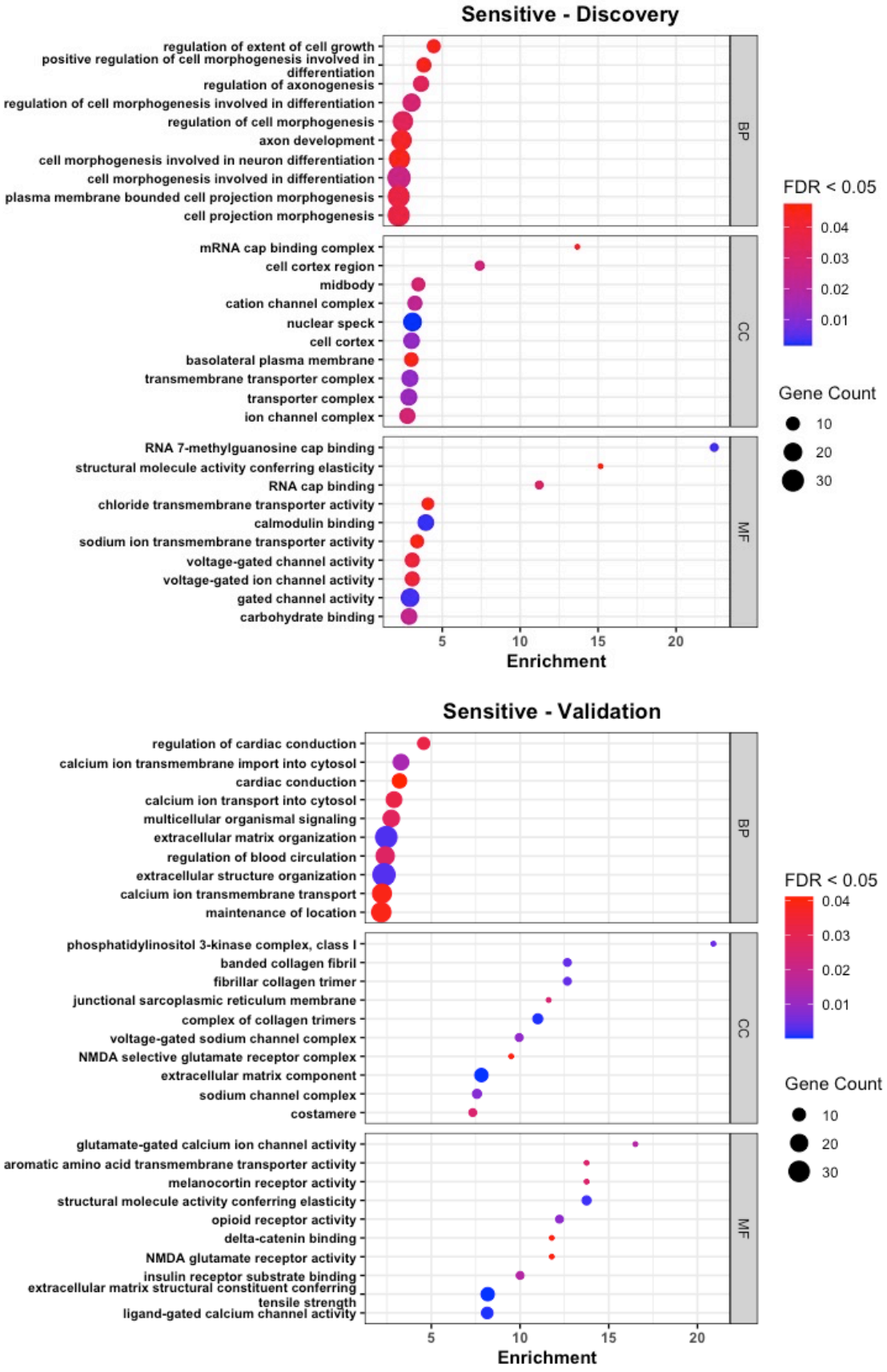


Figure 3-7: Functional enrichment analysis.

Combined Bubble plots illustrating the results of Gene Ontology (GO) function enrichment analysis. The three main categories of GO terms are represented: Biological processes (BP); Cellular components (CC); Molecular function (MF).

Green and Red boxes show common pathways between the cohorts. The y-axis shows pathway terms, while the x-axis denotes degree of enrichment. The size of the bubble determined by the number of genes associated with the term whereas the colour of the bubbles indicates the significance (adj *p* values). Top 10 altered pathways are visualized only for:
A: Sensitive group of variants in the discovery cohort.
B: Sensitive group of variants in the validation cohort.
Combined Bubble plots illustrating the results of Gene Ontology (GO) function enrichment analysis. The three main categories of GO terms are represented: Biological processes (BP); Cellular components (CC); Molecular function (MF). Green and Red boxes show common pathways between the cohorts. The y-axis shows pathway terms, while the x-axis denotes degree of enrichment. The size of the bubble determined by the number of genes associated with the term whereas the colour of the bubbles indicates the significance (adj *p* values). Top 10 altered pathways are visualized only for:
A: Sensitive group of variants in the discovery cohort.
B: Sensitive group of variants in the validation cohort.

3.3.3 Pathway enrichment analysis using PathScore

To determine which pathways may be involved in treatment resistance and vulnerability, I used PathScore (Gaffney and Townsend, 2016), a pathway-level analysis tool that integrates gene-level variant data. After submitting the gene lists (list of all variants) to the PathScore server, variants were processed and labelled as follows:

- **Loaded**: Variants used in the analysis, after filtering. These correspond to variants with valid hugo-entrez pairs for genes that are present in MSigDB.
- **Unused**: Variants with valid hugo-entrez pairs, but the corresponding genes are not present in MSigDB.
- **Rejected**: Variants with invalid hugo-entrez pairs.

PathScore utilized the variants labelled as 'Loaded' in Table 3-3 for pathway analysis. As the analysis progressed, the number of patients initially enrolled was adjusted based on the status of their variants. This led to reduction, or exclusion, of patients from each group indicating that some carried mutations in genes not present in MSigDB. Hence, the number of patients changes from one group to another, caused by their removal from the study.

Table 3-3: Summary of variants accepted for pathway analysis

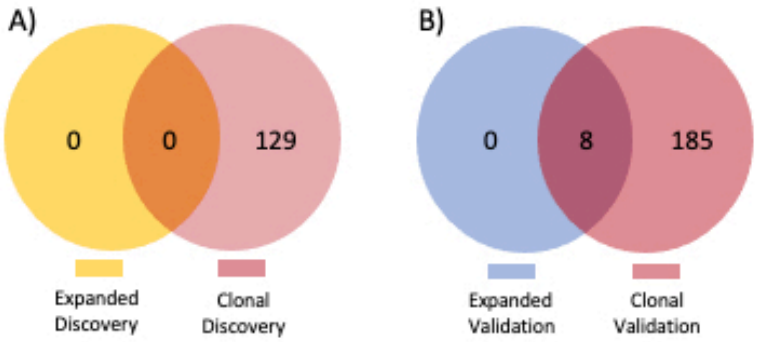
Variant grouping	Cohort	Number of patients	Loaded variants	Unused variants	Rejected variants
Low VAF in primary and recurrent	Discovery	32	4761	5070	199
	Validation	92	1410	1367	0
Declining VAF from primary to recurrent	Discovery	24	207	199	1
	Validation	75	435	381	0
Increasing VAF from primary to recurrent	Discovery	17	248	266	10
	Validation	64	402	389	0
High VAF in primary and recurrent	Discovery	31	265	208	4
	Validation	87	1022	901	1

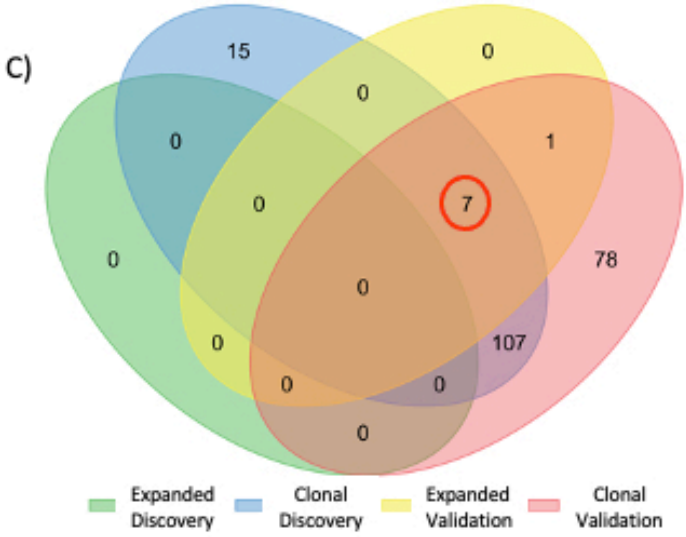
3.3.3.1 Clonally expanded variants

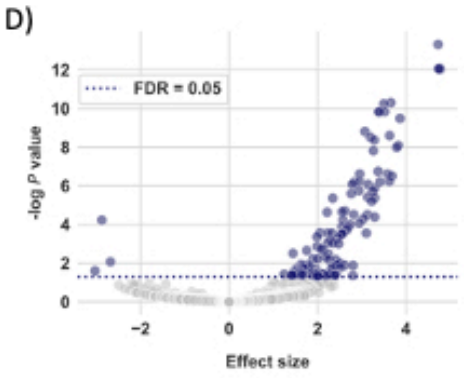
To identify disrupted pathways resulting from these variants, I ran PathScore using the common variants between primary and recurrent tumours that followed a clonal expansion pattern. I applied this analysis separately to my two cohorts, the discovery and validation datasets, treating them as independent sources to increase confidence in the results. After submitting gene lists from each

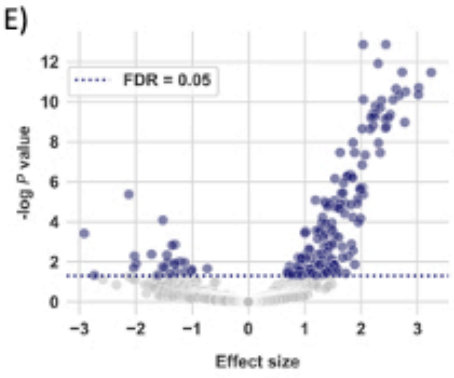
cohort to PathScore, I analysed the results separately and then compared them. Only the overlapping results, which demonstrated a high degree of concordance between the two cohorts, were considered for further analysis. This methodological approach was employed to mitigate potential sources of bias and enhance the robustness of the analysis outcomes. Initially, gene lists of expanded variants group in the discovery cohort did not show any significant pathways while the expanded group of variants in the validation cohort showed eight significant pathways. I then expanded the analysis by comparing the clonal group of variants. Because there is no universally accepted cutoff for distinguishing clonal from subclonal mutations, I defined a practical threshold based on the mean variant allele frequency (VAF). Variants with a VAF equal to or greater than the mean VAF of all variants within each tumour were considered clonal, representing the clonally dominant population. Using this classification, I identified 129 significantly altered pathways in the discovery cohort and 193 significantly altered pathways in the validation cohort. The discovery cohort is smaller than the validation cohort, and with PathScore removing many variants that are not present in the MSigDB, these factors contributed to the lack of significant pathways in the clonal group of expanded variants in the discovery cohort. Additionally, a biological explanation may also underlie this observation: if resistance pathways were already clonal in the discovery cohort, they would not be able to expand further. Therefore, I combined all cohorts to increase the likelihood of identifying significant pathways in the clonal expanded variants group, and seven common pathways were found across the clonal group of variants in the discovery cohort, as well as the expanded and clonal groups of variants in the validation cohort (Figure 3-8A-C). I validated these seven pathways across both cohorts, using them as representatives for the expanded variants in the discovery cohort (Figure 3-8).

From my analysis, I then focused on seven common pathways shared by the two distinct cohorts, summarised in table 3-3.


A)
0
0
129
Expanded Discovery
Clonal Discovery
B)
0
8
185
Expanded Validation
Clonal Validation


C)
15
0
0
0
1
0
7
0
78
0
0
107
0
0
0
0
Expanded Discovery
Clonal Discovery
Expanded Validation
Clonal Validation


D)
FDR = 0.05
-log P value
Effect size


E)
FDR = 0.05
-log P value
Effect size

F)

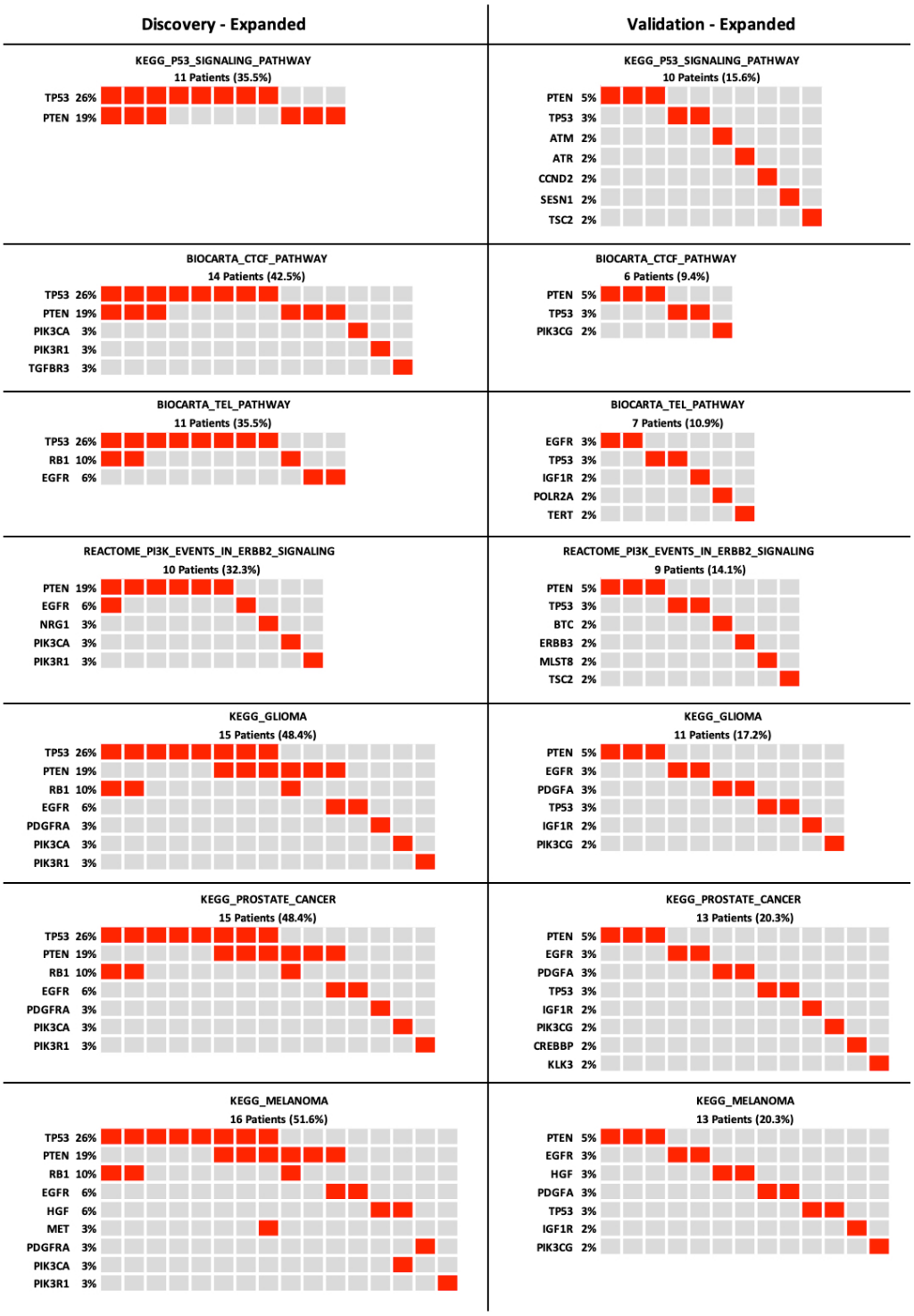


Figure 3-8: PathScore results for expanded variants.

A: Intersection between expanded and clonal variants of discovery cohort.

B: Intersection between expanded and clonal variants of validation cohort.

C: The common pathways across all group of variants.

D-E: Volcano plots of discovery *(D)* and validation *(E)* cohorts, indicating the altered pathways affected by variants that increased in prevalence through treatment or variants of unchanged variant allele frequency. The Effect Size (X-axis) is the PathScore metric that quantifies the magnitude of the pathway alteration (enrichment). It is calculated as the ratio of the estimated effective pathway size to the actual pathway size. The estimated effective pathway size is the maximum likelihood estimate derived from the observed gene harbouring VAF-increasing variants. This ratio measures the relative degree of overburden of the pathway, and is used to rank pathways by the strength of the

observed effect. The −logP-value (Y-axis) represents the statistical significance of the disparity between the actual and effective pathway sizes, calculated using the likelihood ratio test. Pathways that are considered significantly enriched are highlighted in blue based on a threshold of an adjusted P-value (FDR≤0.05).
F: Matrix plot of patient-gene pairs for the common 7 pathways. Each column is a patient, and each row is a gene. Only patients that have the pathway altered are plotted.

This section systematically delineates the involvement of seven crucial signalling pathways (table 3-4), as evidenced in both the discovery and validation cohorts, thereby deepening our understanding of the molecular mechanisms underlying glioblastoma multiforme (GBM) and suggesting potential therapeutic targets.

Table 3-4: Common pathways between expanded variants of discovery and validation cohorts

PATHWAY NAME	No. of genes	% affected discovery patients	% affected validation patients
KEGG_MELANOMA	12/71	51.6	20.3
KEGG_PROSTATE_CANCER	12/89	48.4	20.3
KEGG_P53_SIGNALING_PATHWAY	7/68	35.5	15.6
KEGG_GLIOMA	10/65	48.4	17.2
REACTOME_PI3K_EVENTS_IN_ERBB2_SIGNALING	9/42	32.3	14.1
BIOCARTA_CTCF_PATHWAY	6/23	45.2	9.4
BIOCARTA_TEL_PATHWAY	6/18	35.5	10.9

The exploration of the KEGG Melanoma pathway, which primarily focuses on *BRAF* mutations that enhance the *MAPK*/*ERK* signalling pathways, highlights the broader roles these pathways play across cancers, including GBM. Although *BRAF* mutations are rare in GBM (McNulty et al., 2021, Munjapara et al., 2022), our analysis identified variants in 12 other genes within the 71-gene KEGG Melanoma pathway, present in 51.6% of the discovery cohort and 20.3% of the validation cohort (table 3-4 and fig. 9F). Targeting specific mutations such as BRAF V600E is challenging in GBM due to its low prevalence, limiting the applicability of mutation-specific therapies. Therefore, therapeutic strategies aimed at modulating pathway activity as a whole, rather than focusing on individual mutations, may be more effective (Kaley et al., 2018).

Similarly, the KEGG Prostate Cancer pathway elucidates the importance of androgen receptor signalling and cell cycle control mechanisms that are pivotal for the progression of prostate cancer and potentially relevant for GBM (Zalcman et al., 2018). In my analysis, I identified variants in 12 out of 89 genes within this pathway, with 48.4% of patients in the discovery cohort and 20.3% in the

validation cohort carrying variants in these genes. While these genes are not exclusive to prostate cancer and are involved in multiple biological processes, their presence in GBM suggests that some pathway components may be shared across tumour types. This observation may help inform future therapeutic exploration, particularly in targeting cell cycle and signalling pathways.

The direct relevance of the KEGG Glioma pathway to GBM cannot be overstated, as it encapsulates the genetic alterations typical of gliomas such as mutations in *TP53*, *EGFR*, and *PTEN* (Barthel et al., 2019). The variants were identified in 10 out of 65 genes within this pathway, found in nearly half of patients in the discovery cohort and 17.2% of patients in the validation cohort. While these genes are well-established in glioma biology, the observed differences in frequency between cohorts may reflect underlying biological variation or technical factors such as sample size and filtering thresholds. Nevertheless, the presence of variants in canonical glioma genes supports the biological relevance of this pathway and its potential contribution to treatment response.

Turning to the KEGG P53 Signalling pathway, which is instrumental in regulating cell cycle, DNA repair, and apoptosis, the analysis revealed that 7 out of 68 genes harboured variants, influencing 35.5% of the discovery cohort and 15.6% of the validation cohort. The ubiquitous nature of p53 mutations across various cancers offers a compelling case for exploring therapeutic strategies aimed at restoring p53 function, which could also be applicable to GBM (Zhang et al., 2018).

The BIOCARTA_CTCF_PATHWAY, which focuses on the CCCTC-binding factor (CTCF), a pivotal transcriptional regulator involved in chromatin organization, revealed variants in 6 out of 23 genes (Sese et al., 2021). These variants impacted 45.2% of the discovery cohort and 9.4% of the validation cohort, highlighting the pathway's potential role in the transcriptional dysregulation observed in GBM (Liu et al., 2023).

The BIOCARTA_TEL_PATHWAY, dealing with telomere maintenance, critical for cellular longevity and immortality in cancer cells, showed changes affecting 35.5% and 10.9% of the discovery and validation cohorts, respectively. This pathway's alteration underpins the fundamental role of telomere dynamics in cancer progression and provides a basis for exploring telomerase as a therapeutic target in GBM (Langford et al., 1995, Lotsch et al., 2013, Diplas et al., 2018).

Finally, the analysis identified the Reactome 'PI3K events in ErbB2 signalling' pathway will be discussed in detail in the ErbB signalling pathway section later in this chapter.

The disparities in variant frequencies observed between the discovery and validation cohorts necessitate careful consideration of cohort bias and the inherent heterogeneity of GBM. Such factors are critical in interpreting the genetic data and ensuring the robustness of conclusions drawn from this study. Consequently, this analysis not only enhances our comprehension of the complex genetic architecture of GBM but also aligns with broader efforts to develop targeted therapeutic strategies based on the unique genetic profiles observed in these patients.

In summary, the pathway analysis highlights critical biological mechanisms and potential therapeutic targets within GBM. Addressing the variations observed between cohorts requires careful cohort selection and characterization, particularly given GBM's heterogeneity and complexity. This understanding is crucial for developing more effective therapies tailored to the genomic landscape of GBM.

3.3.3.2 Declining variants

The analysis of the declining group of variants identified 7 significant pathways in the discovery cohort and 220 significant pathways in the validation cohort (Figure 3-9). I then took the overlapping pathways to cross validate between the two cohorts. The pathways examined, number of genes affected, and the percentage of patients affected in both cohorts are summarized in Table 3-5.

Table 3-5: Common pathways between declining variants of discovery and validation cohorts

PATHWAY NAME	No. of genes	% affected discovery patients	% affected validation patients
BIOCARTA_CBL_PATHWAY	4/13	12.5	13.3
BIOCARTA_EGFR_SMRTE_PATHWAY	2/11	12.5	8
PID_ARF6_PATHWAY	5/35	20.8	9.3
PID_PTP1B_PATHWAY	8/52	20.8	14.7
REACTOME_SIGNALING_BY_ERBB4	8/87	20.8	21.3
REACTOME_GAB1_SIGNALSOME	4/36	16.7	20

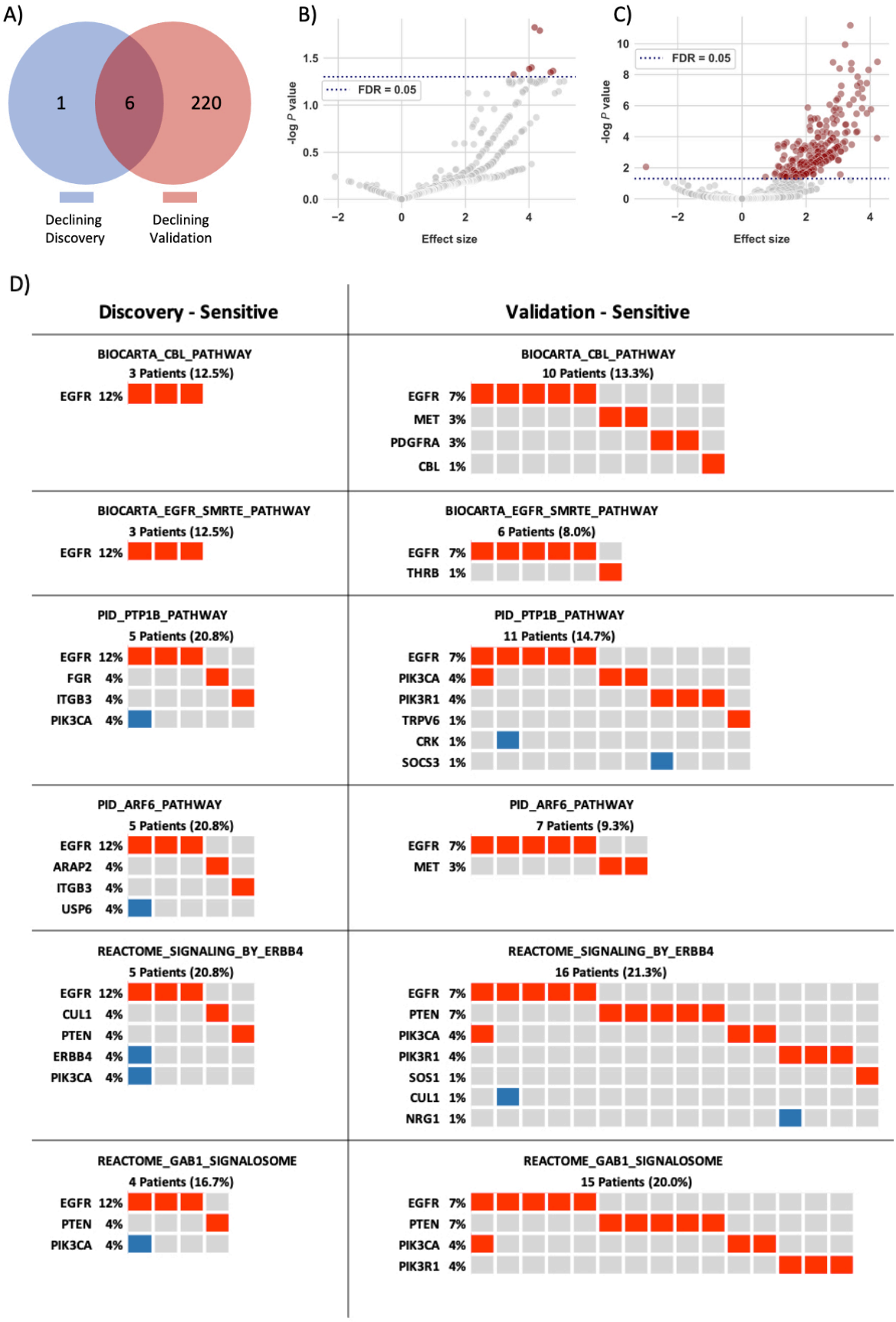


Figure 3-9: PathScore results for declined variants.

A: Overlapping pathways of declined variants between discovery and validation cohorts.

B-C: Volcano plots of discovery *(B)* and validation *(C)* cohorts, showing pathways impacted by variants that decreased in variant allele frequency through treatment. The Effect Size (X-axis) is the PathScore metric that quantifies the magnitude of the pathway alteration (enrichment). It is calculated as the ratio of the estimated effective pathway size to the actual pathway size. The estimated effective pathway size is the maximum likelihood estimate derived from the observed gene harbouring VAF-declining variants. This ratio measures the relative degree of overburden of the pathway, and is used to rank pathways by the strength of the observed effect. The −logP-value (Y-axis) represents the statistical significance of the disparity between the actual and effective pathway sizes, calculated using the likelihood ratio test. Pathways that are considered significantly enriched are highlighted in red based on a threshold of an adjusted P-value (FDR≤0.05).
D: Matrix plot of patient-gene pairs for the common 6 pathways. Each column is a patient, and each row is a gene. Only patients have the pathway altered are plotted. Red boxes are pathway genes that are mutated in one patient at least, where blue boxes indicate co-occurrence.

The comparative pathway analysis between the discovery and validation cohorts revealed several pathways with significant alterations, suggesting their potential roles in sensitizing glioblastoma cells to therapy. Below, I expand upon these findings to explore if and how these pathways could enhance the cancer cells' susceptibility to treatment, and whether targeting these pathways could be disadvantageous for the tumour when exposed to therapy.

The CBL (Casitas B-lineage Lymphoma) pathway involves CBL proteins which are a family of E3 ubiquitin ligases that regulate receptor tyrosine kinase (RTK) signalling through ubiquitination and degradation (Jing et al., 2016). Variations observed in this pathway (12.5% in discovery and 13.3% in validation) suggest its role in modulating growth factor signalling, which is crucial for tumour survival and proliferation. In the literature, it has been established that disruption of CBL function can lead to impaired degradation of RTKs, resulting in altered signalling dynamics that could make tumour cells more vulnerable to therapy. This aligns with our findings, indicating that disrupted BIOCARTA_CBL_PATHWAY enhances therapeutic sensitivity by reducing oncogenic signalling and increasing tumour cell sensitivity to treatments.

Alterations in the EGFR_SMRTE (Silencing Mediator for Retinoid and Thyroid hormone Receptors) pathway (12.5% in discovery and 8% in validation) highlight its significance in glioblastoma. EGFR is often overexpressed or mutated in glioblastoma, leading to enhanced cell proliferation and survival. Targeting this pathway with EGFR inhibitors TKIs or monoclonal antibodies could disrupt these processes (Darré et al., 2024). Previous studies suggest that combining EGFR inhibitors with other treatments can improve efficacy and overcome resistance mechanisms (Chong and Janne, 2013). Our findings support this, as the altered EGFR_SMRTE pathway may represent a vulnerability that can be therapeutically exploited.

The PTP1B (Protein Tyrosine Phosphatase 1B) pathway, with a higher perturbation rate in discovery (20.8%) compared to validation (14.7%), involves Protein Tyrosine Phosphatase 1B, which negatively regulates insulin and leptin signalling. Literature suggests that inhibiting PTP1B can reduce tumour growth and improve sensitivity to chemotherapy, highlighting its potential as a therapeutic target (Bartolome et al., 2020). Our data align with these reports, indicating that targeting the disrupted PID_PTP1B_PATHWAY could enhance glioblastoma cell susceptibility to treatment and reduce therapeutic resistance.

The ARF6 (ADP-Ribosylation Factor 6) pathway is involved in actin cytoskeleton remodelling and membrane trafficking. Alterations observed (20.8% in discovery vs. 9.3% in validation) indicate its role in cancer cell invasion and metastasis. Studies have shown that inhibiting ARF6 can reduce these processes, potentially decreasing the invasive capacity of glioblastoma cells and making them more susceptible to conventional therapies (Yamauchi et al., 2017, Miao et al., 2012). This is consistent with our findings, suggesting that disruption of the ARF6 pathway can enhance the effectiveness of cancer treatments.

The GAB1 (GRB2-Associated-Binding Protein 1) signalsome pathway, with higher alterations in validation (20%) compared to discovery (16.7%), plays a role in signal transduction downstream of RTKs. *Gab1* acts as a docking platform for various signalling molecules, mediating pathways that promote cell survival and proliferation. Disruption of Gab1 signalling could impair these survival pathways, sensitizing glioblastoma cells to apoptosis and enhancing the efficacy of treatments. This aligns with our observations, suggesting that targeting the disrupted GAB1 signalsome pathway could make tumour cells more susceptible to therapy.

Finally, the ErbB4 signalling pathway showed consistent alterations (20.8% in discovery and 21.3% in validation), suggesting its pivotal role in glioblastoma. ErbB4 will be further discussed in the next section that explain in detail the ErbB signalling pathway.

In conclusion, analysis of somatic mutations in primary and recurrent tumours highlights key biochemical pathways associated with declining variants, suggesting potential vulnerabilities that could be exploited therapeutically. Pathways such as *CBL, EGFR, PTP1B, ARF6, and GAB1* were commonly altered in variants that decreased in frequency following treatment, indicating a possible role in sensitizing glioblastoma cells to therapy. These shared insights support the development of pathway-targeted strategies aimed at enhancing treatment response, rather than overcoming

resistance alone. Continued research and clinical validation are essential to translate these findings into effective therapies that improve outcomes for glioblastoma patients.

3.3.3.3 ErbB Signalling Pathways

The pathway analysis revealed two disrupted pathways associated with *ERBB* signalling. The REACTOME_PI3K_EVENTS_IN_ERBB2_SIGNALING pathway was favoured by cancer cells during therapy, contributing to the development of resistance while REACTOME_SIGNALING_BY_ERBB4 pathway was found to be favoured by the treatment, sensitizing cancer cells when exposed to therapy.

ErbB signalling is a critical pathway in the regulation of cellular processes, including proliferation, differentiation, migration, and survival. The ErbB family of receptor tyrosine kinases (RTKs) comprises four members: *EGFR (ErbB1), ErbB2 (HER2/neu), ErbB3,* and *ErbB4*. These receptors are activated by ligand binding, leading to receptor dimerization and autophosphorylation. This, in turn, triggers downstream signalling cascades like the *PI3K/AKT* and *RAS/RAF/MEK/ERK* pathways. Dysregulation of *ErbB* signalling is implicated in various cancers, making it a focal point for targeted cancer therapies(Hynes and Lane, 2005). GBM IDH wild-type (IDHwt) frequently exhibits *EGFR* amplification or mutation, leading to aberrant activation of *ErbB* signalling pathways, which drive tumour growth and survival (Brennan et al., 2013, Mellinghoff et al., 2005).

ErbB2, particularly, is known for its role in various cancers, including glioblastoma. It lacks a direct ligand and instead is activated through heterodimerization with other *ErbB* family members. Overexpression or amplification of *ErbB2* is associated with aggressive tumour behaviour and poor prognosis. Consequently, targeting *ErbB2* with monoclonal antibodies (e.g., trastuzumab) or small molecule inhibitors has become a cornerstone in treating ErbB2-positive cancers (Moasser, 2007).

ErbB4, a less studied but emerging player in cancer biology, including GBM, presents a fascinating complexity. It can undergo proteolytic cleavage, releasing an intracellular domain that can translocate to the nucleus and influence gene expression. The role of *ErbB4* in cancer can vary significantly depending on tumour type, with both tumour-promoting and tumour-suppressing functions reported (Lucas et al., 2022). In GBM, *ErbB4* expression has been correlated with more differentiated tumour phenotypes and better prognosis, suggesting a potential tumour-suppressive role (Donoghue et al., 2018).

The REACTOME_PI3K_EVENTS_IN_ERBB2_SIGNALING pathway was disrupted in 32% of the discovery cohort and 14% of the validation cohort. The PI3K/AKT pathway, often activated by *ERBB2* (*HER2*) signalling, is a well-known mediator of cell growth and survival. Aberrant activation of this pathway can confer resistance to therapies by promoting survival signals that counteract the effects of treatment (Rascio et al., 2021).

In the REACTOME_PI3K_EVENTS_IN_ERBB2_SIGNALING pathway, I focused on deleterious mutations to explore their potential functional impact within critical regions of the genes, which provided insights into their role in tumour progression. By examining these mutations, I identified whether they clustered in key functional domains that could affect gene activity and contribute to treatment resistance. For a detailed view of where these mutations occurred within the gene structures, refer to Figure 3-10, which maps each variant to its corresponding position along the genes. Mutations in *EGFR* were primarily observed in the furin-like domain such as the A289V variant, and the kinase tyrosine domain, potentially altering receptor activation and signal transduction (Miyashita et al., 2020). For *PTEN*, mutations were spread across the gene and mainly in the PTEN_C2 domain. Other genes involved in this pathway had one variant per gene such as *NRG1* which had a splice site mutation. The *PIK3CA* pathway showed a single mutation identified as a start loss variant, while *PIK3R1* had a mutation in the SH2 domain.

One limitation of this approach is that variant effect prediction tools such as SIFT, and PolyPhen-2 tend to prioritise loss-of-function mutations, and may underestimate the impact of gain-of-function missense mutations particularly relevant for oncogenes like *PIK3CA*, and *mTOR*, where activating mutations are known to drive cancer. As a result, some potentially important variants may have been missed by the filtering strategy. Nevertheless, the domain-specific variation patterns observed in the retained variants suggest distinct mechanisms of disruption that may still have therapeutic relevance.

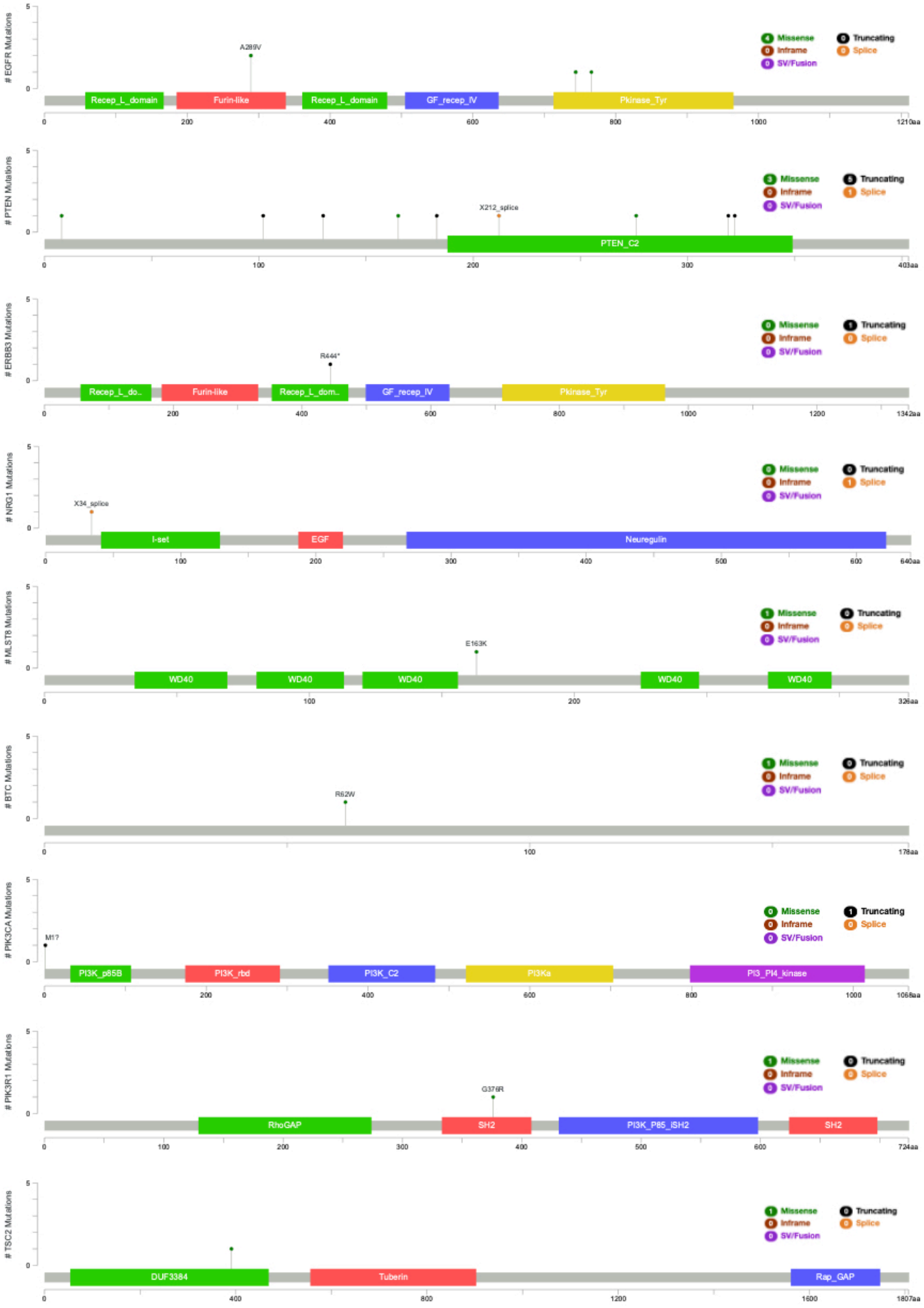

EGFR Mutations
A289V
Recep_L_domain
Furin-like
Recep_L_domain
GF_recep_IV
Pkinase_Tyr
1210aa
PTEN Mutations
X212_splice
PTEN_C2
403aa
ERBB3 Mutations
R444*
Recep_L_do..
Furin-like
Recep_L_dom..
GF_recep_IV
Pkinase_Tyr
1342aa
NRG1 Mutations
X34_splice
I-set
EGF
Neuregulin
640aa
MLST8 Mutations
E163K
WD40
WD40
WD40
WD40
WD40
326aa
BTC Mutations
R62W
178aa
PIK3CA Mutations
M1?
PI3K_p85B
PI3K_rbd
PI3K_C2
PI3Ka
PI3_PI4_kinase
1068aa
PIK3R1 Mutations
G376R
RhoGAP
SH2
PI3K_P85_iSH2
SH2
724aa
TSC2 Mutations
DUF3384
Tuberin
Rap_GAP
1807aa
Missense
Truncating
Inframe
Splice
SV/Fusion

Figure 3-10: Variants associated with REACTOME_PI3K_EVENTS_IN_ERBB2_SIGNALING.

This lollipop plot, generated by MutationMapper (https://www.cbioportal.org/mutation_mapper), visualizes the somatic variants found in the protein structure.

X-axis: Represents the amino acid length of the protein. **Coloured Blocks**: Denote known functional protein domains.

Lollipops: Represent specific amino acid substitution variants:

- The height of the lollipop stem indicates the frequency (number of samples/patients) in which that specific mutation was observed.
- Only the most recurrent or significant hotspot mutations are explicitly labelled with their protein change, unlabelled lollipops represent less frequent variants.

Lollipop Color: Black for Truncating**, Green** for Missense**, Purple** for Structural Variant / Fusion (SV/Fusion)**, Brown** for In-frame Deletion/Insertion and **Gold** for Splice Site Mutation.

Additionally, the REACTOME_SIGNALING_BY_ERBB4 pathway was disrupted in 21% of the patients across both the discovery and validation cohorts. ERBB4, along with its ligands and downstream signalling molecules, can activate pathways such as PI3K/AKT and MAPK, which are crucial for cell survival and proliferation. However, when ERBB4 signalling is altered, it may enhance the efficacy of certain therapies by increasing cellular susceptibility to treatment-induced apoptosis.

In this pathway, mutations in EGFR were found not only in the furin-like and kinase tyrosine domains but also in the receptor-ligand domain, suggesting broader impacts on receptor function. For PTEN, the PTEN_C2 domain was notably free from variants, indicating a more localized impact on its lipid-binding role. NRG1 mutations were located in the EGF domain, hinting at different effects on signalling mechanisms. Regarding PIK3CA, mutations were detected in the PI3K_p85B, PI3K_C2, and PI3Ka domains, indicating broader impacts on the PI3K/AKT pathway. Given that PI3K proteins function primarily as kinases, these mutations may also influence their phosphorylation activity, potentially leading to dysregulated downstream signaling (Roger Belizaire, 2021). PIK3R1 mutations were found in the PI3K_P85_iSH2 domain, potentially affecting PI3K interactions. The variant-gene maps to protein domains are illustrated in Figure 3-11.

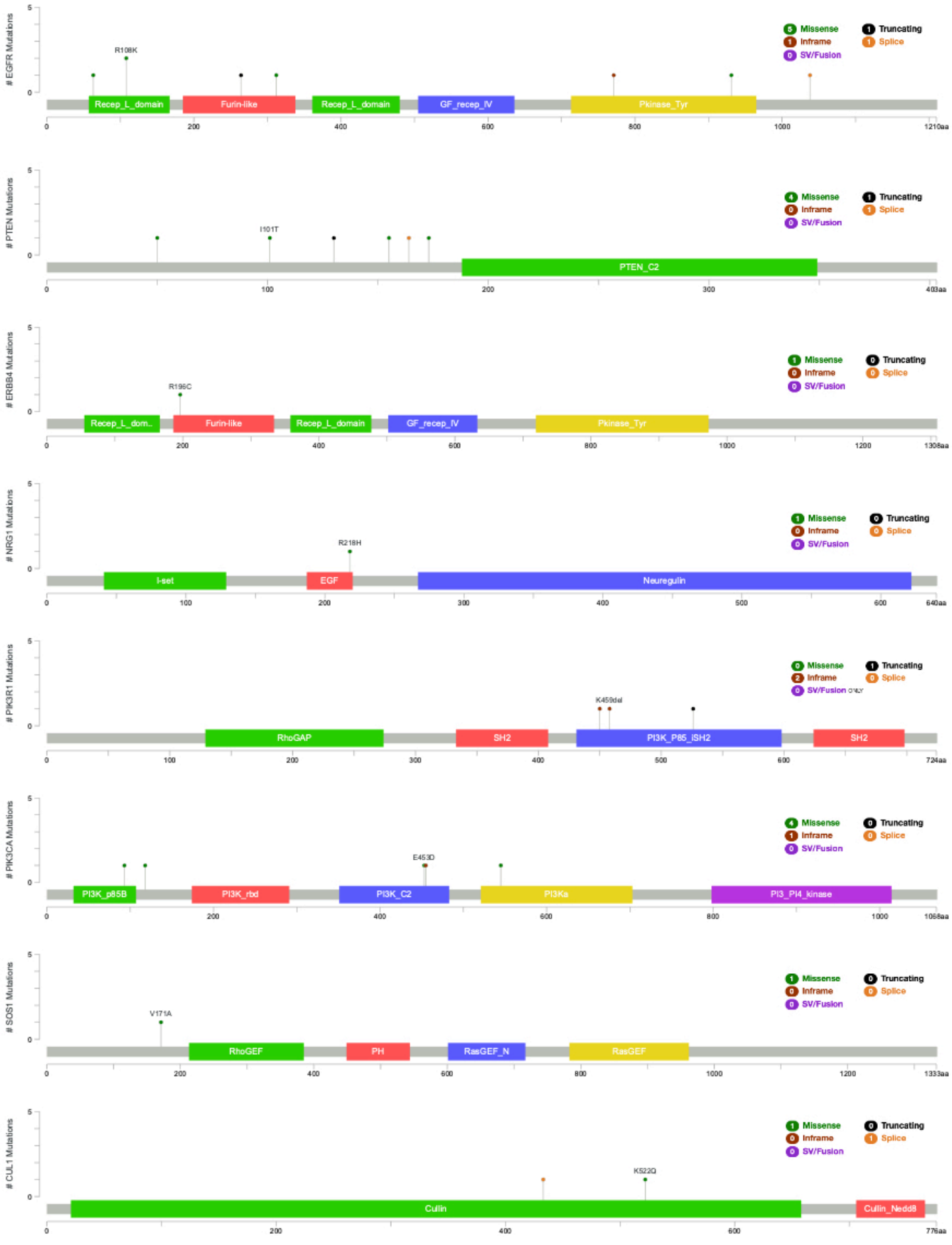

EGFR Mutations
R108K
Recep_L_domain
Furin-like
Recep_L_domain
GF_recep_IV
Pkinase_Tyr
5 Missense
1 Truncating
1 Inframe
1 Splice
0 SV/Fusion
0
200
400
600
800
1000
1210aa
PTEN Mutations
I101T
PTEN_C2
4 Missense
1 Truncating
0 Inframe
1 Splice
0 SV/Fusion
100
200
300
403aa
ERBB4 Mutations
R195C
Recep_L_dom..
Furin-like
Recep_L_domain
GF_recep_IV
Pkinase_Tyr
1 Missense
0 Truncating
0 Inframe
0 Splice
0 SV/Fusion
1000
1200
1308aa
NRG1 Mutations
R218H
I-set
EGF
Neuregulin
1 Missense
0 Truncating
0 Inframe
0 Splice
0 SV/Fusion
100
200
300
400
500
600
640aa
PIK3R1 Mutations
K459del
RhoGAP
SH2
PI3K_P85_iSH2
SH2
0 Missense
1 Truncating
2 Inframe
0 Splice
0 SV/Fusion
100
200
300
400
500
600
724aa
PIK3CA Mutations
E453D
PI3K_p85B
PI3K_rbd
PI3K_C2
PI3Ka
PI3_PI4_kinase
4 Missense
0 Truncating
1 Inframe
0 Splice
0 SV/Fusion
200
400
600
800
1000
1068aa
SOS1 Mutations
V171A
RhoGEF
PH
RasGEF_N
RasGEF
1 Missense
0 Truncating
0 Inframe
0 Splice
0 SV/Fusion
200
400
600
800
1000
1200
1333aa
CUL1 Mutations
K522Q
Cullin
Cullin_Nedd8
1 Missense
0 Truncating
0 Inframe
1 Splice
0 SV/Fusion
200
400
600
776aa

Figure 3-11: Variants associated with REACTOME_SIGNALING_BY_ERBB4

This lollipop plot, generated by MutationMapper (https://www.cbioportal.org/mutation_mapper), visualizes the somatic variants found in the protein structure.

X-axis: Represents the amino acid length of the protein. **Coloured Blocks**: Denote known functional protein domains.

Lollipops: Represent specific amino acid substitution variants:

- The height of the lollipop stem indicates the frequency (number of samples/patients) in which that specific mutation was observed.
- Only the most recurrent or significant hotspot mutations are explicitly labelled with their protein change, unlabelled lollipops represent less frequent variants.

Lollipop Color: Black for Truncating**, Green** for Missense**, Purple** for Structural Variant / Fusion (SV/Fusion)**, Brown** for In-frame Deletion/Insertion and **Gold** for Splice Site Mutation.

These findings suggest that while some ERBB pathway disruptions may enhance treatment sensitivity, others contribute to treatment resistance, highlighting the dual role of ERBB signalling in cancer therapy outcomes. Furthermore, the findings highlight the importance of pathway-specific genetic alterations in predicting and optimising cancer treatment responses. The variability in mutational landscapes across patients further underscores the need for personalized therapeutic strategies, rather than uniform treatment approaches, to ensure patients receive the most effective interventions for their individual GBM profiles.

3.3.4 Copy Number Results

3.3.4.1 Pathways corrected for copy number

In this chapter, I highlight the advantages of Cancer Cell Fraction (CCF) over Variant Allele Frequency (VAF) in defining the clonal architecture of tumours. CCF estimates the proportion of cancer cells harbouring specific genetic alterations, offering a clearer view of tumour heterogeneity than VAF, which is confounded by factors like copy number variations and tumour purity. VAF's limitation is particularly evident in regions with variable copy numbers, where it fails to accurately reflect clonal status. In regions of copy number stability, VAF and CCF are more closely aligned, making VAF a more reliable proxy for clonal fraction. However, in regions with copy number gains or losses, VAF can misrepresent the proportion of cancer cells with the variant.

Given the requirement for high-depth sequencing data (around 200X coverage) to reliably estimate CCF (Tanner et al., 2021), our dataset's lower coverage necessitated alternative approaches. I compared copy number profiles of primary and recurrent tumours, using primary samples as a baseline to track changes post-treatment. Traditional copy number callers were unsuitable due to requiring normal samples for this kind of assessment, so copy number data were converted into BED

files to assess changes at specific loci, allowing for a detailed analysis of copy number alterations and the filtering of variants in regions with variable copy numbers.

After filtering out variants from regions with variable CNAs in the selected-for group of variants, the reanalysis of significant pathways revealed that all pathways, except the REACTOME_PI3K_EVENTS_IN_ERBB2_SIGNALING pathway, were detected in both cohorts. This pathway was absent in the discovery cohort but present in the validation cohort. The discrepancy arises from the pathscore's dependence on patient numbers for pathway significance; removing variants from regions with variable CNAs effectively reduces the cohort size, as shown in table 3-6 where the CNA P>R column illustrates the status of copy number of each variant.

Applying the same protocol to the group of genes selected-against variants, it was found that the same pathways were deemed significant in the validation cohort, whereas the discovery cohort showed no significant pathways. This disparity is again likely due to the smaller size of the discovery cohort, which diminishes its statistical power to detect significant pathways. However, REACTOME_SIGNALING_BY_ERBB4 pathway persisted in the validation cohort, demonstrating that its significance is not solely linked to regions affected by CNAs as indicated in table 3-7.

The exclusion of variants in regions with variable CNAs was intended to refine the analysis, as VAF alone does not account for the complexities introduced by CNAs. This filtering improved pathway analysis by focusing on regions with stable copy numbers, where VAF more accurately reflects the mutation's clonal fraction. While variable CNAs can obscure certain pathways, both ERBB signalling pathways remained significant in the validation cohort, demonstrating robustness to the filtering process. This indicates that the presence of ERBB signalling is not merely an artifact of VAF misinterpretation due to CNAs, but rather reflects genuine biological relevance. Therefore, the analysis confirms that both ERBB signalling pathways are not redundant and remain pathways of interest, even after accounting for copy number variability.

Overall, these findings demonstrate that the identified pathways remain robust and significant after accounting for copy number changes, highlighting the importance of adequate sample sizes in genomic studies to capture critical biological insights. Refining pathway analysis to consider copy number stability provides a more accurate framework for understanding tumour genetics and potential therapeutic targets.

Table 3-6: REACTOME_PI3K_EVENTS_IN_ERBB2_SIGNALING

Chr	Position	VAF	CNA	Variant effect	Protein	AA change
		P>R	P>R			
chr4	75681166	0.121 > 0.463	2 > 2	MISSENSE	BTC	R62W
chr7	55154129	0.923 > 0.88	87 > 86	MISSENSE	EGFR	A289V
chr7	55154129	0.209 > 0.196	2 > 2	MISSENSE	EGFR	A289V
chr7	55249000	0.0004094 > 0.294	132 > 3	MISSENSE	EGFR	M766I
chr7	55242461	0.0005139 > 0.345	2 > 2	MISSENSE	EGFR	I744T
chr12	56487184	0.024 > 0.211	2 > 2	NONSENSE	ERBB3	R444*
chr16	2257260	0.183 > 0.355	2 > 2	MISSENSE	MLST8	E163K
chr8	32595826	0.129 > 0.181	2 > 2	SPLICE_SITE	NRG1	X34_splice
chr3	179198826	0.444 > 0.444	2 > 2	start_lost	PIK3CA	M1?
chr5	68293310	0.196 > 0.334	6 > 5	MISSENSE	PIK3R1	G376R
chr10	87864492	0.282 > 0.77	1 > 2	MISSENSE	PTEN	I8S
chr10	87952260	0.424 > 0.889	2 > 2	SPLICE_SITE	PTEN	X212_splice
chr10	87933147	0.497 > 0.25	0 > 1	NONSENSE	PTEN	R130*
chr10	87933063	0.52 > 0.753	2 > 2	FRAME_SHIFT	PTEN	K102Nfs*11
chr10	87961056	0.609 > 0.486	2 > 2	NONSENSE	PTEN	K322*
chr10	87961042	0.432 > 0.467	1 > 2	FRAME_SHIFT	PTEN	T319*
chr10	89711929	0.174 > 0.305	2 > 2	FRAME_SHIFT	PTEN	K183Tfs*5
chr10	89720676	0.061 > 0.349	2 > 2	MISSENSE	PTEN	N276S
chr10	89711875	0.075 > 0.303	2 > 2	SPLICE_SITE	PTEN	G165R
chr16	2111923	0.002543 > 0.29	2 > 2	MISSENSE	TSC2	V391M

Table 3-7: REACTOME_SIGNALING_BY_ERBB4

Chr	Position	VAF	CNA	Variant effect	Protein	AA change
		P>R	P>R			
chr7	148786549	0.229 > 0.117	5 > 4	SPLICE_SITE	CUL1	X433_splice
chr7	148485733	0.341 > 0.078	4 > 3	MISSENSE	CUL1	K522Q
chr7	55143387	0.633 > 0.063	50 > 5	MISSENSE	EGFR	R108K
chr7	55143387	0.834 > 0.003232	2 > 2	MISSENSE	EGFR	R108K
chr7	55154055	0.236 > 0.032	2 > 2	FRAME_SHIFT	EGFR	P266Hfs*14
chr7	55210077	0.85 > 0.0003778	49 > 64	MISSENSE	EGFR	G63R
chr7	55269049	0.9 > 0.002786	2 > 2	SPLICE_SITE	EGFR	
chr7	55249010	0.218 > 0.179	3 > 2	IN_FRAME_INS	EGFR	N771_H773dup
chr7	55266500	0.292 > 0.124	4 > 3	MISSENSE	EGFR	E931G
chr7	55223567	0.948 > 0.0004922	2 > 2	MISSENSE	EGFR	G312W
chr2	211750675	0.323 > 0.03	2 > 2	MISSENSE	ERBB4	R196C
chr8	31498153	0.358 > 0.029	2 > 2	MISSENSE	NRG1	R218H
chr3	179199102	0.344 > 0.052	2 > 2	MISSENSE	PIK3CA	R93W
chr3	178917478	0.594 > 0.18	2 > 2	SPLICE_SITE	PIK3CA	G118D
chr3	178936091	0.219 > 0.109	2 > 2	MISSENSE	PIK3CA	E545K
chr3	178928081	0.247 > 0.151	2 > 2	MISSENSE	PIK3CA	E453D
chr3	178928086	0.222 > 0.12	2 > 2	IN_FRAME_DEL	PIK3CA	L455_G460delinsF

chr5	67590984	0.328 > 0.156	2 > 2	FRAME_SHIFT	PIK3R1	H526QfsTer6
chr5	67589609	0.389 > 0.044	2 > 2	IN_FRAME_DEL	PIK3R1	K459del
chr5	67589585	0.268 > 0.152	2 > 2	IN_FRAME_DEL	PIK3R1	E451_Y452del
chr10	87933223	0.268 > 0.031	2 > 2	MISSENSE	PTEN	Y155S
chr10	89653851	0.26 > 0.197	2 > 2	MISSENSE	PTEN	I50T
chr10	89692904	0.74 > 0.146	2 > 2	NONSENSE	PTEN	R130*
chr10	89692818	0.412 > 0.125	2 > 2	MISSENSE	PTEN	I101T
chr10	89711899	0.471 > 0.124	2 > 2	MISSENSE	PTEN	R173C
chr10	89693009	0.233 > 0.192	2 > 2	SPLICE_SITE	PTEN	
chr2	39281963	0.413 > 0.147	1 > 3	MISSENSE	SOS1	V171A

3.3.4.2 Copy Number Changes from Primary to Recurrent

To identify the copy number changes through treatment, I used the primary tumour profiles as baseline and looked at the recurrent tumour profiles. The strategy was to track copy number changes at the subclonal level by identifying regions that were selected for or selected against during therapy. For the discovery cohort, I identified significant deletions at loci 7p11.2 and 15q11.2 (Figure 3-12A) however no amplifications were identified which could be attributed to the small sample size of the discovery cohort. In comparison with the validation cohort, the same deletion loci were also identified (Figure 3-12B). Investigating the genes involved in these loci, the *EGFR* on 7p11.2 was deleted in both cohorts whereas 15q11.2 had deletions of *OR4M2*. Looking at the validation cohort as it is larger than the discovery cohort, more significant regions have been detected. Interestingly, same regions that were deleted were also amplified (Figure 3-12B), however, those regions were

distinct and not shared by same patients. *OR4M2*, and *EGFR* are among the genes that were amplified or deleted in the validation cohort.

EGFR is commonly amplified in glioblastoma (GBM), with deletions and amplifications observed across multiple patients, which aligns with studies associating this phenomenon with extrachromosomal DNA (ecDNA/eccDNA). Extrachromosomal DNA refers to circular DNA fragments that exist independently of the main chromosomal DNA within cells. EcDNA are large molecules, spanning millions of base pairs, and frequently contain entire or partial oncogenes, regulatory elements, and other sequences critical for cell proliferation and survival. In cancers like GBM, ecDNA carrying amplified *EGFR* and other oncogenes contribute to tumour heterogeneity, driving rapid progression and resistance to standard therapies (Liao et al., 2020, Noer et al., 2022, Yang et al., 2023, Zhao et al., 2022, Verhaak et al., 2019).
EGFR amplification occurs in about 50% of glioblastomas (Lassman et al., 2019), typically within small circular ecDNA fragments. A frequent mutation associated with this amplification is *EGFRvIII*, which involves an in-frame deletion of exons 2-7 and is found in approximately 50% of EGFR-amplified GBM patients (Lassman et al., 2019, French et al., 2019, Gan et al., 2013, Hoogstrate et al., 2022). *EGFRvIII* is a variant of *EGFR* that remains constitutively active at low levels without the need for ligand binding, likely due to the partial deletion of the extracellular ligand-binding domain. This mutation arises from a genomic deletion rather than alternative or aberrant splicing. *EGFRvIII* is generally considered a subclonal event, arising after chromosome 7 and EGFR amplifications (French et al., 2019). Although subclonal, *EGFRvIII* significantly impacts tumour biology by further enhancing genetic variability and complicating targeted therapeutic approaches.

OR4M2 is part of the olfactory receptor proteins, which are members of the G-protein-coupled receptors (Malnic et al., 2004). G-protein-coupled receptors (GPCRs) are a large family of proteins that detect molecules outside the cell and activate internal signal transduction pathways and cellular responses (Rosenbaum et al., 2009). Although GPCRs have been studied as therapeutic targets for GBM (Stephan et al., 2021, Byrne et al., 2021), there is currently no study specifically associating olfactory receptors with GBM.

Other amplified or deleted loci in the validation cohort are indicated in Figure 3-12B, with the genes involved in these variable copy number regions listed in Table 3-8. Notably, most of these regions are associated with single genes, suggesting that patients could potentially be stratified by their copy number profiles. However, applying various clustering methods to explore this possibility did not yield significant results (Figure 3-13). The absence of distinct patient clusters demonstrates that the

copy number changes at these loci are highly heterogeneous and not shared across the cohort. This suggests that the alterations represent individualized, subclonal events likely driven by selective pressure during therapeutic intervention rather than common drivers suitable for patient stratification. It is possible that these genes have also been affected by ecDNA amplifications or deletions, which could contribute to the observed variability.

Table 3-8: The loci with the associated genes and fraction of impacted patients

Locus	Gene - Amp	Gene - Del	Amp % (n =94)	Del % (n =94)
1p36.13	CROCC	FAM131C	21	19
1q21.1	SEC22B, NOTCH2NL	NOTCH2NL	26	31
1q21.3	FLG, multiple genes	FLG	19	27
1q44	OR2T3	OR2T27	21	22
3q29	MUC20	MUC20	27	23
6p22.1	HLA-A	HLA-A	26	31
6p21.33	HLA genes	HLA genes	21	26
6p21.32	HLA genes	HLA genes	22	29
7p11.2	EGFR	EGFR	17	37
9p21.3	multiple genes	multiple genes	12	17
14q11.2	OR4K1	OR4N2, OR4M1	21	32
15q11.2	OR4M2	OR4M2, multiple genes	27	24
16q12.2	CES1	CES1	22	29
16q22.1	PDPR	PDPR	19	21
19p13.3	PLIN4	PLIN4	18	22
19p12	ZNF676	ZNF676	28	24

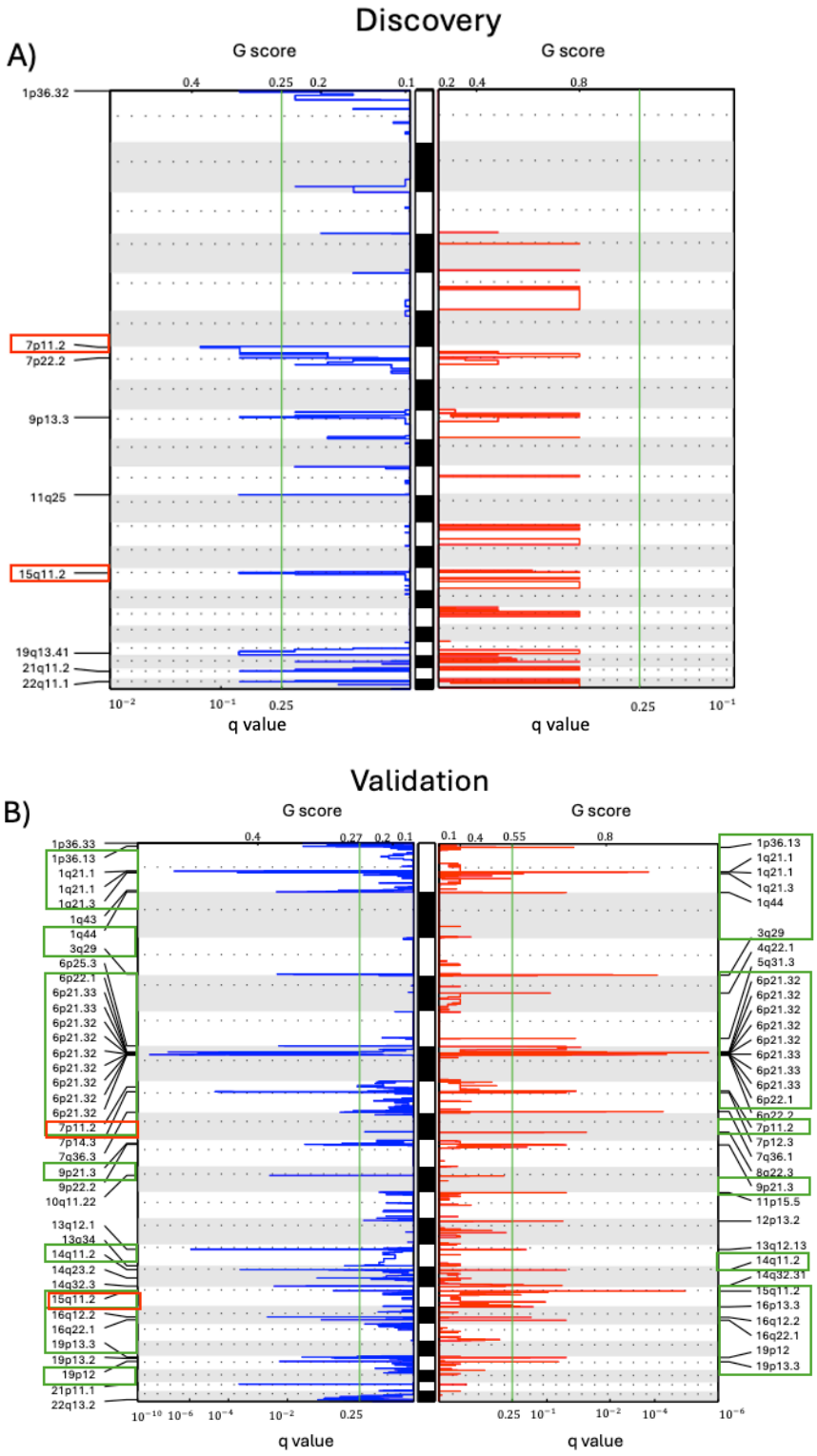


Figure 3-12: Copy number plots of discovery cohort (A) and validation cohort (B).

Blue peaks indicate the deletions, and the red peaks indicate the amplifications. The green line represents the significance threshold, the G score represents the aberration (amplitude X frequency) and q-value represents the false discovery rate. The red boxes are loci shared between the cohorts and the green boxes are the significant amplified and deleted loci in the validation cohort.

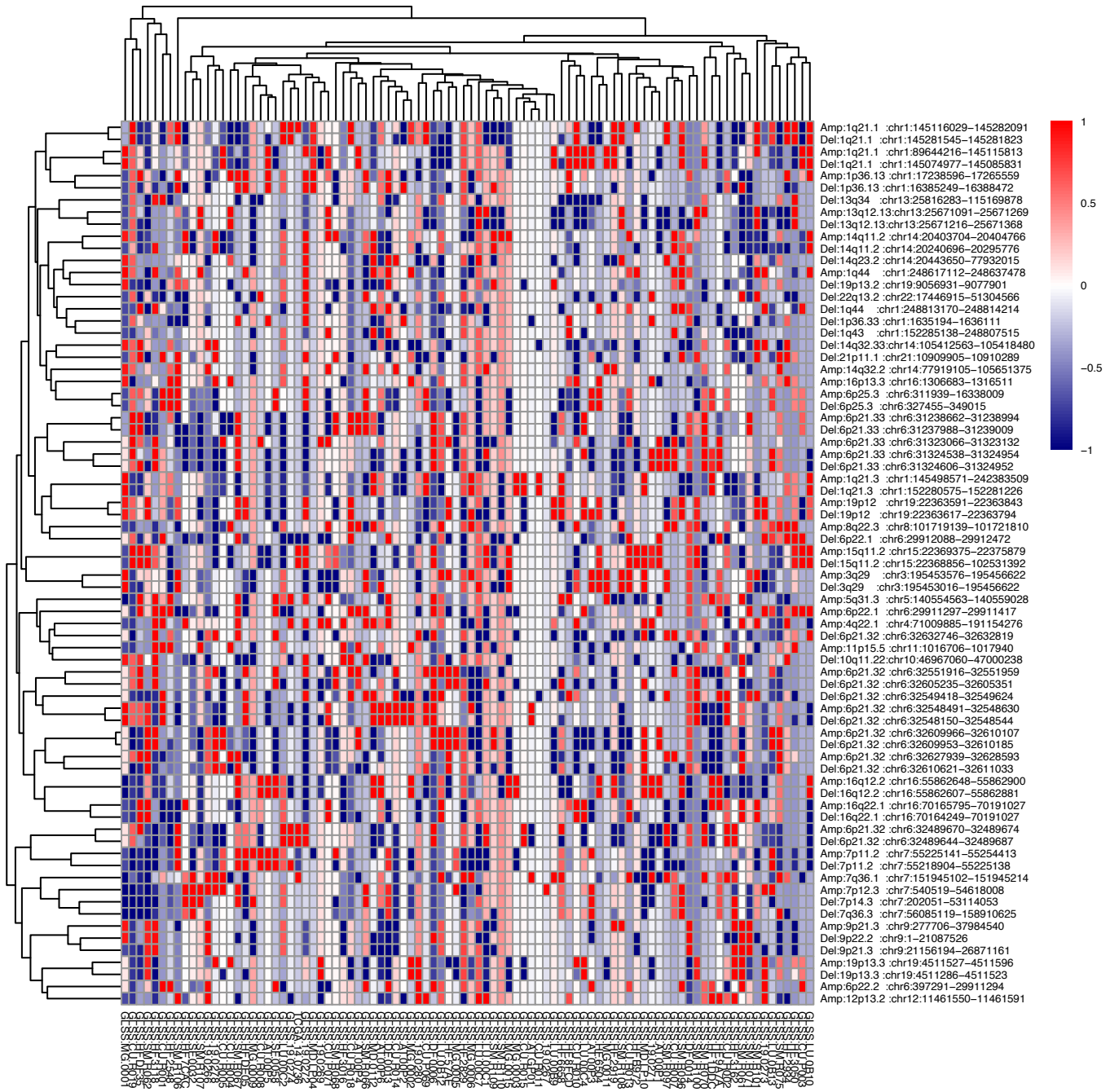


Figure 3-13: Heatmap of Copy Number Profiles Across the Validation Cohort Following Clustering Analysis.

This heatmap visualizes the somatic copy number alteration (SCNA) data for genomic loci with variable copy number across the patients (columns) in the validation cohort. The values are expressed as the Log 2 (Ratio) of the tumour copy number relative to the normal reference. Hierarchical clustering was applied to both rows and columns to identify common copy number profiles for patient stratification.

Rows: Genomic loci showing variable copy number regions.

Columns: Individual patient/sample in the validation cohort.

Colour Scale: The colour indicates the magnitude and direction of the copy number change. Blue for deletion and red for amplification

3.4 DISCUSSION

3.4.1 Variants Conferring Treatment Resistance

Investigating the genes involved in GBM progression by identifying mutations that are likely selected for during treatment, I have identified several key biological processes, cellular components, and molecular functions that are enriched in both our discovery and validation cohorts, suggesting a possible mechanistic foundation in GBM.

The gene set enrichment analysis of variants expanding from primary to recurrent tumours in both discovery and validation cohorts reveals enrichment in GO biological processes related to sinoatrial (SA) and atrioventricular (AV) node functions. The discovery cohort includes terms like "SA node cell action potential" "SA node cell to atrial cardiac muscle cell signalling" and "SA node cell to atrial cardiac muscle cell communication". In the validation cohort, we observed enrichment in processes such as "membrane depolarization during SA node cell action potential" "membrane depolarization during AV node cell action potential" and "AV node cell action potential". These terms suggest a potential role for rhythmic and coordinated signalling in the cells that harboured the expanded variants. Interestingly, these enriched processes in our cohorts reflect a phenomenon observed in a study by (Hausmann et al., 2023), who found that a subpopulation of glioblastoma cells exhibits periodic calcium (Ca^{2+}) activity, acting as network hubs within the tumour. This rhythmic activity was crucial for generating intercellular Ca^{2+} waves that activate key pathways like MAPK and NF-κB, driving tumour growth. This aligns with our findings of enrichment in SA and AV node-related processes, suggesting that rhythmic signalling mechanisms may be a shared feature between cardiac pace-making and glioblastoma progression. Targeting these pacemaker-like mechanisms, particularly in glioblastoma, could potentially disrupt the tumour's progression and offer a potential therapy.

Extending beyond the rhythmic signalling observed in these biological processes, the findings also point to the involvement of specific cellular components, such as the laminin complex, which may further elucidate the tumour's invasive characteristics. The persistence of certain cellular components, such as laminin complexes, has emerged as a significant factor in both discovery and validation cohorts of glioblastoma patients. This finding aligns with the known involvement of laminins in the glioma microenvironment and suggests a specific role in glioblastoma pathophysiology. Laminins, due to their integral role in cell adhesion, migration, and differentiation, may influence tumour progression and the invasive behaviour of glioblastoma cells. Notably, laminin has also been shown to support the growth of glioblastoma stem-like cells (Lathia et al., 2012), and contributes to GBM cell migration and invasion (Kawataki et al., 2007), further underscoring its

multifaceted role in disease progression. Variants of genes involved in this pathway have expanded through therapy, implying that these genes may confer treatment resistance. In a study by (Tanner et al., 2024), RNA sequencing data from IDHwt GBM longitudinal tumours was analysed, and patients were stratified based on changes in gene expression into two types: Up-responders and Down-responders. The Up-responders were characterized by a proneural signature, while the Down-responders were associated with mesenchymal transitions. The study found that in the Up-responder subtype, oligodendrocyte progenitor cell-like (OPC) cancer cells increased from primary to recurrent stages. Preclinical models included cell lines grown in laminin-treated environments that mimic the extracellular matrix (ECM) of tissues. The development and function of oligodendrocytes are regulated by various molecules, including laminin, a major component of the ECM. Consistent with this finding, laminin complexes were found to be enriched in genes that expanded through treatment, suggesting that cells harbouring these genes might belong to the proneural signature and Up-responder subtype.

In addition to the structural implications of laminin complexes within the tumour microenvironment, the findings also highlight the significance of molecular functions, such as ATP-dependent microtubule motor activities, in contributing to therapy resistance.

ATP-dependent microtubule motor activity, minus-end-directed, is a molecular function associated with motor proteins that move along microtubules in a direction towards the minus end, using energy derived from ATP hydrolysis (Ambrose et al., 2005, Ali and Yang, 2020). The minus-end-directed movement typically involves dynein motor proteins, which play critical roles in various cellular processes including intracellular transport, positioning of organelles, and mitotic spindle assembly during cell division (Wadsworth and Lee, 2013). By analysing the molecular function GO terms, I found that this pathway was present in both the discovery and validation cohorts. Moreover, this term was found among the group of variants that expanded from primary to recurrent tumours, further implying a role in resistance mechanisms. A study by (Wang et al., 2016a) identified the *DHC2* gene, also known as *DYNC2H1*, as being associated with resistance to temozolomide (TMZ), and this gene was mutated in the validation cohort. Other dynein family genes linked with this molecular function term were also mutated and shared between the cohorts, suggesting they may contribute to TMZ resistance. Interestingly, germline mutations in *DYNC2H1* are known to cause Jeune syndrome, a ciliopathy characterised by defects in primary cilia—cellular structures that act as key signalling hubs (Higgins et al., 2019). In the context of GBM, such mutations may disrupt primary cilia formation or function, potentially impairing cell-environment signalling. Additionally, cells lacking primary cilia tend to arrest in G0 phase of the cell cycle, which may allow them to evade treatment, as many therapies, including TMZ, target actively dividing cells. This suggests that mutations in

DYNC2H1 may contribute to therapy resistance not only through direct mechanisms, but also by altering cell cycle dynamics and signalling sensitivity.

Simultaneously, the significant disruptions observed in the androgen receptor signalling genes within the KEGG Prostate Cancer pathway suggest that anti-androgen therapies, effective in prostate cancer, could be adapted for GBM treatment. By targeting similar oncogenic mechanisms in GBM, hormone therapy strategies might provide a novel therapeutic avenue, further expanding the arsenal against GBM. Combining these approaches could offer a multifaceted strategy that capitalizes on the mechanistic similarities between these diverse cancers, opening up new possibilities for tailored and effective GBM therapies.

Mutations in the KEGG *P53* Signalling pathway often result in the loss of tumour suppressor functions, enabling GBM cells to evade apoptosis, a common mechanism of action for many anticancer drugs. Similarly, the aberrations in the *PI3K/Akt* pathway, as seen in the Reactome *PI3K* events in *ERBB2* signalling, can lead to unchecked cellular proliferation and survival, providing a biological basis for resistance to therapies that target these growth signalling pathways.

The alterations observed in the BIOCARTA_TEL_PATHWAY are particularly noteworthy. The maintenance and elongation of telomeres in cancer cells, typically via the activation of telomerase, allow these cells to replicate indefinitely, thus contributing to their immortal phenotype. Targeting telomerase has been proposed as a therapeutic strategy, yet resistance often develops through alternative lengthening of telomeres (ALT) mechanisms, showcasing the adaptive nature of GBM.

The variants identified in the BIOCARTA_CTCF_PATHWAY suggest a role in genomic instability, which is a hallmark of cancer that contributes to both the heterogeneity of tumour cells and their ability to resist multiple drug mechanisms. Thus, the disruption of CTCF-dependent chromatin remodelling, and gene expression regulation could provide a fertile ground for the emergence of drug-resistant cancer cell variants.

These observations show the complexity of treating GBM and highlight the necessity for a targeted approach that considers the genetic makeup of individual tumours. Developing therapies that can effectively target these altered pathways may provide a means to circumvent or overcome the resistance mechanisms, offering hope for improved treatment outcomes in GBM patients.

The analysis I conducted in this study highlights the robustness of the significant pathways detected, even after careful adjustments for copy number variations. My method involved revisiting gene lists and excluding variants located in regions with variable copy numbers, ensuring that the subsequent

pathway analysis accurately reflected true genetic signals. This rigorous reanalysis confirmed the presence of critical pathways across both cohorts, demonstrating their intrinsic biological significance in the disease process. The consistent identification of significant pathways in the larger validation cohort confirms that these pathways are fundamental to the underlying biology of the disease, independent of genomic alterations like copy number changes. This discovery is crucial for reinforcing the validity of the pathways involved in the pathogenesis and progression of the disease, indicating that these pathways are not merely incidental but likely central to the disease mechanisms.

3.4.2 Variants Conferring Treatment Sensitivity

The combined analysis of enriched Gene Ontology (GO) terms and altered pathways in treatment-responsive glioblastoma patients revealed a multifaceted picture of biological processes and molecular functions associated with treatment susceptibility.

GO enrichment analysis of the discovery cohort displayed enrichment in neuronal development pathways, suggesting therapies might target cells reliant on these processes for survival. The recurrent presence of the *BDNF* gene in key pathways further supports this notion, as BDNF is known to play a crucial role in neuroplasticity and survival (Taylor et al., 2023).

In contrast, the validation cohort highlighted enrichment in pathways related to cardiac function and calcium signalling. This could point to shared pathways between cardiac function and GBM progression. The enrichment of "regulation of cardiac conduction" strengthens the cardiotoxicity hypothesis, potentially explaining some of the side effects observed during chemoradiotherapy (Griffin et al., 2020, Pei et al., 2020). Alternatively, the shared enrichment of calcium signalling pathways suggests potential links between cardiac health and GBM development or progression, warranting further investigation.

The analysis of cellular components also yielded intriguing insights. Mutations in the discovery cohort impacted mRNA cap binding complex and other membrane-associated structures, suggesting the importance of mRNA processing and ion channel function under therapeutic pressure. These findings align with the established role of mRNA processing in cellular stress responses (Dutertre et al., 2014). The validation cohort further strengthens this concept with the enrichment of the phosphatidylinositol 3-kinase complex and the voltage-gated sodium channel complex, both crucial

for signal transduction and ion homeostasis (Tsai et al., 2020). These observations suggest that therapies might exploit vulnerabilities in cellular signalling and ion regulation to eliminate GBM cells.

The analysis of molecular functions provided further clues about the impact of therapy. In the discovery cohort, enrichment for "RNA 7-methylguanosine cap binding" and "structural molecule activity conferring elasticity" suggests that the stability of RNA and cellular structure might be particularly vulnerable to therapeutic stresses. This aligns with the notion that therapies can disrupt cellular processes essential for GBM cell survival. The validation cohort, with its enrichment in "glutamate-gated calcium ion channel activity" and "structural molecules conferring elasticity," underscores the importance of neurotransmission and structural stability in the context of therapeutic response (Pei et al., 2020). Interestingly, the enrichment of "melanocortin receptor activity" and "opioid receptor activity" suggests a potential role for sensory signalling pathways in treatment effectiveness, warranting further exploration (Pasqualetti et al., 2018, Zhou et al., 2013).

Expanding the investigation by conducting pathway analysis led to identifying promising targets potentially contributing to therapeutic effectiveness. One prominent pathway exhibiting disruption was the *CBL* pathway. *CBL* proteins act as cellular janitors, regulating growth factor signalling by tagging receptor tyrosine kinases (RTKs) for degradation (Liyasova et al., 2015). Variations observed in this pathway suggest compromised *CBL* function, likely leading to aberrantly high and persistent growth factor signalling. This disruption weakens oncogenic signals and makes GBM cells more susceptible to treatment, aligning with previous reports demonstrating the efficacy of disrupting CBL pathway dysfunction in cancer (Roger Belizaire, 2021).

Similarly, alterations in the EGFR_SMRTE pathway reinforce the rationale for targeting *EGFR*, a protein that is frequently mutated or overexpressed in glioblastoma (Brennan et al., 2013). SMRTE, a variant of the SMRT corepressor with an extended N-terminal sequence that shares similarities with N-CoR, appears to play a role in regulating *EGFR* signalling (Mottis et al., 2013). Disruptions within the EGFR_SMRTe pathway could therefore be beneficial, as they potentially make glioblastoma cells more responsive to EGFR-targeted therapies (Xu et al., 2017).

The *PTP1B* pathway, a negative regulator of insulin and leptin signalling, emerged as another promising target. *PTP1B* functions as a molecular brake on cell growth by removing phosphate groups from proteins involved in insulin and leptin signalling pathways (Liu et al., 2022). Abnormally high PTP1B activity, potentially reflected by the pathway disruptions observed in our study, can lead to constantly active signalling through these pathways, promoting uncontrolled growth and survival of

cancer cells (Cheng and Guo, 2019). Inhibiting *PTP1B* may offer several potential therapeutic benefits: reduced cancer cell growth by blocking insulin and leptin signalling, enhanced treatment sensitivity by disrupting survival signals, and potentially overcoming chemotherapeutic resistance mechanisms linked to *PTP1B* activity.

The *ARF6* pathway, involved in actin cytoskeleton remodelling, also displayed consistent alterations. The actin cytoskeleton provides structure and allows cells to move. *ARF6*, a small GTPase protein, acts as a molecular switch that orchestrates actin polymerization and organization by recruiting and activating various proteins (Sun et al., 2023). The observed pathway disruptions suggest potential ARF6 dysregulation in glioblastoma. Disruption of ARF6 signalling could hinder the formation of actin protrusions necessary for cell movement and invasion, thereby reducing the ability of glioblastoma cells to migrate and spread to other parts of the brain, a hallmark of glioblastoma aggressiveness (Li et al., 2009). Our data aligns with the notion that targeting the ARF6 pathway might be beneficial by hindering the invasive potential of glioblastoma cells, potentially improving treatment efficacy, and reducing the risk of metastasis. Further research is needed to elucidate the specific mechanisms of ARF6 dysregulation in glioblastoma and develop therapeutic strategies to target this pathway.

Finally, consistent alterations were observed in the *GAB1* signalsome pathway. *GAB1*, a scaffolding protein, acts as a signalling hub, forming a complex with various proteins to transduce signals involved in cell growth, survival, proliferation, and migration (Mattoon et al., 2004). In glioblastoma, *GAB1* may contribute to tumorigenesis by promoting these processes (Singh et al., 2017). Disruption of the *GAB1* signalsome pathway, as suggested by the findings, could impair these pro-survival signalling pathways, sensitizing glioblastoma cells to apoptosis and enhancing the efficacy of treatments.

In conclusion, this study has identified several promising pathways whose disruption is associated with enhanced therapeutic sensitivity in glioblastoma. These findings provide a valuable foundation for further investigation into targeted therapies that capitalize on these vulnerabilities. By targeting these pathways, we may improve treatment outcomes for glioblastoma patients.

3.4.3 ERBB Signalling Pathways

In the investigation of the *ERBB* signaling pathways, distinct variants in *EGFR*, *PTEN*, and *PIK3CA* exhibit a nuanced interplay that influences cellular behavior and treatment responses in cancer. *EGFR* mutations in different domains affect its interaction with ERBB family members; for instance,

mutations in the Recept_L_domain and Pkinase_Tyr domain in the ERBB4 and ERBB2 pathways, respectively, modulate the receptor's dimerization and signalling efficacy. Altered EGFR-ERBB dimerization can either impair or enhance downstream signalling, depending on the nature of the mutations, potentially influencing the cellular response to therapies targeting these receptors (Andersson et al., 2010, Donoghue et al., 2018, Lucas et al., 2022).

PTEN variants, particularly in the PTEN_C2 domain in the *ERBB2* pathway, disrupt its critical lipid phosphatase activity and membrane localization. By dephosphorylating *PIP3, PTEN* directly opposes the PI3K-induced signalling, thus regulating cell survival and proliferation. When *PTEN* function is compromised due to loss of function mutations, this leads to unregulated *PI3K/AKT* signalling, which enhances cell survival and resistance to apoptosis, presenting major challenges in cancer treatment, especially in tumours with hyperactive *ERBB2* signalling (Matsuoka and Ueda, 2018, Yehia et al., 2019, Jang et al., 2021).

PIK3CA mutations introduce significant complexities into the cellular signalling dynamics within the *ERBB* signalling pathways. In the *ERBB2* pathway, a start loss variant in PIK3CA suggests a potential reduction in the functional activity of the PI3K catalytic subunit p110α, which might nominally decrease PI3K/AKT pathway activity. However, the impact of PIK3CA mutations is markedly different in the ERBB4 pathway. Here, mutations in the *PI3K_p85B, PI3K_C2*, and *PI3Kα* domains likely impair the regulatory and catalytic functions of PI3K. This impairment weakens the *PI3K/AKT* signalling, reducing cell growth and survival capabilities.
As a result, cells harbouring these mutations may exhibit diminished proliferative potential or enter a quiescent (G0) state, making them less responsive to therapies that target actively dividing cells. Alternatively, some cells may become more susceptible to treatment-induced apoptosis due to reduced survival signalling, indicating that *PIK3CA* mutations could contribute to therapy response through multiple, context-dependent mechanisms (Okkenhaug et al., 2016, Liu et al., 2018).

Understanding these complex interactions is crucial for developing targeted therapies. By delineating how specific domain mutations in *EGFR, PTEN*, and *PIK3CA* influence *ERBB* signalling, researchers can better predict treatment outcomes and refine therapeutic approaches to exploit vulnerabilities in cancer signalling networks. This approach emphasizes the need for a personalized medicine strategy, tailoring treatments based on detailed genetic and molecular profiles to optimize efficacy and overcome resistance in cancer therapies.

3.5 REFERENCES

AALTONEN, L. A., ABASCAL, F., ABESHOUSE, A., ABURATANI, H., ADAMS, D. J., AGRAWAL, N., AHN, K. S., AHN, S.-M., AIKATA, H., AKBANI, R., AKDEMIR, K. C., AL-AHMADIE, H., AL-SEDAIRY, S. T., AL-SHAHROUR, F., ALAWI, M., ALBERT, M., ALDAPE, K., ALEXANDROV, L. B., ALLY, A., ALSOP, K., ALVAREZ, E. G., AMARY, F., AMIN, S. B., AMINOU, B., AMMERPOHL, O., ANDERSON, M. J., ANG, Y., ANTONELLO, D., ANUR, P., APARICIO, S., APPELBAUM, E. L., ARAI, Y., ARETZ, A., ARIHIRO, K., ARIIZUMI, S.-I., ARMENIA, J., ARNOULD, L., ASA, S., ASSENOV, Y., ATWAL, G., AUKEMA, S., AUMAN, J. T., AURE, M. R. R., AWADALLA, P., AYMERICH, M., BADER, G. D., BAEZ-ORTEGA, A., BAILEY, M. H., BAILEY, P. J., BALASUNDARAM, M., BALU, S., BANDOPADHAYAY, P., BANKS, R. E., BARBI, S., BARBOUR, A. P., BARENBOIM, J., BARNHOLTZ-SLOAN, J., BARR, H., BARRERA, E., BARTLETT, J., BARTOLOME, J., BASSI, C., BATHE, O. F., BAUMHOER, D., BAVI, P., BAYLIN, S. B., BAZANT, W., BEARDSMORE, D., BECK, T. A., BEHJATI, S., BEHREN, A., NIU, B., BELL, C., BELTRAN, S., BENZ, C., BERCHUCK, A., BERGMANN, A. K., BERGSTROM, E. N., BERMAN, B. P., BERNEY, D. M., BERNHART, S. H., BEROUKHIM, R., BERRIOS, M., BERSANI, S., BERTL, J., BETANCOURT, M., BHANDARI, V., BHOSLE, S. G., BIANKIN, A. V., BIEG, M., BIGNER, D., BINDER, H., BIRNEY, E., BIRRER, M., BISWAS, N. K., BJERKEHAGEN, B., BODENHEIMER, T., BOICE, L., BONIZZATO, G., DE BONO, J. S., et al. 2020. Pan-cancer analysis of whole genomes. *Nature,* 578**,** 82-93.

ALI, I. & YANG, W. C. 2020. The functions of kinesin and kinesin-related proteins in eukaryotes. *Cell Adh Migr,* 14**,** 139-152.

AMBROSE, J. C., LI, W., MARCUS, A., MA, H. & CYR, R. 2005. A minus-end-directed kinesin with plus-end tracking protein activity is involved in spindle morphogenesis. *Mol Biol Cell,* 16**,** 1584-92.

ANDERSSON, U., SCHWARTZBAUM, J., WIKLUND, F., SJOSTROM, S., LIU, Y., TSAVACHIDIS, S., AHLBOM, A., AUVINEN, A., COLLATZ-LAIER, H., FEYCHTING, M., JOHANSEN, C., KIURU, A., LONN, S., SCHOEMAKER, M. J., SWERDLOW, A. J., HENRIKSSON, R., BONDY, M. & MELIN, B. 2010. A comprehensive study of the association between the EGFR and ERBB2 genes and glioma risk. *Acta Oncol,* 49**,** 767-75.

ASHBURNER, M., BALL, C. A., BLAKE, J. A., BOTSTEIN, D., BUTLER, H., CHERRY, J. M., DAVIS, A. P., DOLINSKI, K., DWIGHT, S. S., EPPIG, J. T., HARRIS, M. A., HILL, D. P., ISSEL-TARVER, L., KASARSKIS, A., LEWIS, S., MATESE, J. C., RICHARDSON, J. E., RINGWALD, M., RUBIN, G. M. & SHERLOCK, G. 2000. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet,* 25**,** 25-9.

BAILEY, M. H., TOKHEIM, C., PORTA-PARDO, E., SENGUPTA, S., BERTRAND, D., WEERASINGHE, A., COLAPRICO, A., WENDL, M. C., KIM, J., REARDON, B., NG, P. K., JEONG, K. J., CAO, S., WANG, Z., GAO, J., GAO, Q., WANG, F., LIU, E. M., MULARONI, L., RUBIO-PEREZ, C., NAGARAJAN, N., CORTES-CIRIANO, I., ZHOU, D. C., LIANG, W. W., HESS, J. M., YELLAPANTULA, V. D., TAMBORERO, D., GONZALEZ-PEREZ, A., SUPHAVILAI, C., KO, J. Y., KHURANA, E., PARK, P. J., VAN ALLEN, E. M., LIANG, H., GROUP, M. C. W., CANCER GENOME ATLAS RESEARCH, N., LAWRENCE, M. S., GODZIK, A., LOPEZ-BIGAS, N., STUART, J., WHEELER, D., GETZ, G., CHEN, K., LAZAR, A. J., MILLS, G. B., KARCHIN, R. & DING, L. 2018. Comprehensive Characterization of Cancer Driver Genes and Mutations. *Cell,* 173**,** 371-385 e18.

BARTHEL, F. P., JOHNSON, K. C., VARN, F. S., MOSKALIK, A. D., TANNER, G., KOCAKAVUK, E., ANDERSON, K. J., ABIOLA, O., ALDAPE, K., ALFARO, K. D., ALPAR, D., AMIN, S. B., ASHLEY, D. M., BANDOPADHAYAY, P., BARNHOLTZ-SLOAN, J. S., BEROUKHIM, R., BOCK, C., BRASTIANOS, P. K., BRAT, D. J., BRODBELT, A. R., BRUNS, A. F., BULSARA, K. R., CHAKRABARTY, A., CHAKRAVARTI, A., CHUANG, J. H., CLAUS, E. B., COCHRAN, E. J., CONNELLY, J., COSTELLO, J. F., FINOCCHIARO, G., FLETCHER, M. N., FRENCH, P. J., GAN, H. K., GILBERT, M. R., GOULD, P. V., GRIMMER, M. R., IAVARONE, A., ISMAIL, A., JENKINSON, M. D., KHASRAW, M., KIM, H., KOUWENHOVEN, M. C. M., LAVIOLETTE, P. S., LI, M., LICHTER, P., LIGON, K. L., LOWMAN, A. K., MALTA, T. M., MAZOR, T., MCDONALD, K. L., MOLINARO, A. M., NAM, D. H., NAYYAR, N., NG, H. K., NGAN, C. Y., NICLOU, S. P., NIERS, J. M., NOUSHMEHR, H., NOORBAKHSH, J., ORMOND, D. R., PARK, C. K., POISSON, L. M., RABADAN, R., RADLWIMMER, B., RAO, G., REIFENBERGER, G., SA, J. K., SCHUSTER, M., SHAW, B. L., SHORT, S. C., SMITT, P. A. S., SLOAN, A. E., SMITS, M., SUZUKI, H., TABATABAI, G., VAN MEIR, E. G., WATTS, C., WELLER, M., WESSELING, P., WESTERMAN, B. A., WIDHALM, G., WOEHRER, A., YUNG, W. K. A., ZADEH, G., HUSE, J. T., DE GROOT, J. F., STEAD, L. F., VERHAAK, R. G. W. & CONSORTIUM, G. 2019. Longitudinal molecular trajectories of diffuse glioma in adults. *Nature,* 576**,** 112-120.

BARTOLOME, R. A., MARTIN-REGALADO, A., JAEN, M., ZANNIKOU, M., ZHANG, P., DE LOS RIOS, V., BALYASNIKOVA, I. V. & CASAL, J. I. 2020. Protein Tyrosine Phosphatase-1B Inhibition Disrupts IL13Ralpha2-Promoted Invasion and Metastasis in Cancer Cells. *Cancers (Basel),* 12.

BENJAMIN, D., SATO, T., CIBULSKIS, K., GETZ, G., STEWART, C. & LICHTENSTEIN, L. 2019.

BRENNAN, C. W., VERHAAK, R. G., MCKENNA, A., CAMPOS, B., NOUSHMEHR, H., SALAMA, S. R., ZHENG, S., CHAKRAVARTY, D., SANBORN, J. Z., BERMAN, S. H., BEROUKHIM, R., BERNARD, B., WU, C. J., GENOVESE, G., SHMULEVICH, I., BARNHOLTZ-SLOAN, J., ZOU, L., VEGESNA, R., SHUKLA, S. A., CIRIELLO, G., YUNG, W. K., ZHANG, W., SOUGNEZ, C., MIKKELSEN, T., ALDAPE, K., BIGNER, D. D., VAN MEIR, E. G., PRADOS, M., SLOAN, A., BLACK, K. L., ESCHBACHER, J., FINOCCHIARO, G., FRIEDMAN, W., ANDREWS, D. W., GUHA, A., IACOCCA, M., O'NEILL, B. P., FOLTZ, G., MYERS, J., WEISENBERGER, D. J., PENNY, R., KUCHERLAPATI, R., PEROU, C. M., HAYES, D. N., GIBBS, R., MARRA, M., MILLS, G. B., LANDER, E., SPELLMAN, P., WILSON, R., SANDER, C., WEINSTEIN, J., MEYERSON, M., GABRIEL, S., LAIRD, P. W., HAUSSLER, D., GETZ, G., CHIN, L. & NETWORK, T. R. 2013. The somatic genomic landscape of glioblastoma. *Cell,* 155**,** 462-77.

BYRNE, K. F., PAL, A., CURTIN, J. F., STEPHENS, J. C. & KINSELLA, G. K. 2021. G-protein-coupled receptors as therapeutic targets for glioblastoma. *Drug Discov Today,* 26**,** 2858-2870.

CHENG, F. & GUO, D. 2019. MET in glioma: signaling pathways and targeted therapies. *J Exp Clin Cancer Res,* 38**,** 270.

CHONG, C. R. & JANNE, P. A. 2013. The quest to overcome resistance to EGFR-targeted therapies in cancer. *Nat Med,* 19**,** 1389-400.

COLOGNATO, H. & YURCHENCO, P. D. 2000. Form and function: The laminin family of heterotrimers. *Developmental Dynamics,* 218**,** 213-234.

CONSORTIUM, G. 2018. Glioma through the looking GLASS: molecular evolution of diffuse gliomas and the Glioma Longitudinal Analysis Consortium. *Neuro-Oncology,* 20**,** 873-884.

CONSORTIUM, P.-C. A. O. W. G. 2020. Pan-cancer analysis of whole genomes. *Nature,* 578**,** 82-93.

DARRÉ, H., MASSON, P., NATIVEL, A., VILLAIN, L., LEFAUDEUX, D., COUTY, C., MARTIN, B., JACOB, E., DURUISSEAUX, M., PALGEN, J.-L., MONTEIRO, C. & L'HOSTIS, A. 2024. Comparing the Efficacy of Two Generations of EGFR-TKIs: An Integrated Drug–Disease Mechanistic Model Approach in EGFR-Mutated Lung Adenocarcinoma. *Biomedicines,* 12.

DIPLAS, B. H., HE, X., BROSNAN-CASHMAN, J. A., LIU, H., CHEN, L. H., WANG, Z., MOURE, C. J., KILLELA, P. J., LORIAUX, D. B., LIPP, E. S., GREER, P. K., YANG, R., RIZZO, A. J., RODRIGUEZ, F. J., FRIEDMAN, A. H., FRIEDMAN, H. S., WANG, S., HE, Y., MCLENDON, R. E., BIGNER, D. D., JIAO, Y., WAITKUS, M. S., MEEKER, A. K. & YAN, H. 2018. The genomic landscape of TERT promoter wildtype-IDH wildtype glioblastoma. *Nat Commun,* 9**,** 2087.

DONOGHUE, J. F., KERR, L. T., ALEXANDER, N. W., GREENALL, S. A., LONGANO, A. B., GOTTARDO, N. G., WANG, R., TABAR, V., ADAMS, T. E., MISCHEL, P. S. & JOHNS, T. G. 2018. Activation of ERBB4 in Glioblastoma Can Contribute to Increased Tumorigenicity and Influence Therapeutic Response. *Cancers (Basel),* 10.

DUTERTRE, M., SANCHEZ, G., BARBIER, J., CORCOS, L. & AUBOEUF, D. 2014. The emerging role of pre-messenger RNA splicing in stress responses: Sending alternative messages and silent messengers. *RNA Biology,* 8**,** 740-747.

FRENCH, P. J., EOLI, M., SEPULVEDA, J. M., DE HEER, I., KROS, J. M., WALENKAMP, A., FRENEL, J. S., FRANCESCHI, E., CLEMENT, P. M., WELLER, M., ANSELL, P., LOOMAN, J., BAIN, E., MORFOUACE, M., GORLIA, T. & VAN DEN BENT, M. 2019. Defining EGFR amplification status for clinical trial inclusion. *Neuro Oncol,* 21**,** 1263-1272.

GAFFNEY, S. G. & TOWNSEND, J. P. 2016. PathScore: a web tool for identifying altered pathways in cancer data. *Bioinformatics,* 32**,** 3688-3690.

GAN, H. K., CVRLJEVIC, A. N. & JOHNS, T. G. 2013. The epidermal growth factor receptor variant III (EGFRvIII): where wild things are altered. *FEBS J,* 280**,** 5350-70.

GENE ONTOLOGY, C., ALEKSANDER, S. A., BALHOFF, J., CARBON, S., CHERRY, J. M., DRABKIN, H. J., EBERT, D., FEUERMANN, M., GAUDET, P., HARRIS, N. L., HILL, D. P., LEE, R., MI, H., MOXON, S., MUNGALL, C. J., MURUGANUGAN, A., MUSHAYAHAMA, T., STERNBERG, P. W., THOMAS, P. D., VAN AUKEN, K., RAMSEY, J., SIEGELE, D. A., CHISHOLM, R. L., FEY, P., ASPROMONTE, M. C., NUGNES, M. V., QUAGLIA, F., TOSATTO, S., GIGLIO, M., NADENDLA, S., ANTONAZZO, G., ATTRILL, H., DOS SANTOS, G., MARYGOLD, S., STRELETS, V., TABONE, C. J., THURMOND, J., ZHOU, P., AHMED, S. H., ASANITTHONG, P., LUNA BUITRAGO, D., ERDOL, M. N., GAGE, M. C., ALI KADHUM, M., LI, K. Y. C., LONG, M., MICHALAK, A., PESALA, A., PRITAZAHRA, A., SAVERIMUTTU, S. C. C., SU, R., THURLOW, K. E., LOVERING, R. C., LOGIE, C., OLIFERENKO, S., BLAKE, J., CHRISTIE, K., CORBANI, L., DOLAN, M. E., DRABKIN, H. J., HILL, D. P., NI, L., SITNIKOV, D., SMITH, C., CUZICK, A., SEAGER, J., COOPER, L., ELSER, J., JAISWAL, P., GUPTA, P., JAISWAL, P., NAITHANI, S., LERA-RAMIREZ, M., RUTHERFORD, K., WOOD, V., DE PONS, J. L., DWINELL, M. R., HAYMAN, G. T., KALDUNSKI, M. L., KWITEK, A. E., LAULEDERKIND, S. J. F., TUTAJ, M. A., VEDI, M., WANG, S. J., D'EUSTACHIO, P., AIMO, L., AXELSEN, K., BRIDGE, A., HYKA-NOUSPIKEL, N., MORGAT, A., ALEKSANDER, S. A., CHERRY, J. M., ENGEL, S. R., KARRA, K., MIYASATO, S. R., NASH, R. S., SKRZYPEK, M. S., WENG, S., WONG, E. D., BAKKER, E., et al. 2023. The Gene Ontology knowledgebase in 2023. *Genetics,* 224.

GRIFFIN, M., KHAN, R., BASU, S. & SMITH, S. 2020. Ion Channels as Therapeutic Targets in High Grade Gliomas. *Cancers (Basel),* 12.

HAUSMANN, D., HOFFMANN, D. C., VENKATARAMANI, V., JUNG, E., HORSCHITZ, S., TETZLAFF, S. K., JABALI, A., HAI, L., KESSLER, T., AZORIN, D. D., WEIL, S., KOURTESAKIS, A., SIEVERS, P., HABEL, A., BRECKWOLDT, M. O., KARREMAN, M. A., RATLIFF, M., MESSMER, J. M., YANG, Y., REYHAN, E., WENDLER, S., LOB, C., MAYER, C., FIGARELLA, K., OSSWALD, M., SOLECKI, G., SAHM, F., GARASCHUK, O., KUNER, T., KOCH, P., SCHLESNER, M., WICK, W. & WINKLER, F. 2023. Autonomous rhythmic activity in glioma networks drives brain tumour growth. *Nature,* 613**,** 179-186.

HIGGINS, M., OBAIDI, I. & MCMORROW, T. 2019. Primary cilia and their role in cancer. *Oncol Lett,* 17**,** 3041-3047.

HOOGSTRATE, Y., GHISAI, S. A., DE WIT, M., DE HEER, I., DRAAISMA, K., VAN RIET, J., VAN DE WERKEN, H. J. G., BOURS, V., BUTER, J., VANDEN BEMPT, I., EOLI, M., FRANCESCHI, E., FRENEL, J. S., GORLIA, T., HANSE, M. C., HOEBEN, A., KERKHOF, M., KROS, J. M., LEENSTRA, S., LOMBARDI, G., LUKACOVA, S., ROBE, P. A., SEPULVEDA, J. M., TAAL, W., TAPHOORN, M., VERNHOUT, R. M., WALENKAMP, A. M. E., WATTS, C., WELLER, M., DE VOS, F. Y. F., JENSTER, G. W., VAN DEN BENT, M. & FRENCH, P. J. 2022. The EGFRvIII transcriptome in glioblastoma: A meta-omics analysis. *Neuro Oncol,* 24**,** 429-441.

HYNES, N. E. & LANE, H. A. 2005. ERBB receptors and cancer: the complexity of targeted inhibitors. *Nat Rev Cancer,* 5**,** 341-54.

JALIFE, J. 1984. Mutual entrainment and electrical coupling as mechanisms for synchronous firing of rabbit sino-atrial pacemaker cells. *The Journal of Physiology,* 356**,** 221-243.

JANG, H., SMITH, I. N., ENG, C. & NUSSINOV, R. 2021. The mechanism of full activation of tumor suppressor PTEN at the phosphoinositide-enriched membrane. *iScience,* 24**,** 102438.

JING, Z., LI, L., WANG, X., WANG, M., CAI, Y., JIN, Z. I. & ZHANG, Y. E. 2016. High c-Cbl expression in gliomas is associated with tumor progression and poor prognosis. *Oncol Lett,* 11**,** 2787-2791.

KAWATAKI, T., YAMANE, T., NAGANUMA, H., ROUSSELLE, P., ANDUREN, I., TRYGGVASON, K. & PATARROYO, M. 2007. Laminin isoforms and their integrin receptors in glioma cell migration and invasiveness: Evidence for a role of alpha5-laminin(s) and alpha3beta1 integrin. *Exp Cell Res,* 313**,** 3819-31.

KORBER, V., YANG, J., BARAH, P., WU, Y., STICHEL, D., GU, Z., FLETCHER, M. N. C., JONES, D., HENTSCHEL, B., LAMSZUS, K., TONN, J. C., SCHACKERT, G., SABEL, M., FELSBERG, J., ZACHER, A., KAULICH, K., HUBSCHMANN, D., HEROLD-MENDE, C., VON DEIMLING, A., WELLER, M., RADLWIMMER, B., SCHLESNER, M., REIFENBERGER, G., HOFER, T. & LICHTER, P. 2019. Evolutionary Trajectories of IDH(WT) Glioblastomas Reveal a Common Path of Early Tumorigenesis Instigated Years ahead of Initial Diagnosis. *Cancer Cell,* 35**,** 692-704 e12.

LAKATTA, E. G., MALTSEV, V. A. & VINOGRADOVA, T. M. 2010. A coupled SYSTEM of intracellular Ca2+ clocks and surface membrane voltage clocks controls the timekeeping mechanism of the heart's pacemaker. *Circ Res,* 106**,** 659-73.

LANGFORD, L. A., PIATYSZEK, M. A., XU, R., SCHOLD, S. C., JR. & SHAY, J. W. 1995. Telomerase activity in human brain tumours. *Lancet,* 346**,** 1267-8.

LASSMAN, A. B., ALDAPE, K. D., ANSELL, P. J., BAIN, E., CURRAN, W. J., EOLI, M., FRENCH, P. J., KINOSHITA, M., LOOMAN, J., MEHTA, M., MURAGAKI, Y., NARITA, Y., OCAMPO, C., ROBERTS-RAPP, L., SONG, M., VOGELBAUM, M. A., WALENKAMP, A. M. E., WANG, T. J. C., ZHANG, P. & VAN DEN BENT, M. J. 2019. Epidermal growth factor receptor (EGFR) amplification rates observed in screening patients for randomized trials in glioblastoma. *J Neurooncol,* 144**,** 205-210.

LATHIA, J. D., LI, M., HALL, P. E., GALLAGHER, J., HALE, J. S., WU, Q., VENERE, M., LEVY, E., RANI, M. R., HUANG, P., BAE, E., SELFRIDGE, J., CHENG, L., GUVENC, H., MCLENDON, R. E., NAKANO, I., SLOAN, A. E., PHILLIPS, H. S., LAI, A., GLADSON, C. L., BREDEL, M., BAO, S., HJELMELAND, A. B. & RICH, J. N. 2012. Laminin alpha 2 enables glioblastoma stem cell growth. *Ann Neurol,* 72**,** 766-78.

LI, M., WANG, J., NG, S. S., CHAN, C. Y., HE, M. L., YU, F., LAI, L., SHI, C., CHEN, Y., YEW, D. T., KUNG, H. F. & LIN, M. C. 2009. Adenosine diphosphate-ribosylation factor 6 is required for epidermal growth factor-induced glioblastoma cell proliferation. *Cancer,* 115**,** 4959-72.

LIAO, Z., JIANG, W., YE, L., LI, T., YU, X. & LIU, L. 2020. Classification of extrachromosomal circular DNA with a focus on the role of extrachromosomal DNA (ecDNA) in tumor heterogeneity and progression. *Biochim Biophys Acta Rev Cancer,* 1874**,** 188392.

LIU, M., WANG, W., ZHANG, H., BI, J., ZHANG, B., SHI, T., SU, G., ZHENG, Y., FAN, S., HUANG, X., CHEN, B., SONG, Y., ZHAO, Z., SHI, J., LI, P., LU, W. & ZHANG, L. 2023. Three-Dimensional Gene Regulation Network in Glioblastoma Ferroptosis. *Int J Mol Sci,* 24.

LIU, R., MATHIEU, C., BERTHELET, J., ZHANG, W., DUPRET, J. M. & RODRIGUES LIMA, F. 2022. Human Protein Tyrosine Phosphatase 1B (PTP1B): From Structure to Clinical Inhibitor Perspectives. *Int J Mol Sci,* 23.

LIU, S., TANG, Y., YAN, M. & JIANG, W. 2018. PIK3CA mutation sensitizes breast cancer cells to synergistic therapy of PI3K inhibition and AMPK activation. *Invest New Drugs,* 36**,** 763-772.

LIYASOVA, M. S., MA, K. & LIPKOWITZ, S. 2015. Molecular Pathways: Cbl Proteins in Tumorigenesis and Antitumor Immunity—Opportunities for Cancer Treatment. *Clinical Cancer Research,* 21**,** 1789-1794.

LOTSCH, D., GHANIM, B., LAABER, M., WURM, G., WEIS, S., LENZ, S., WEBERSINKE, G., PICHLER, J., BERGER, W. & SPIEGL-KREINECKER, S. 2013. Prognostic significance of telomerase-associated parameters in glioblastoma: effect of patient age. *Neuro Oncol,* 15**,** 423-32.

LUCAS, L. M., DWIVEDI, V., SENFELD, J. I., CULLUM, R. L., MILL, C. P., PIAZZA, J. T., BRYANT, I. N., COOK, L. J., MILLER, S. T., LOTT, J. H. T., KELLEY, C. M., KNERR, E. L., MARKHAM, J. A., KAUFMANN, D. P., JACOBI, M. A., SHEN, J. & RIESE, D. J., 2ND 2022. The Yin and Yang of ERBB4: Tumor Suppressor and Oncoprotein. *Pharmacol Rev,* 74**,** 18-47.

MACDONALD, E. A., MADL, J., GREINER, J., RAMADAN, A. F., WELLS, S. M., TORRENTE, A. G., KOHL, P., ROG-ZIELINSKA, E. A. & QUINN, T. A. 2020. Sinoatrial Node Structure, Mechanics, Electrophysiology and the Chronotropic Response to Stretch in Rabbit and Mouse. *Front Physiol,* 11**,** 809.

MALNIC, B., GODFREY, P. A. & BUCK, L. B. 2004. The human olfactory receptor gene family. *Proc Natl Acad Sci U S A,* 101**,** 2584-9.

MARINO, S., MENNA, G., DI BONAVENTURA, R., LISI, L., MATTOGNO, P., FIGA, F., BILGIN, L., D'ALESSANDRIS, Q. G., OLIVI, A. & DELLA PEPA, G. M. 2023. The Extracellular Matrix in Glioblastomas: A Glance at Its Structural Modifications in Shaping the Tumoral Microenvironment-A Systematic Review. *Cancers (Basel),* 15.

MARTINEZ-JIMENEZ, F., MUINOS, F., SENTIS, I., DEU-PONS, J., REYES-SALAZAR, I., ARNEDO-PAC, C., MULARONI, L., PICH, O., BONET, J., KRANAS, H., GONZALEZ-PEREZ, A. & LOPEZ-BIGAS, N. 2020. A compendium of mutational cancer driver genes. *Nat Rev Cancer,* 20**,** 555-572.

MATSUOKA, S. & UEDA, M. 2018. Mutual inhibition between PTEN and PIP3 generates bistability for polarity in motile cells. *Nat Commun,* 9**,** 4481.

MATTOON, D. R., LAMOTHE, B., LAX, I. & SCHLESSINGER, J. 2004. *BMC Biology,* 2.

MCNULTY, S. N., SCHWETYE, K. E., FERGUSON, C., STORER, C. E., ANSSTAS, G., KIM, A. H., GUTMANN, D. H., RUBIN, J. B., HEAD, R. D. & DAHIYA, S. 2021. BRAF mutations may identify a clinically distinct subset of glioblastoma. *Sci Rep,* 11**,** 19999.

MELLINGHOFF, I. K., WANG, M. Y., VIVANCO, I., HAAS-KOGAN, D. A., ZHU, S., DIA, E. Q., LU, K. V., YOSHIMOTO, K., HUANG, J. H., CHUTE, D. J., RIGGS, B. L., HORVATH, S., LIAU, L. M., CAVENEE, W. K., RAO, P. N., BEROUKHIM, R., PECK, T. C., LEE, J. C., SELLERS, W. R., STOKOE, D., PRADOS, M., CLOUGHESY, T. F., SAWYERS, C. L. & MISCHEL, P. S. 2005. Molecular determinants of the response of glioblastomas to EGFR kinase inhibitors. *N Engl J Med,* 353**,** 2012-24.

MIAO, B., SKIDAN, I., YANG, J., YOU, Z., FU, X., FAMULOK, M., SCHAFFHAUSEN, B., TORCHILIN, V., YUAN, J. & DEGTEREV, A. 2012. Inhibition of cell migration by PITENINs: the role of ARF6. *Oncogene,* 31**,** 4317-32.

MIURA, S., GOMEZ, K., MURILLO, O., HUUKI, L. A., VU, T., BUTURLA, T. & KUMAR, S. 2018. Predicting clone genotypes from tumor bulk sequencing of multiple samples. *Bioinformatics,* 34**,** 4017-4026.

MIYASHITA, Y., KO, R., SHIMADA, N., MITSUISHI, Y., MIURA, K., MATSUMOTO, N., ASAO, T., SHUKUYA, T., SHIBAYAMA, R., KOYAMA, R. & TAKAHASHI, K. 2020. Impact of the generation of EGFR-TKIs administered as prior therapy on the efficacy of osimertinib in patients with non-small cell lung cancer harboring EGFR T790M mutation. *Thoracic Cancer,* 12**,** 329-338.

MOASSER, M. M. 2007. The oncogene HER2: its signaling and transforming functions and its role in human cancer pathogenesis. *Oncogene,* 26**,** 6469-87.

MOTTIS, A., MOUCHIROUD, L. & AUWERX, J. 2013. Emerging roles of the corepressors NCoR1 and SMRT in homeostasis. *Genes Dev,* 27**,** 819-35.

MUNJAPARA, V., HEUMANN, T., SCHRECK, K. C., GROSS, J. M., PEREZ-HEYDRICH, C., GUJAR, S. K., EBERHART, C. G. & HOLDHOFF, M. 2022. BRAF V600E-Mutant Glioblastoma with Extracranial Metastases Responsive to Combined BRAF and MEK Targeted Inhibition: A Case Report. *Case Rep Oncol,* 15**,** 909-917.

NADEU, F., DELGADO, J., ROYO, C., BAUMANN, T., STANKOVIC, T., PINYOL, M., JARES, P., NAVARRO, A., MARTIN-GARCIA, D., BEA, S., SALAVERRIA, I., OLDREIVE, C., AYMERICH, M., SUAREZ-CISNEROS, H., ROZMAN, M., VILLAMOR, N., COLOMER, D., LOPEZ-GUILLERMO, A., GONZALEZ, M., ALCOCEBA, M., TEROL, M. J., COLADO, E., PUENTE, X. S., LOPEZ-OTIN, C., ENJUANES, A. & CAMPO, E. 2016. Clinical impact of clonal and subclonal TP53, SF3B1, BIRC3, NOTCH1, and ATM mutations in chronic lymphocytic leukemia. *Blood,* 127**,** 2122-30.

NOER, J. B., HORSDAL, O. K., XIANG, X., LUO, Y. & REGENBERG, B. 2022. Extrachromosomal circular DNA in cancer: history, current knowledge, and methods. *Trends Genet,* 38**,** 766-781.

OKKENHAUG, K., GRAUPERA, M. & VANHAESEBROECK, B. 2016. Targeting PI3K in Cancer: Impact on Tumor Cells, Their Protective Stroma, Angiogenesis, and Immunotherapy. *Cancer Discov,* 6**,** 1090-1105.

PASQUALETTI, F., ORLANDI, P., SIMEON, V., CANTARELLA, M., GIULIANI, D., DI DESIDERO, T., GONNELLI, A., DELISHAJ, D., LOMBARDI, G., SECHI, A., SANSON, M., ZAGONEL, V., PAIAR, F., DANESI, R., GUARINI, S. & BOCCI, G. 2018. Melanocortin Receptor-4 Gene Polymorphisms in Glioblastoma Patients Treated with Concomitant Radio-Chemotherapy. *Mol Neurobiol,* 55**,** 1396-1404.

PEI, Z., LEE, K. C., KHAN, A., ERISNOR, G. & WANG, H. Y. 2020. Pathway analysis of glutamate-mediated, calcium-related signaling in glioma progression. *Biochem Pharmacol,* 176**,** 113814.

QAZI, M. A., SALIM, S. K., BROWN, K. R., MIKOLAJEWICZ, N., SAVAGE, N., HAN, H., SUBAPANDITHA, M. K., BAKHSHINYAN, D., NIXON, A., VORA, P., DESMOND, K., CHOKSHI, C., SINGH, M., KHOO, A., MACKLIN, A., KHAN, S., TATARI, N., WINEGARDEN, N., RICHARDS, L., PUGH, T., BOCK, N., MANSOURI, A., VENUGOPAL, C., KISLINGER, T., GOYAL, S., MOFFAT, J. & SINGH, S. K. 2022. Characterization of the minimal residual disease state reveals distinct evolutionary trajectories of human glioblastoma. *Cell Rep,* 40**,** 111420.

QUINLAN, A. R. & HALL, I. M. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics,* 26**,** 841-2.

RASCIO, F., SPADACCINO, F., ROCCHETTI, M. T., CASTELLANO, G., STALLONE, G., NETTI, G. S. & RANIERI, E. 2021. The Pathogenic Role of PI3K/AKT Pathway in Cancer Onset and Drug Resistance: An Updated Review. *Cancers,* 13.

ROGER BELIZAIRE, S. H. J. K., NAMRATA D. UDESHI, ALEXIS VEDDER, LEI SUN, TANYA SVINKINA, CHRISTINA HARTIGAN, MARIE MCCONKEY, VERONICA KOVALCIK, AMANUEL BIZUAYEHU, CAROLINE STANCLIFT, MONICA SCHENONE, STEVEN A.

CARR, ERIC PADRON, AND BENJAMIN L. EBERT 2021. CBL mutations drive PI3K/AKT signaling via increased interaction with LYN and PIK3R1.

ROSENBAUM, D. M., RASMUSSEN, S. G. & KOBILKA, B. K. 2009. The structure and function of G-protein-coupled receptors. *Nature,* 459**,** 356-63.

SESE, B., ENSENYAT-MENDEZ, M., INIGUEZ, S., LLINAS-ARIAS, P. & MARZESE, D. M. 2021. Chromatin insulation dynamics in glioblastoma: challenges and future perspectives of precision oncology. *Clin Epigenetics,* 13**,** 150.

SHOMALI, W. & GOTLIB, J. 2018. The new tool "KIT" in advanced systemic mastocytosis. *Hematology Am Soc Hematol Educ Program,* 2018**,** 127-136.

SINGH, D. K., KOLLIPARA, R. K., VEMIREDDY, V., YANG, X. L., SUN, Y., REGMI, N., KLINGLER, S., HATANPAA, K. J., RAISANEN, J., CHO, S. K., SIRASANAGANDLA, S., NANNEPAGA, S., PICCIRILLO, S., MASHIMO, T., WANG, S., HUMPHRIES, C. G., MICKEY, B., MAHER, E. A., ZHENG, H., KIM, R. S., KITTLER, R. & BACHOO, R. M. 2017. Oncogenes Activate an Autonomous Transcriptional Regulatory Circuit That Drives Glioblastoma. *Cell Rep,* 18**,** 961-976.

STEPHAN, G., RAVN-BOESS, N. & PLACANTONAKIS, D. G. 2021. Adhesion G protein-coupled receptors in glioblastoma. *Neurooncol Adv,* 3**,** vdab046.

SUN, D., GUO, Y., TANG, P., LI, H. & CHEN, L. 2023. Arf6 as a therapeutic target: Structure, mechanism, and inhibitors. *Acta Pharm Sin B,* 13**,** 4089-4104.

TANNER, G., BARROW, R., AJAIB, S., AL-JABRI, M., AHMED, N., POLLOCK, S., FINETTI, M., RIPPAUS, N., BRUNS, A. F., SYED, K., POULTER, J. A., MATTHEWS, L., HUGHES, T., WILSON, E., JOHNSON, C., VARN, F. S., BRUNING-RICHARDSON, A., HOGG, C., DROOP, A., GUSNANTO, A., CARE, M. A., CUTILLO, L., WESTHEAD, D. R., SHORT, S. C., JENKINSON, M. D., BRODBELT, A., CHAKRABARTY, A., ISMAIL, A., VERHAAK, R. G. W. & STEAD, L. F. 2024. IDHwt glioblastomas can be stratified by their transcriptional response to standard treatment, with implications for targeted therapy. *Genome Biol,* 25**,** 45.

TANNER, G., WESTHEAD, D. R., DROOP, A. & STEAD, L. F. 2021. Benchmarking pipelines for subclonal deconvolution of bulk tumour sequencing data. *Nat Commun,* 12**,** 6396.

TAYLOR, K. R., BARRON, T., HUI, A., SPITZER, A., YALCIN, B., IVEC, A. E., GERAGHTY, A. C., HARTMANN, G. G., ARZT, M., GILLESPIE, S. M., KIM, Y. S., MALEKI JAHAN, S., ZHANG, H., SHAMARDANI, K., SU, M., NI, L., DU, P. P., WOO, P. J., SILVA-TORRES, A., VENKATESH, H. S., MANCUSI, R., PONNUSWAMI, A., MULINYAWE, S., KEOUGH, M. B., CHAU, I., AZIZ-BOSE, R., TIROSH, I., SUVA, M. L. & MONJE, M. 2023. Glioma synapses recruit mechanisms of adaptive plasticity. *Nature*.

TSAI, H. F., C, I. J. & SHEN, A. Q. 2020. Voltage-gated ion channels mediate the electrotaxis of glioblastoma cells in a hybrid PMMA/PDMS microdevice. *APL Bioeng,* 4**,** 036102.

VERHAAK, R. G. W., BAFNA, V. & MISCHEL, P. S. 2019. Extrachromosomal oncogene amplification in tumour pathogenesis and evolution. *Nat Rev Cancer,* 19**,** 283-288.

WADSWORTH, P. & LEE, W. L. 2013. Microtubule motors: doin' it without dynactin. *Curr Biol,* 23**,** R563-5.

WANG, H., FENG, W., LU, Y., LI, H., XIANG, W., CHEN, Z., HE, M., ZHAO, L., SUN, X., LEI, B., QI, S. & LIU, Y. 2016a. Expression of dynein, cytoplasmic 2, heavy chain 1 (DHC2) associated with glioblastoma cell resistance to temozolomide. *Sci Rep,* 6**,** 28948.

WANG, J., CAZZATO, E., LADEWIG, E., FRATTINI, V., ROSENBLOOM, D. I., 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. 2016b. Clonal evolution of glioblastoma under therapy. *Nat Genet,* 48**,** 768-76.

WONG, S. Q., LI, J., TAN, A. Y., VEDURURU, R., PANG, J. M., DO, H., ELLUL, J., DOIG, K., BELL, A., MACARTHUR, G. A., FOX, S. B., THOMAS, D. M., FELLOWES, A., PARISOT, J. P., DOBROVIC, A. & COHORT, C. 2014. Sequence artefacts in a prospective series of formalin-fixed tumours tested for mutations in hotspot regions by massively parallel sequencing. *BMC Med Genomics,* 7**,** 23.

XU, L., CHEN, Y., DUTRA-CLARKE, M., MAYAKONDA, A., HAZAWA, M., SAVINOFF, S. E., DOAN, N., SAID, J. W., YONG, W. H., WATKINS, A., YANG, H., DING, L. W., JIANG, Y. Y., TYNER, J. W., CHING, J., KOVALIK, J. P., MADAN, V., CHAN, S. L., MUSCHEN, M., BREUNIG, J. J., LIN, D. C. & KOEFFLER, H. P. 2017. BCL6 promotes glioma and serves as a therapeutic target. *Proc Natl Acad Sci U S A,* 114**,** 3981-3986.

YAMAUCHI, Y., MIURA, Y. & KANAHO, Y. 2017. Machineries regulating the activity of the small GTPase Arf6 in cancer cells are potential targets for developing innovative anti-cancer drugs. *Adv Biol Regul,* 63**,** 115-121.

YANG, M., ZHANG, S., JIANG, R., CHEN, S. & HUANG, M. 2023. Circlehunter: a tool to identify extrachromosomal circular DNA from ATAC-Seq data. *Oncogenesis,* 12**,** 28.

YEHIA, L., NGEOW, J. & ENG, C. 2019. PTEN-opathies: from biological insights to evidence-based precision medicine. *J Clin Invest,* 129**,** 452-464.

ZALCMAN, N., CANELLO, T., OVADIA, H., CHARBIT, H., ZELIKOVITCH, B., MORDECHAI, A., FELLIG, Y., RABANI, S., SHAHAR, T., LOSSOS, A. & LAVON, I. 2018. Androgen receptor: a potential therapeutic target for glioblastoma. *Oncotarget,* 9**,** 19980-19993.

ZHANG, Y., DUBE, C., GIBERT, M., JR., 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 (Basel),* 10.

ZHAO, Y., YU, L., ZHANG, S., SU, X. & ZHOU, X. 2022. Extrachromosomal circular DNA: Current status and future prospects. *Elife,* 11.

ZHOU, L., GUO, X., CHEN, M., FU, S., ZHOU, J., REN, G., YANG, Z. & FAN, W. 2013. Inhibition of delta-opioid receptors induces brain glioma cell apoptosis through the mitochondrial and protein kinase C pathways. *Oncol Lett,* 6**,** 1351-1357.

CHAPTER 4

4.1 INTRODUCTION

4.1.1 Epigenetics in GBM

While much of the research in GBM has traditionally focused on identifying mutations in specific genes, it has become increasingly clear that epigenetic changes also play a significant role in tumour behaviour (Maleszewska and Kaminska, 2013, Wu et al., 2021, Tanner et al., 2024, Amirmahani et al., 2025, Meleiro and Henrique, 2025). Epigenetics refers to modifications in gene activity that do not involve changes to the actual DNA sequence but can still influence how genes are expressed. These changes are especially relevant in cancer biology, where they can cooperate with genetic mutations to promote tumour initiation, progression, and treatment resistance.

The word "epigenetics" was first introduced by the biologist Conrad Waddington in the early 1940s (Waddington, 2012). Initially, the term was used in the context of development to describe how different cell types arise from the same genetic code based on gene regulation. Over the years, the term has evolved and now refers to heritable changes in gene expression that occur without any alteration to the DNA sequence itself (Bird, 2007). These changes are mediated by chemical modifications to DNA or to the histone proteins that surround it. These modifications include DNA methylation, histone acetylation or methylation, and regulatory effects by non-coding RNAs.

In GBM, several studies have shown that epigenetic alterations, such as DNA methylation, are involved in silencing tumour suppressor genes, influencing how cells respond to treatments like radiotherapy and chemotherapy (Hegi et al., 2005, Esteller, 2008). These epigenetic mechanisms do not act in isolation but interact with the genetic background of the tumour, creating a complex network of regulation that ultimately determines tumour phenotype.

4.1.2 DNA methylation in GBM

Among the various forms of epigenetic regulation, DNA methylation is one of the most widely studied and best understood. It involves the chemical addition of a methyl group (-CH3) to the fifth carbon of the cytosine ring in DNA, forming 5-methylcytosine (5mC) (Wilson et al., 2007). This reaction is catalysed by enzymes known as DNA methyltransferases (DNMTs), primarily *DNMT1, DNMT3A*, and *DNMT3B* (Robert et al., 2003). In mammals, DNA methylation primarily occurs at CpG dinucleotides, which are regions where a cytosine nucleotide is followed by a guanine nucleotide. These CpG sites are not uniformly distributed throughout the genome. Instead, they tend to cluster in regions known as CpG islands, which are typically located in the promoter regions of genes. In normal cells, CpG islands at the promoters of housekeeping or essential genes are typically unmethylated, allowing these genes to remain active. When these CpG islands become

methylated, especially in promoter regions, it often results in the transcriptional silencing of the associated gene (Robert et al., 2003).

Once established, DNA methylation marks can be reliably maintained during cell division by *DNMT1*, which copies the methylation pattern onto the new DNA strand (Robert et al., 2003). This renders DNA methylation a semi-stable and heritable form of gene regulation. However, methylation can also be removed through passive dilution over successive cell divisions or by active demethylation mechanisms involving enzymes such as TET (Ten-Eleven Translocation) proteins.

In GBM and other cancers, certain regions undergo abnormal methylation (hypermethylation), particularly in the promoters of tumour suppressor genes, resulting in their silencing (Etcheverry et al., 2010). Simultaneously, other regions, such as large repetitive sequences and intergenic areas, may become hypomethylated, potentially leading to genomic instability or the activation of usually silent regions. This combined pattern of hyper- and hypomethylation is typical of many cancers, including GBM (Mulholland et al., 2012, Etcheverry et al., 2010).

4.1.3 Types of methylation

DNA methylation occurs in specific regions of the genome, each playing a unique role in gene regulation and cellular function.

4.1.3.1 CpG Islands and Related Regions

CpG islands are short stretches of DNA rich in CpG sites, often located at gene promoters. In normal cells, these regions are usually unmethylated, allowing gene expression. Aberrant methylation of CpG islands, especially in promoters, can silence important genes, including tumour suppressors (Jones and Baylin, 2002). Surrounding these islands are CpG shores, shelves, and open sea regions, where methylation changes can also impact gene regulation in ways that may not be immediately obvious (Irizarry et al., 2009).

4.1.3.2 Promoter Methylation and Gene Silencing

Methylation at promoter regions typically represses gene expression by preventing transcription initiation. This is a common regulatory mechanism in development but can also contribute to disease (Bird, 2002). In glioblastoma, for example, methylation of specific promoters can influence treatment response (Hegi et al., 2005).

4.1.3.3 Gene Body Methylation

Methylation within the gene body is often linked to active transcription. It may help regulate splicing or prevent unwanted transcription starts. The impact varies depending on the precise location and gene context (Jones, 2012).

4.1.3.4 Enhancer Methylation and Gene Regulation

Enhancers regulate gene expression from a distance. Their activity is often controlled by methylation: unmethylated enhancers are active, while methylated ones are usually repressed. In cancer, changes in enhancer methylation may alter the expression of critical genes and affect tumour progression or therapy response (Bell et al., 2016, Alajem et al., 2021).

4.1.4 MGMT promoter methylation and clinical relevance

One of the most clinically important applications of DNA methylation analysis in GBM is the evaluation of the *MGMT* gene (O6-methylguanine-DNA methyltransferase). *MGMT* encodes a DNA repair enzyme that removes alkyl groups from the O6 position of guanine, a common site of damage caused by alkylating chemotherapy agents such as temozolomide (TMZ), which is part of standard-of-care for treating GBM (Stupp et al., 2005).

The significance of *MGMT* in glioblastoma lies in the fact that when the *MGMT* promoter is methylated, transcription of the gene is suppressed. As a result, the tumour has a reduced ability to repair DNA damage caused by chemotherapy, making it more sensitive to treatment. Conversely, in tumours where the MGMT promoter is unmethylated, the gene is active, and the tumour may be more resistant to alkylating drugs (Lee, 2016).

Clinically, MGMT promoter methylation status has become an important predictive and prognostic biomarker. It is routinely assessed to help determine whether a patient is likely to benefit from temozolomide therapy. Several large-scale studies and clinical trials have demonstrated that GBM patients with methylated *MGMT* promoters tend to have longer progression-free and overall survival following treatment with TMZ compared to those with unmethylated promoters (Dunn et al., 2009).

The method used to assess *MGMT* methylation status can differ, but most clinical laboratories employ either methylation-specific PCR or pyrosequencing. More recently, array-based techniques and methylation classifiers, such as MGMT-STP27, have been utilised in research environments to offer a more reliable and standardised readout (van den Bent et al., 2013, Bady et al., 2016). In this study, *MGMT* promoter methylation was examined alongside global methylation patterns to understand its behaviour over time in paired primary and recurrent GBM samples.

4.1.5 Differential Methylation Analysis of Longitudinal GBM Samples

Differential methylation analysis involves comparing DNA methylation patterns between different biological states, such as primary versus recurrent tumours. This approach is crucial for identifying specific epigenetic alterations that contribute to disease pathogenesis, predict outcomes, or serve as potential therapeutic targets.

Recent work from our group has significantly advanced the understanding of glioblastoma (GBM) recurrence following standard treatment. Specifically, Tanner et al. (2024) identified two distinct patient-specific transcriptional responses in isocitrate dehydrogenase wild-type (IDHwt) GBM, categorising patients into 'Up' and 'Down' responder subtypes. The 'Up' and 'Down' nomenclature is derived directly from the opposing transcriptional reprogramming trajectories of a core set of genes, specifically the 'Leading Edge' (LE) genes (LE50 and LE70), in response to standard GBM treatment (Tanner et al., 2024).

- Up Responders: Exhibit upregulation of the LE gene set (e.g., genes involved in mesenchymal transition and inflammation) in the recurrent tumour compared to the primary tumour.
- Down Responders: Exhibit downregulation of the LE gene set in the recurrent tumour compared to the primary tumour.

This seminal work revealed that these responder classifications are intrinsic to cancer cells, suggesting different adaptive resistance mechanisms. While their study highlighted the potential role of Polycomb-based chromatin remodelling in driving these transcriptional shifts, it could not definitely rule out that that global differential DNA methylation might contribute to these particular transcriptional changes.

The primary goal of this study was to investigate whether specific differential DNA methylation patterns correlate with the 'Up'/'Down' transcriptional classification, thereby testing the central hypothesis that epigenetic alterations accompany these distinct resistance trajectories. While Tanner et al. (2024) strongly suggested a role for Polycomb-based chromatin changes, this current work directly investigates the extent to which DNA methylation correlates with and potentially contributes to the observed 'Up' and 'Down' transcriptional reprogramming. The following analysis specifically targets methylation changes in genomic regions functionally linked to the LE gene set and other differentially regulated pathways identified in those subtypes.

While differential methylation analysis has been a common approach in numerous cancer studies, including those on GBM, longitudinal investigations have frequently reported a remarkable stability in overall GBM methylation profiles over time (Malta et al., 2024). This consistent observation suggests that broad, genome-wide methylation changes may not be the primary dynamic drivers of certain tumour responses or

therapeutic resistance mechanisms. Recognising this, my approach deviated from a broad differential analysis to instead investigate specific, functional epigenetic landscapes linked to defined transcriptional behaviours, offering a distinct perspective on methylation's role.

Building upon my group's critical findings, specifically, the finding that IDHwt GBM recurrence involves two distinct, patient-intrinsic transcriptional trajectories ('Up' and 'Down' subtypes), which are associated with differences in Polycomb Repressive Complex (PRC) member expression (Tanner et al., 2024). This chapter explains a novel analysis that I performed to comprehensively characterise the accompanying alterations in DNA methylation patterns within 'Up' and 'Down' responder subtypes. By using established patient groups, I investigated the role of DNA methylation in GBM recurrence and treatment resistance. The goal was to identify unique epigenetic patterns associated with each patient's response trajectory, thereby revealing how DNA methylation contributes to or reflects the distinct biological processes in these patient subtypes. This investigation was conducted with the understanding that DNA methylation may not be the sole driver of transcriptional reprogramming.

4.1.6 Methylation Profiling Techniques

Several experimental methods are available to measure DNA methylation across the genome. These techniques vary in terms of resolution, coverage, cost, and computational complexity (Bock, 2012). The most commonly used approaches can be grouped into sequencing-based and array-based technologies.

4.1.6.1 Reduced Representation Bisulfite Sequencing (RRBS)

RRBS is a targeted bisulfite sequencing method developed to focus on CpG-rich regions of the genome, particularly CpG islands and promoters (Meissner et al., 2005). It begins with the digestion of genomic DNA using a restriction enzyme like MspI, which cuts at CCGG sites. The resulting fragments are size-selected (typically 40–220 bp), then treated with bisulfite to convert unmethylated cytosines to uracil, leaving methylated cytosines unchanged. Finally, the fragments are sequenced (Meissner et al., 2005).

One of the key advantages of RRBS is its high resolution at CpG sites and relatively low cost compared to whole-genome bisulfite sequencing. It is especially useful for identifying differentially methylated promoters and CpG islands. However, because RRBS only captures about 5–10% of the genome, it may miss important regulatory elements in enhancer or intergenic regions. This trade-off makes RRBS an efficient choice for focused studies, but less suitable for mapping the full methylome (Bock, 2012).

4.1.6.2 Whole Genome Bisulfite Sequencing (WGBS)

WGBS is the gold standard for methylation analysis, providing base-pair resolution across nearly all CpG sites in the genome (Lister et al., 2009, Sakthikumar et al., 2020). The method involves random fragmentation of genomic DNA, addition of sequencing adapters, bisulfite conversion, and high-throughput sequencing.

WGBS can detect both CpG and non-CpG methylation, offering the most comprehensive view of the methylome. It is particularly useful in developmental studies, stem cell research, and cancer epigenomics where non-CpG methylation plays a role. The major limitations are its high cost, large data volumes, and the need for advanced bioinformatics pipelines to process and interpret the data. In GBM research, WGBS has been used to identify novel methylation patterns and regulatory elements that are missed by targeted or array-based methods (Hovestadt et al., 2014).

4.1.6.3 Illumina Methylation Arrays

The Infinium BeadChip array platform developed by Illumina has been an essential tool for analysing DNA methylation across the genome. Although the original BeadChip arrays were primarily designed for genotyping, the technology was adapted to target specific CpG sites and measure DNA methylation using bisulfite-treated DNA (Bibikova et al., 2011). These methylation arrays provided a practical and scalable alternative to earlier methods, such as genome-wide bisulfite sequencing or tiling microarrays, which were more costly and less accessible (Bibikova et al., 2011).

The first widely adopted methylation array in this family was the HumanMethylation27 BeadChip (27K), introduced in 2008. This array targeted 27,578 CpG sites, mostly within promoter regions of over 14,000 human genes (Bibikova et al., 2011). One of its main advantages was the small amount of input DNA needed — about 1 μg — which enabled researchers to analyse samples even when only limited material was available, such as formalin-fixed tissues. The array could process 12 samples simultaneously, making it suitable for high-throughput studies with relatively low technical complexity (Bibikova et al., 2011).

The 27K array depends on Infinium chemistry, which starts with bisulfite conversion of DNA. This process converts unmethylated cytosines into uracils, while methylated cytosines stay the (Bibikova et al., 2011). After bisulfite treatment, the DNA is amplified, fragmented, and hybridised to allele-specific probes attached to microscopic beads. These probes are meant to detect either methylated or unmethylated sequences at each CpG site.

In the 27K platform, two types of beads are used per CpG locus: one for the methylated version of the site and another for the unmethylated version. After probe hybridisation, a single-base extension reaction is performed using fluorescently labelled dideoxynucleotides (ddNTPs), often tagged with Biotin or DNP. The choice of nucleotide incorporated depends on the base immediately upstream of the cytosine of interest. Importantly, both methylated and unmethylated probes can utilise the same dye channel, and signal intensity is measured during array scanning using fluorescence detection. (Barrera and Peinado, 2012)The methylation level at each site is expressed as a β-value, ranging from 0 (completely unmethylated) to 1

(completely methylated), based on the ratio of the methylated signal to the total signal intensity at that probe (Barrera and Peinado, 2012).

Although the 27K array was a significant advance in simplifying genome-wide methylation studies, its coverage was relatively limited, mainly concentrating on promoter-associated CpGs. To improve this, Illumina launched the HumanMethylation450K (450K) array in 2011. This newer platform examined over 450,000 CpG sites, covering 99% of RefSeq genes and including a substantial representation of CpG islands, shores, shelves, gene bodies, and regulatory regions, such as miRNA promoters (Morris and Beck, 2015). It signified a shift from solely promoter-focused studies to more comprehensive methylome profiling.

The 450K array combines both Infinium I and Infinium II probe types. Infinium I probes use two separate beads (one for methylated and one for unmethylated) and work with both red and green channels. Conversely, Infinium II probes employ a single bead and one base extension site for both methylation states, allowing for a more compact probe design and greater density on the array. About 30% of the probes on the 450K platform are Infinium I, while the rest are Infinium II (Morris and Beck, 2015). However, the design differences between these probe types create a type I/type II bias, which can affect downstream analysis. To address this, normalisation methods such as Beta Mixture Quantile (BMIQ) normalisation have been developed to correct for technical differences between probe types (Teschendorff et al., 2013). Figure 4-1 shows how both probe types detect methylation based on bisulfite conversion and labelled nucleotide incorporation.

While the 450K array significantly increased coverage, especially over gene bodies and CpG shores, it lacked sufficient probes for distal regulatory elements like enhancers and DNase I hypersensitive sites. This gap prompted the development of the next-generation platform — the MethylationEPIC v1.0 (850K) BeadChip, launched in 2015. The EPICv1 array extended total CpG coverage to more than 850,000 sites, with a strong focus on non-promoter regulatory elements. It includes probes covering >90% of the original 450K sites and adds substantial content from enhancer regions identified in large epigenomic projects such as ENCODE, FANTOM5, and BLUEPRINT.

The EPICv1 array was developed in response to user feedback requesting improved representation of enhancers and other distal elements that play a role in transcriptional regulation and cancer epigenetics. A more recent development in this series is the Infinium MethylationEPIC v2.0 BeadChip (EPICv2), which further enhances genomic coverage. This updated platform includes over 935,000 CpG sites, offering expanded probe content with improved targeting of distal regulatory elements, especially in non-coding regions and immune-related genes. Importantly, EPICv2 retains compatibility with most sites present in the original EPICv1 array, enabling direct comparisons across datasets generated with the two versions (Zhuang et al., 2025).

Although this newer platform offers extended insight into the methylome, it has not yet been widely adopted in large-scale studies due to its recent release. All Infinium arrays include a range of control probes embedded within the platform. These controls monitor bisulfite conversion efficiency, staining and hybridisation

performance, extension success, and background signal levels. Some controls are sample-specific, while others are independent of sample DNA and ensure overall platform reliability.

The preparation and analysis workflow for all Infinium methylation arrays typically takes several days to one week (Bibikova et al., 2011). Due to their accessibility, cost-effectiveness, and compatibility with standard bioinformatics pipelines, these arrays remain one of the most widely used technologies for epigenome-wide association studies, including cancer profiling, developmental biology, and clinical biomarker research.

In this chapter, all methylation data analysed were obtained using either the 450K or EPIC arrays, depending on the cohort. These platforms offered the resolution, coverage, and sample scalability required to perform differential methylation analysis on longitudinal glioblastoma samples. The raw data generated by these arrays are stored in IDAT files, which serve as the entry point for downstream processing and analysis (Smith et al., 2013).

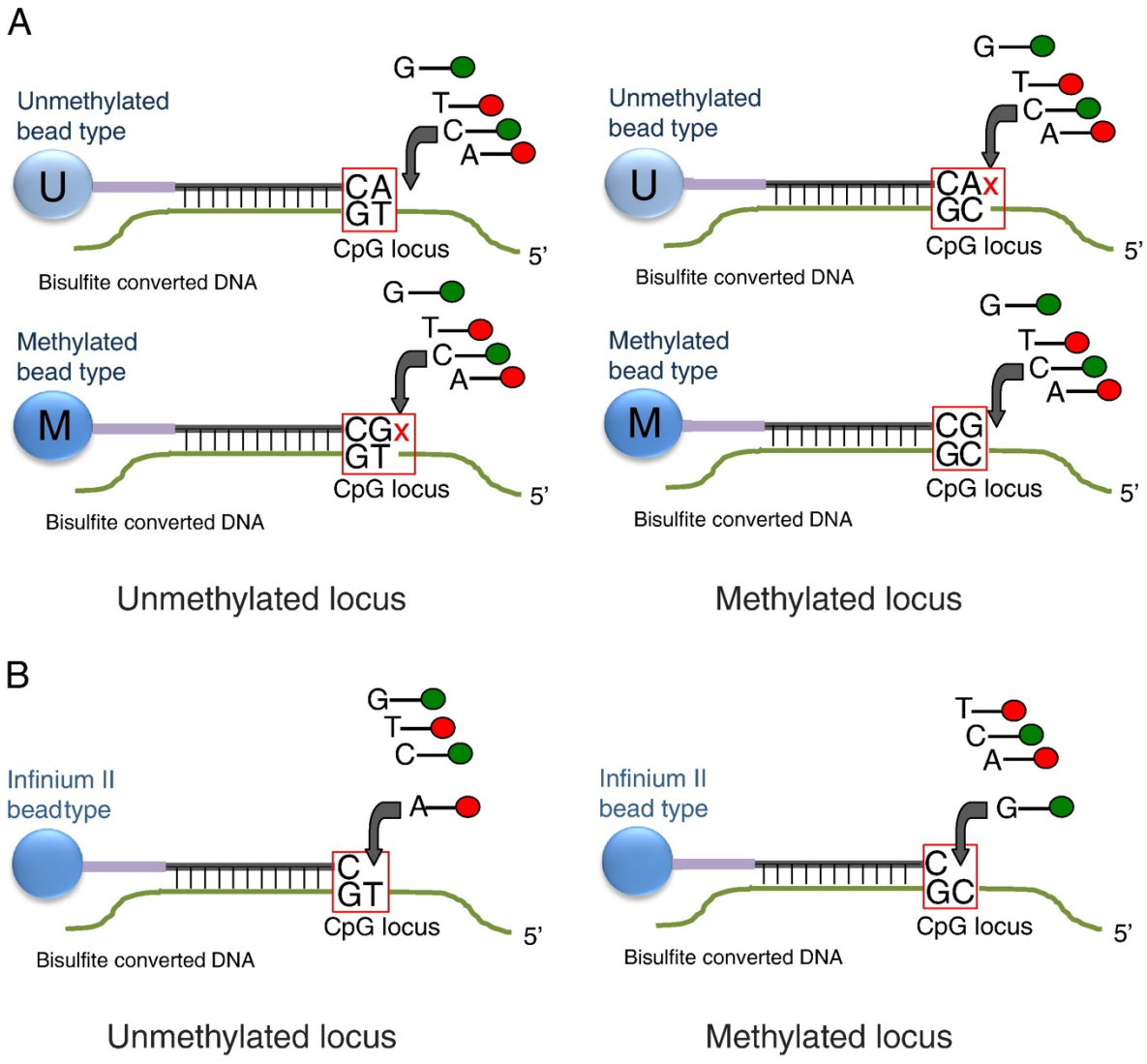


Figure 4-1: Illustration of Infinium methylation assay probe designs.

(A) The Infinium I design uses two separate bead types per CpG site: one to detect the methylated state (C) and one for the unmethylated state (T). Both probes incorporate the same fluorescently labelled nucleotide, depending on the base just upstream of the target site, and are read in the same colour channel.

(B) The Infinium II design uses a single bead for each CpG locus. Methylation is detected by single-base extension: an "A" is added when the site is unmethylated (original C converted to T), and a "G" is added when it is methylated (C remains). Each state is distinguished by a different fluorescent dye.
*adapted with permission from (Bibikova et al., 2011).

4.1.7 Analytical Tools and Pipelines

The raw output from Illumina methylation arrays comes in the form of IDAT files, which contain fluorescence intensity measurements for each probe on the array. Each sample is associated with two IDAT files — one for the red channel and one for the green channel — capturing the signal intensities of methylated and unmethylated probes. In addition to raw intensity values, these files store data on control probes, background levels, array barcodes, and scanning parameters.

Working directly with IDAT files offers the advantage of full control over preprocessing steps, including background correction, normalisation, and filtering. These files are widely supported by open-source tools developed in R, which allow researchers to customise workflows based on study design, array platform, and sample quality. This chapter draws on several such tools to process and analyse the 450K and EPIC array data from GBM samples.

Several R packages exist for handling Illumina methylation array data. These tools vary in their level of automation, customizability, and scope. The following tools were the primary tools utilised for the differential methylation analysis. They were selected based on their flexibility, compatibility with raw IDAT files, and suitability for large-scale or longitudinal designs, such as the one employed in this chapter.

4.1.7.1 RnBeads

RnBeads is a widely used and freely available R-based package designed for the comprehensive analysis of genome-wide DNA methylation data (Muller et al., 2019). It provides robust support for Illumina's 450K and EPIC BeadChip arrays and has also been extended to handle data from sequencing-based platforms such as RRBS and WGBS. As a pipeline-driven tool, RnBeads enables users to perform all essential steps of methylation data processing within a single framework, making it particularly useful for large-scale studies or projects requiring reproducibility.

One of the core strengths of RnBeads lies in its integrated architecture. It bundles key components of a standard analysis pipeline — including quality control, probe filtering, normalisation, exploratory analysis, and differential methylation testing — into a single function (rnb.run.analysis), which can be easily customised depending on the dataset and research objectives. This modularity allows users to tailor the workflow to their specific needs, whether analysing paired tumour samples or comparing distinct biological groups.

The package includes a variety of normalisation methods, each suited to particular array designs or experimental conditions. Among these are BMIQ (Beta Mixture Quantile normalisation) (Teschendorff et al.,

2013), SWAN (Subset-quantile Within Array Normalisation) (Maksimovic et al., 2012), watermelon's dasen method (Pidsley et al., 2013), and noob (normal-exponential out-of-band), which is implemented via the methylumi package (Sean Davis, 2025). These methods are essential for correcting systematic technical variation, such as probe type bias, and ensuring accurate downstream comparisons.

In this study, RnBeads was applied to the discovery cohort, where raw IDAT files were available for paired primary and recurrent GBM samples. The platform's ability to incorporate metadata, stratify samples by experimental conditions, and annotate results by genomic regions (e.g., promoters, gene bodies, enhancers) made it particularly suitable for identifying biologically meaningful methylation changes in longitudinal tumour progression.

4.1.7.2 Minfi

The minfi package is a widely used R/Bioconductor tool designed for analysing Illumina Infinium DNA methylation arrays, including the 450K and EPIC (850K) platforms (Aryee et al., 2014). It offers a flexible and modular framework that enables researchers to perform detailed preprocessing, quality control, normalisation, and differential methylation analysis, supporting both standard and customised workflows.

Minfi starts by importing raw intensity data from IDAT files using functions that read the methylated and unmethylated signal intensities for each probe across samples. The data is stored in structured objects such as RGChannelSet, which can then be further processed into MethylSet or RatioSet objects, enabling downstream analyses.

In this chapter, minfi was utilised to process and analyse methylation data from the validation cohort. Its flexible architecture allowed for customised preprocessing and normalisation steps, ensuring compatibility with the discovery cohort processed using RnBeads. The integration of quality control measures and differential methylation analysis tools within minfi facilitated a robust and reproducible analysis pipeline.

The package provides various preprocessing and normalisation methods and includes detailed quality control tools to evaluate data integrity. Unlike RnBeads, which offers an integrated, automated pipeline with built-in reporting, minfi allows more detailed control over each analysis step, making it particularly useful when workflows need to be manually customised or adjusted.

4.2 METHODS

In this section, I describe two complementary approaches I used to perform differential methylation analysis on DNA methylation array data. Both methods were implemented in R but followed different pipelines for quality control, normalisation, annotation, and statistical testing. The nature of the available data from the discovery and validation cohorts influenced the decision to use two methods. The discovery cohort included raw IDAT files, which enabled the use of the RnBeads package—an end-to-end pipeline that integrates preprocessing, quality control, and region-level differential methylation analysis. In contrast, the validation cohort consisted of pre-processed beta value matrices, which are incompatible with RnBeads. Therefore, I used a second approach based on the minfi and limma packages, which allowed greater flexibility in working with pre-normalised data and performing customised statistical modelling. Applying both methods not only accommodated the structure of the available data but also provided complementary analytical perspectives and an opportunity to cross-validate findings across cohorts.

4.2.1 Analysis Using RnBeads

For the first approach, I used the RnBeads package, which provides a comprehensive and automated pipeline for the analysis of DNA methylation array data. This method includes integrated steps for quality control, normalization, annotation, and differential methylation analysis, all with detailed reports and visualizations. I chose this approach as a complementary strategy to validate and expand upon the results generated using minfi and limma.

I began by preparing a sample annotation file in CSV format, which included sample identifiers, file paths (when applicable), array platform information, and metadata such as patient ID and sample status (e.g., primary or recurrent tumor). I then ran the analysis using the rnb.run.analysis() function, specifying my project directory, sample annotation, and chosen analysis options.

For quality control and filtering, I used the rnb.run.qc() function with several important parameters enabled:

- filtering.greedycut = TRUE was set to iteratively remove probes with the highest proportion of unreliable measurements, based on detection p-values. This approach helps clean the data by excluding probes that fail quality thresholds in multiple samples, ensuring that the remaining data points are consistently reliable.
- filtering.sex.chromosomes.removal = TRUE was used to eliminate probes located on the X and Y chromosomes. This step helps avoid sex-related methylation variability, especially in mixed-sex cohorts, and is particularly important when sex is not a primary variable of interest.
- filtering.cross.reactive = TRUE excluded probes known to bind to multiple genomic locations due to sequence homology, which can introduce misleading methylation signals. Removing

these cross-reactive probes improves specificity and reduces the risk of artifacts in downstream analysis.

- filtering.snp = "3" was chosen to remove probes that contain single nucleotide polymorphisms (SNPs) either at the targeted CpG site or at the single-base extension site. SNPs at these positions can affect hybridization efficiency or interfere with probe design, potentially leading to spurious methylation calls. Using level "3" applies the strictest filtering criteria based on known dbSNP annotations.

Together, these filters were applied to ensure that only high-confidence methylation measurements were included in downstream analysis, thereby increasing the robustness and biological reliability of the results.

Normalization was performed using the BMIQ method by setting normalization.method = "bmiq", which adjusts for probe-type bias between Infinium type I and type II probes. This step is essential for achieving comparability across probe types and preventing artificial technical differences from influencing the analysis.

Although the original 450K and EPIC1 arrays were designed for the hg19 genome build, I conducted the analysis using hg38, as RnBeads provides updated annotation packages for these older arrays mapped to the newer genome version. This allowed me to take advantage of improved genomic coordinates and region definitions.

For differential methylation analysis, I used the rnb.run.differential() function and selected the paired analysis mode, since my comparison involved matched samples—specifically, primary versus recurrent tumours from the same patients. This paired design increased the statistical power by accounting for inter-patient variability.

The RnBeads pipeline generated comprehensive HTML reports that included summaries of the filtering and normalisation steps, interactive plots (e.g., PCA, clustering, and methylation profiles), and tables of differentially methylated positions (DMPs) and regions (DMRs). These results allowed me to examine not only individual CpG sites but also broader patterns across genomic features, including promoters, gene bodies, CpG islands, and other regulatory elements. In addition to the default region types provided by RnBeads, I also incorporated an in-house curated list of enhancers, which I added as a custom region set to assess methylation changes in biologically relevant regulatory domains.

This automated workflow provided a robust complementary analysis to the more hands-on minfi/limma pipeline, contributing to a more comprehensive understanding of methylation differences between primary and recurrent tumours.

4.2.2 Analysis Using minfi and limma

As a second approach of the analysis, I used the R packages minfi (*version 1.54.1*) and limma (*version 3.63.0*) to process and analyse DNA methylation data obtained from Illumina 450K and EPIC arrays. This analysis was

conducted on the discovery dataset, aiming to identify differentially methylated probes (DMPs) between sample groups and to understand how these changes were distributed across various genomic regions and features.

I began by importing the raw IDAT files using the read.metharray.exp() function from the minfi package. This produced an RGChannelSet object, which stores the raw methylated and unmethylated signal intensities from the array. I then performed initial quality control using the qcReport() function to check for low-quality samples or technical issues. In addition, I examined density plots and conducted principal component analysis (PCA) to visually inspect for outliers and batch effects that could influence the results.

Across all datasets, I applied consistent filtering steps to remove problematic probes. These included probes located on the X and Y chromosomes, those containing known single nucleotide polymorphisms (SNPs) at the CpG or extension sites, and probes known to cross-react with multiple genomic locations. I also filtered out probes with poor detection p-values (>0.01), as these may reflect background noise. These steps helped ensure that only reliable and biologically meaningful probes were retained.

To measure DNA methylation levels, I extracted beta values using the getBeta() function. Beta values range between 0 and 1, representing the proportion of methylation at each CpG site. While they are easy to interpret biologically, they are not ideal for statistical modeling due to their unequal variability across the range. Therefore, for differential methylation analysis, I converted the beta values into M-values, which are calculated as the log2 ratio of methylated to unmethylated intensities. M-values offer better statistical properties and are more suitable for linear modeling.

Illumina arrays contain two types of probes (type I and type II), which have different signal distributions. To correct for this design bias, I applied BMIQ (Beta MIxture Quantile) normalization using the wateRmelon package (version, ref). BMIQ adjusts type II probe values to match the distribution of type I probes, improving the overall comparability of methylation values.

The discovery dataset included samples from both 450K and EPIC arrays. The 450K and EPIC1 arrays were aligned to the hg19 reference genome, while EPIC2 arrays used hg38. I accounted for these differences in genome build throughout the analysis to ensure consistent probe annotation.

To add biological context, I used annotation resources from the RnBeads framework, which provide mappings of CpG probes to genomic features such as promoters, genes, CpG islands, and tiling regions. These annotations were later used to explore the distribution of significant DMPs across different genomic regions.

Differential methylation analysis was performed using the limma package. I created a design matrix to describe the comparison groups (e.g. primary vs recurrent) and fitted a linear model using lmFit(), followed by empirical Bayes moderation with eBayes(). Probes with an FDR-adjusted p-value below 0.05 were considered significant. In some cases, I applied an additional threshold based on delta-beta to highlight CpGs with both statistical and biological significance. Results were visualised with volcano plots and explored further by feature type enrichment.

For the validation cohort, I used publicly available data from the GLASS consortium. Raw IDAT files were not available for this dataset, but preprocessed beta values were provided (processed with preprocessNoob from minfi). I filtered this cohort to retain only IDH-wildtype (IDHwt) cases to match the discovery dataset. I also applied the same filtering and BMIQ normalization steps used in the discovery cohort to maintain consistency. These samples came from 450K and EPIC1 arrays and were aligned to the hg19 genome build.

4.2.3 Trends in Up and Down Responder Subtypes

To identify genomic regions with altered DNA methylation during glioblastoma recurrence, I calculated the ratio of average methylation levels in recurrent tumours relative to matched primary tumours. This was done for multiple region types, including gene bodies, promoters, CpG islands, enhancers, and tiling regions. For each region, I calculated two separate methylation ratios:

REC2PRIM_UP = (mean methylation in Up responder recurrent tumours + 0.01) / (mean methylation in Up responder primary tumours + 0.01)

REC2PRIM_DOWN = (mean methylation in Down responder recurrent tumours + 0.01) / (mean methylation in Down responder primary tumours + 0.01)

These values were derived independently from two differential methylation comparisons representing distinct patient response groups (referred to as "Up" and "Down" responders). The small constant (0.01) was added to each value to avoid division by zero and to stabilise the ratio when methylation levels were very low.

I then merged the data using a common region identifier and constructed two-dimensional scatter plots where REC2PRIM_DOWN values were plotted on the x-axis and REC2PRIM_UP values on the y-axis. Each point in the scatter plot represents a genomic feature, and its position reflects the relative change in methylation between recurrent and primary tumours in both comparison groups. To further investigate the functional implications of these methylation changes, genes previously identified as JARID2 binding site genes (JBSgenes), including those classified as LE50 or LE70 from our group's published work, were specifically highlighted on these two-dimensional scatter plots to examine their distribution within the quadrant analysis. The specific highlighting of JARID2 Binding Site genes (JBSgenes) and Leading Edge (LE50/LE70) genes is driven by the critical findings of Tanner et al. (2024), which established a plausible link between these genes, transcriptional reprogramming. The 'Leading Edge' (LE) refers to the subset of genes within a gene set that contributes most to the enrichment score in a Gene Set Enrichment Analysis (GSEA), indicating their consistent and significant alteration; LE50 genes are those found in the leading edge in at least 50% of patients, while LE70 genes are found in at least 70% of patients.

To identify outlier regions, I calculated the mean and standard deviation (SD) of methylation ratios across all features. I defined a 99% confidence interval using the formula ±2.576 × SD, and used these thresholds to delineate the following categories:

- Top_Right: high methylation ratios in both Up and Down groups
- Top_Left: high in Up, low in Down
- Bottom_Right: low in Up, high in Down
- Bottom_Left: low in both groups
- Right_Extreme and Left_Extreme: extreme values in Down only
- Top_Extreme and Bottom_Extreme: extreme values in Up only
- Within_CI99: values within 99% CI for both axes

These categories allowed me to classify genomic regions based on how consistently and strongly their methylation changed across patient subtypes. I extracted the gene IDs, coordinates, and annotations for gene body and promoter features that fell outside the 99% CI, creating region-specific gene lists. These gene lists were subsequently submitted to clusterProfiler (*version 4.14.3*) to perform Gene Ontology (GO) enrichment analysis using the Over-Representation Analysis (ORA) method, examining Biological Process (BP), Cellular Component (CC), and Molecular Function (MF) terms. I applied the following parameters (Gene = entrez_ids, OrgDb = org.Hs.eg.db, keyType = "ENTREZID", ont = ont, pAdjustMethod = "BH", pvalueCutoff = 0.05, qvalueCutoff = 0.2). The results for each gene list were then grouped by GO category (BP, CC, and MF) to generate multi-faceted bubble charts for visualization of the enriched terms.

The scripts used for analyses in this chapter is available at:
(https://github.com/umyma1/thesis_appendix/tree/main/chapter4).

4.3 RESULTS

4.3.1 Cohort description

This chapter presents an analysis of methylation profiles across two cohorts: a discovery cohort and a validation cohort. Both cohorts comprise longitudinal glioblastoma (GBM) samples profiled using Illumina methylation arrays, which utilise either the 450K or EPIC platforms. For each patient, samples from the primary tumour and a corresponding local recurrence were included when available.

Tables 4-1 and -4-2 detail the samples groupings. The discovery cohort consists of 56 samples, representing 27 matched primary–recurrent pairs. These include 24 samples processed on the 450K array (12 pairs), 20 samples on the EPIC1 array (10 pairs), and 12 samples on the EPIC2 array (5 pairs). The validation cohort includes 114 samples from 57 matched primary–recurrent pairs, of which 34 samples (17 pairs) were profiled using the 450K array and 80 samples (40 pairs) using the EPIC1 array.

In a subset of samples, stratification based on responder subtype was performed, dividing patients into UP and DOWN responders. This classification was only possible for a proportion of the dataset, as the classification relies on matched RNAseq data which was not available for all samples. Within the discovery cohort, stratified samples included 8 samples UP and 8 samples DOWN responders on the 450K array (4 pairs each), 4 samples UP and 2 samples DOWN responders on EPIC1 (2 and 1 pair, respectively), and 4 samples UP and 6 samples DOWN responders on EPIC2 (2 and 3 pairs, respectively). In the validation cohort, stratified samples consisted of 2 samples UP responders (1 pair) from the 450K array, and 16 samples UP and 4 samples DOWN responders from the EPIC1 platform (8 and 2 pairs, respectively). No DOWN responders were identified on the 450K validation set.

Table 4-1: The discovery cohort – sample breakdown

	No. Patients	No. Samples	Array	UP	DOWN	Unstratified
Discovery	12	24	450K	4 pairs	4 pairs	4 pairs
	10	20	EPICv1	2 pairs	1 pairs	7 pairs
	5	12	EPICv2	2 pairs	3 pairs	0 pairs
Total	27	56	-	8 pairs	8 pairs	11 pairs

Table 4-2: The validation cohort – sample breakdown

	No. Patients	No. Samples	Array	UP	DOWN	Unstratified
Validation	17	34	450K	1 pair	0 pairs	17 pairs
	40	80	EPICv1	8 pairs	2 pairs	29 pairs
Total	57	114	-	9 pairs	2 pairs	46 pairs

This collection of longitudinally sampled GBM tumours forms the basis for two analyses presented in this chapter. First, differential methylation analysis (from primary to matched recurrence) was performed across the full cohort, including all patients regardless of responder subtype annotation. Second, cohorts were stratified and analysed separately for UP and DOWN responders. This stratified analysis allowed for the exploration of epigenetic differences associated with each subtype, specifically, and their potential role in tumour progression.

4.3.2 Quality Contol (QC)

For the discovery cohort, raw methylation data were available in the form of IDAT files and were processed using the minfi R package with the preprocessNoob function. This method applies background correction and dye-bias adjustment using a normal-exponential convolution model. In parallel, the same samples were also analysed using the QC functions built into the RnBeads package.

For the validation cohort, only pre-processed beta values were available, having already been normalised using the minfi pipeline with preprocessNoob (Consortium, 2018). As the raw IDAT files were not accessible, these samples were not processed through RnBeads.

In all workflows, background correction was applied, and standard quality control steps were implemented to exclude unreliable probes. This included the removal of probes with poor detection p-values, probes overlapping known single nucleotide polymorphisms (SNPs), and those reported to be cross-reactive. For the RnBeads pipeline, default parameters were used. These included the exclusion of probes on sex chromosomes to avoid potential confounding due to gender-specific methylation patterns, and the activation of “greedycut” filtering, which removes probes failing QC in any sample across the dataset. SNP filtering was applied to ensure exclusion of probes with SNPs at or near the target CpG site.

Two recurrent samples in the EPIC2 set, which were extra replicates, did not pass QC and were excluded from downstream analysis.

The number of probes retained after QC varied depending on the array platform and preprocessing approach, highlighting inherent differences between minfi and RnBeads. This resulted in a final dataset comprising 54 samples for the discovery cohort and 114 samples for the validation cohort, which were subsequently used for downstream methylation analysis (Tables 4-3 and 4-4).

Table 4-3: Number of samples in the discovery cohort after QC

	Patient	Samples	Array	UP	DOWN	Unstratified
Discovery	12	24	450K	4 pairs	4 pairs	4 pairs
	10	20	EPICv1	2 pairs	1 pairs	7 pairs
	5	10	EPICv2	2 pairs	3 pairs	0 pairs
Total	27	54	-	8 pairs	8 pairs	11 pairs

Table 4-4: Number of samples in the validation cohort after QC

	Patient	Samples	Array	UP	DOWN	Unstratified
Validation	17	34	450K	1 pair	0 pairs	17 pairs
	40	80	EPICv1	8 pairs	2 pairs	29 pairs
Total	57	114	-	9 pairs	2 pairs	46 pairs

4.3.3 MGMT methylation

To evaluate MGMT promoter methylation dynamics during glioblastoma progression, I used the MGMT-STP27 classifier. This tool infers methylation status from Illumina array data by combining the signals from two CpG probes (cg12434587 and cg12981137) located in the MGMT promoter. Methylation at these sites is strongly linked to the transcriptional silencing of MGMT, which predicts increased sensitivity to alkylating agents such as temozolomide. The classifier provides a continuous score that is divided into a binary status Methylated (M) or Unmethylated (U) using a validated cutoff. Unlike the Cancer Cell Fraction (CCF) estimates described in the previous chapters, which quantify the proportion of tumour cells carrying a given genetic alteration, the MGMT-STP27 score reflects the average methylation level across all cells in the sample. Therefore, it does not represent clonal fractions but rather the bulk methylation state of the MGMT promoter region.

In the discovery cohort (n = 27), most patients retained the same MGMT methylation status between primary and recurrent tumours. Specifically, 24 out of 27 patients (88.9%) showed no change: 10 remained methylated (M>M) and 14 remained unmethylated (U>U). Three patients (11.1%) exhibited a change in MGMT status at recurrence—two gained methylation (U>M) and one lost methylation (M>U).

To investigate whether response subtype might be linked to MGMT methylation changes, I divided the cohort into UP (n = 8) and DOWN (n = 8) responders based on longitudinal tumour progression data. Among UP responders, 7 out of 8 exhibited stable MGMT status (3 M>M, 4 U>U), while one switched from U>M. Similarly, among DOWN responders, 7 out of 8 were also stable (3 M>M, 4 U>U), with one switching from M>U. A chi-square test showed no significant association between response subtype and MGMT switching ($p = 0.5724$).

Although MGMT methylation is a well-established prognostic marker, my focus here was not on survival prediction but on examining how MGMT status may evolve in relation to treatment response subtype.

In the larger validation cohort (n = 57), MGMT methylation changes between primary and recurrent tumours were observed in eight patients (14%). Of these, five patients lost methylation (M>U), while three gained methylation (U>M). The remaining 49 patients (86%) maintained a stable status across all time points.

When stratified by response subtype, nine patients were UP responders and two were DOWN responders. Among the UP group, eight showed stable MGMT status (2 M>M, 6 U>U), while one patient switched from M>U. Both DOWN responders maintained a stable, unmethylated status (U > U). Again, no statistically significant association was found between response subtype and MGMT switching ($p = 0.6323$).

These results suggest that while MGMT methylation changes can occur during disease progression, they do not appear to be associated with responder subtype (when known). Nevertheless, the presence of switching events—particularly the more frequent M>U loss—may reflect underlying tumour evolution or treatment-induced selective pressure. Prior studies (Choi et al., 2021, Brandes et al., 2017, Birzu et al., 2020) have

reported MGMT switching in approximately 22% of cases, a rate similar to our findings in the validation cohort and twice that of the discovery cohort.

4.3.4 Differential methylation analysis

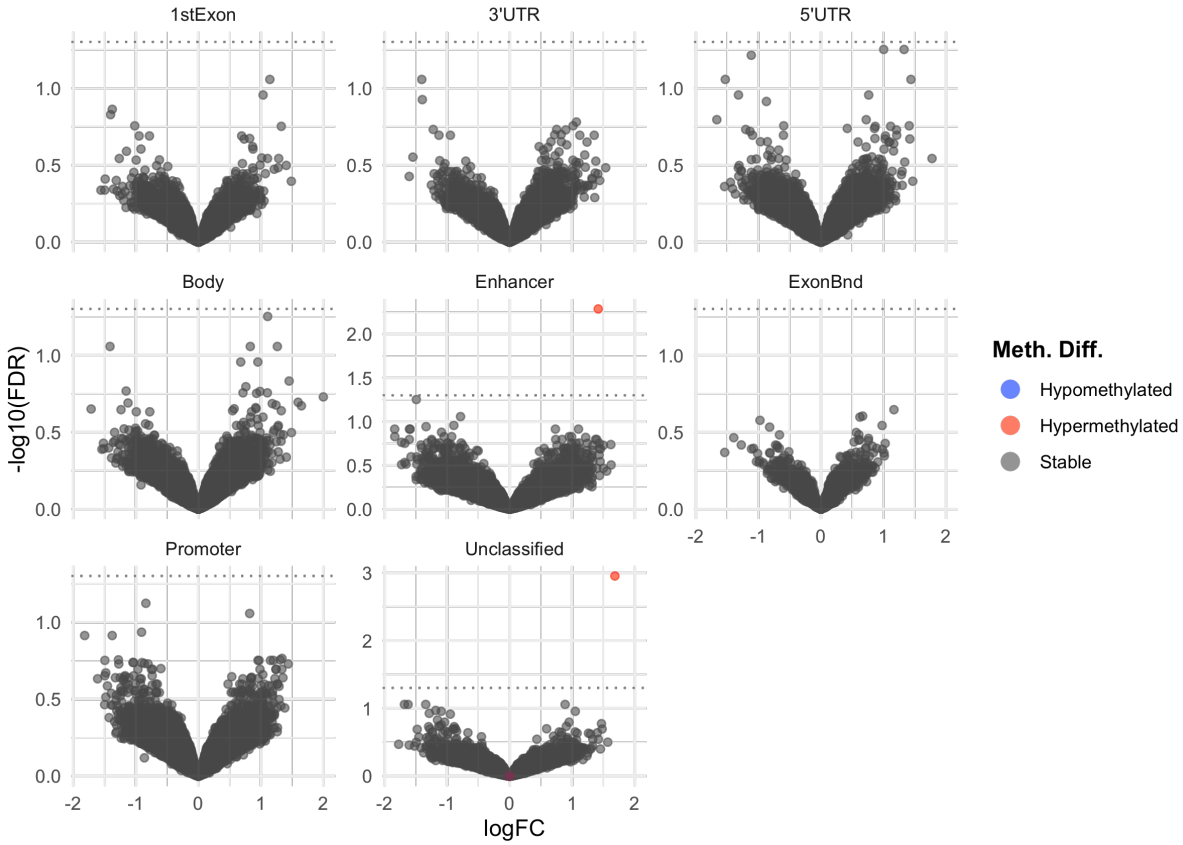
In this part of the analysis, I explored DNA methylation changes between primary and recurrent glioblastoma (GBM) tumours using longitudinal samples from matched patients. GBM is a highly heterogeneous and aggressive brain tumour, and while its genetic drivers have been extensively studied, the role of epigenetic changes—particularly DNA methylation—in tumour recurrence remains less well understood. To address this, I conducted differential methylation analysis to determine whether there are consistent changes in the methylation landscape during tumour progression.

4.3.4.1 Identification of Differentially Methylated Probes (DMPs)

As a first step, I focused on identifying differentially methylated positions (DMPs) using the limma package. This was done independently for the discovery cohort and the validation cohort, as they were processed differently and derived from different array platforms. The analysis compared methylation levels between matched primary and recurrent tumour samples.

In the discovery cohort, no significant DMPs were identified (adjusted p-value < 0.05), whereas the validation cohort yielded 2,332 significant DMPs under the same threshold. These significant sites were then grouped according to functional genomic regions to assess where changes were most concentrated, such as promoters, gene bodies, and enhancer regions (Figure 4-2). This regional annotation provided preliminary insight into which genomic region might be more susceptible to epigenetic alterations during the recurrence of GBM.

A


1stExon
3'UTR
5'UTR
Body
Enhancer
ExonBnd
Promoter
Unclassified
-log10(FDR)
logFC
Meth. Diff.
Hypomethylated
Hypermethylated
Stable

B

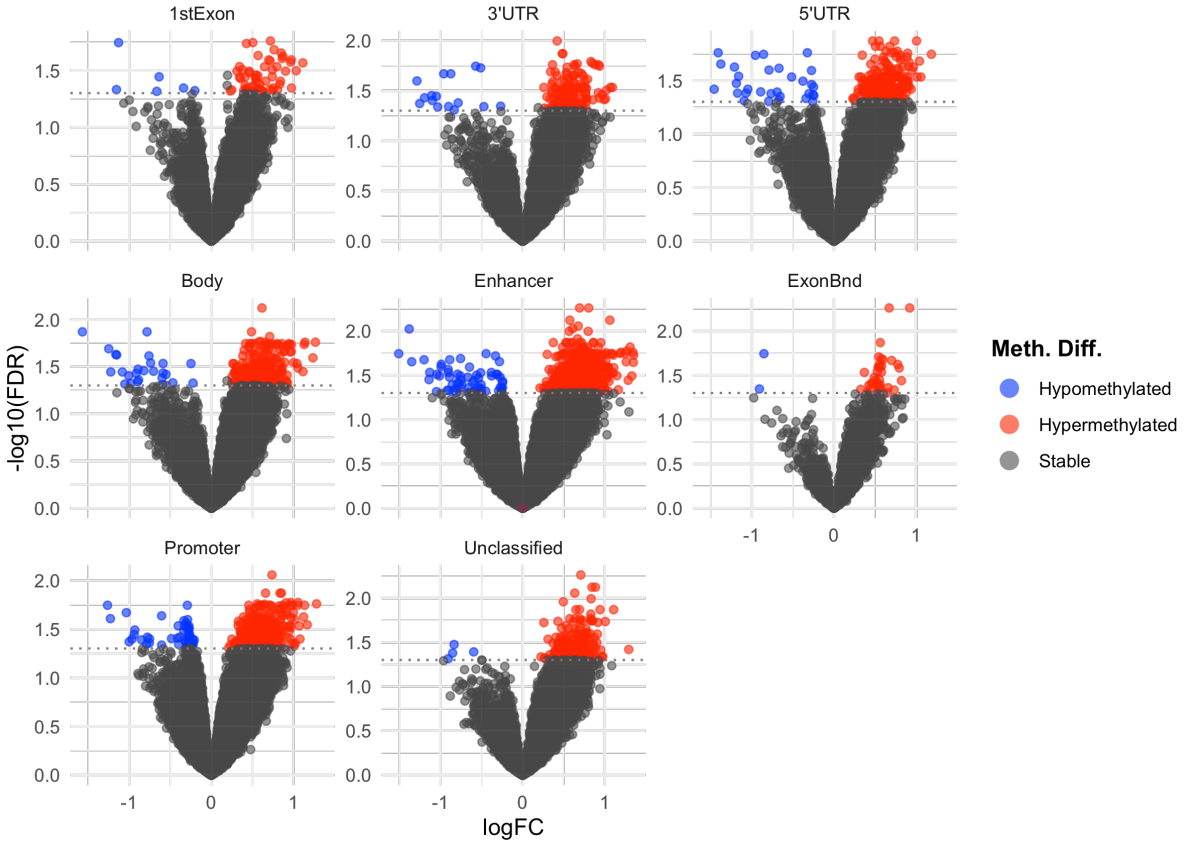

1stExon
3'UTR
5'UTR
Body
Enhancer
ExonBnd
Promoter
Unclassified
-log10(FDR)
logFC
Meth. Diff.
Hypomethylated
Hypermethylated
Stable

Figure 4-2: Volcano plots showing log-fold changes in DNA methylation between recurrent and primary GBM samples across different genomic regions.

Each panel corresponds to a specific genomic annotation. The x-axis represents the LogFC in average methylation between recurrent and primary tumours, with positive values indicating hypermethylation in recurrence and negative values indicating hypomethylation. The y-axis shows the $-\log_{10}$ of the false discovery rate (FDR), reflecting statistical significance. The dashed horizontal line marks the FDR threshold for significance. Points above this line are coloured by the direction of change: red for significantly increased methylation over time, blue for significantly reduced methylation over time, and grey for non-significant changes in methylation from primary to recurrence.
A- Discovery Cohort
B- Validation Cohort

4.3.4.2 Region-Level Analysis and DMR Calling

Following the identification of DMPs, I utilised two region-based packages, DMRcate and Bumphunter, to identify differentially methylated regions (DMRs). These tools aggregate neighbouring probes to detect coordinated changes in broader genomic regions, which is often more biologically meaningful than isolated single-site changes.

However, despite the presence of significant DMPs in the validation cohort, no significant DMRs were detected by either method in either cohort. This apparent discrepancy suggests that while individual CpG sites may undergo minor methylation shifts, these changes do not cluster tightly enough within functional elements to meet the statistical criteria for DMRs. Another possibility is that inter-patient variability, small effect sizes, or limited sample numbers reduced the statistical power to detect regional methylation changes with confidence.

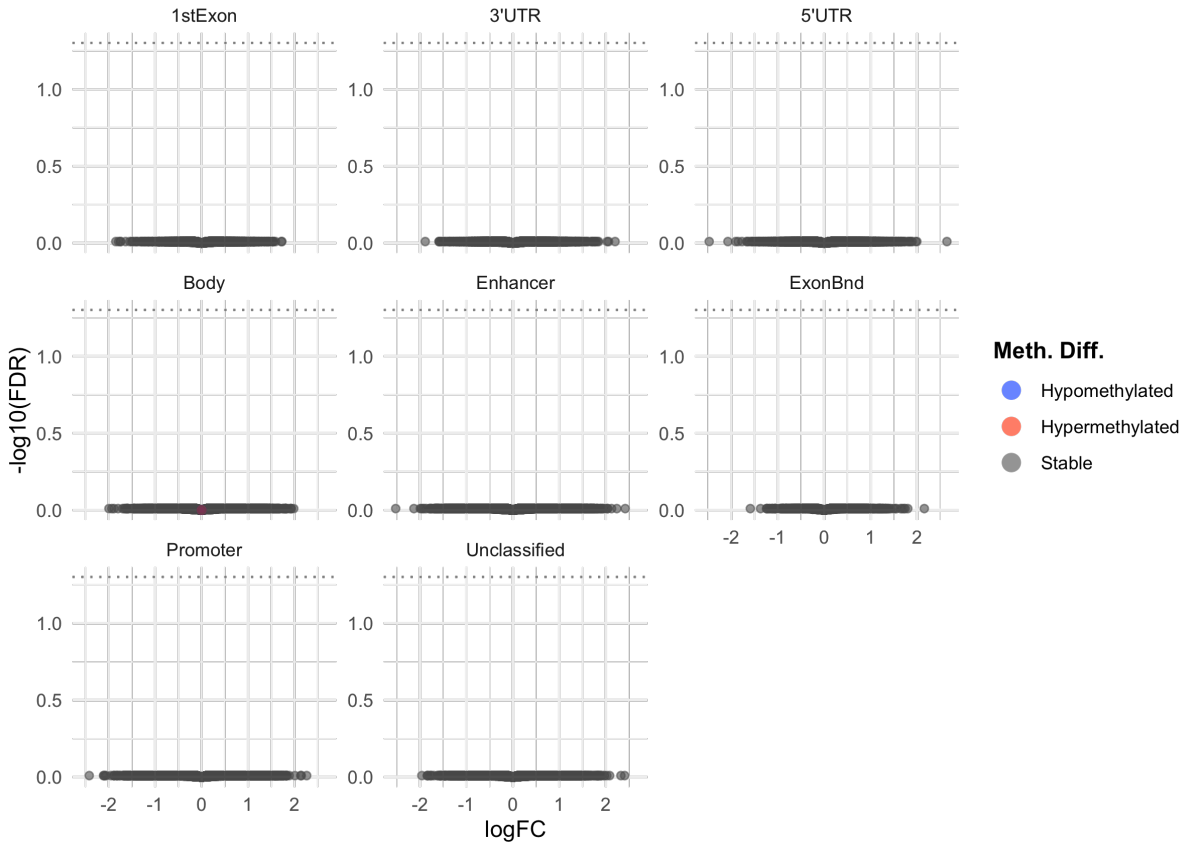
4.3.4.3 Biological Stratification by Responder Subtype

To refine the analysis further and test whether epigenetic changes might be more prominent in specific biological subgroups, I repeated the full DMP and DMR analysis using responder subtype-based stratification. Patients were divided into UP and DOWN responders based on the changes in expression of genes that have a JARID2 binding site in their promoter between primary and recurrent tumours.

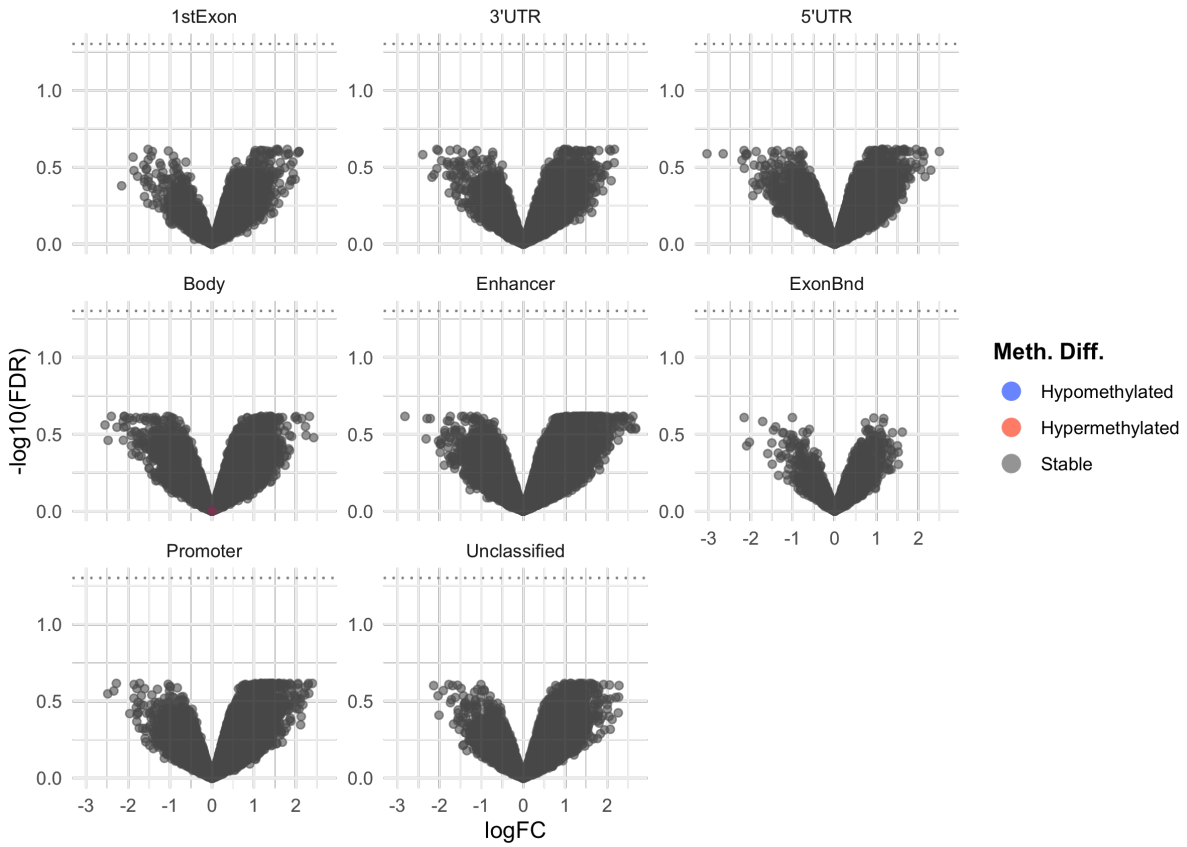
I performed the analysis this time by separating the samples into UP and DOWN responders. Our group (Tanner et al., 2024) has proposed that JARID2 plays a role in promoting GBM recurrence after treatment by enabling transcriptional reprogramming in surviving tumour cells, thereby helping to restore the phenotypic heterogeneity required for tumour regrowth. This reprogramming mechanism may represent a therapeutic vulnerability in GBM, and as such, exploring its epigenetic underpinnings through methylation profiling is of particular interest. The goal of this stratified analysis was therefore to determine whether DNA methylation changes between primary and recurrent tumours might be driving the differential changes we see in gene expression, from primary to recurrence, across the response subgroups, potentially shedding light on the epigenetic basis of this proposed reprogramming mechanism.

The same R packages, limma, DMRcate, and Bumphunter, were used for the stratified groups. As with the unstratified patients, the analysis revealed no significant DMPs from primary to recurrence in either the UP or DOWN responder subtypes (Figure 4-3).

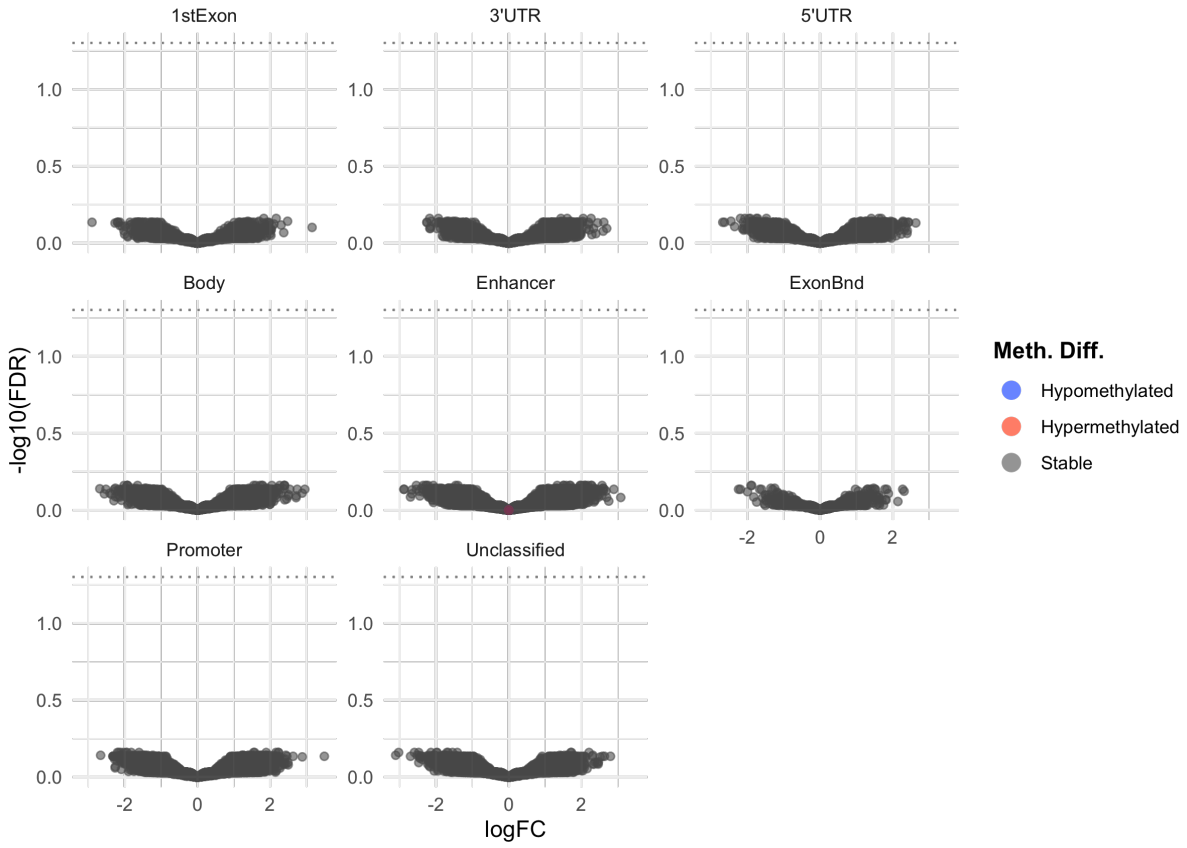
A



B


1stExon
3'UTR
5'UTR
Body
Enhancer
ExonBnd
Promoter
Unclassified
-log10(FDR)
logFC
Meth. Diff.
Hypomethylated
Hypermethylated
Stable

C


1stExon
3'UTR
5'UTR
Body
Enhancer
ExonBnd
Promoter
Unclassified
-log10(FDR)
logFC
Meth. Diff.
Hypomethylated
Hypermethylated
Stable

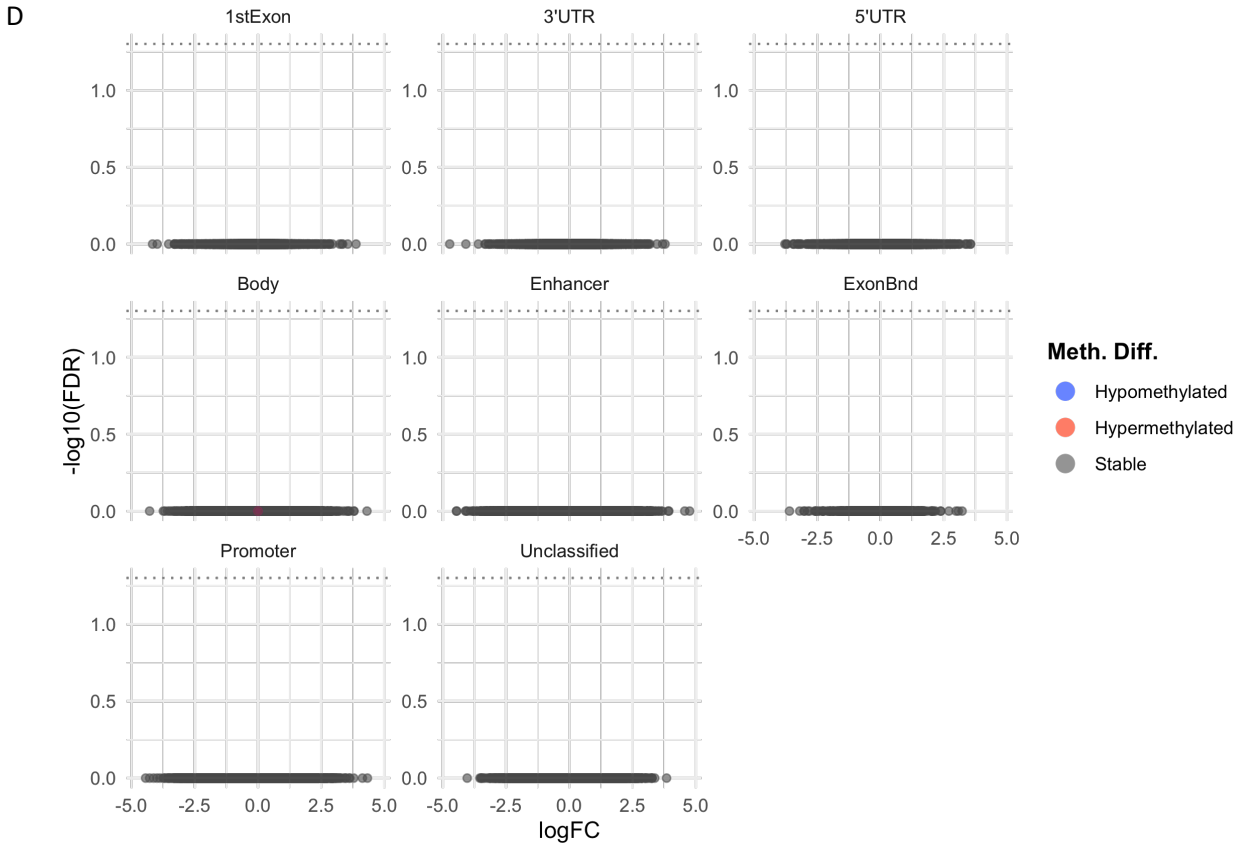


Figure 4-3: Volcano plots showing log-fold changes in DNA methylation between recurrent and primary GBM samples, stratified by response group and cohort.

Each panel corresponds to a specific genomic annotation. The x-axis represents the LogFC in average methylation between recurrent and primary tumours, with positive values indicating hypermethylation in recurrence and negative values indicating hypomethylation. The y-axis displays the $-\log_{10}$ of the false discovery rate (FDR), representing statistical significance. The dashed horizontal line marks the FDR threshold used to identify significant DMPs. Points above the threshold are coloured by the direction of change: red indicates significantly hypermethylated regions in recurrence, blue indicates significantly hypomethylated regions and grey indicates non-significant changes.

- Panels A and B represent UP responders, with A corresponding to the discovery cohort and B to the validation cohort.
- Panels C and D represent DOWN responders, with C corresponding to the discovery cohort and D to the validation cohort.

4.3.4.4 Re-analysis Using RnBeads

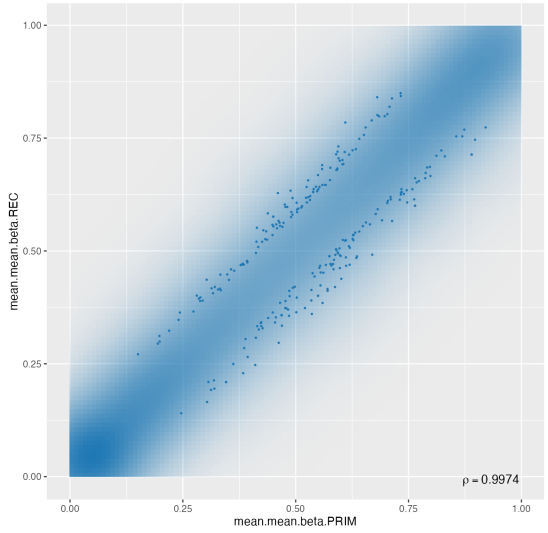
Given that the earlier approach relied on using separate tools for different parts of the analysis, and that raw IDAT files were not available for the validation cohort, I then repeated the analysis using RnBeads on the discovery cohort only. RnBeads provides an end-to-end pipeline for preprocessing, normalisation, annotation, and differential analysis, and helps ensure consistency across analytical stages.

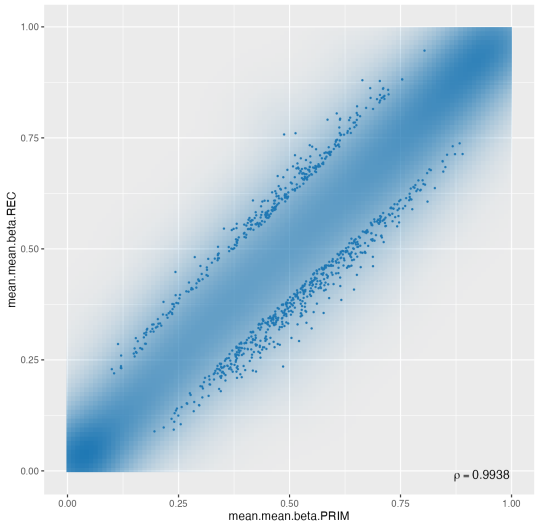
The RnBeads workflow was applied separately for each platform (450K, EPIC1, and EPIC2) within the discovery cohort. Again, no significant DMRs were detected between primary and recurrent samples (Figures 4-4,4-5,4-6). This confirmed that the earlier results were not due to inconsistencies between packages or preprocessing steps.

Promoter

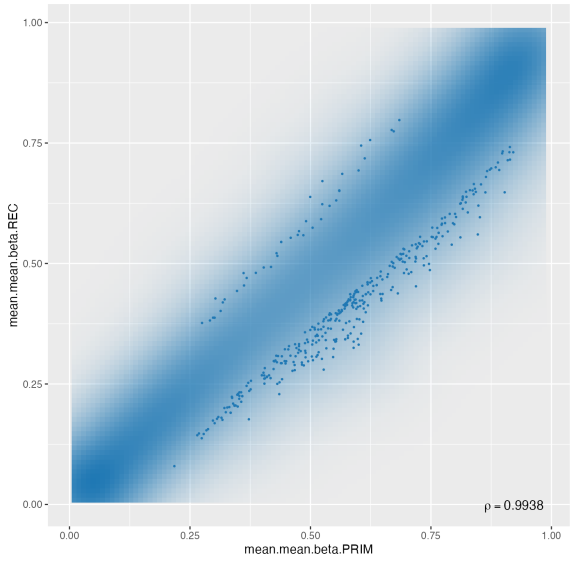
Enhancer

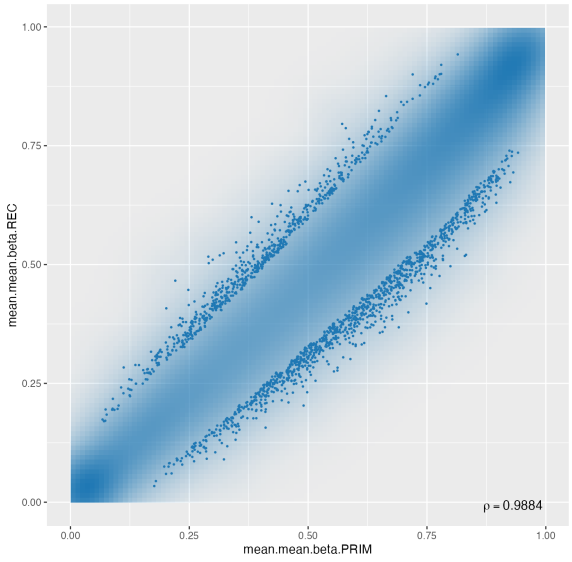
450K array



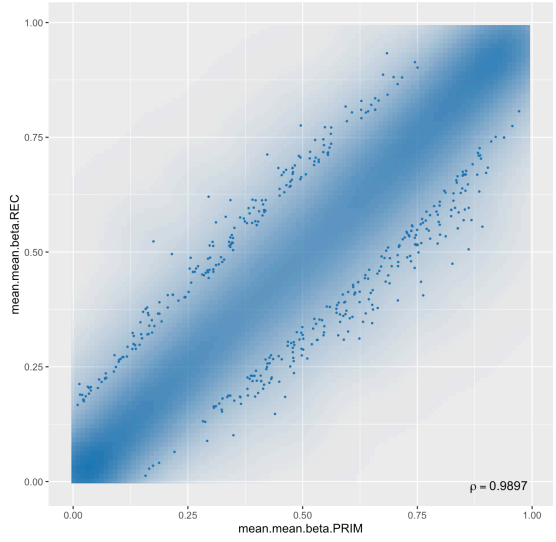


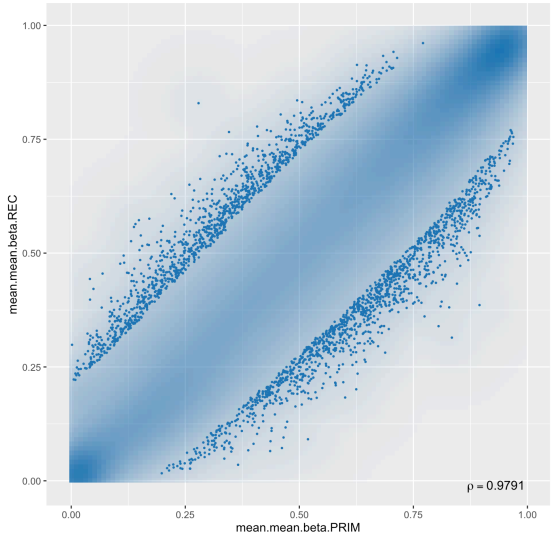
EPICv1 array

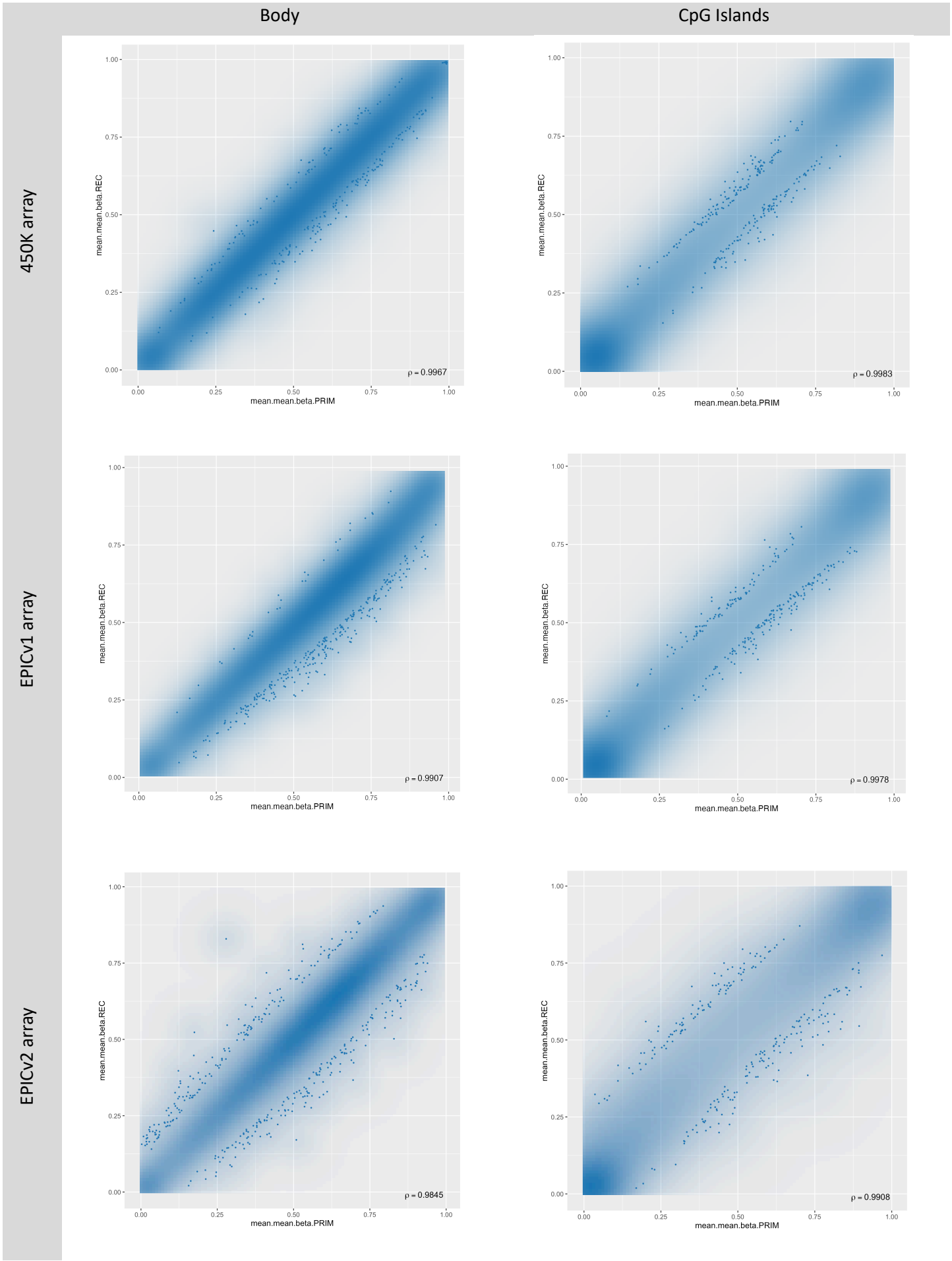




EPICv2 array






Body
CpG Islands
450K array
EPICv1 array
EPICv2 array
mean.mean.beta.REC
mean.mean.beta.PRIM
ρ = 0.9967
ρ = 0.9983
ρ = 0.9907
ρ = 0.9978
ρ = 0.9845
ρ = 0.9908

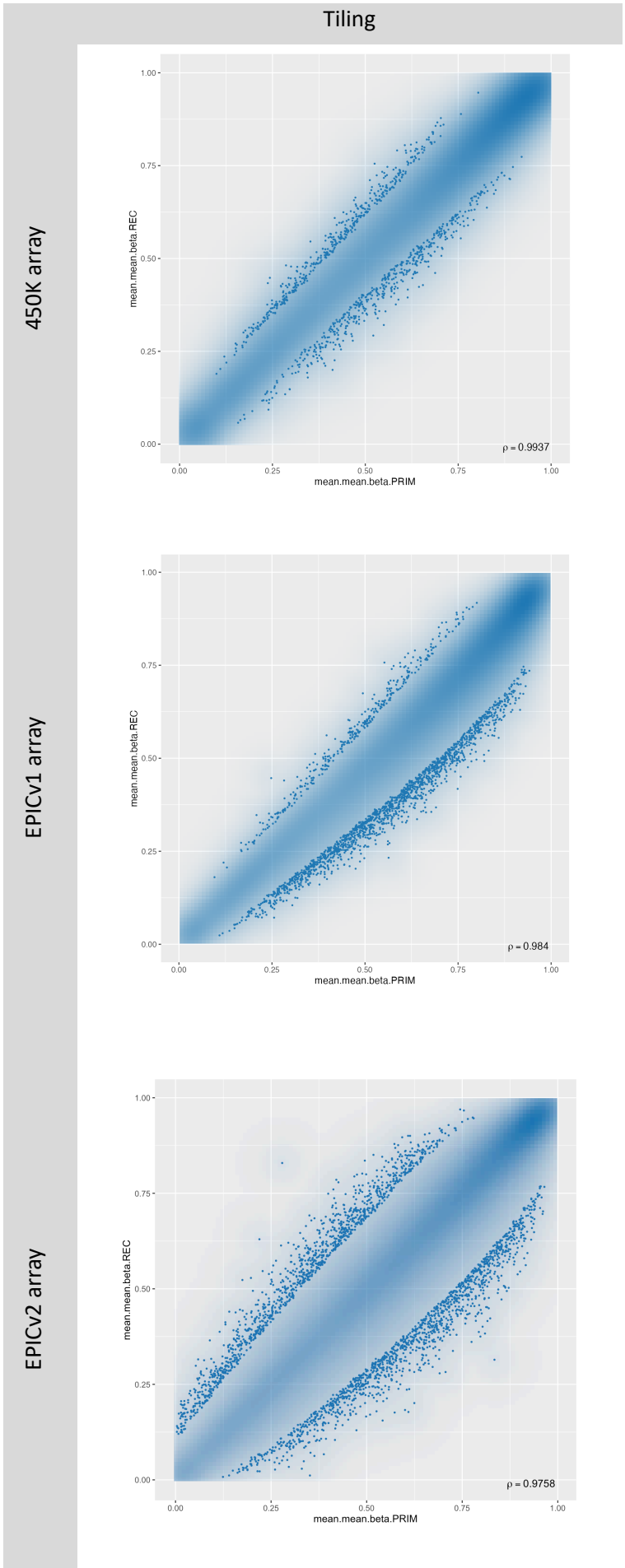


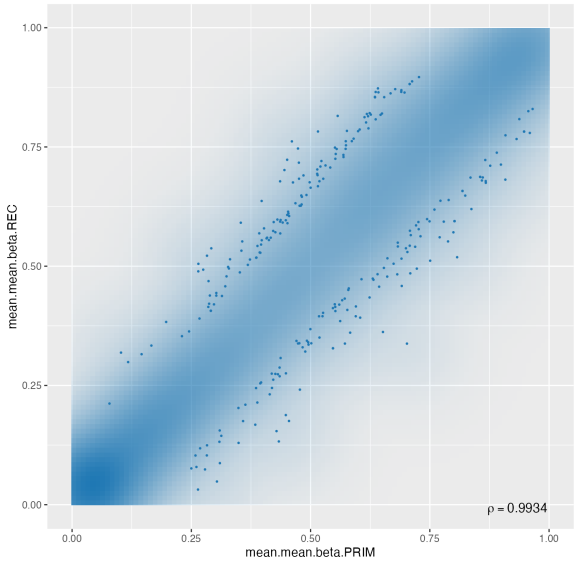
Figure 4-4: Scatter plots showing unstratified primary vs recurrent methylation across different arrays and region types.

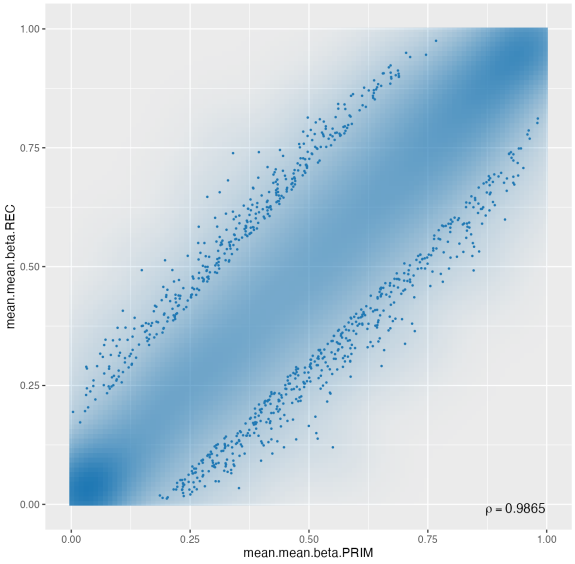
Each point represents a genomic region, with average beta value in primary tumours (x-axis) and recurrent tumours (y-axis). Arrays (450K, EPICv1, EPICv2) are arranged in rows, and region types (e.g., promoters, gene bodies, tiling, CpG islands) in columns. A diagonal trend is included in each plot, and the Pearson correlation coefficient (ρ) quantifies the linear association. Point colour reflects local density, with warmer tones indicating higher regional concentration. Statistically significant differentially methylated regions (DMRs) are highlighted in red.

Promoters

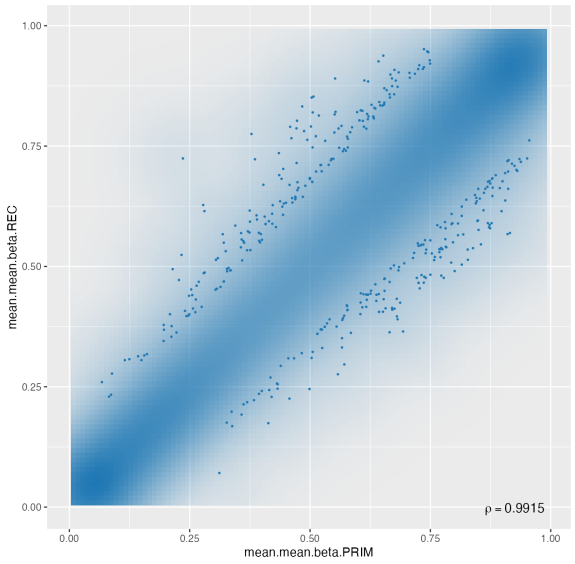
Enhancers

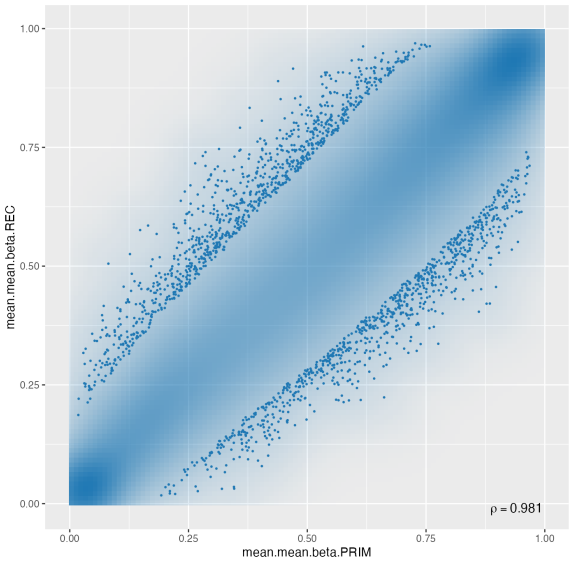
450K array



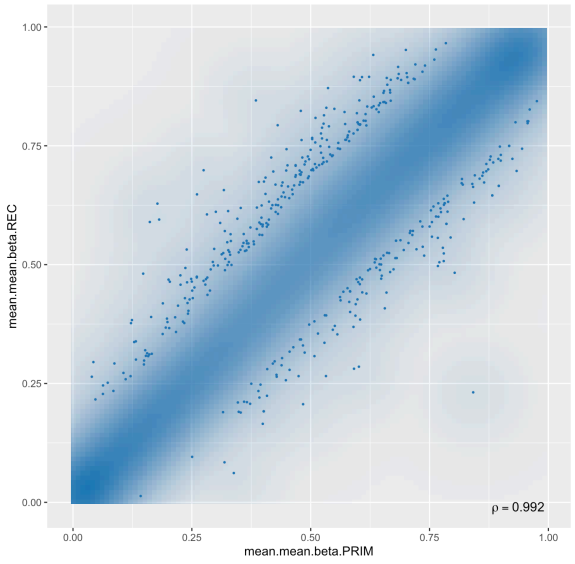


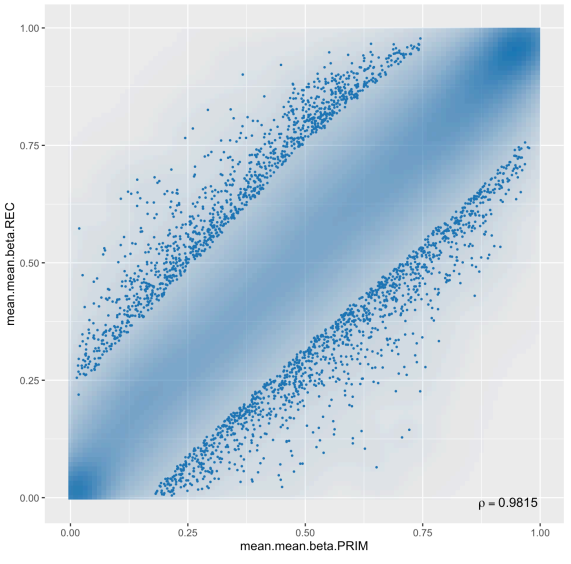
EPICv1 array

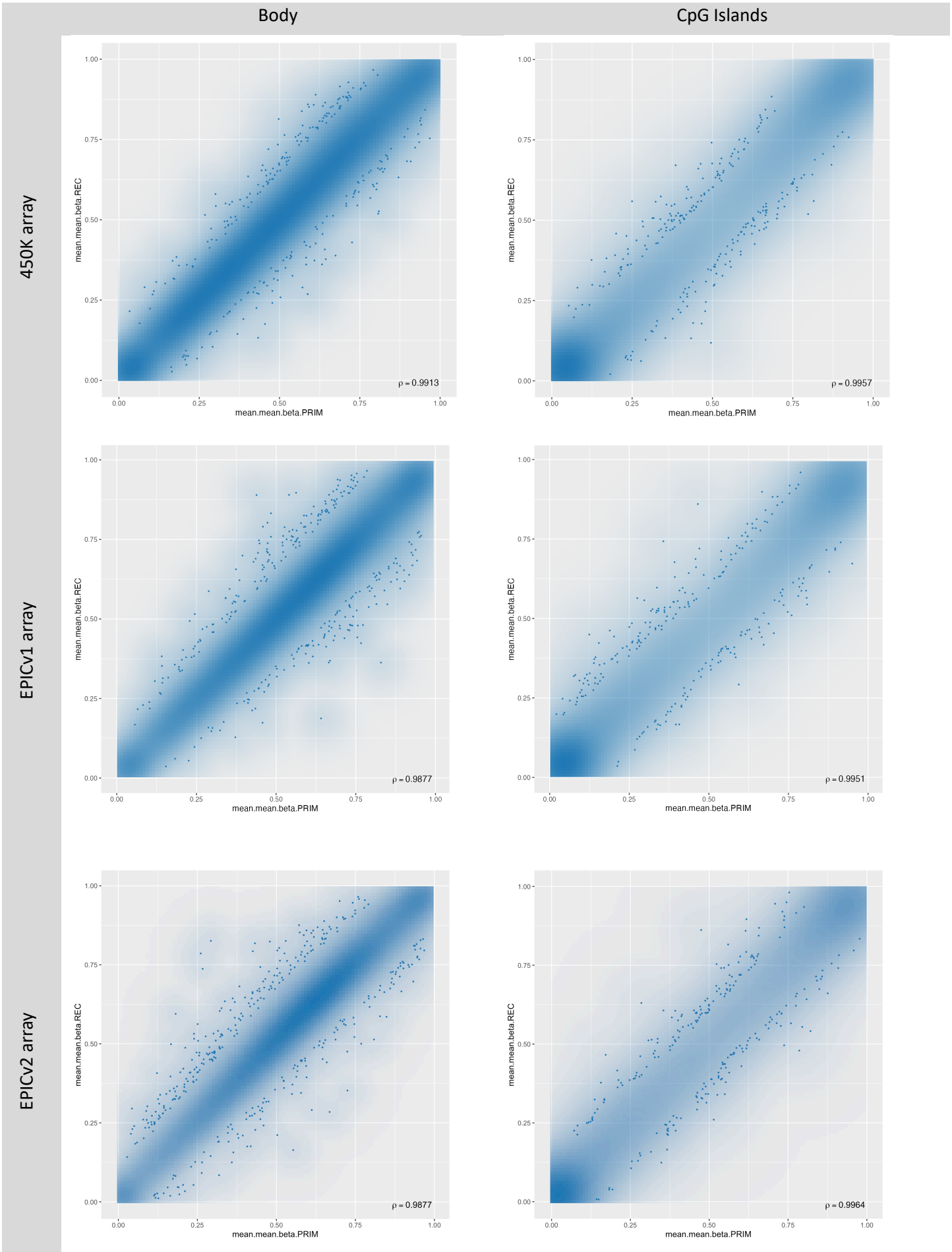




EPICv2 array






Body
CpG Islands
450K array
EPICv1 array
EPICv2 array
mean.mean.beta.REC
mean.mean.beta.PRIM
ρ = 0.9913
ρ = 0.9957
ρ = 0.9877
ρ = 0.9951
ρ = 0.9877
ρ = 0.9964

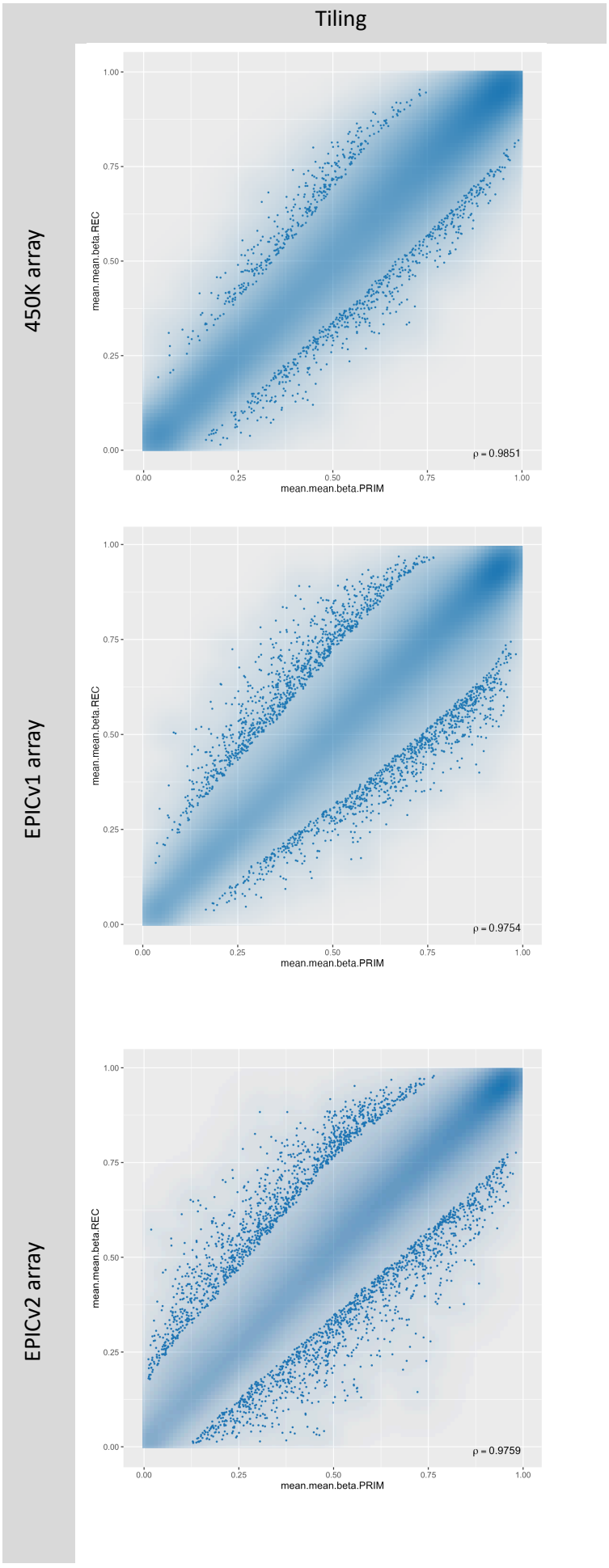
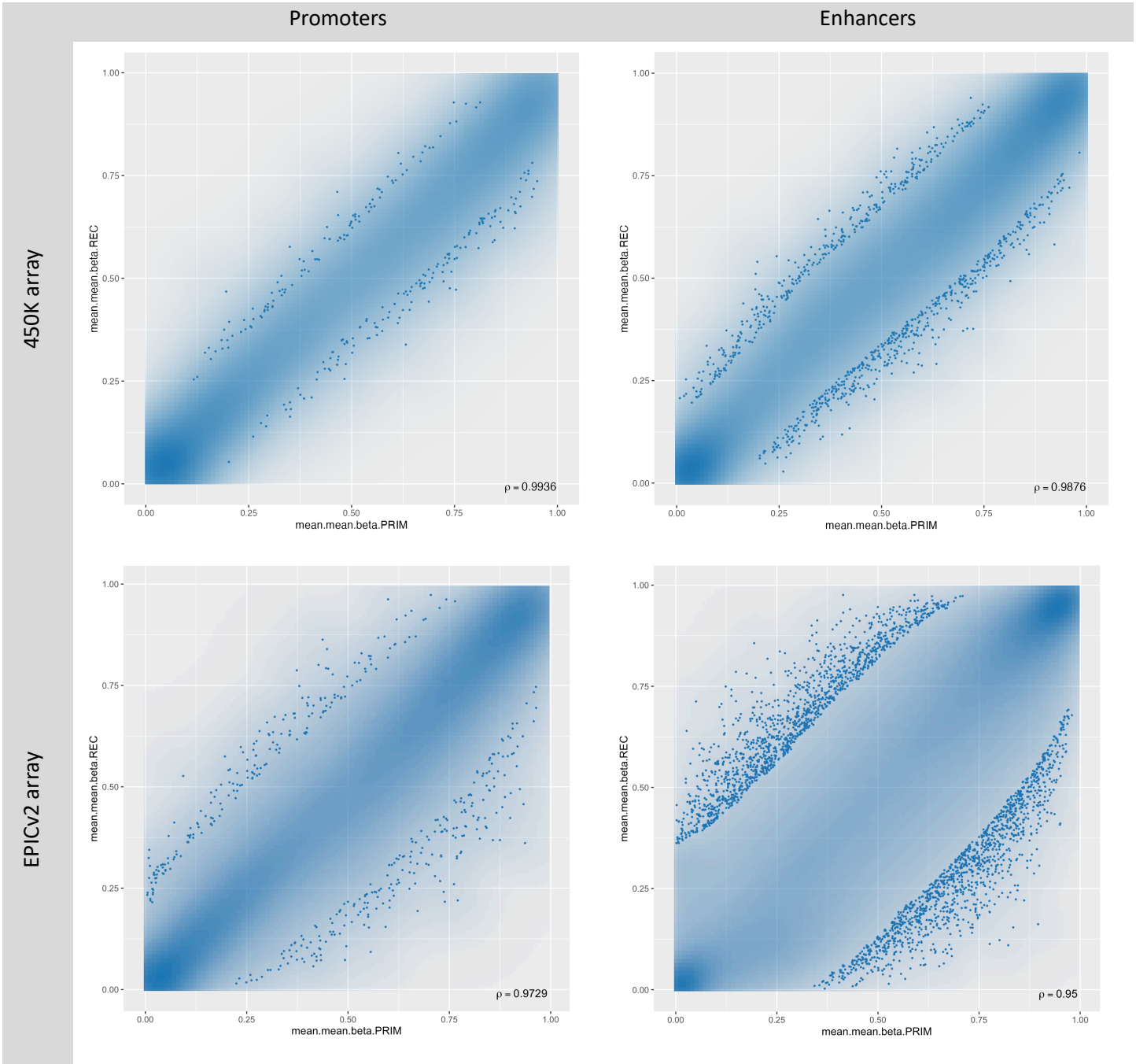
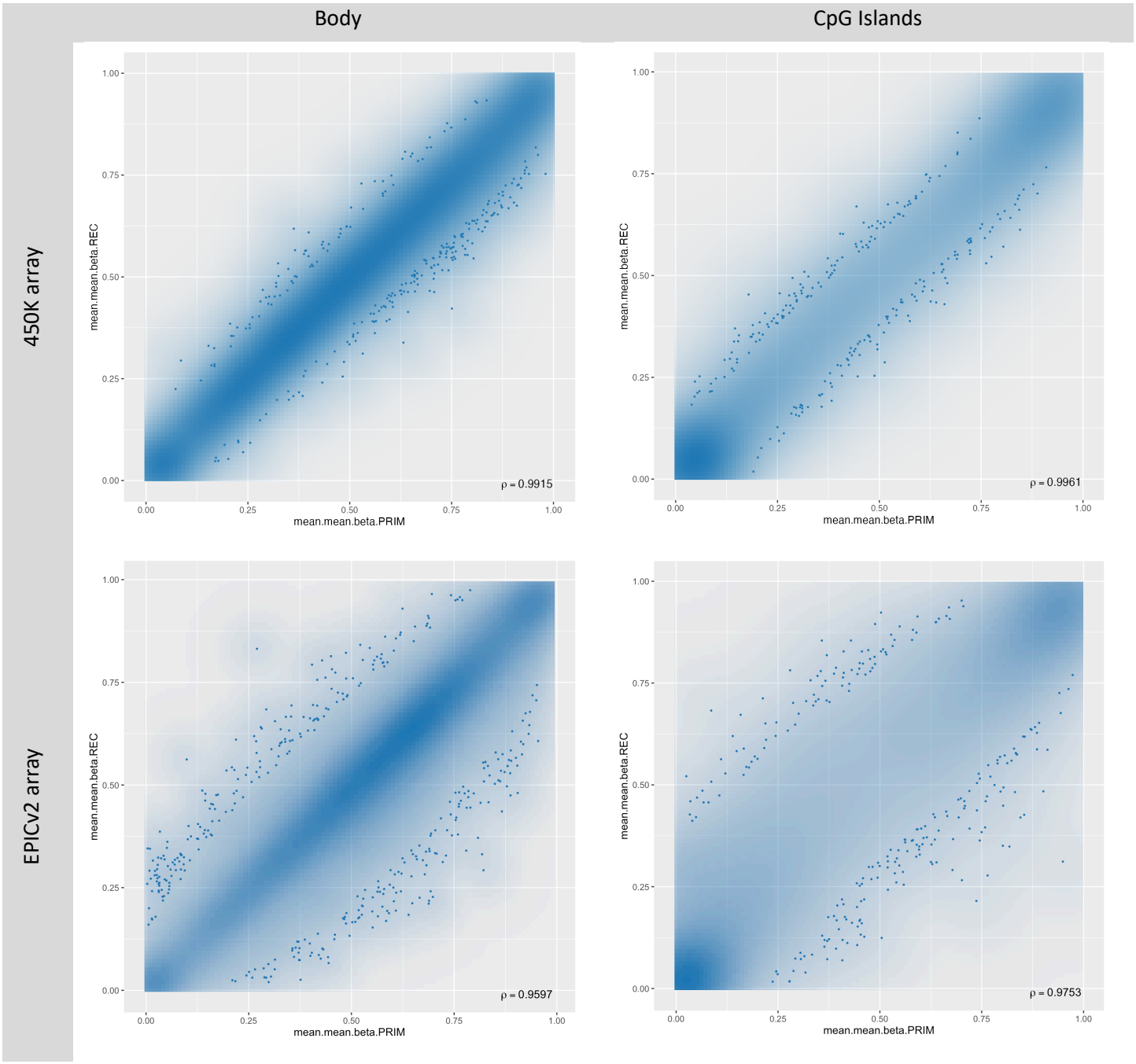


Figure 4-5: Scatter plots comparing methylation profiles in UP responders across arrays and region types.

Each point represents a genomic region's average beta value in primary (x-axis) and recurrent (y-axis) GBM samples. Rows correspond to arrays (450K, EPICv1, EPICv2), and columns to region types. Diagonal trends and Pearson correlation coefficients (ρ) are shown in each panel. Point colour reflects density, and significant DMRs are marked in red.


Promoters
Enhancers
450K array
EPICv2 array
mean.mean.beta.REC
mean.mean.beta.PRIM
ρ = 0.9936
ρ = 0.9876
ρ = 0.9729
ρ = 0.95
0.00
0.25
0.50
0.75
1.00


Body
CpG Islands
450K array
EPICv2 array
mean.mean.beta.REC
mean.mean.beta.PRIM
ρ = 0.9915
ρ = 0.9961
ρ = 0.9597
ρ = 0.9753
0.00
0.25
0.50
0.75
1.00

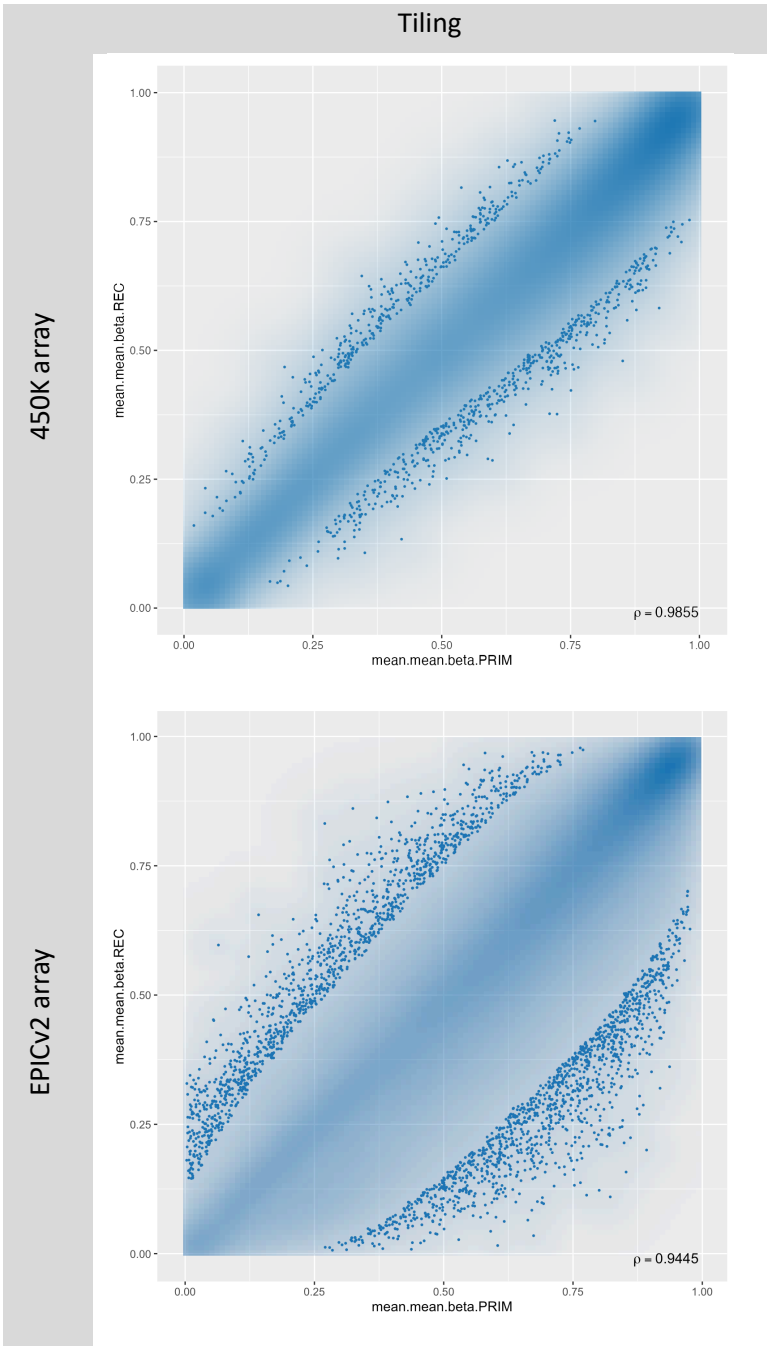


Figure 4-6: Scatter plots comparing methylation in DOWN responders across arrays and region types.

Each point represents the average methylation of a genomic region in primary (x-axis) and recurrent (y-axis) tumours. Arrays (450K and EPICv2) are shown in rows; region types are in columns. Diagonal trends and Pearson correlation coefficients (ρ) are displayed. Colour intensity reflects point density. Note: EPICv1 was excluded due to insufficient sample size, which prevented RnBeads from completing the analysis. Red points indicate significant DMRs.

4.3.4.5 Combined Platform Analysis

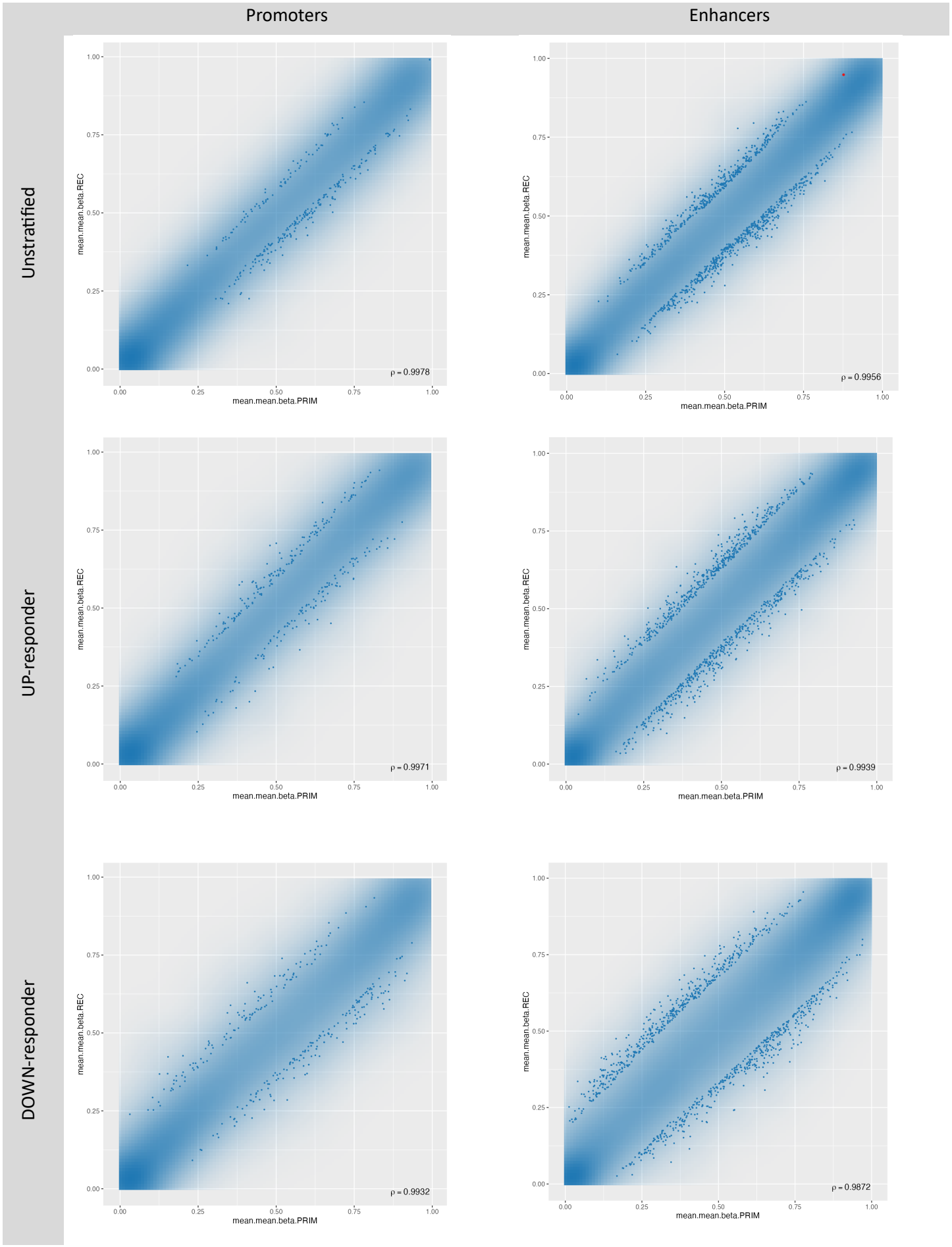
To address the possibility that the lack of significant findings was due to underpowered subgroup sizes, I next combined the datasets from all platforms to maximise sample size. Given that each methylation array platform (450K, EPIC1, and EPIC2) contains a different number and distribution of probes, I limited the analysis to probes that were common across all three platforms (n = 347430 probes) to ensure comparability. The combined dataset was analysed again using RnBeads. Although no DMRs reached statistical significance, some trends were observed in enhancer and promoter regions, particularly within the stratified responder subgroups. In some loci, consistent directional shifts in methylation were observed across multiple samples, particularly in DOWN responders, which may hint at underlying biological changes that are subtle and cohort-specific. However, these did not meet formal criteria for statistical significance, likely due to sample size and heterogeneity.

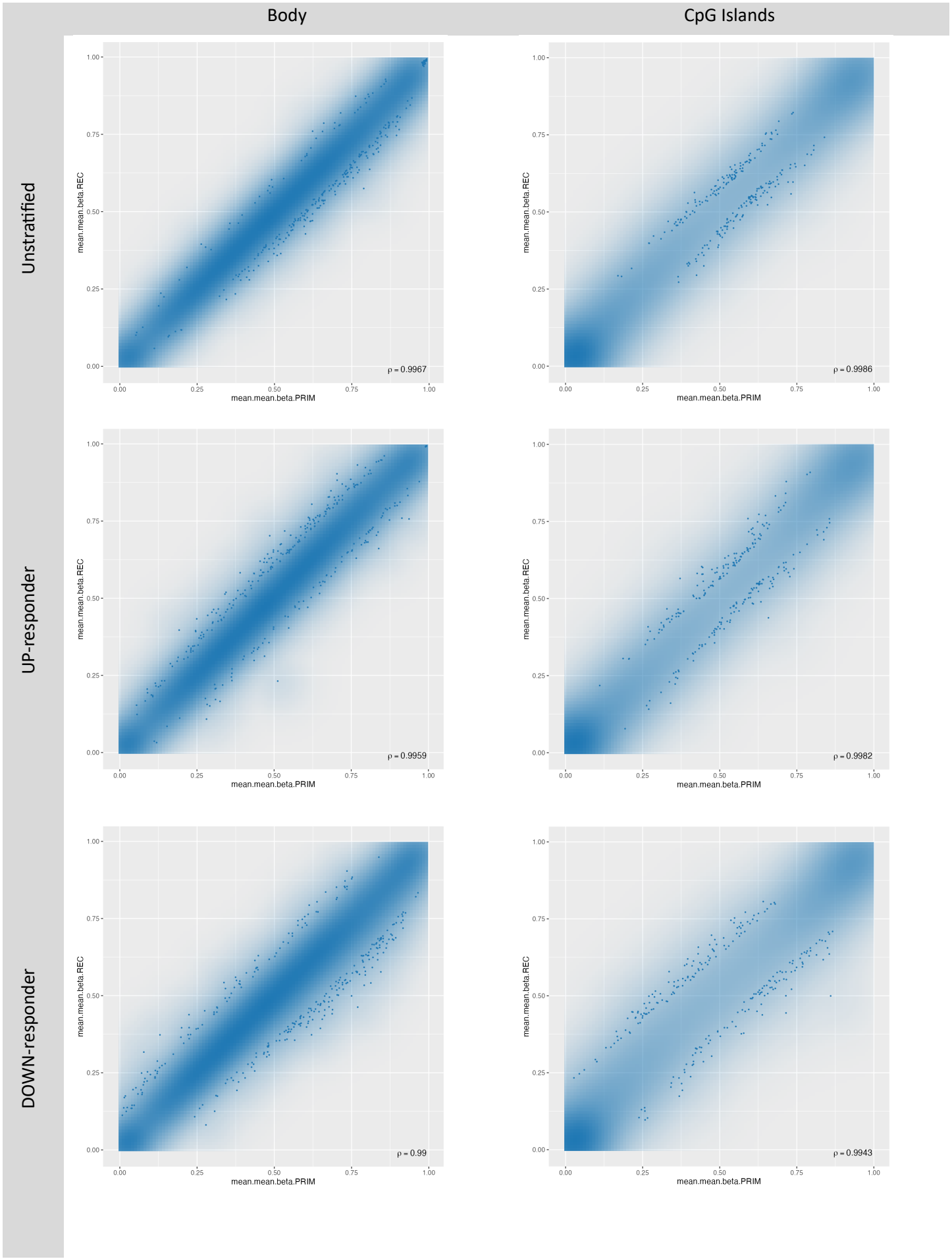
4.3.4.6 Exploratory Stratification and Visual Divergence in Methylation Changes

Although no statistically significant differentially methylated regions (DMRs) were identified across the combined dataset, the visual outputs from RnBeads suggested potential biological variation worth further investigation. Specifically, I examined the correlation plots that RnBeads generates to compare primary and recurrent tumour methylation profiles within each patient.

In the analysis of the entire cohort, which included all 27 matched pairs regardless of subtype, the plots consistently showed narrow, concentrated regression bands across all major genomic annotation types e.g., promoters, gene bodies, CpG islands, enhancers, and tiling windows (Figure 4-4). This indicated a high degree of similarity between primary and recurrent methylation profiles across patients, consistent with the lack of statistically significant DMRs.

However, when the same plots were examined within biologically stratified subgroups—UP and DOWN responders—the pattern changed (Figures 4-5 and 4-6). In both UP and DOWN groups (each with n=8 pairs), the regression bands appeared visibly broader and more diffuse, suggesting increased variability between primary and recurrent profiles (Figure 4-7). While the changes were not statistically significant, the wider distribution of data points implied potentially greater divergence in methylation patterns following recurrence, depending on the response subtype.


Promoters
Enhancers
Unstratified
UP-responder
DOWN-responder
mean.mean.beta.REC
mean.mean.beta.PRIM
ρ = 0.9978
ρ = 0.9956
ρ = 0.9971
ρ = 0.9939
ρ = 0.9932
ρ = 0.9872
0.00
0.25
0.50
0.75
1.00


Body
CpG Islands
Unstratified
UP-responder
DOWN-responder
mean.mean.beta.REC
mean.mean.beta.PRIM
ρ = 0.9967
ρ = 0.9986
ρ = 0.9959
ρ = 0.9982
ρ = 0.99
ρ = 0.9943

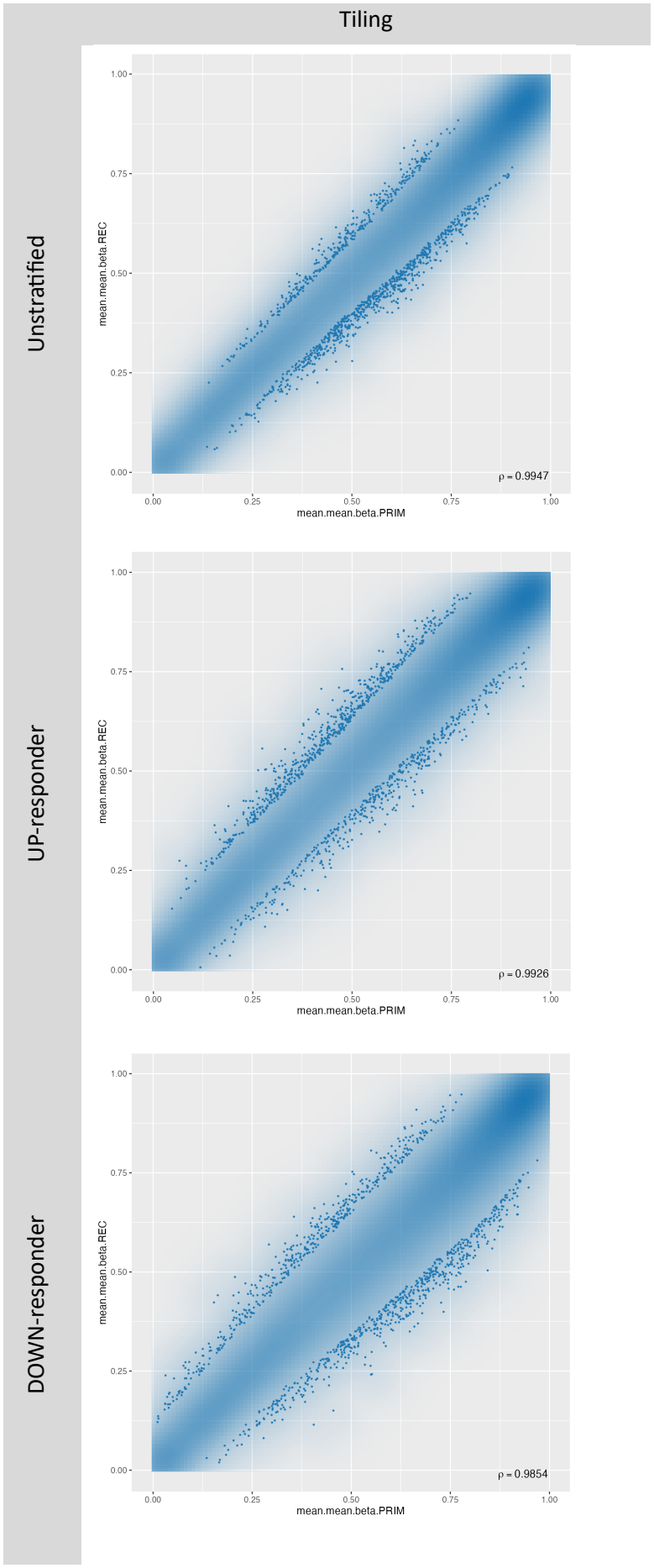
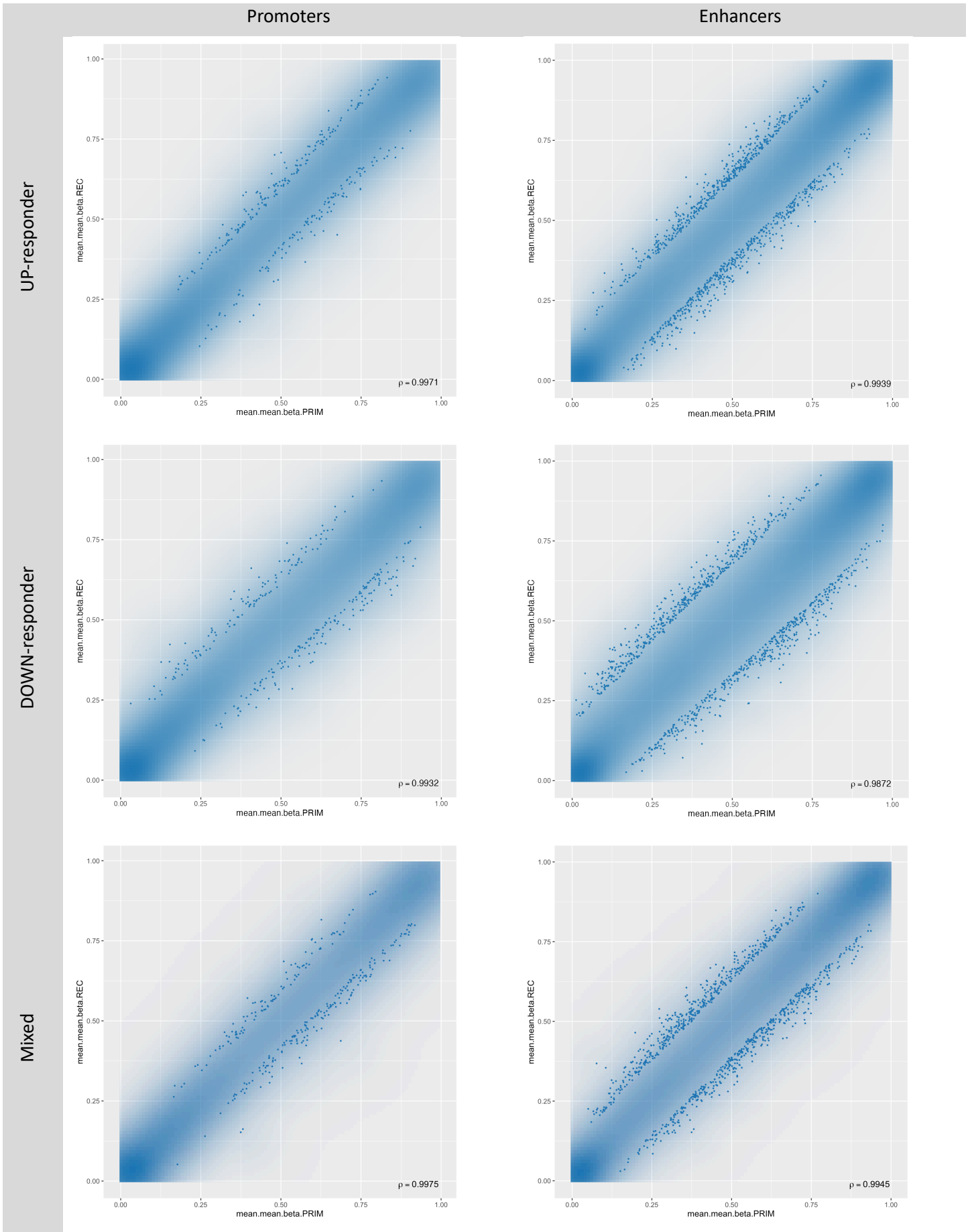


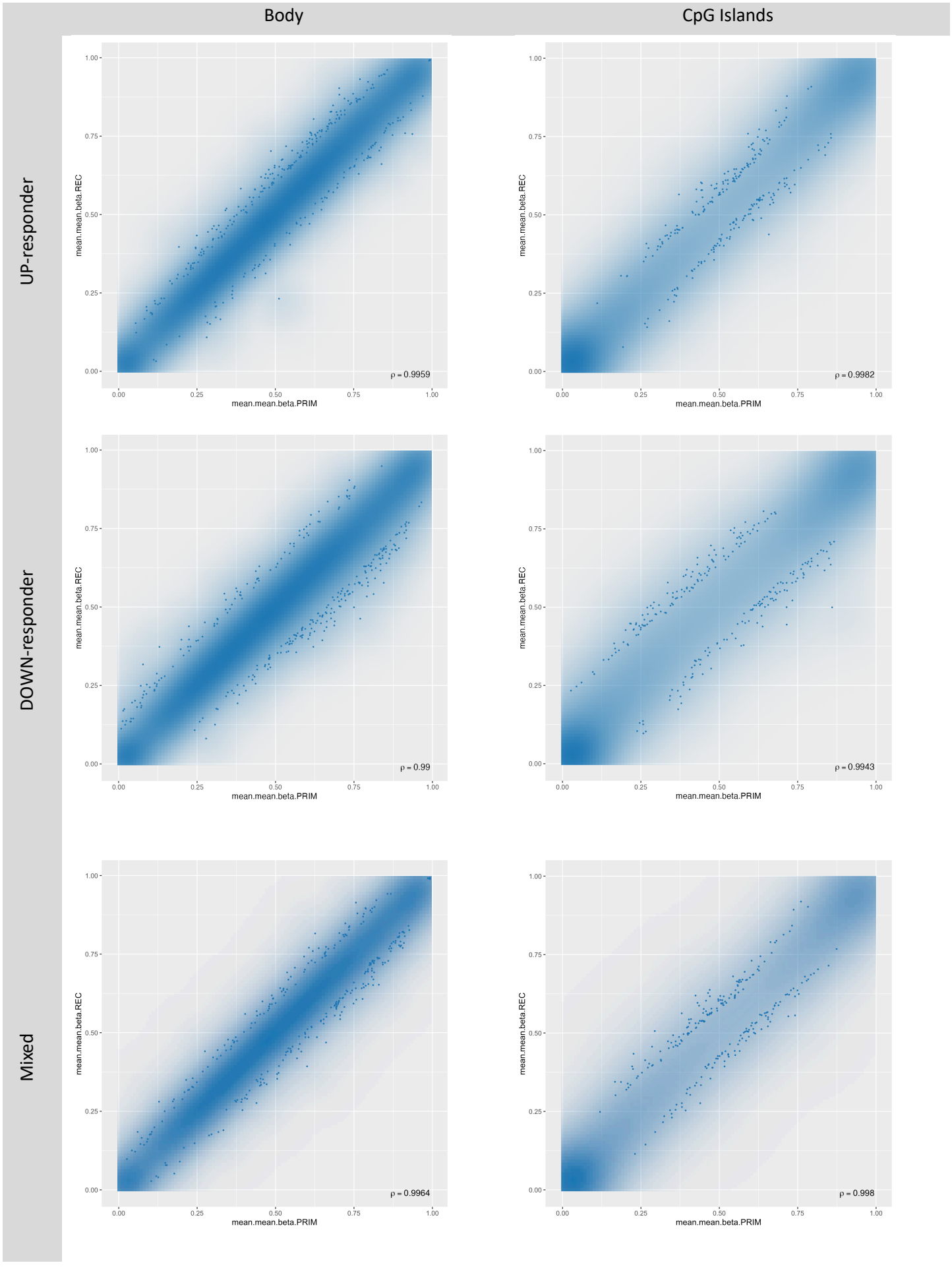
Figure 4-7: Scatter plots based on combined arrays, stratified by responder subtype and region type.

Data from 450K, EPICv1, and EPICv2 arrays were merged. Rows indicate stratification (unstratified, UP, DOWN), and columns represent genomic region types. Each point shows a region's mean beta value in primary (x-axis) vs recurrent (y-axis) samples. Diagonal trends and Pearson correlation coefficients (ρ) are included. Point density is colour-mapped, and red indicates significant DMRs.

This observation draws a hypothesis that the increased spread could either reflect biological differences emerging after stratification or could simply be an artefact caused by smaller sample sizes. To investigate this further, I created a mixed group composed of four randomly selected UP and four randomly selected DOWN pairs, matched in size (n = 8) to the individual stratified subgroups. This group was used to test whether the broadened spread was purely due to fewer samples.

After re-running the RnBeads analysis on the mixed cohort (4 UP + 4 DOWN pairs), the resulting correlation plots showed narrower regression bands overall, and the methylation profiles between primary and recurrent samples appeared more tightly correlated than in the fully stratified DOWN group (Figure 4-8). In some regions, particularly promoters and enhancers, the MIXED group plots more closely resembled those seen in the UP responder group, where primary–recurrent correlation remained relatively high. This suggests that the broader variation seen in the stratified groups—especially in DOWN responders—is unlikely to be solely due to reduced sample size and instead may reflect biologically distinct patterns of epigenetic change related to patient subtype. The resemblance between MIXED and UP further supports the idea that UP responders may retain more stable methylation profiles, while DOWN responders show greater divergence at recurrence.


Promoters
Enhancers
UP-responder
DOWN-responder
Mixed
mean.mean.beta.REC
mean.mean.beta.PRIM
ρ = 0.9971
ρ = 0.9939
ρ = 0.9932
ρ = 0.9872
ρ = 0.9975
ρ = 0.9945


Body
CpG Islands
UP-responder
DOWN-responder
Mixed
mean.mean.beta.REC
mean.mean.beta.PRIM
ρ = 0.9959
ρ = 0.9982
ρ = 0.99
ρ = 0.9943
ρ = 0.9964
ρ = 0.998

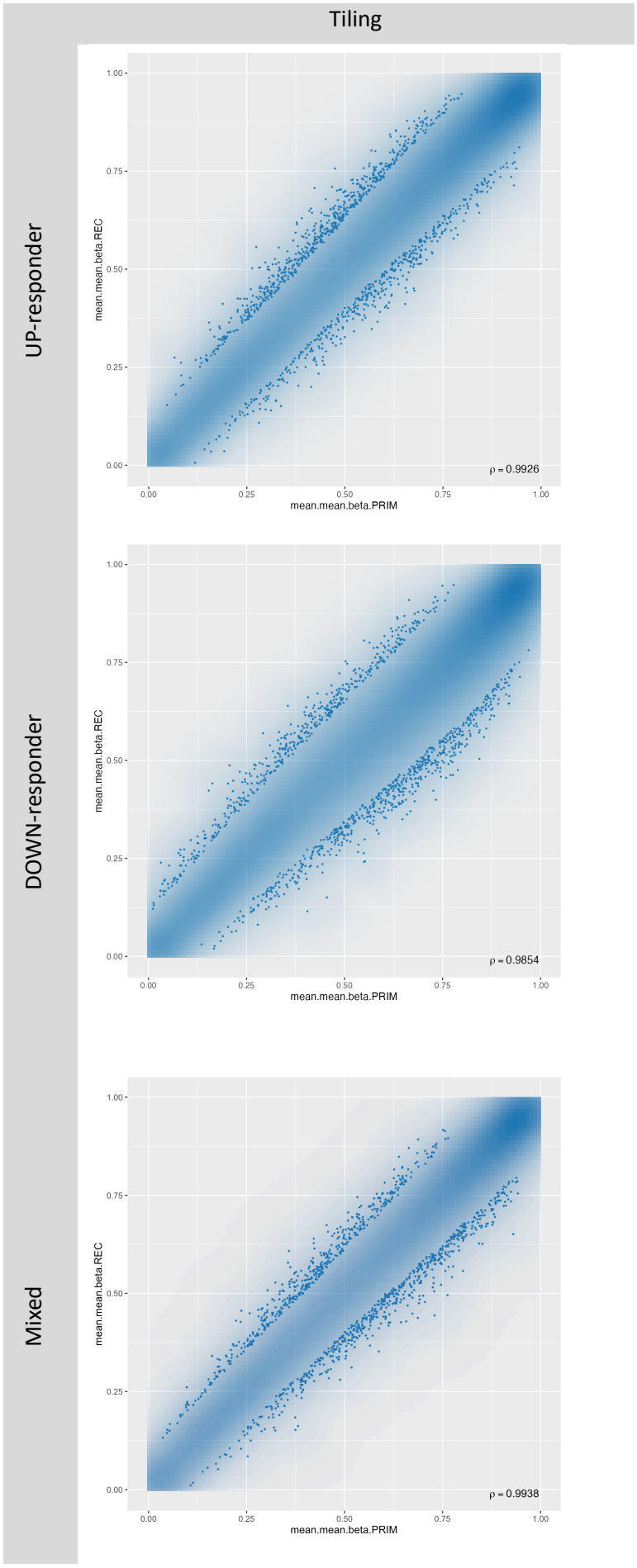


Figure 4-8: Scatter plots assessing methylation divergence across UP, DOWN, and mixed responders.

Rows display different responder groups; columns represent region types. Each point reflects a genomic region's average methylation in primary (x-axis) and recurrent (y-axis) samples. A diagonal trend is shown in each panel, with Pearson correlation coefficient (ρ) indicating the degree of methylation conservation. Point colour encodes local density. Statistically significant DMRs are highlighted in red.

4.3.4.7 Quadrant Analysis of Directional Methylation Shifts

I wished to explore whether the UP and DOWN responder subtypes exhibited directionally opposing DNA methylation changes during glioblastoma recurrence, which could indicate that differential methylation contributes to the differential transcriptional reprogramming. To this end, I conducted a quadrant-based comparison, plotting the longitudinal changes in average methylation levels across annotated genomic regions in Up versus Down responders. The regions in questions were promoters, CpG islands, gene bodies, GBM-specific enhancers, and tiling windows. For each region, I calculated a recurrence/primary methylation ratio per region separately for each subtype. These ratios were plotted on a two-dimensional scatter plot, with DOWN responders on the x-axis and UP responders on the y-axis. This design allowed me to distinguish concordant changes (top right – longitudinal increases in methylation for both responders - and bottom left - longitudinal decreases in both- quadrants) and discordant changes (top left and bottom right, where methylation increased in one subgroup over time but ecreased in the other). Although global changes were modest and not statistically significant overall, I observed that several features fell into the discordant quadrants, suggesting that UP and DOWN responders may undergo some opposing epigenetic shifts during recurrence. Given the hypothesis that responder subtypes follow distinct epigenetic trajectories, this was worth further investigation to see if opposing changes in methylation are contributing to transcriptional reprogramming mechanisms.

Among the five region types analysed, I now focused downstream interpretation on promoters and gene bodies (Figure 4-9), as these provide the most direct insight into potential transcriptional regulation and have the most comprehensive biological annotation. To further assess the biological significance of observed methylation shifts, I annotated the plots with LE50 and LE70 gene sets, which represent genes found in the leading edge of treatment-driven dysregulated gene signatures in at least 50% and 70% of patients, respectively. These gene sets are, thus, those most consistently and significantly changed in expression from glioblastoma primary to recurrence. However, in my analysis, both LE50 and LE70 genes clustered primarily within the 99% confidence interval (CI99%) around zero, indicating that their methylation changes were modest and not strongly subtype-specific. This suggests that although these genes are transcriptionally relevant in recurrence, they may not be regulated by differential methylation in a divergent manner across subtypes.

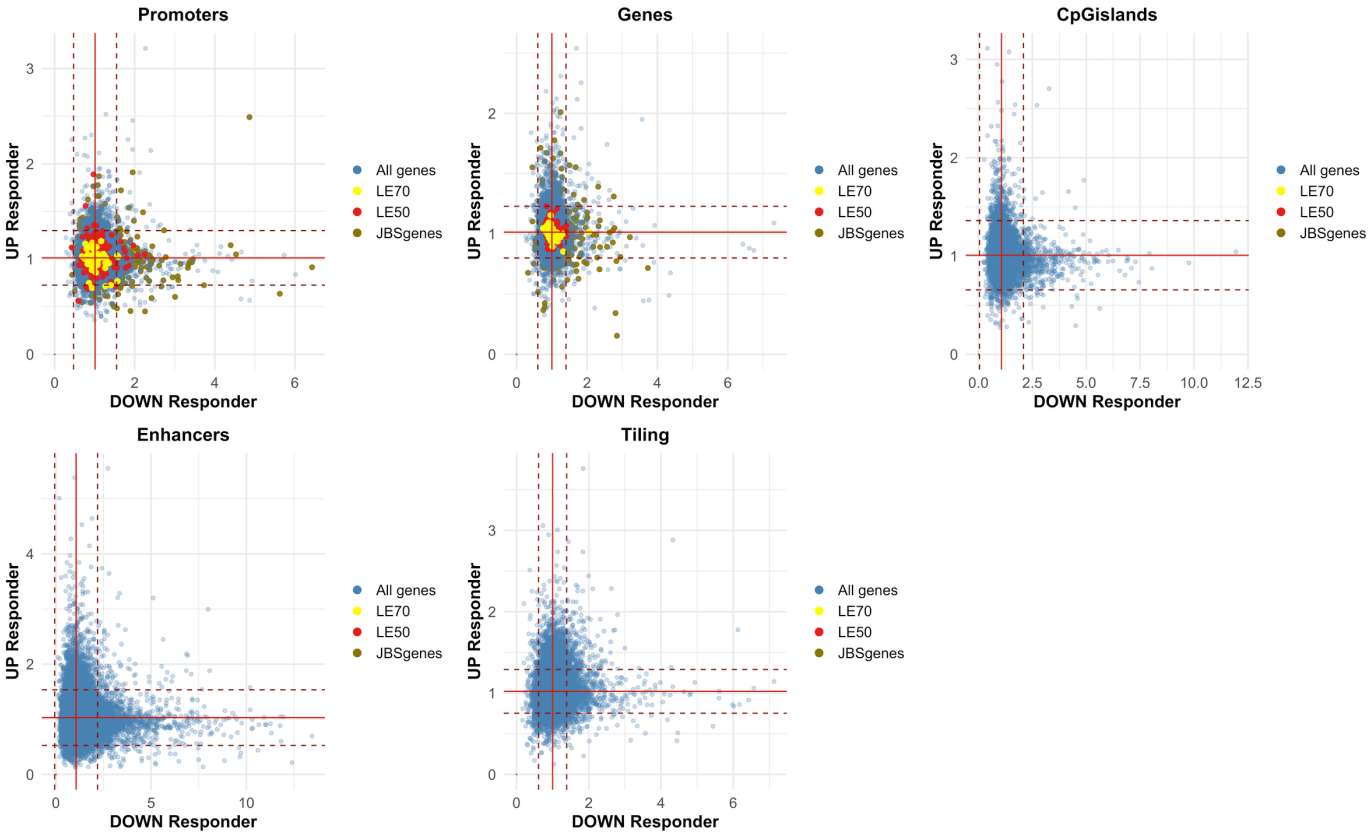


Figure 4-9: Quadrant scatter plots of DNA methylation ratios across genomic regions in UP and DOWN responders.

Scatter plots display the recurrence-to-primary methylation ratios for matched glioblastoma samples, stratified by UP and DOWN responder subtypes, across five genomic annotation categories: promoters, gene bodies, CpG islands, enhancers, and tiling regions. Each point represents a single annotated region, with the x-axis indicating the methylation ratio in DOWN responders and the y-axis in UP responders. Methylation ratios were calculated using the formula:

mean methylation in recurrence / mean methylation in primary.

Solid red lines represent the mean methylation ratio for each group and define the central point of the plot. Dashed red lines indicate the boundaries of the 99% confidence interval (CI99%) along both axes. The central box defined by the dashed lines represents regions where methylation changes fall within expected variation for both subtypes. Regions located outside the CI99% on either or both axes were considered to exhibit significant directional methylation changes and were classified into quadrant-based gene sets for downstream analysis.

Coloured points indicate annotated genes:

- All genes (blue).
- LE50 genes (red): genes found in the leading edge of GSEA-enriched gene sets in ≥50% of patients.
- LE70 genes (yellow): genes found in ≥70% of patients' leading edges.
- JBSgenes (brown): genes identified as JARID2 binding site targets.

These gene sets were overlaid only on promoter and gene body plots, as only these regions included gene name annotations. CpG islands, enhancers, and tiling regions did not have gene identifiers and were therefore not annotated with LE50, LE70, or JBSgenes.

Nonetheless, to investigate further, I grouped all data points into eight categories based on their deviation from the 99% confidence interval around zero i.e. no change in methylation over time: Figure 4-10. I then conducted Gene Ontology (GO) enrichment analysis using the Over-Representation Analysis (ORA) method

via clusterProfiler. Of the eight categories, only Gene Set 6 (Right_Extreme: increased methylation over time in Down responders only) (Figures 4-11A-B) and Gene Set 4 (Bottom_Extreme: decreased methylation over time in Up responders only) produced biologically meaningful enrichment results (Figures 4-11C). The number of genes per quadrant submitted to the clusterProfiler tool is summarised in Table 4-5.

Table 4-5: Gene lists extracted from promoters and gene bodies

Quadrant	Promoter	Gene body
Bottom_Left	4	18
Top_Left	6	18
Top_Right	45	50
Bottom_Right	34	47
Left_Extreme:	13	42
Right_Extreme	445	273
Top_Extreme	489	499
Bottom_Extreme	209	232

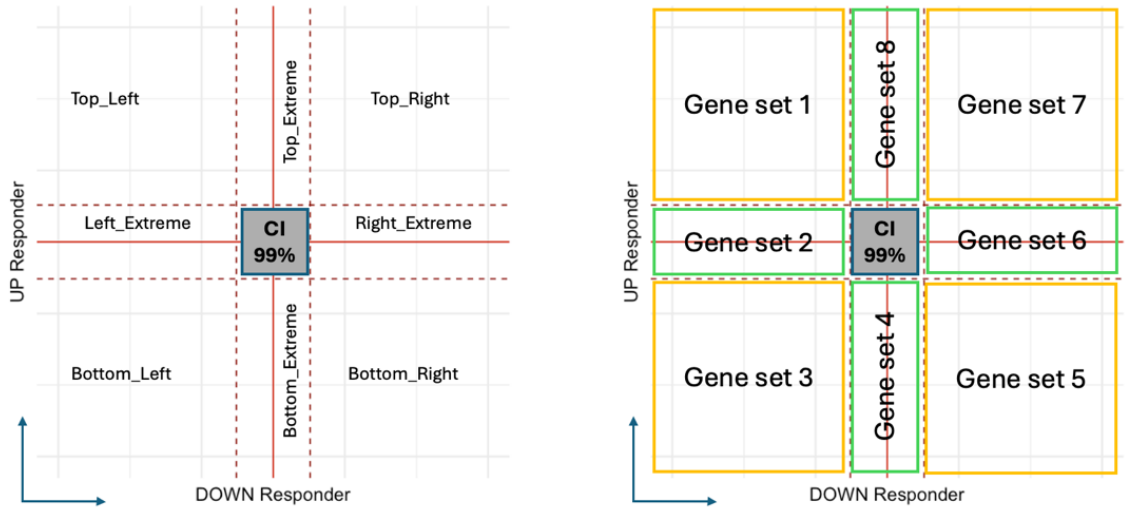


Figure 4-10: Schematic Overview of Quadrant-Based Classification and Gene Set Assignment for Subtype-Specific Methylation Shifts.

Schematic illustration of quadrant-based classification and gene set assignment in recurrence-to-primary methylation analysis. This schematic depicts how genomic regions were categorised based on their recurrence-to-primary methylation ratio in DOWN responders (x-axis) and UP responders (y-axis).

Left panel: Conceptual layout showing eight quadrant zones surrounding the central 99% confidence interval (CI99%), shaded in grey. Solid red lines represent the mean recurrence-to-primary ratio for each axis, and dashed red lines indicate the CI99% boundaries. Each quadrant reflects a specific directional methylation change:

- Top_Left: increased methylation in UP responders and decreased in DOWN responders.
- Top_Right: increased methylation in both UP and DOWN responders.
- Bottom_Right: decreased methylation in UP responders and increased in DOWN responders.

- Bottom_Left: decreased methylation in both subtypes.
- Left_Extreme: decreased methylation in DOWN responders only.
- Right_Extreme: increased methylation in DOWN responders only.
- Top_Extreme: increased methylation in UP responders only.
- Bottom_Extreme: decreased methylation in UP responders only.

Right panel: Mapping of the eight gene sets (Gene sets 1–8) to their corresponding quadrants. Gene sets were defined as regions falling outside the CI99% in at least one direction:

- Gene sets 1, 3, 5, and 7 (corner quadrants) represent features with concordant (3 and 7) or opposing (1 and 5) methylation changes in both subtypes.
- Gene sets 2, 4, 6, and 8 (horizontal/vertical extremes) represent features with directional changes in only one responder group.

The central CI99% region contains genes with no significant change in either group, and was excluded from GO analysis.

4.3.4.7.1 Elevated Methylation in Down Responders (Right_Extreme - Gene Set 6)

This category consisted of genomic features that showed increased methylation specifically in Down responders at recurrence. I interpreted this as an indication of potential transcriptional repression occurring selectively in this subgroup.

GO enrichment analysis revealed significant enrichment in biological processes related to development and differentiation. For gene body-associated regions, I identified enrichment in terms such as “pattern specification process” and “connective tissue development”, while promoter-associated regions were enriched for “mesenchyme development” (Figure 11A–B). These findings implicate genes involved in spatial and lineage-specific cell fate decisions during recurrence in Down responders.

Importantly, while promoter methylation is generally associated with gene silencing, gene body methylation is more commonly linked to active transcription. Therefore, the increased gene body methylation observed in recurrence may reflect the upregulation or maintenance of expression of genes related to differentiation and mesenchymal development. This interpretation is consistent with the notion that Down responders acquire a more mesenchymal-like phenotype over time (Tanner et al., 2024).

Rather than suppressing differentiation, the methylation changes may facilitate the activation of mesenchymal lineage pathways, supporting cellular plasticity and recurrence. This is in line with broader literature suggesting that epigenetic mechanisms, including gene body methylation, can be involved in sustaining transcription of lineage-defining genes, potentially helping tumour cells maintain a stem-like, therapy-resistant state within a mesenchymal trajectory.

4.3.4.7.2 Reduced Methylation in Up Responders (Bottom_Extreme - Gene Set 4)

This category incorporates genomic regions that exhibited a decrease in methylation in Up responders only, suggesting that these genes might be transcriptionally activated during recurrence in this subgroup.

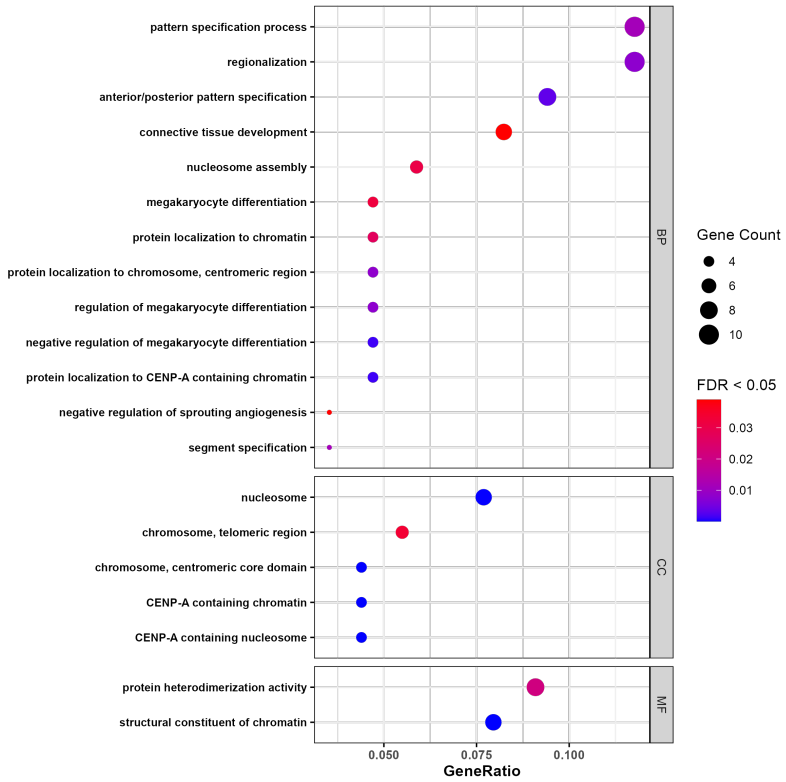
In this gene set, only gene body-associated regions produced significant enrichment results; no promoter-derived features yielded any enriched GO terms. The gene body list showed strong enrichment in immune-related pathways, particularly the “defense response to virus” (Figure 4-11C). This term encompasses a range

of innate and adaptive immune mechanisms activated in response to viral infection and aimed at limiting viral replication and spread.

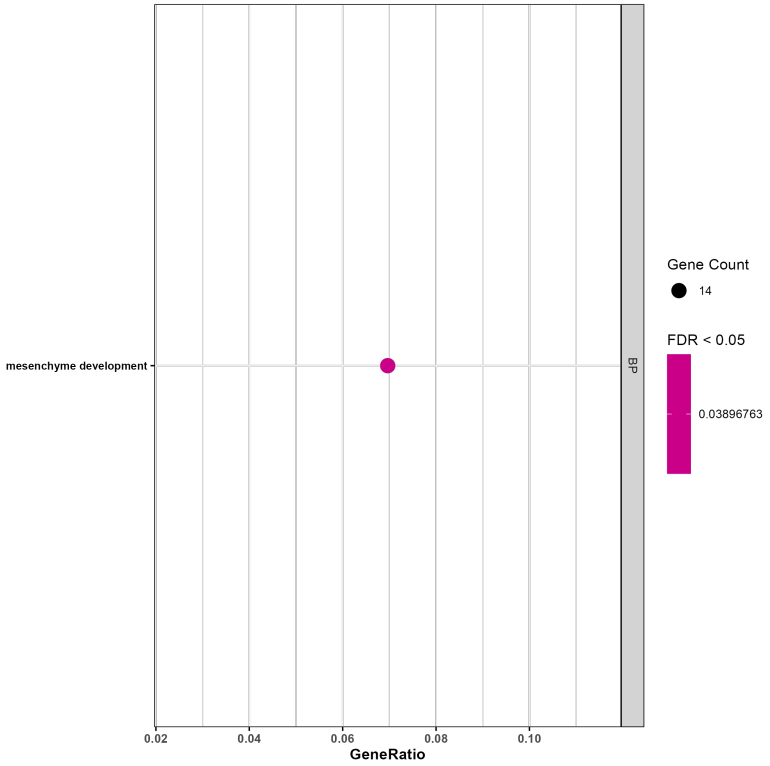
Based on this enrichment, it suggests that immune-related genes become epigenetically activated in Up responders during recurrence, potentially contributing to an immunologically primed tumour microenvironment.

This observation may have therapeutic relevance. The apparent activation of antiviral response genes suggests that Up responders could be more responsive to immunotherapies, possibly due to enhanced immune activity or surveillance.

A



B


mesenchyme development
BP
0.02
0.04
0.06
0.08
0.10
GeneRatio
Gene Count
14
FDR < 0.05
0.03896763

C

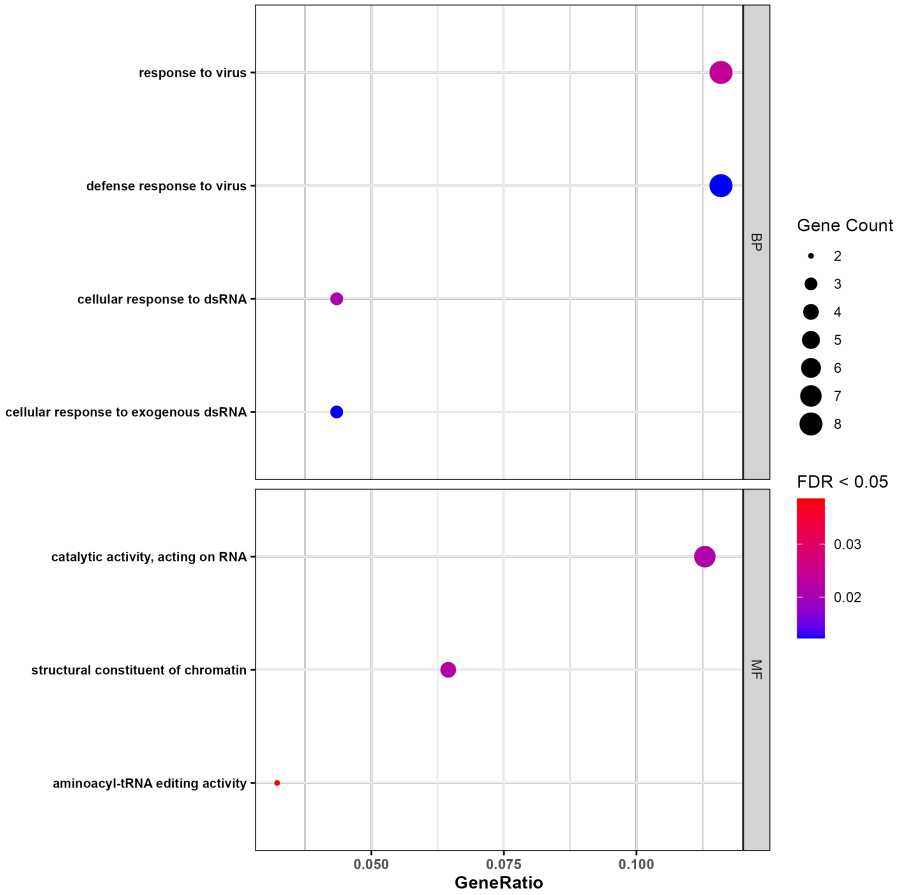

response to virus
defense response to virus
cellular response to dsRNA
cellular response to exogenous dsRNA
BP
catalytic activity, acting on RNA
structural constituent of chromatin
aminoacyl-tRNA editing activity
MF
0.050
0.075
0.100
GeneRatio
Gene Count
2
3
4
5
6
7
8
FDR < 0.05
0.03
0.02

Figure 4-11: Functional Enrichment of Genes with Directional Methylation Changes in UP and DOWN Responders.

Dot plots show significantly enriched Gene Ontology (GO) terms identified using Over-Representation Analysis (ORA) performed via clusterProfiler on differentially methylated regions extracted from quadrant-based gene sets. Terms are shown for Biological Process (BP), Cellular Component (CC), and Molecular Function (MF) when exists.

- **A:** GO enrichment results for gene body–associated regions showing increased methylation in DOWN responders only (Right_Extreme, Gene Set 6 – Genes).
- **B:** GO enrichment results for promoter-associated regions in the gene promoters set (Right_Extreme**,** Gene Set 6 – Promoters).
- **C:** GO enrichment for gene body–associated regions showing decreased methylation in UP responders only (Bottom_Extreme, Gene Set 4 – Genes).

In all panels, the x-axis represents the GeneRatio, defined as the number of input genes mapped to each term divided by the total number of genes in the set. Dot size reflects the number of genes contributing to each term (Gene Count), while colour indicates the adjusted p-value (FDR < 0.05), with blue shades representing stronger statistical significance.

4.3.4.7.3 The Remaining Quadrants

Gene Set 7 (Top_Right) included relatively few features that showed increased methylation in both Up and Down responders. Although this gene set produced statistically significant GO terms, the small number of genes involved limited the interpretive value of the findings. Nonetheless, the enrichment for processes related to nucleosome organisation suggests that shared chromatin remodelling mechanisms may be active during recurrence in both subtypes.

Gene Sets 1, 2, 5, 7, and 8 also showed statistically significant enrichment results; however, the small gene counts in each case limited the biological insights that could be reliably drawn. Therefore, although these gene sets were technically significant, I opted not to discuss them in detail. I have included the relevant plots in the appendix.

In contrast, Gene Set 3 (Bottom_Left), which comprises features with decreased methylation in both up- and down-regulated responders, did not produce any significant GO terms from either gene body or promoter regions.

Across the entire analysis, I found that gene body-associated DMRs were the most consistent contributors to biologically meaningful enrichment results. Both Gene Set 6 (Right_Extreme) and Gene Set 4 (Bottom_Extreme) yielded significant GO terms from gene body-derived lists, while only Gene Set 6 showed additional enrichment from promoter-associated DMRs.

This pattern suggests that gene body methylation changes may play a particularly important role in transcriptional regulation during glioblastoma recurrence, contributing to processes such as cell differentiation suppression (in Down responders) and immune activation (in Up responders). Although promoter methylation is classically associated with gene regulation, the findings highlight that intragenic methylation, especially when analysed in the context of tumour subtypes, may also serve as a critical regulatory mechanism influencing tumour evolution.

4.4 DISCUSSION

4.4.1 Assessment of MGMT Methylation

This study examined the methylation status of the MGMT promoter in paired glioblastoma samples at both primary and recurrent stages, using the MGMT-STP27 classifier. This classifier integrates signal from two promoter CpG sites to estimate methylation status and is widely used in clinical practice to predict response to alkylating therapy.

In the discovery cohort (n = 27), MGMT methylation status was largely stable between the two timepoints. Specifically, 24 out of 27 patients (89%) retained the same status, while three patients (11%) exhibited changes—two gained methylation (U→M) and one lost it (M→U). In the validation cohort (n = 57), a similar pattern was observed: 49 patients (86%) showed no change, while 8 patients (14%) experienced switching, with five losing methylation (M→U) and three gaining it (U→M). These results are consistent with prior studies indicating that MGMT methylation is generally stable over the disease course but can change in a subset of patients.

Such switching, although relatively uncommon, is clinically relevant. Previous studies, such as PMID: 33632732, reported MGMT methylation status changes in about 22% of glioblastoma cases, particularly after temozolomide treatment. The slightly lower rate observed in our validation cohort (14%) may still reflect treatment-related clonal selection or tumour evolution. Notably, most of the observed changes in our cohorts involved loss of methylation at recurrence, a pattern that may be associated with acquired resistance mechanisms.

While MGMT methylation is a well-established prognostic and predictive biomarker, our analysis also explored whether changes in MGMT status were associated with tumour progression behaviour—specifically, stratified by responder subtypes (UP and DOWN). Across both cohorts, MGMT status switching occurred at similarly low frequencies in both groups, and chi-square testing revealed no significant differences in switching rates between UP and DOWN responders. This suggests that MGMT switching is not uniquely enriched in either progression subtype and may instead reflect broader tumour-intrinsic or treatment-related factors.

In summary, our study supports the view that MGMT methylation is a largely stable biomarker in glioblastoma, with clear prognostic relevance at the primary stage. However, methylation changes do occur in a minority of patients and may impact clinical decision-making at recurrence, particularly when considering second-line therapies.

4.4.2 DMRs analysis

In this chapter, I examined DNA methylation differences between primary and recurrent glioblastoma (GBM) tumours using matched longitudinal samples from both discovery and validation cohorts. Various methods

were employed to evaluate methylation variations, including probe-level (DMP) analysis with the limma package, and region-level (DMR) analysis using DMRcate, Bumphunter, and the integrated RnBeads pipeline. Although several significantly differentially methylated positions (DMPs) were identified, especially within promoter and enhancer regions, no statistically significant differentially methylated regions (DMRs) were found at any stage of the analysis — whether unstratified, stratified by responder subtype, or across platform-integrated data. This underscores an important point: while individual CpG sites may show slight shifts, these changes do not appear to cluster strongly enough to meet regional significance thresholds, indicating that epigenetic alterations during GBM recurrence are either subtle, dispersed, or vary between patients.

To better understand whether patient subtypes might reveal underlying epigenetic differences, the analysis was repeated using a biologically informed stratification based on treatment driven changes in expression of a subset of genes (shown to have a JARID2 bindig site in their promoter), dividing patients into UP and DOWN responders. This stratification did not lead to the identification of statistically significant DMRs; however, visual inspection of RnBeads-generated correlation plots indicated broader variation between primary and recurrent samples in both UP and DOWN responder groups, compared to the full unstratified cohort. These broader regression bands were most notable in enhancer and promoter regions, suggesting potentially increased divergence in methylation profiles following tumour recurrence within these subgroups. This observation raised the possibility that the JARID2 subtypes may follow distinct epigenetic trajectories, even if these changes fall below the threshold of statistical detection in this sample size.

To understand whether the increased variation observed in stratified groups was due to biological effects or simply a smaller sample size, a controlled “MIXED” cohort of 4 UP and 4 DOWN responder pairs was created. Rerunning the RnBeads analysis on this cohort yielded narrower, more compact regression bands, resembling those in the full, unstratified cohort. Interestingly, in some regions—particularly promoters and enhancers—the MIXED group resembled the UP-responder group more closely, which generally showed stronger primary–recurrent correlation than the DOWN group. This suggests that the wider spread observed in the DOWN responder group is unlikely to be a sample size artefact and may instead reflect true biological divergence associated with treatment response patterns.

To explore this further, a quadrant analysis was performed to compare the directional change in methylation between UP and DOWN responders. While no regions showed significant differences, several genomic features displayed opposite trends between the two subtypes, with a number falling into discordant quadrants. This supports the hypothesis that UP and DOWN responders may undergo distinct and possibly opposing methylation changes during recurrence, adding nuance to the idea that epigenetic reprogramming in GBM may be context-dependent and subtype-specific.

These findings align with prior literature suggesting that while genetic and transcriptomic changes are frequently observed in recurrent GBM, DNA methylation profiles are generally more stable. However, this study suggests that stratifying patients based on biological characteristics, such as responder subtype, may

reveal otherwise hidden epigenetic variations. The subtlety of the changes observed here, combined with the consistent presence of DMPs and diverging correlation trends, suggests that epigenetic evolution in GBM may occur on a smaller or more heterogeneous scale than previously appreciated.

A significant limitation of this study is the small sample size, especially within stratified groups, which reduces the statistical power to detect small or region-specific changes. Additionally, using different array platforms introduced variability in probe coverage, requiring analysis to be limited to common probes, which may potentially exclude biologically informative sites. The dependence on array-based bulk tissue methylation also restricts resolution, potentially masking methylation changes in subclonal populations or tumour microenvironments.

Future studies could address these limitations by employing higher-resolution methods such as whole-genome bisulfite sequencing (WGBS) or single-cell methylation profiling. Additionally, the development of a methylation-based classifier for JARID2-bound gene activity could enhance stratification in cohorts lacking transcriptomic data, thereby allowing for a broader application of this approach. Finally, integrating methylation data with other molecular layers, such as gene expression, chromatin accessibility, or mutational burden, may provide a more comprehensive view of tumour evolution and clarify the role of epigenetics in recurrence and therapy resistance in GBM.

4.4.3 Quadrant Analysis of Directional Methylation Shifts

In this chapter, I investigated DNA methylation changes between matched primary and recurrent GBM tumours by comparing recurrence/primary methylation ratios across two clinical response subgroups: Up and Down responders. By classifying genes into eight quadrant-based categories based on methylation changes in each group, I identified two categories—Gene Set 6 (Right_Extreme) and Gene Set 4 (Bottom_Extreme)—that yielded biologically meaningful GO enrichment results. Gene Set 6 included features with increased methylation specifically in Down responders, while Gene Set 4 contained features with decreased methylation exclusively in Up responders. These patterns suggest distinct regulatory trajectories during GBM recurrence.

In Down responders, I observed enrichment for developmental processes such as pattern specification, connective tissue development, and mesenchyme development, derived from both gene body and promoter DMRs. Given that gene body methylation is generally associated with active transcription, these findings suggest that genes involved in differentiation and tissue organisation may be epigenetically upregulated during recurrence. I interpret this as an adaptive mechanism supporting the acquisition or maintenance of a mesenchymal stem-like phenotype. Rather than silencing differentiation pathways, the observed methylation changes may help sustain expression of lineage-specific programs while preventing full differentiation. This

would preserve tumour plasticity, a hallmark of treatment-resistant and recurrent glioblastoma, by maintaining cells in a dynamic, partially differentiated state that can adapt to therapy-induced stress.

In contrast, Up responders showed reduced methylation over time in the gene body regions of immune-related genes, especially those involved in the defence response to viruses. This demethylation might facilitate the transcriptional repression of immune programmes during recurrence. It suggests that these changes reflect an epigenetic shift towards immune readiness or reactivation in this subgroup. Importantly, promoter regions did not display notable enrichment for these genes, emphasising the importance of gene body methylation in this context. The activation of antiviral pathways could serve as a mechanism for enhanced immune surveillance, suggesting a potential increased susceptibility to immune-based therapies in the UP-responder subgroup.

The fact that most LE50 and LE70 genes fall within the 99% CI suggests that, although these genes are consistently transcriptionally altered across patients, they are unlikely to be regulated by subtype-specific DNA methylation changes. This aligns with findings from our group's recent work (Tanner et al., 2024), which showed that transcriptional reprogramming during glioblastoma recurrence is not primarily driven by DNA methylation but is instead linked to dynamic changes in histone modifications. However, my analysis introduces a new dimension by demonstrating that DNA methylation may still have a regulatory role, especially when examined in a subtype-specific context. While Tanner's study focused on broad recurrence-associated changes, my quadrant-based approach highlights how divergent methylation patterns between Up and Down responders may contribute to their distinct biological phenotypes. These findings suggest that histone modifications are not the only epigenetic mechanism involved and that DNA methylation, although more subtle, may also influence transcriptional landscapes in a subgroup-specific manner.

These findings are consistent with prior studies highlighting transcriptional and epigenetic plasticity in GBM, particularly during recurrence. Work by (Neftel et al., 2019) has demonstrated that GBM cells dynamically transition between distinct cell states—including mesenchymal and progenitor-like identities in response to treatment. My observations in Down responders align with this model, suggesting that repression of differentiation programs may be an epigenetic mechanism enabling state switching. The immune activation observed in Up responders also fits with emerging evidence that, while GBM is typically considered an "immune cold" tumour, certain subtypes or recurrent cases may exhibit epigenetically mediated immune competence. The recent review by (Lin et al., 2024) supports this view, describing how DNA methylation, histone modification, and non-coding RNA regulation influence immune activation and T cell infiltration in GBM. Their discussion of epigenetic silencing and reactivation mechanisms, including interferon pathway genes and immune checkpoints, parallels the trends observed in the UP-responder subtype.

The implications of these findings are twofold. First, they suggest that subtype-specific epigenetic changes may underlie differential clinical responses and should be considered in therapeutic stratification. For Down responders, the silencing of differentiation programs may render them vulnerable to epigenetic therapies

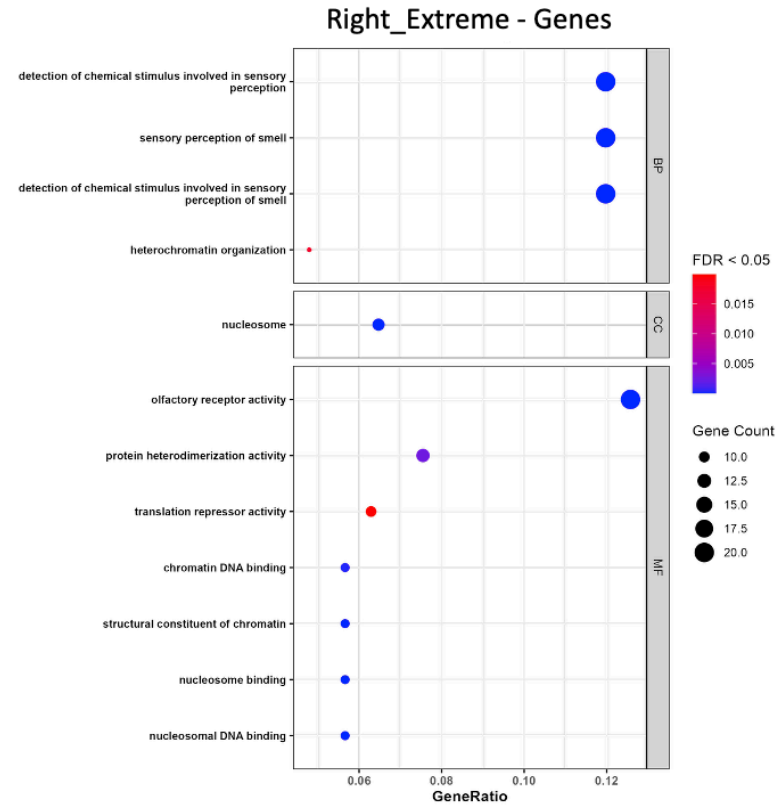
aimed at reversing stemness or restoring cell identity, such as inhibitors of PRC2 or histone deacetylases. For Up responders, the apparent activation of immune pathways may support the use of immunotherapy, especially in recurrence, when epigenetic reactivation may enhance immune engagement. Second, these findings provide a rationale for integrating epigenomic profiling into GBM monitoring and treatment design, as changes in methylation may precede or reflect critical biological transitions during disease progression.

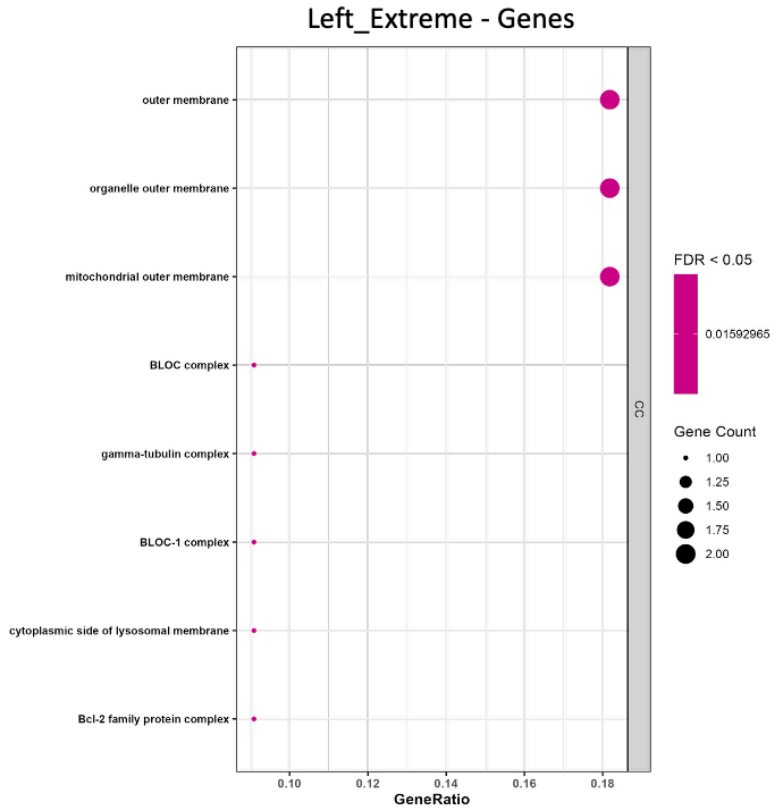
Nonetheless, several limitations should be acknowledged. The quadrant-based classification yielded small gene lists in most categories, limiting the statistical power of enrichment analysis outside Gene Sets 6 and 4. Moreover, although promoter methylation is classically linked to gene silencing, the functional consequences of gene body methylation are more complex and remain context-dependent. Without transcriptomic validation, it is challenging to infer whether hypomethylation directly corresponds to gene upregulation. Additionally, this analysis was based on bulk tumour methylation data, which cannot resolve cell-type-specific contributions. It remains possible that some observed methylation changes reflect alterations in the tumour microenvironment rather than tumour-intrinsic changes. Finally, while GO analysis provides functional clues, it does not establish causality; further mechanistic studies are needed to confirm the regulatory roles of these epigenetic changes.

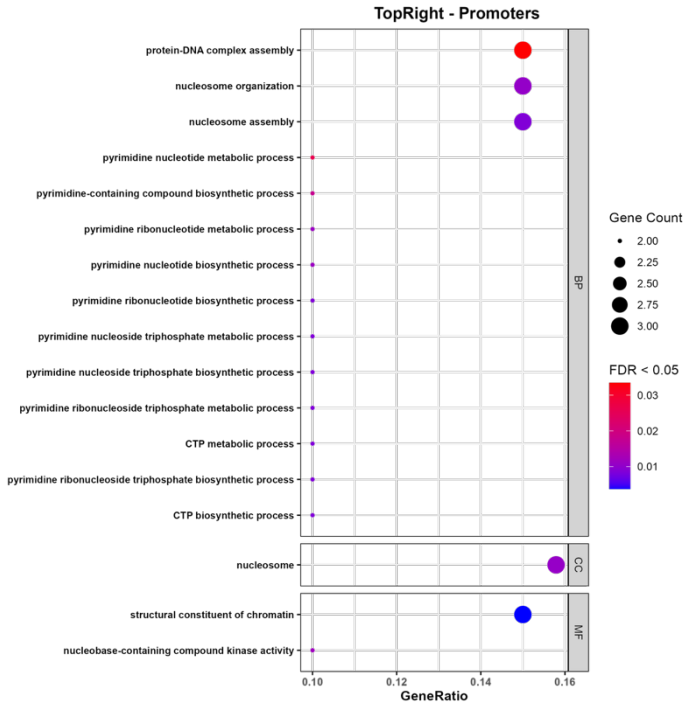
In summary, my analysis reveals that GBM recurrence is accompanied by divergent epigenetic remodelling programs in different subgroups. In Down responders, recurrence is characterised by increased methylation and repression of developmental genes, supporting a shift toward a stem-like mesenchymal state. In contrast, Up responders exhibit decreased methylation in immune-related gene bodies, indicating immune activation. These findings highlight the importance of integrating epigenetic data with clinical and molecular stratification to enhance understanding of recurrence biology and identify new opportunities for personalised therapy in glioblastoma

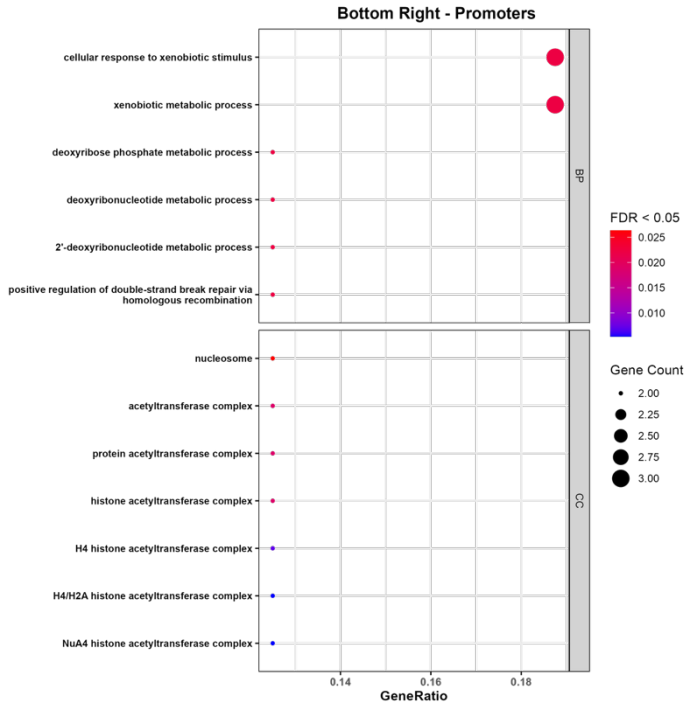
4.5 APPENDIX

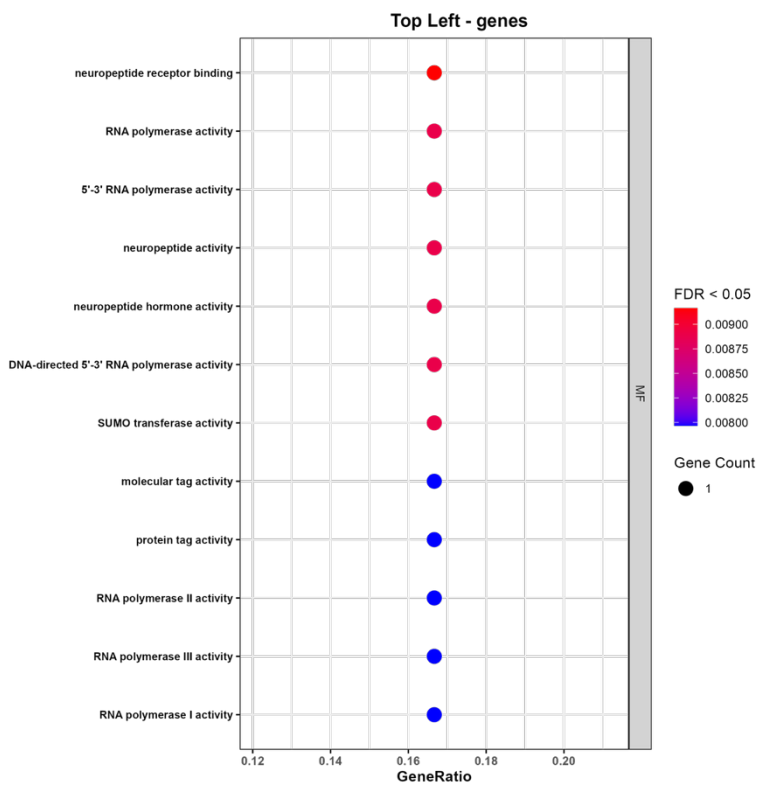
Top_Left:	increasing methylation in up responder and decreasing in down responder
Top_Right:	increasing methylation in both up and down responder
Right_Extreme:	increasing methylation in up responder only
Left_Extreme:	decreasing methylation in down responder only
Bottom_Extreme:	decreased methylation in UP responders only.
Bottom_Right:	increasing methylation in down responder and decreasing in up responder

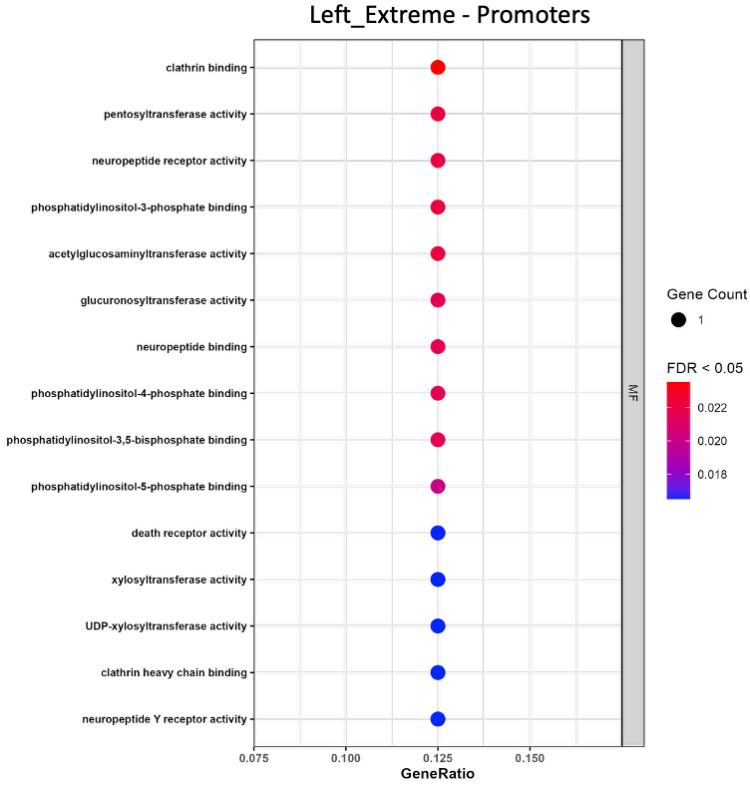





TopRight - Promoters
protein-DNA complex assembly
nucleosome organization
nucleosome assembly
pyrimidine nucleotide metabolic process
pyrimidine-containing compound biosynthetic process
pyrimidine ribonucleotide metabolic process
pyrimidine nucleotide biosynthetic process
pyrimidine ribonucleotide biosynthetic process
pyrimidine nucleoside triphosphate metabolic process
pyrimidine nucleoside triphosphate biosynthetic process
pyrimidine ribonucleoside triphosphate metabolic process
CTP metabolic process
pyrimidine ribonucleoside triphosphate biosynthetic process
CTP biosynthetic process
nucleosome
structural constituent of chromatin
nucleobase-containing compound kinase activity
BP
CC
MF
0.10
0.12
0.14
0.16
GeneRatio
Gene Count
2.00
2.25
2.50
2.75
3.00
FDR < 0.05
0.03
0.02
0.01


Bottom Right - Promoters
cellular response to xenobiotic stimulus
xenobiotic metabolic process
deoxyribose phosphate metabolic process
deoxyribonucleotide metabolic process
2'-deoxyribonucleotide metabolic process
positive regulation of double-strand break repair via homologous recombination
nucleosome
acetyltransferase complex
protein acetyltransferase complex
histone acetyltransferase complex
H4 histone acetyltransferase complex
H4/H2A histone acetyltransferase complex
NuA4 histone acetyltransferase complex
BP
CC
0.14
0.16
0.18
GeneRatio
FDR < 0.05
0.025
0.020
0.015
0.010
Gene Count
2.00
2.25
2.50
2.75
3.00


Top Left - genes
neuropeptide receptor binding
RNA polymerase activity
5'-3' RNA polymerase activity
neuropeptide activity
neuropeptide hormone activity
DNA-directed 5'-3' RNA polymerase activity
SUMO transferase activity
molecular tag activity
protein tag activity
RNA polymerase II activity
RNA polymerase III activity
RNA polymerase I activity
MF
0.12
0.14
0.16
0.18
0.20
GeneRatio
FDR < 0.05
0.00900
0.00875
0.00850
0.00825
0.00800
Gene Count
1


Left_Extreme - Promoters
clathrin binding
pentosyltransferase activity
neuropeptide receptor activity
phosphatidylinositol-3-phosphate binding
acetylglucosaminyltransferase activity
glucuronosyltransferase activity
neuropeptide binding
phosphatidylinositol-4-phosphate binding
phosphatidylinositol-3,5-bisphosphate binding
phosphatidylinositol-5-phosphate binding
death receptor activity
xylosyltransferase activity
UDP-xylosyltransferase activity
clathrin heavy chain binding
neuropeptide Y receptor activity
MF
0.075
0.100
0.125
0.150
GeneRatio
Gene Count
1
FDR < 0.05
0.022
0.020
0.018

4.6 REFERENCES

AALTONEN, L. A., ABASCAL, F., ABESHOUSE, A., ABURATANI, H., ADAMS, D. J., AGRAWAL, N., AHN, K. S., AHN, S.-M., AIKATA, H., AKBANI, R., AKDEMIR, K. C., AL-AHMADIE, H., AL-SEDAIRY, S. T., AL-SHAHROUR, F., ALAWI, M., ALBERT, M., ALDAPE, K., ALEXANDROV, L. B., ALLY, A., ALSOP, K., ALVAREZ, E. G., AMARY, F., AMIN, S. B., AMINOU, B., AMMERPOHL, O., ANDERSON, M. J., ANG, Y., ANTONELLO, D., ANUR, P., APARICIO, S., APPELBAUM, E. L., ARAI, Y., ARETZ, A., ARIHIRO, K., ARIIZUMI, S.-I., ARMENIA, J., ARNOULD, L., ASA, S., ASSENOV, Y., ATWAL, G., AUKEMA, S., AUMAN, J. T., AURE, M. R. R., AWADALLA, P., AYMERICH, M., BADER, G. D., BAEZ-ORTEGA, A., BAILEY, M. H., BAILEY, P. J., BALASUNDARAM, M., BALU, S., BANDOPADHAYAY, P., BANKS, R. E., BARBI, S., BARBOUR, A. P., BARENBOIM, J., BARNHOLTZ-SLOAN, J., BARR, H., BARRERA, E., BARTLETT, J., BARTOLOME, J., BASSI, C., BATHE, O. F., BAUMHOER, D., BAVI, P., BAYLIN, S. B., BAZANT, W., BEARDSMORE, D., BECK, T. A., BEHJATI, S., BEHREN, A., NIU, B., BELL, C., BELTRAN, S., BENZ, C., BERCHUCK, A., BERGMANN, A. K., BERGSTROM, E. N., BERMAN, B. P., BERNEY, D. M., BERNHART, S. H., BEROUKHIM, R., BERRIOS, M., BERSANI, S., BERTL, J., BETANCOURT, M., BHANDARI, V., BHOSLE, S. G., BIANKIN, A. V., BIEG, M., BIGNER, D., BINDER, H., BIRNEY, E., BIRRER, M., BISWAS, N. K., BJERKEHAGEN, B., BODENHEIMER, T., BOICE, L., BONIZZATO, G., DE BONO, J. S., et al. 2020. Pan-cancer analysis of whole genomes. *Nature,* 578**,** 82-93.

ALAJEM, A., ROTH, H., RATGAUZER, S., BAVLI, D., MOTZIK, A., LAHAV, S., PELED, I. & RAM, O. 2021. DNA methylation patterns expose variations in enhancer-chromatin modifications during embryonic stem cell differentiation. *PLoS Genet,* 17**,** e1009498.

ALI, I. & YANG, W. C. 2020. The functions of kinesin and kinesin-related proteins in eukaryotes. *Cell Adh Migr,* 14**,** 139-152.

ALIREZA MANSOURI, J. K., SUNIT DAS 2017. *Molecular Genetics of Secondary Glioblastoma,* Brisbane (AU), Codon Publications; Chapter 2.

AMBROSE, J. C., LI, W., MARCUS, A., MA, H. & CYR, R. 2005. A minus-end-directed kinesin with plus-end tracking protein activity is involved in spindle morphogenesis. *Mol Biol Cell,* 16**,** 1584-92.

AMIRMAHANI, F., KUMAR, S. & MUTHUKRISHNAN, S. D. 2025. Epigenetic mechanisms of plasticity and resistance in glioblastoma: therapeutic targets and implications. *Frontiers in Epigenetics and Epigenomics,* 3.

ANDERSSON, U., SCHWARTZBAUM, J., WIKLUND, F., SJOSTROM, S., LIU, Y., TSAVACHIDIS, S., AHLBOM, A., AUVINEN, A., COLLATZ-LAIER, H., FEYCHTING, M., JOHANSEN, C., KIURU, A., LONN, S., SCHOEMAKER, M. J., SWERDLOW, A. J., HENRIKSSON, R., BONDY, M. & MELIN, B. 2010. A comprehensive study of the association between the EGFR and ERBB2 genes and glioma risk. *Acta Oncol,* 49**,** 767-75.

ANDREWS, S. 2010. FastQC: A Quality Control Tool for High Throughput Sequence Data.

ARORA, A. & SOMASUNDARAM, K. 2019. Glioblastoma vs temozolomide: can the red queen race be won? *Cancer Biol Ther,* 20**,** 1083-1090.

ARYEE, M. J., JAFFE, A. E., CORRADA-BRAVO, H., LADD-ACOSTA, C., FEINBERG, A. P., HANSEN, K. D. & IRIZARRY, R. A. 2014. Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. *Bioinformatics,* 30**,** 1363-9.

ASHBURNER, M., BALL, C. A., BLAKE, J. A., BOTSTEIN, D., BUTLER, H., CHERRY, J. M., DAVIS, A. P., DOLINSKI, K., DWIGHT, S. S., EPPIG, J. T., HARRIS, M. A., HILL, D. P., ISSEL-TARVER, L., KASARSKIS, A., LEWIS, S., MATESE, J. C., RICHARDSON, J. E., RINGWALD, M., RUBIN, G. M. & SHERLOCK, G. 2000. Gene ontology: tool for the unification of biology. The Gene Ontology Consortium. *Nat Genet,* 25**,** 25-9.

BADY, P., DELORENZI, M. & HEGI, M. E. 2016. Sensitivity Analysis of the MGMT-STP27 Model and Impact of Genetic and Epigenetic Context to Predict the MGMT Methylation Status in Gliomas and Other Tumors. *J Mol Diagn,* 18**,** 350-361.

BAGLEY, S. J., KOTHARI, S., RAHMAN, R., LEE, E. Q., DUNN, G. P., GALANIS, E., CHANG, S. M., NABORS, L. B., AHLUWALIA, M. S., STUPP, R., MEHTA, M. P., REARDON, D. A., GROSSMAN, S. A., SULMAN, E. P., SAMPSON, J. H., KHAGI, S., WELLER, M., CLOUGHESY, T. F., WEN, P. Y. & KHASRAW, M. 2022. Glioblastoma Clinical Trials: Current Landscape and Opportunities for Improvement. *Clin Cancer Res,* 28**,** 594-602.

BAILEY, M. H., TOKHEIM, C., PORTA-PARDO, E., SENGUPTA, S., BERTRAND, D., WEERASINGHE, A., COLAPRICO, A., WENDL, M. C., KIM, J., REARDON, B., NG, P. K., JEONG, K. J., CAO, S., WANG, Z., GAO, J., GAO, Q., WANG, F., LIU, E. M., MULARONI, L., RUBIO-PEREZ, C., NAGARAJAN, N., CORTES-CIRIANO, I., ZHOU, D. C., LIANG, W. W., HESS, J. M., YELLAPANTULA, V. D., TAMBORERO, D., GONZALEZ-PEREZ, A., SUPHAVILAI, C., KO, J. Y., KHURANA, E., PARK, P. J., VAN ALLEN, E. M., LIANG, H., GROUP, M. C. W., CANCER GENOME ATLAS RESEARCH, N., LAWRENCE, M. S., GODZIK, A., LOPEZ-BIGAS, N., STUART, J., WHEELER, D., GETZ, G., CHEN, K., LAZAR, A. J., MILLS, G. B., KARCHIN, R. & DING, L. 2018. Comprehensive Characterization of Cancer Driver Genes and Mutations. *Cell,* 173**,** 371-385 e18.

BAKER, S. D., WIRTH, M., STATKEVICH, P., REIDENBERG, P., ALTON, K., SARTORIUS, S. E., DUGAN, M., CUTLER, D., BATRA, V., GROCHOW, L. B., DONEHOWER, R. C. & ROWINSKY, E. K. 1999. Absorption, metabolism, and excretion of 14C-temozolomide following oral administration to patients with advanced cancer. *Clin Cancer Res,* 5**,** 309-17.

BANELLI, B., CARRA, E., BARBIERI, F., WURTH, R., PARODI, F., PATTAROZZI, A., CAROSIO, R., FORLANI, A., ALLEMANNI, G., MARUBBI, D., FLORIO, T., DAGA, A. & ROMANI, M. 2015. The histone demethylase KDM5A is a key factor for the resistance to temozolomide in glioblastoma. *Cell Cycle,* 14**,** 3418-29.

BARBAGALLO, G. M., JENKINSON, M. D. & BRODBELT, A. R. 2008. 'Recurrent' glioblastoma multiforme, when should we reoperate? *Br J Neurosurg,* 22**,** 452-5.

BARRERA, V. & PEINADO, M. A. 2012. Evaluation of single CpG sites as proxies of CpG island methylation states at the genome scale. *Nucleic Acids Res,* 40**,** 11490-8.

BARTHEL, F. P., JOHNSON, K. C., VARN, F. S., MOSKALIK, A. D., TANNER, G., KOCAKAVUK, E., ANDERSON, K. J., ABIOLA, O., ALDAPE, K., ALFARO, K. D., ALPAR, D., AMIN, S. B., ASHLEY, D. M., BANDOPADHAYAY, P., BARNHOLTZ-SLOAN, J. S., BEROUKHIM, R., BOCK, C., BRASTIANOS, P. K., BRAT, D. J., BRODBELT, A. R., BRUNS, A. F., BULSARA, K. R., CHAKRABARTY, A., CHAKRAVARTI, A., CHUANG, J. H., CLAUS, E. B., COCHRAN, E. J., CONNELLY, J., COSTELLO, J. F., FINOCCHIARO, G., FLETCHER, M. N., FRENCH, P. J., GAN, H. K., GILBERT, M. R., GOULD, P. V., GRIMMER, M. R., IAVARONE, A., ISMAIL, A., JENKINSON, M. D., KHASRAW, M., KIM, H., KOUWENHOVEN, M. C. M., LAVIOLETTE, P. S., LI, M., LICHTER, P., LIGON, K. L., LOWMAN, A. K., MALTA, T. M., MAZOR, T., MCDONALD, K. L., MOLINARO, A. M., NAM, D. H., NAYYAR, N., NG, H. K., NGAN, C. Y., NICLOU, S. P., NIERS, J. M., NOUSHMEHR, H., NOORBAKHSH, J., ORMOND, D. R., PARK, C. K., POISSON, L. M., RABADAN, R., RADLWIMMER, B., RAO, G., REIFENBERGER, G., SA, J. K., SCHUSTER, M., SHAW, B. L., SHORT, S. C., SMITT, P. A. S., SLOAN, A. E., SMITS, M., SUZUKI, H., TABATABAI, G., VAN MEIR, E. G., WATTS, C., WELLER, M., WESSELING, P., WESTERMAN, B. A., WIDHALM, G., WOEHRER, A., YUNG, W. K. A., ZADEH, G., HUSE, J. T., DE GROOT, J. F., STEAD, L. F., VERHAAK, R. G. W. & CONSORTIUM, G. 2019. Longitudinal molecular trajectories of diffuse glioma in adults. *Nature,* 576**,** 112-120.

BARTOLOME, R. A., MARTIN-REGALADO, A., JAEN, M., ZANNIKOU, M., ZHANG, P., DE LOS RIOS, V., BALYASNIKOVA, I. V. & CASAL, J. I. 2020. Protein Tyrosine Phosphatase-1B Inhibition Disrupts IL13Ralpha2-Promoted Invasion and Metastasis in Cancer Cells. *Cancers (Basel),* 12.

BECKER, A. P., SELLS, B. E., HAQUE, S. J. & CHAKRAVARTI, A. 2021. Tumor Heterogeneity in Glioblastomas: From Light Microscopy to Molecular Pathology. *Cancers (Basel),* 13.

BELL, R. E., GOLAN, T., SHEINBOIM, D., MALCOV, H., AMAR, D., SALAMON, A., LIRON, T., GELFMAN, S., GABET, Y., SHAMIR, R. & LEVY, C. 2016. Enhancer methylation dynamics contribute to cancer plasticity and patient mortality. *Genome Res,* 26**,** 601-11.

BENJAMIN, D., SATO, T., CIBULSKIS, K., GETZ, G., STEWART, C. & LICHTENSTEIN, L. 2019.

BERENS, M. E. & GIESE, A. 1999. "...those left behind." Biology and oncology of invasive glioma cells. *Neoplasia,* 1**,** 208-19.

BERGMANN, N., DELBRIDGE, C., GEMPT, J., FEUCHTINGER, A., WALCH, A., SCHIRMER, L., BUNK, W., ASCHENBRENNER, T., LIESCHE-STARNECKER, F. & SCHLEGEL, J. 2020. The Intratumoral Heterogeneity Reflects the Intertumoral Subtypes of Glioblastoma Multiforme: A Regional Immunohistochemistry Analysis. *Front Oncol,* 10**,** 494.

BIBIKOVA, M., BARNES, B., TSAN, C., HO, V., KLOTZLE, B., LE, J. M., DELANO, D., ZHANG, L., SCHROTH, G. P., GUNDERSON, K. L., FAN, J. B. & SHEN, R. 2011. High density DNA methylation array with single CpG site resolution. *Genomics,* 98**,** 288-95.

BIRD, A. 2002. DNA methylation patterns and epigenetic memory. *Genes Dev,* 16**,** 6-21.

BIRD, A. 2007. Perceptions of epigenetics. *Nature,* 447**,** 396-8.

BIRZU, C., FRENCH, P., CACCESE, M., CERRETTI, G., IDBAIH, A., ZAGONEL, V. & LOMBARDI, G. 2020. Recurrent Glioblastoma: From Molecular Landscape to New Treatment Perspectives. *Cancers (Basel),* 13.

BJORLAND, L. S., FLUGE, O., GILJE, B., MAHESPARAN, R. & FARBU, E. 2021. Treatment approach and survival from glioblastoma: results from a population-based retrospective cohort study from Western Norway. *BMJ Open,* 11**,** e043208.

BOCK, C. 2012. Analysing and interpreting DNA methylation data. *Nat Rev Genet,* 13**,** 705-19.

BRANDES, A. A., FRANCESCHI, E., PACCAPELO, A., TALLINI, G., DE BIASE, D., GHIMENTON, C., DANIELI, D., ZUNARELLI, E., LANZA, G., SILINI, E. M., STURIALE, C., VOLPIN, L., SERVADEI, F., TALACCHI, A., FIORAVANTI, A., PIA FOSCHINI, M., BARTOLINI, S., PESSION, A. & ERMANI, M. 2017. Role of MGMT Methylation Status at Time of Diagnosis and Recurrence for Patients with Glioblastoma: Clinical Implications. *Oncologist,* 22**,** 432-437.

BRASTIANOS, P. K., NAYYAR, N., ROSEBROCK, D., LESHCHINER, I., GILL, C. M., LIVITZ, D., BERTALAN, M. S., D'ANDREA, M., HOANG, K., AQUILANTI, E., CHUKWUEKE, U. N., KANEB, A., CHI, A., PLOTKIN, S., GERSTNER, E. R., FROSCH, M. P., SUVA, M. L., CAHILL, D. P., GETZ, G. & BATCHELOR, T. T. 2017. Resolving the phylogenetic origin of glioblastoma via multifocal genomic analysis of pre-treatment and treatment-resistant autopsy specimens. *NPJ Precis Oncol,* 1**,** 33.

BRENNAN, C. W., VERHAAK, R. G., MCKENNA, A., CAMPOS, B., NOUSHMEHR, H., SALAMA, S. R., ZHENG, S., CHAKRAVARTY, D., SANBORN, J. Z., BERMAN, S. H., BEROUKHIM, R., BERNARD, B., WU, C. J., GENOVESE, G., SHMULEVICH, I., BARNHOLTZ-SLOAN, J., ZOU, L., VEGESNA, R., SHUKLA, S. A., CIRIELLO, G., YUNG, W. K., ZHANG, W., SOUGNEZ, C., MIKKELSEN, T., ALDAPE, K., BIGNER, D. D., VAN MEIR, E. G., PRADOS, M., SLOAN, A., BLACK, K. L., ESCHBACHER, J., FINOCCHIARO, G., FRIEDMAN, W., ANDREWS, D. W., GUHA, A., IACOCCA, M., O'NEILL, B. P., FOLTZ, G., MYERS, J., WEISENBERGER, D. J., PENNY, R., KUCHERLAPATI, R., PEROU, C. M., HAYES, D. N., GIBBS, R., MARRA, M., MILLS, G. B., LANDER, E., SPELLMAN, P., WILSON, R., SANDER, C., WEINSTEIN, J., MEYERSON, M., GABRIEL, S., LAIRD, P. W., HAUSSLER, D., GETZ, G., CHIN, L. & NETWORK, T. R. 2013. The somatic genomic landscape of glioblastoma. *Cell,* 155**,** 462-77.

BRODBELT, A., GREENBERG, D., WINTERS, T., WILLIAMS, M., VERNON, S., COLLINS, V. P. & NATIONAL CANCER INFORMATION NETWORK BRAIN TUMOUR, G. 2015. Glioblastoma in England: 2007-2011. *Eur J Cancer,* 51**,** 533-542.

BUSHNELL, B., ROOD, J. & SINGER, E. 2017. BBMerge - Accurate paired shotgun read merging via overlap. *PLoS One,* 12**,** e0185056.

BYRNE, K. F., PAL, A., CURTIN, J. F., STEPHENS, J. C. & KINSELLA, G. K. 2021. G-protein-coupled receptors as therapeutic targets for glioblastoma. *Drug Discov Today,* 26**,** 2858-2870.

CANCER GENOME ATLAS RESEARCH, N. 2008. Comprehensive genomic characterization defines human glioblastoma genes and core pathways. *Nature,* 455**,** 1061-8.

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., HOLSKEN, 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., BRUCK, W., HOFER, S., BREHMER, S., SEIZ-ROSENHAGEN, M., HANGGI, 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., MUHLEISEN, H., MUELLER, W., POHL, U., PRINZ, M., BENNER, A., ZAPATKA, M., GOTTARDO, N. G., DRIEVER, P. H., KRAMM, C. M., MULLER, H. L., RUTKOWSKI, S., VON HOFF, K., FRUHWALD, 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.

CHANG, S. M., PARNEY, I. F., HUANG, W., ANDERSON, F. A., JR., ASHER, A. L., BERNSTEIN, M., LILLEHEI, K. O., BREM, H., BERGER, M. S., LAWS, E. R. & GLIOMA OUTCOMES PROJECT, I. 2005. Patterns of care for adults with newly diagnosed malignant glioma. *JAMA,* 293**,** 557-64.

CHENG, F. & GUO, D. 2019. MET in glioma: signaling pathways and targeted therapies. *J Exp Clin Cancer Res,* 38**,** 270.

CHOI, H. J., CHOI, S. H., YOU, S. H., YOO, R. E., KANG, K. M., YUN, T. J., KIM, J. H., SOHN, C. H., PARK, C. K. & PARK, S. H. 2021. MGMT Promoter Methylation Status in Initial and Recurrent Glioblastoma: Correlation Study with DWI and DSC PWI Features. *AJNR Am J Neuroradiol,* 42**,** 853-860.

CHONG, C. R. & JANNE, P. A. 2013. The quest to overcome resistance to EGFR-targeted therapies in cancer. *Nat Med,* 19**,** 1389-400.

COLOGNATO, H. & YURCHENCO, P. D. 2000. Form and function: The laminin family of heterotrimers. *Developmental Dynamics,* 218**,** 213-234.

COMBA, A., FAISAL, S. M., VARELA, M. L., HOLLON, T., AL-HOLOU, W. N., UMEMURA, Y., NUNEZ, F. J., MOTSCH, S., CASTRO, M. G. & LOWENSTEIN, P. R. 2021. Uncovering Spatiotemporal Heterogeneity of High-Grade Gliomas: From Disease Biology to Therapeutic Implications. *Front Oncol,* 11**,** 703764.

CONSORTIUM, G. 2018. Glioma through the looking GLASS: molecular evolution of diffuse gliomas and the Glioma Longitudinal Analysis Consortium. *Neuro-Oncology,* 20**,** 873-884.

CONSORTIUM, P.-C. A. O. W. G. 2020. Pan-cancer analysis of whole genomes. *Nature,* 578**,** 82-93.

COSTELLO, M., PUGH, T. J., FENNELL, T. J., STEWART, C., LICHTENSTEIN, L., MELDRIM, J. C., FOSTEL, J. L., FRIEDRICH, D. C., PERRIN, D., DIONNE, D., KIM, S., GABRIEL, S. B., LANDER, E. S., FISHER, S. & GETZ, G. 2013. Discovery and characterization of artifactual mutations in deep coverage targeted capture sequencing data due to oxidative DNA damage during sample preparation. *Nucleic Acids Res,* 41**,** e67.

CRESPO, I., VITAL, A. L., GONZALEZ-TABLAS, M., PATINO MDEL, C., OTERO, A., LOPES, M. C., DE OLIVEIRA, C., DOMINGUES, P., ORFAO, A. & TABERNERO, M. D. 2015. Molecular and Genomic Alterations in Glioblastoma Multiforme. *Am J Pathol,* 185**,** 1820-33.

DARRÉ, H., MASSON, P., NATIVEL, A., VILLAIN, L., LEFAUDEUX, D., COUTY, C., MARTIN, B., JACOB, E., DURUISSEAUX, M., PALGEN, J.-L., MONTEIRO, C. & L'HOSTIS, A. 2024. Comparing the Efficacy of Two Generations of EGFR-TKIs: An Integrated Drug–Disease Mechanistic Model Approach in EGFR-Mutated Lung Adenocarcinoma. *Biomedicines,* 12.

DE CONTI, G., DIAS, M. H. & BERNARDS, R. 2021. Fighting Drug Resistance through the Targeting of Drug-Tolerant Persister Cells. *Cancers (Basel),* 13.

DI TOMMASO, P., CHATZOU, M., FLODEN, E. W., BARJA, P. P., PALUMBO, E. & NOTREDAME, C. 2017. Nextflow enables reproducible computational workflows. *Nat Biotechnol,* 35**,** 316-319.

DIPLAS, B. H., HE, X., BROSNAN-CASHMAN, J. A., LIU, H., CHEN, L. H., WANG, Z., MOURE, C. J., KILLELA, P. J., LORIAUX, D. B., LIPP, E. S., GREER, P. K., YANG, R., RIZZO, A. J., RODRIGUEZ, F. J., FRIEDMAN, A. H., FRIEDMAN, H. S., WANG, S., HE, Y., MCLENDON, R. E., BIGNER, D. D., JIAO, Y., WAITKUS, M. S., MEEKER, A. K. & YAN, H. 2018. The genomic landscape of TERT promoter wildtype-IDH wildtype glioblastoma. *Nat Commun,* 9**,** 2087.

DONOGHUE, J. F., KERR, L. T., ALEXANDER, N. W., GREENALL, S. A., LONGANO, A. B., GOTTARDO, N. G., WANG, R., TABAR, V., ADAMS, T. E., MISCHEL, P. S. & JOHNS, T. G. 2018. Activation of ERBB4 in Glioblastoma Can Contribute to Increased Tumorigenicity and Influence Therapeutic Response. *Cancers (Basel),* 10.

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. *Br J Cancer,* 101**,** 124-31.

DUTERTRE, M., SANCHEZ, G., BARBIER, J., CORCOS, L. & AUBOEUF, D. 2014. The emerging role of pre-messenger RNA splicing in stress responses: Sending alternative messages and silent messengers. *RNA Biology,* 8**,** 740-747.

DYMOVA, M. A., KULIGINA, E. V. & RICHTER, V. A. 2021. Molecular Mechanisms of Drug Resistance in Glioblastoma. *Int J Mol Sci,* 22.

EDER, K. & KALMAN, B. 2014. Molecular heterogeneity of glioblastoma and its clinical relevance. *Pathol Oncol Res,* 20**,** 777-87.

EINAGA, N., YOSHIDA, A., NODA, H., SUEMITSU, M., NAKAYAMA, Y., SAKURADA, A., KAWAJI, Y., YAMAGUCHI, H., SASAKI, Y., TOKINO, T. & ESUMI, M. 2017. Assessment of the quality of DNA from various formalin-fixed paraffin-embedded (FFPE) tissues and the use of this DNA for next-generation sequencing (NGS) with no artifactual mutation. *PLoS One,* 12**,** e0176280.

ESTELLER, M. 2008. Epigenetics in cancer. *N Engl J Med,* 358**,** 1148-59.

ETCHEVERRY, A., AUBRY, M., DE TAYRAC, M., VAULEON, E., BONIFACE, R., GUENOT, F., SAIKALI, S., HAMLAT, A., RIFFAUD, L., MENEI, P., QUILLIEN, V. & MOSSER, J. 2010. DNA methylation in glioblastoma: impact on gene expression and clinical outcome. *BMC Genomics,* 11**,** 701.

EYLER, C. E., MATSUNAGA, H., HOVESTADT, V., VANTINE, S. J., VAN GALEN, P. & BERNSTEIN, B. E. 2020. Single-cell lineage analysis reveals genetic and epigenetic interplay in glioblastoma drug resistance. *Genome Biol,* 21**,** 174.

FERRER, V. P. 2022. MUC16 mutation is associated with tumor grade, clinical features, and prognosis in glioma patients.

FRENCH, P. J., EOLI, M., SEPULVEDA, J. M., DE HEER, I., KROS, J. M., WALENKAMP, A., FRENEL, J. S., FRANCESCHI, E., CLEMENT, P. M., WELLER, M., ANSELL, P., LOOMAN, J., BAIN, E., MORFOUACE, M., GORLIA, T. & VAN DEN BENT, M. 2019. Defining EGFR amplification status for clinical trial inclusion. *Neuro Oncol,* 21**,** 1263-1272.

FRIEDMANN-MORVINSKI, D. 2014. Glioblastoma heterogeneity and cancer cell plasticity. *Crit Rev Oncog,* 19**,** 327-36.

GAFFNEY, S. G. & TOWNSEND, J. P. 2016. PathScore: a web tool for identifying altered pathways in cancer data. *Bioinformatics,* 32**,** 3688-3690.

GAILLARD, F. 2024. Temozolomide mechanism of action and resistance. (Case study). *Radiopaedia.org*.

GAN, H. K., CVRLJEVIC, A. N. & JOHNS, T. G. 2013. The epidermal growth factor receptor variant III (EGFRvIII): where wild things are altered. *FEBS J,* 280**,** 5350-70.

GARCIA, J., HURWITZ, H. I., SANDLER, A. B., MILES, D., COLEMAN, R. L., DEURLOO, R. & CHINOT, O. L. 2020. Bevacizumab (Avastin(R)) in cancer treatment: A review of 15 years of clinical experience and future outlook. *Cancer Treat Rev,* 86**,** 102017.

GENE ONTOLOGY, C., ALEKSANDER, S. A., BALHOFF, J., CARBON, S., CHERRY, J. M., DRABKIN, H. J., EBERT, D., FEUERMANN, M., GAUDET, P., HARRIS, N. L., HILL, D. P., LEE, R., MI, H., MOXON, S., MUNGALL, C. J., MURUGANUGAN, A., MUSHAYAHAMA, T., STERNBERG, P. W., THOMAS, P. D., VAN AUKEN, K., RAMSEY, J., SIEGELE, D. A., CHISHOLM, R. L., FEY, P., ASPROMONTE, M. C., NUGNES, M. V., QUAGLIA, F., TOSATTO, S., GIGLIO, M., NADENDLA, S., ANTONAZZO, G., ATTRILL, H., DOS SANTOS, G., MARYGOLD, S., STRELETS, V., TABONE, C. J., THURMOND, J., ZHOU, P., AHMED, S. H., ASANITTHONG, P., LUNA BUITRAGO, D., ERDOL, M. N., GAGE, M. C., ALI KADHUM, M., LI, K. Y. C., LONG, M., MICHALAK, A., PESALA, A., PRITAZAHRA, A., SAVERIMUTTU, S. C. C., SU, R., THURLOW, K. E., LOVERING, R. C., LOGIE, C., OLIFERENKO, S., BLAKE, J., CHRISTIE, K., CORBANI, L., DOLAN, M. E., DRABKIN, H. J., HILL, D. P., NI, L., SITNIKOV, D., SMITH, C., CUZICK, A., SEAGER, J., COOPER, L., ELSER, J., JAISWAL, P., GUPTA, P., JAISWAL, P., NAITHANI, S., LERA-RAMIREZ, M., RUTHERFORD, K., WOOD, V., DE PONS, J. L., DWINELL, M. R., HAYMAN, G. T., KALDUNSKI, M. L., KWITEK, A. E., LAULEDERKIND, S. J. F., TUTAJ, M. A., VEDI, M., WANG, S. J., D'EUSTACHIO, P., AIMO, L., AXELSEN, K., BRIDGE, A., HYKA-NOUSPIKEL, N., MORGAT, A., ALEKSANDER, S. A., CHERRY, J. M., ENGEL, S. R., KARRA, K., MIYASATO, S. R., NASH, R. S., SKRZYPEK, M. S., WENG, S., WONG, E. D., BAKKER, E., et al. 2023. The Gene Ontology knowledgebase in 2023. *Genetics,* 224.

GILARD, V., TEBANI, A., DABAJ, I., LAQUERRIERE, A., FONTANILLES, M., DERREY, S., MARRET, S. & BEKRI, S. 2021. Diagnosis and Management of Glioblastoma: A Comprehensive Perspective. *J Pers Med,* 11.

GILBERT, M. R. 2011. Recurrent glioblastoma: a fresh look at current therapies and emerging novel approaches. *Semin Oncol,* 38 Suppl 4**,** S21-33.

GRIFFIN, M., KHAN, R., BASU, S. & SMITH, S. 2020. Ion Channels as Therapeutic Targets in High Grade Gliomas. *Cancers (Basel),* 12.

GROCHANS, S., CYBULSKA, A. M., SIMINSKA, D., KORBECKI, J., KOJDER, K., CHLUBEK, D. & BARANOWSKA-BOSIACKA, I. 2022. Epidemiology of Glioblastoma Multiforme-Literature Review. *Cancers (Basel),* 14.

HADJIPANAYIS, C. G. & STUMMER, W. 2019. 5-ALA and FDA approval for glioma surgery. *J Neurooncol,* 141**,** 479-486.

HANIF, F., MUZAFFAR, K., PERVEEN, K., MALHI, S. M. & SIMJEE SH, U. 2017. Glioblastoma Multiforme: A Review of its Epidemiology and Pathogenesis through Clinical Presentation and Treatment. *Asian Pac J Cancer Prev,* 18**,** 3-9.

HAUSMANN, D., HOFFMANN, D. C., VENKATARAMANI, V., JUNG, E., HORSCHITZ, S., TETZLAFF, S. K., JABALI, A., HAI, L., KESSLER, T., AZORIN, D. D., WEIL, S., KOURTESAKIS, A., SIEVERS, P., HABEL, A., BRECKWOLDT, M. O., KARREMAN, M. A., RATLIFF, M., MESSMER, J. M., YANG, Y., REYHAN, E., WENDLER, S., LOB, C., MAYER, C., FIGARELLA, K., OSSWALD, M., SOLECKI, G., SAHM, F., GARASCHUK, O., KUNER, T., KOCH, P., SCHLESNER, M., WICK, W. & WINKLER, F. 2023. Autonomous rhythmic activity in glioma networks drives brain tumour growth. *Nature,* 613**,** 179-186.

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.

HIGGINS, M., OBAIDI, I. & MCMORROW, T. 2019. Primary cilia and their role in cancer. *Oncol Lett,* 17**,** 3041-3047.

HOOGSTRATE, Y., GHISAI, S. A., DE WIT, M., DE HEER, I., DRAAISMA, K., VAN RIET, J., VAN DE WERKEN, H. J. G., BOURS, V., BUTER, J., VANDEN BEMPT, I., EOLI, M., FRANCESCHI, E., FRENEL, J. S., GORLIA, T., HANSE, M. C., HOEBEN, A., KERKHOF, M., KROS, J. M., LEENSTRA, S., LOMBARDI, G., LUKACOVA, S., ROBE, P. A., SEPULVEDA, J. M., TAAL, W., TAPHOORN, M., VERNHOUT, R. M., WALENKAMP, A. M. E., WATTS, C., WELLER, M., DE VOS, F. Y. F., JENSTER, G. W., VAN DEN BENT, M. & FRENCH, P. J. 2022. The EGFRvIII transcriptome in glioblastoma: A meta-omics analysis. *Neuro Oncol,* 24**,** 429-441.

HOVESTADT, V., JONES, D. T., PICELLI, S., WANG, W., KOOL, M., NORTHCOTT, P. A., SULTAN, M., STACHURSKI, K., RYZHOVA, M., WARNATZ, H. J., RALSER, M., BRUN, S., BUNT, J., JAGER, N., KLEINHEINZ, K., ERKEK, S., WEBER, U. D., BARTHOLOMAE, C. C., VON KALLE, C., LAWERENZ, C., EILS, J., KOSTER, J., VERSTEEG, R., MILDE, T., WITT, O., SCHMIDT, S., WOLF, S., PIETSCH, T., RUTKOWSKI, S., SCHEURLEN, W., TAYLOR, M. D., BRORS, B., FELSBERG, J., REIFENBERGER, G., BORKHARDT, A., LEHRACH, H., WECHSLER-REYA, R. J., EILS, R., YASPO, M. L., LANDGRAF, P., KORSHUNOV, A., ZAPATKA, M., RADLWIMMER, B., PFISTER, S. M. & LICHTER, P. 2014. Decoding the regulatory landscape of medulloblastoma using DNA methylation sequencing. *Nature,* 510**,** 537-41.

HUANG, R. X. & ZHOU, P. K. 2020. DNA damage response signaling pathways and targets for radiotherapy sensitization in cancer. *Signal Transduct Target Ther,* 5**,** 60.

HYNES, N. E. & LANE, H. A. 2005. ERBB receptors and cancer: the complexity of targeted inhibitors. *Nat Rev Cancer,* 5**,** 341-54.

IRIZARRY, R. A., LADD-ACOSTA, C., WEN, B., WU, Z., MONTANO, C., ONYANGO, P., CUI, H., GABO, K., RONGIONE, M., WEBSTER, M., JI, H., POTASH, J., SABUNCIYAN, S. & FEINBERG, A. P. 2009. The human colon cancer methylome shows similar hypo- and hypermethylation at conserved tissue-specific CpG island shores. *Nat Genet,* 41**,** 178-186.

JALIFE, J. 1984. Mutual entrainment and electrical coupling as mechanisms for synchronous firing of rabbit sino-atrial pacemaker cells. *The Journal of Physiology,* 356**,** 221-243.

JANG, H., SMITH, I. N., ENG, C. & NUSSINOV, R. 2021. The mechanism of full activation of tumor suppressor PTEN at the phosphoinositide-enriched membrane. *iScience,* 24**,** 102438.

JING, Z., LI, L., WANG, X., WANG, M., CAI, Y., JIN, Z. I. & ZHANG, Y. E. 2016. High c-Cbl expression in gliomas is associated with tumor progression and poor prognosis. *Oncol Lett,* 11**,** 2787-2791.

JOHNSON, D. R. & O'NEILL, B. P. 2012. Glioblastoma survival in the United States before and during the temozolomide era. *J Neurooncol,* 107**,** 359-64.

JONES, P. A. 2012. Functions of DNA methylation: islands, start sites, gene bodies and beyond. *Nat Rev Genet,* 13**,** 484-92.

JONES, P. A. & BAYLIN, S. B. 2002. The fundamental role of epigenetic events in cancer. *Nat Rev Genet,* 3**,** 415-28.

KAINA, B. 2023. Temozolomide, Procarbazine and Nitrosoureas in the Therapy of Malignant Gliomas: Update of Mechanisms, Drug Resistance and Therapeutic Implications. *J Clin Med,* 12.

KAWATAKI, T., YAMANE, T., NAGANUMA, H., ROUSSELLE, P., ANDUREN, I., TRYGGVASON, K. & PATARROYO, M. 2007. Laminin isoforms and their integrin receptors in glioma cell migration and invasiveness: Evidence for a role of alpha5-laminin(s) and alpha3beta1 integrin. *Exp Cell Res,* 313**,** 3819-31.

KHANNA, A., LARSON, D., SRIVATSAN, S., MOSIOR, M., ABBOTT, T., KIWALA, S., LEY, T., DUNCAVAGE, E., WALTER, M., WALKER, J., GRIFFITH, O., GRIFFITH, M. & MILLER, C. 2022. Bam-readcount - rapid generation of basepair-resolution sequence metrics. *Journal of Open Source Software,* 7.

KIM, S., PARK, C., JI, Y., KIM, D. G., BAE, H., VAN VRANCKEN, M., KIM, D. H. & KIM, K. M. 2017. Deamination Effects in Formalin-Fixed, Paraffin-Embedded Tissue Samples in the Era of Precision Medicine. *J Mol Diagn,* 19**,** 137-146.

KITANGE, G. J., CARLSON, B. L., SCHROEDER, M. A., GROGAN, P. T., LAMONT, J. D., DECKER, P. A., WU, W., JAMES, C. D. & SARKARIA, J. N. 2009. Induction of MGMT expression is associated with temozolomide resistance in glioblastoma xenografts. *Neuro Oncol,* 11**,** 281-91.

KORBER, V., YANG, J., BARAH, P., WU, Y., STICHEL, D., GU, Z., FLETCHER, M. N. C., JONES, D., HENTSCHEL, B., LAMSZUS, K., TONN, J. C., SCHACKERT, G., SABEL, M., FELSBERG, J., ZACHER, A., KAULICH, K., HUBSCHMANN, D., HEROLD-MENDE, C., VON DEIMLING, A., WELLER, M., RADLWIMMER, B., SCHLESNER, M., REIFENBERGER, G., HOFER, T. & LICHTER, P. 2019. Evolutionary Trajectories of IDH(WT) Glioblastomas Reveal a Common Path of Early Tumorigenesis Instigated Years ahead of Initial Diagnosis. *Cancer Cell,* 35**,** 692-704 e12.

LAKATTA, E. G., MALTSEV, V. A. & VINOGRADOVA, T. M. 2010. A coupled SYSTEM of intracellular Ca2+ clocks and surface membrane voltage clocks controls the timekeeping mechanism of the heart's pacemaker. *Circ Res,* 106**,** 659-73.

LANGFORD, L. A., PIATYSZEK, M. A., XU, R., SCHOLD, S. C., JR. & SHAY, J. W. 1995. Telomerase activity in human brain tumours. *Lancet,* 346**,** 1267-8.

LASSMAN, A. B., ALDAPE, K. D., ANSELL, P. J., BAIN, E., CURRAN, W. J., EOLI, M., FRENCH, P. J., KINOSHITA, M., LOOMAN, J., MEHTA, M., MURAGAKI, Y., NARITA, Y., OCAMPO, C., ROBERTS-RAPP, L., SONG, M., VOGELBAUM, M. A., WALENKAMP, A. M. E., WANG, T. J. C., ZHANG, P. & VAN DEN BENT, M. J. 2019. Epidermal growth factor receptor (EGFR) amplification rates observed in screening patients for randomized trials in glioblastoma. *J Neurooncol,* 144**,** 205-210.

LATHIA, J. D., LI, M., HALL, P. E., GALLAGHER, J., HALE, J. S., WU, Q., VENERE, M., LEVY, E., RANI, M. R., HUANG, P., BAE, E., SELFRIDGE, J., CHENG, L., GUVENC, H., MCLENDON, R. E., NAKANO, I., SLOAN, A. E., PHILLIPS, H. S., LAI, A., GLADSON, C. L., BREDEL, M., BAO, S., HJELMELAND, A. B. & RICH, J. N. 2012. Laminin alpha 2 enables glioblastoma stem cell growth. *Ann Neurol,* 72**,** 766-78.

LEE, S. Y. 2016. Temozolomide resistance in glioblastoma multiforme. *Genes Dis,* 3**,** 198-210.

LI, H. & DURBIN, R. 2009. Fast and accurate short read alignment with Burrows-Wheeler transform. *Bioinformatics,* 25**,** 1754-60.

LI, M., WANG, J., NG, S. S., CHAN, C. Y., HE, M. L., YU, F., LAI, L., SHI, C., CHEN, Y., YEW, D. T., KUNG, H. F. & LIN, M. C. 2009. Adenosine diphosphate-ribosylation factor 6 is required for epidermal growth factor-induced glioblastoma cell proliferation. *Cancer,* 115**,** 4959-72.

LIAO, Z., JIANG, W., YE, L., LI, T., YU, X. & LIU, L. 2020. Classification of extrachromosomal circular DNA with a focus on the role of extrachromosomal DNA (ecDNA) in tumor heterogeneity and progression. *Biochim Biophys Acta Rev Cancer,* 1874**,** 188392.

LIN, H., LIU, C., HU, A., ZHANG, D., YANG, H. & MAO, Y. 2024. Understanding the immunosuppressive microenvironment of glioma: mechanistic insights and clinical perspectives. *J Hematol Oncol,* 17**,** 31.

LISTER, R., PELIZZOLA, M., DOWEN, R. H., HAWKINS, R. D., HON, G., TONTI-FILIPPINI, J., NERY, J. R., LEE, L., YE, Z., NGO, Q. M., EDSALL, L., ANTOSIEWICZ-BOURGET, J., STEWART, R., RUOTTI, V., MILLAR, A. H., THOMSON, J. A., REN, B. & ECKER, J. R. 2009. Human DNA methylomes at base resolution show widespread epigenomic differences. *Nature,* 462**,** 315-22.

LIU, M., WANG, W., ZHANG, H., BI, J., ZHANG, B., SHI, T., SU, G., ZHENG, Y., FAN, S., HUANG, X., CHEN, B., SONG, Y., ZHAO, Z., SHI, J., LI, P., LU, W. & ZHANG, L. 2023. Three-Dimensional Gene Regulation Network in Glioblastoma Ferroptosis. *Int J Mol Sci,* 24.

LIU, R., MATHIEU, C., BERTHELET, J., ZHANG, W., DUPRET, J. M. & RODRIGUES LIMA, F. 2022. Human Protein Tyrosine Phosphatase 1B (PTP1B): From Structure to Clinical Inhibitor Perspectives. *Int J Mol Sci,* 23.

LIU, S., TANG, Y., YAN, M. & JIANG, W. 2018. PIK3CA mutation sensitizes breast cancer cells to synergistic therapy of PI3K inhibition and AMPK activation. *Invest New Drugs,* 36**,** 763-772.

LIYASOVA, M. S., MA, K. & LIPKOWITZ, S. 2015. Molecular Pathways: Cbl Proteins in Tumorigenesis and Antitumor Immunity—Opportunities for Cancer Treatment. *Clinical Cancer Research,* 21**,** 1789-1794.

LOMBARDI, G., GIUNCO, S., CAVALLIN, F., ANGELINI, C., CACCESE, M., CERRETTI, G., BONIS, P. D., ROSSI, A. D. & ZAGONEL, V. 2021. The clinical significance of telomerase reverse transcriptase (TERT) promoter mutations, telomere length and O6-methylguanine DNA methyltransferase (MGMT) promoter methylation status in newly diagnosed and recurrent IDH-wildtype glioblastoma (GBM) patients (PTS): A large mono-institutional study. *Journal of Clinical Oncology,* 39**,** 2053-2053.

LOTSCH, D., GHANIM, B., LAABER, M., WURM, G., WEIS, S., LENZ, S., WEBERSINKE, G., PICHLER, J., BERGER, W. & SPIEGL-KREINECKER, S. 2013. Prognostic significance of telomerase-associated parameters in glioblastoma: effect of patient age. *Neuro Oncol,* 15**,** 423-32.

LOUIS, D. N., PERRY, A., WESSELING, P., BRAT, D. J., CREE, I. A., FIGARELLA-BRANGER, D., HAWKINS, C., NG, H. K., PFISTER, S. M., REIFENBERGER, G., SOFFIETTI, R., VON DEIMLING, A. & ELLISON, D. W. 2021. The 2021 WHO Classification of Tumors of the Central Nervous System: a summary. *Neuro Oncol,* 23**,** 1231-1251.

LUCAS, L. M., DWIVEDI, V., SENFELD, J. I., CULLUM, R. L., MILL, C. P., PIAZZA, J. T., BRYANT, I. N., COOK, L. J., MILLER, S. T., LOTT, J. H. T., KELLEY, C. M., KNERR, E. L., MARKHAM, J. A., KAUFMANN, D. P., JACOBI, M. A., SHEN, J. & RIESE, D. J., 2ND 2022. The Yin and Yang of ERBB4: Tumor Suppressor and Oncoprotein. *Pharmacol Rev,* 74**,** 18-47.

MACDONALD, E. A., MADL, J., GREINER, J., RAMADAN, A. F., WELLS, S. M., TORRENTE, A. G., KOHL, P., ROG-ZIELINSKA, E. A. & QUINN, T. A. 2020. Sinoatrial Node Structure, Mechanics, Electrophysiology and the Chronotropic Response to Stretch in Rabbit and Mouse. *Front Physiol,* 11**,** 809.

MAKSIMOVIC, J., GORDON, L. & OSHLACK, A. 2012. SWAN: Subset-quantile within array normalization for illumina infinium HumanMethylation450 BeadChips. *Genome Biol,* 13**,** R44.

MALESZEWSKA, M. & KAMINSKA, B. 2013. Is glioblastoma an epigenetic malignancy? *Cancers (Basel),* 5**,** 1120-39.

MALNIC, B., GODFREY, P. A. & BUCK, L. B. 2004. The human olfactory receptor gene family. *Proc Natl Acad Sci U S A,* 101**,** 2584-9.

MALTA, T. M., SABEDOT, T. S., MOROSINI, N. S., DATTA, I., GAROFANO, L., VALLENTGOED, W., VARN, F. S., ALDAPE, K., D'ANGELO, F., BAKAS, S., BARNHOLTZ-SLOAN, J. S., GAN, H. K., HASANAIN, M., HAU, A. C., JOHNSON, K. C., CAZACU, S., DECARVALHO, A. C., KHASRAW, M., KOCAKAVUK, E., KOUWENHOVEN, M. C. M., MIGLIOZZI, S., NICLOU, S. P., NIERS, J. M., ORMOND, D. R., PAEK, S. H., REIFENBERGER, G., SILLEVIS SMITT, P. A., SMITS, M., STEAD, L. F., VAN DEN BENT, M. J., VAN MEIR, E. G., WALENKAMP, A., WEISS, T., WELLER, M., WESTERMAN, B. A., YLSTRA, B., WESSELING, P., LASORELLA, A., FRENCH, P. J., POISSON, L. M., CONSORTIUM THE, G., VERHAAK, R. G. W., IAVARONE, A. & NOUSHMEHR, H. 2024. The Epigenetic Evolution of Glioma Is Determined by the IDH1 Mutation Status and Treatment Regimen. *Cancer Res,* 84**,** 741-756.

MANDEL, J. J., YOUSSEF, M., LUDMIR, E., YUST-KATZ, S., PATEL, A. J. & DE GROOT, J. F. 2018. Highlighting the need for reliable clinical trials in glioblastoma. *Expert Rev Anticancer Ther,* 18**,** 1031-1040.

MARINO, S., MENNA, G., DI BONAVENTURA, R., LISI, L., MATTOGNO, P., FIGA, F., BILGIN, L., D'ALESSANDRIS, Q. G., OLIVI, A. & DELLA PEPA, G. M. 2023. The Extracellular Matrix in Glioblastomas: A Glance at Its Structural Modifications in Shaping the Tumoral Microenvironment-A Systematic Review. *Cancers (Basel),* 15.

MARTIN, M. 2011. Cutadapt removes adapter sequences from high-throughput sequencing reads. *EMBnet.journal,* 17.

MARTINEZ-JIMENEZ, F., MUINOS, F., SENTIS, I., DEU-PONS, J., REYES-SALAZAR, I., ARNEDO-PAC, C., MULARONI, L., PICH, O., BONET, J., KRANAS, H., GONZALEZ-PEREZ, A. & LOPEZ-BIGAS, N. 2020. A compendium of mutational cancer driver genes. *Nat Rev Cancer,* 20**,** 555-572.

MATSUOKA, S. & UEDA, M. 2018. Mutual inhibition between PTEN and PIP3 generates bistability for polarity in motile cells. *Nat Commun,* 9**,** 4481.

MATTOON, D. R., LAMOTHE, B., LAX, I. & SCHLESSINGER, J. 2004. *BMC Biology,* 2.

MAYAKONDA, A., LIN, D. C., ASSENOV, Y., PLASS, C. & KOEFFLER, H. P. 2018. Maftools: efficient and comprehensive analysis of somatic variants in cancer. *Genome Res,* 28**,** 1747-1756.

MCLAREN, W., GIL, L., HUNT, S. E., RIAT, H. S., RITCHIE, G. R., THORMANN, A., FLICEK, P. & CUNNINGHAM, F. 2016. The Ensembl Variant Effect Predictor. *Genome Biol,* 17**,** 122.

MCNULTY, S. N., SCHWETYE, K. E., FERGUSON, C., STORER, C. E., ANSSTAS, G., KIM, A. H., GUTMANN, D. H., RUBIN, J. B., HEAD, R. D. & DAHIYA, S. 2021. BRAF mutations may identify a clinically distinct subset of glioblastoma. *Sci Rep,* 11**,** 19999.

MEISSNER, A., GNIRKE, A., BELL, G. W., RAMSAHOYE, B., LANDER, E. S. & JAENISCH, R. 2005. Reduced representation bisulfite sequencing for comparative high-resolution DNA methylation analysis. *Nucleic Acids Res,* 33**,** 5868-77.

MELEIRO, M. & HENRIQUE, R. 2025. Epigenetic Alterations in Glioblastoma Multiforme as Novel Therapeutic Targets: A Scoping Review. *Int J Mol Sci,* 26.

MELLINGHOFF, I. K., WANG, M. Y., VIVANCO, I., HAAS-KOGAN, D. A., ZHU, S., DIA, E. Q., LU, K. V., YOSHIMOTO, K., HUANG, J. H., CHUTE, D. J., RIGGS, B. L., HORVATH, S., LIAU, L. M., CAVENEE, W. K., RAO, P. N., BEROUKHIM, R., PECK, T. C., LEE, J. C., SELLERS, W. R., STOKOE, D., PRADOS, M., CLOUGHESY, T. F., SAWYERS, C. L. & MISCHEL, P. S. 2005. Molecular determinants of the response of glioblastomas to EGFR kinase inhibitors. *N Engl J Med,* 353**,** 2012-24.

MERMEL, C. H., SCHUMACHER, S. E., HILL, B., MEYERSON, M. L., BEROUKHIM, R. & GETZ, G. 2011. GISTIC2.0 facilitates sensitive and confident localization of the targets of focal somatic copy-number alteration in human cancers. *Genome Biol,* 12**,** R41.

MIAO, B., SKIDAN, I., YANG, J., YOU, Z., FU, X., FAMULOK, M., SCHAFFHAUSEN, B., TORCHILIN, V., YUAN, J. & DEGTEREV, A. 2012. Inhibition of cell migration by PITENINs: the role of ARF6. *Oncogene,* 31**,** 4317-32.

MINATA, M., AUDIA, A., SHI, J., LU, S., BERNSTOCK, J., PAVLYUKOV, M. S., DAS, A., KIM, S. H., SHIN, Y. J., LEE, Y., KOO, H., SNIGDHA, K., WAGHMARE, I., GUO, X., MOHYELDIN, A., GALLEGO-PEREZ, D., WANG, J., CHEN, D., CHENG, P.,

MUKHEEF, F., CONTRERAS, M., REYES, J. F., VAILLANT, B., SULMAN, E. P., CHENG, S. Y., MARKERT, J. M., TANNOUS, B. A., LU, X., KANGO-SINGH, M., LEE, L. J., NAM, D. H., NAKANO, I. & BHAT, K. P. 2019. Phenotypic Plasticity of Invasive Edge Glioma Stem-like Cells in Response to Ionizing Radiation. *Cell Rep,* 26**,** 1893-1905 e7.

MIURA, S., GOMEZ, K., MURILLO, O., HUUKI, L. A., VU, T., BUTURLA, T. & KUMAR, S. 2018. Predicting clone genotypes from tumor bulk sequencing of multiple samples. *Bioinformatics,* 34**,** 4017-4026.

MIYASHITA, Y., KO, R., SHIMADA, N., MITSUISHI, Y., MIURA, K., MATSUMOTO, N., ASAO, T., SHUKUYA, T., SHIBAYAMA, R., KOYAMA, R. & TAKAHASHI, K. 2020. Impact of the generation of EGFR-TKIs administered as prior therapy on the efficacy of osimertinib in patients with non-small cell lung cancer harboring EGFR T790M mutation. *Thoracic Cancer,* 12**,** 329-338.

MOASSER, M. M. 2007. The oncogene HER2: its signaling and transforming functions and its role in human cancer pathogenesis. *Oncogene,* 26**,** 6469-87.

MONTECUCCO, A., ZANETTA, F. & BIAMONTI, G. 2015. Molecular mechanisms of etoposide. *EXCLI J,* 14**,** 95-108.

MORRIS, T. J. & BECK, S. 2015. Analysis pipelines and packages for Infinium HumanMethylation450 BeadChip (450k) data. *Methods,* 72**,** 3-8.

MOTTIS, A., MOUCHIROUD, L. & AUWERX, J. 2013. Emerging roles of the corepressors NCoR1 and SMRT in homeostasis. *Genes Dev,* 27**,** 819-35.

MULHOLLAND, S., PEARSON, D. M., HAMOUDI, R. A., MALLEY, D. S., SMITH, C. M., WEAVER, J. M., JONES, D. T., KOCIALKOWSKI, S., BACKLUND, L. M., COLLINS, V. P. & ICHIMURA, K. 2012. MGMT CpG island is invariably methylated in adult astrocytic and oligodendroglial tumors with IDH1 or IDH2 mutations. *Int J Cancer,* 131**,** 1104-13.

MULLER, D. M. J., DE SWART, M. E., ARDON, H., BARKHOF, F., BELLO, L., BERGER, M. S., BOUWKNEGT, W., VAN DEN BRINK, W. A., CONTI NIBALI, M., EIJGELAAR, R. S., FURTNER, J., HAN, S. J., HERVEY-JUMPER, S., IDEMA, A. J. S., KIESEL, B., KLOET, A., MANDONNET, E., ROBE, P., ROSSI, M., SCIORTINO, T., VANDERTOP, W. P., VISSER, M., WAGEMAKERS, M., WIDHALM, G., WITTE, M. G. & DE WITT HAMER, P. C. 2021. Timing of glioblastoma surgery and patient outcomes: a multicenter cohort study. *Neurooncol Adv,* 3**,** vdab053.

MULLER, F., SCHERER, M., ASSENOV, Y., LUTSIK, P., WALTER, J., LENGAUER, T. & BOCK, C. 2019. RnBeads 2.0: comprehensive analysis of DNA methylation data. *Genome Biol,* 20**,** 55.

MUNJAPARA, V., HEUMANN, T., SCHRECK, K. C., GROSS, J. M., PEREZ-HEYDRICH, C., GUJAR, S. K., EBERHART, C. G. & HOLDHOFF, M. 2022. BRAF V600E-Mutant Glioblastoma with Extracranial Metastases Responsive to Combined BRAF and MEK Targeted Inhibition: A Case Report. *Case Rep Oncol,* 15**,** 909-917.

NADEU, F., DELGADO, J., ROYO, C., BAUMANN, T., STANKOVIC, T., PINYOL, M., JARES, P., NAVARRO, A., MARTIN-GARCIA, D., BEA, S., SALAVERRIA, I., OLDREIVE, C., AYMERICH, M., SUAREZ-CISNEROS, H., ROZMAN, M., VILLAMOR, N., COLOMER, D., LOPEZ-GUILLERMO, A., GONZALEZ, M., ALCOCEBA, M., TEROL, M. J., COLADO, E., PUENTE, X. S., LOPEZ-OTIN, C., ENJUANES, A. & CAMPO, E. 2016. Clinical impact of clonal and subclonal TP53, SF3B1, BIRC3, NOTCH1, and ATM mutations in chronic lymphocytic leukemia. *Blood,* 127**,** 2122-30.

NAGASAKA, T., GOEL, A., NOTOHARA, K., TAKAHATA, T., SASAMOTO, H., UCHIDA, T., NISHIDA, N., TANAKA, N., BOLAND, C. R. & MATSUBARA, N. 2008. Methylation pattern of the O6-methylguanine-DNA methyltransferase gene in colon during progressive colorectal tumorigenesis. *Int J Cancer,* 122**,** 2429-36.

NATHANSON, D. A., GINI, B., MOTTAHEDEH, J., VISNYEI, K., KOGA, T., GOMEZ, G., ESKIN, A., HWANG, K., WANG, J., MASUI, K., PAUCAR, A., YANG, H., OHASHI, M., ZHU, S., WYKOSKY, J., REED, R., NELSON, S. F., CLOUGHESY, T. F., JAMES, C. D., RAO, P. N., KORNBLUM, H. I., HEATH, J. R., CAVENEE, W. K., FURNARI, F. B. & MISCHEL, P. S. 2014. Targeted therapy resistance mediated by dynamic regulation of extrachromosomal mutant EGFR DNA. *Science,* 343**,** 72-6.

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. & SUVA, M. L. 2019. An Integrative Model of Cellular States, Plasticity, and Genetics for Glioblastoma. *Cell,* 178**,** 835-849 e21.

NOCH, E. K., RAMAKRISHNA, R. & MAGGE, R. 2018. Challenges in the Treatment of Glioblastoma: Multisystem Mechanisms of Therapeutic Resistance. *World Neurosurg,* 116**,** 505-517.

NOER, J. B., HORSDAL, O. K., XIANG, X., LUO, Y. & REGENBERG, B. 2022. Extrachromosomal circular DNA in cancer: history, current knowledge, and methods. *Trends Genet,* 38**,** 766-781.

NOMURA, M., SPITZER, A., JOHNSON, K. C., GAROFANO, L., NEHAR-BELAID, D., GALILI DARNELL, N., GREENWALD, A. C., BUSSEMA, L., OH, Y. T., VARN, F. S., D'ANGELO, F., GRITSCH, S., ANDERSON, K. J., MIGLIOZZI, S., GONZALEZ CASTRO, L. N., CHOWDHFURY, T., ROBINE, N., REEVES, C., PARK, J. B., LIPSA, A., HERTEL, F., GOLEBIEWSKA, A., NICLOU, S. P., NUSRAT, L., KELLET, S., DAS, S., MOON, H. E., PAEK, S. H., BIELLE, F., LAURENGE, A., DI STEFANO, A. L., MATHON, B., PICCA, A., SANSON, M., TANAKA, S., SAITO, N., ASHLEY, D. M., KEIR, S. T., LIGON, K. L., HUSE, J. T., YUNG, W. K. A., LASORELLA, A., VERHAAK, R. G. W., IAVARONE, A., SUVA, M. L. & TIROSH, I. 2025. The multilayered transcriptional architecture of glioblastoma ecosystems. *Nat Genet,* 57**,** 1155-1167.

O'ROURKE, D. M., NASRALLAH, M. P., DESAI, A., MELENHORST, J. J., MANSFIELD, K., MORRISSETTE, J. J. D., MARTINEZ-LAGE, M., BREM, S., MALONEY, E., SHEN, A., ISAACS, R., MOHAN, S., PLESA, G., LACEY, S. F., NAVENOT, J. M., ZHENG, Z., LEVINE, B. L., OKADA, H., JUNE, C. H., BROGDON, J. L. & MAUS, M. V. 2017. A single dose of peripherally infused EGFRvIII-directed CAR T cells mediates antigen loss and induces adaptive resistance in patients with recurrent glioblastoma. *Sci Transl Med,* 9.

OH, J. H., JANG, S. J., KIM, J., SOHN, I., LEE, J. Y., CHO, E. J., CHUN, S. M. & SUNG, C. O. 2020. Spontaneous mutations in the single TTN gene represent high tumor mutation burden. *NPJ Genom Med,* 5**,** 33.

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

OHGAKI, H. & KLEIHUES, P. 2013. The definition of primary and secondary glioblastoma. *Clin Cancer Res,* 19**,** 764-72.

OKKENHAUG, K., GRAUPERA, M. & VANHAESEBROECK, B. 2016. Targeting PI3K in Cancer: Impact on Tumor Cells, Their Protective Stroma, Angiogenesis, and Immunotherapy. *Cancer Discov,* 6**,** 1090-1105.

OLDRINI, B., VAQUERO-SIGUERO, N., MU, Q., KROON, P., ZHANG, Y., GALAN-GANGA, M., BAO, Z., WANG, Z., LIU, H., SA, J. K., ZHAO, J., KIM, H., RODRIGUEZ-PERALES, S., NAM, D. H., VERHAAK, R. G. W., RABADAN, R., JIANG, T., WANG, J. & SQUATRITO, M. 2020. MGMT genomic rearrangements contribute to chemotherapy resistance in gliomas. *Nat Commun,* 11**,** 3883.

OLYMPIOS, N., GILARD, V., MARGUET, F., CLATOT, F., DI FIORE, F. & FONTANILLES, M. 2021. TERT Promoter Alterations in Glioblastoma: A Systematic Review. *Cancers (Basel),* 13.

OSTROM, Q. T., BAUCHET, L., DAVIS, F. G., DELTOUR, I., FISHER, J. L., LANGER, C. E., PEKMEZCI, M., SCHWARTZBAUM, J. A., TURNER, M. C., WALSH, K. M., WRENSCH, M. R. & BARNHOLTZ-SLOAN, J. S. 2014. The epidemiology of glioma in adults: a "state of the science" review. *Neuro Oncol,* 16**,** 896-913.

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.

PASQUALETTI, F., ORLANDI, P., SIMEON, V., CANTARELLA, M., GIULIANI, D., DI DESIDERO, T., GONNELLI, A., DELISHAJ, D., LOMBARDI, G., SECHI, A., SANSON, M., ZAGONEL, V., PAIAR, F., DANESI, R., GUARINI, S. & BOCCI, G. 2018. Melanocortin Receptor-4 Gene Polymorphisms in Glioblastoma Patients Treated with Concomitant Radio-Chemotherapy. *Mol Neurobiol,* 55**,** 1396-1404.

PEI, Z., LEE, K. C., KHAN, A., ERISNOR, G. & WANG, H. Y. 2020. Pathway analysis of glutamate-mediated, calcium-related signaling in glioma progression. *Biochem Pharmacol,* 176**,** 113814.

PERRY, A. & WESSELING, P. 2016. Histologic classification of gliomas. *Handb Clin Neurol,* 134**,** 71-95.

PHILIPS, A., HENSHAW, D. L., LAMBURN, G. & O'CARROLL, M. J. 2018. Brain Tumours: Rise in Glioblastoma Multiforme Incidence in England 1995-2015 Suggests an Adverse Environmental or Lifestyle Factor. *J Environ Public Health,* 2018**,** 7910754.

PIDSLEY, R., CC, Y. W., VOLTA, M., LUNNON, K., MILL, J. & SCHALKWYK, L. C. 2013. A data-driven approach to preprocessing Illumina 450K methylation array data. *BMC Genomics,* 14**,** 293.

POPE, B. J., NGUYEN-DUMONT, T., HAMMET, F. & PARK, D. J. 2014. ROVER variant caller: read-pair overlap considerate variant-calling software applied to PCR-based massively parallel sequencing datasets. *Source Code Biol Med,* 9**,** 3.

QAZI, M. A., SALIM, S. K., BROWN, K. R., MIKOLAJEWICZ, N., SAVAGE, N., HAN, H., SUBAPANDITHA, M. K., BAKHSHINYAN, D., NIXON, A., VORA, P., DESMOND, K., CHOKSHI, C., SINGH, M., KHOO, A., MACKLIN, A., KHAN, S., TATARI, N., WINEGARDEN, N., RICHARDS, L., PUGH, T., BOCK, N., MANSOURI, A., VENUGOPAL, C., KISLINGER, T., GOYAL, S., MOFFAT, J. & SINGH, S. K. 2022. Characterization of the minimal residual disease state reveals distinct evolutionary trajectories of human glioblastoma. *Cell Rep,* 40**,** 111420.

QUACH, N., GOODMAN, M. F. & SHIBATA, D. 2004. In vitro mutation artifacts after formalin fixation and error prone translesion synthesis during PCR. *BMC Clin Pathol,* 4**,** 1.

QUINLAN, A. R. & HALL, I. M. 2010. BEDTools: a flexible suite of utilities for comparing genomic features. *Bioinformatics,* 26**,** 841-2.

RABE, M., DUMONT, S., ALVAREZ-ARENAS, A., JANATI, H., BELMONTE-BEITIA, J., CALVO, G. F., THIBAULT-CARPENTIER, C., SERY, Q., CHAUVIN, C., JOALLAND, N., BRIAND, F., BLANDIN, S., SCOTET, E., PECQUEUR, C., CLAIRAMBAULT, J., OLIVER, L., PEREZ-GARCIA, V., NADARADJANE, A., CARTRON, P. F., GRATAS, C. & VALLETTE, F. M. 2020. Identification of a transient state during the acquisition of temozolomide resistance in glioblastoma. *Cell Death Dis,* 11**,** 19.

RAMIREZ, Y. P., WEATHERBEE, J. L., WHEELHOUSE, R. T. & ROSS, A. H. 2013. Glioblastoma multiforme therapy and mechanisms of resistance. *Pharmaceuticals (Basel),* 6**,** 1475-506.

RAMÓN Y CAJAL, S., SESÉ, M., CAPDEVILA, C., AASEN, T., DE MATTOS-ARRUDA, L., DIAZ-CANO, S. J., HERNÁNDEZ-LOSA, J. & CASTELLVÍ, J. 2020. Clinical implications of intratumor heterogeneity: challenges and opportunities. *Journal of Molecular Medicine,* 98**,** 161-177.

RASCIO, F., SPADACCINO, F., ROCCHETTI, M. T., CASTELLANO, G., STALLONE, G., NETTI, G. S. & RANIERI, E. 2021. The Pathogenic Role of PI3K/AKT Pathway in Cancer Onset and Drug Resistance: An Updated Review. *Cancers,* 13.

RIGANTI, C., SALAROGLIO, I. C., PINZON-DAZA, M. L., CALDERA, V., CAMPIA, I., KOPECKA, J., MELLAI, M., ANNOVAZZI, L., COURAUD, P. O., BOSIA, A., GHIGO, D. & SCHIFFER, D. 2014. Temozolomide down-regulates P-glycoprotein in human blood-brain barrier cells by disrupting Wnt3 signaling. *Cell Mol Life Sci,* 71**,** 499-516.

RINGEL, F., PAPE, H., SABEL, M., KREX, D., BOCK, H. C., MISCH, M., WEYERBROCK, A., WESTERMAIER, T., SENFT, C., SCHUCHT, P., MEYER, B., SIMON, M. & GROUP, S. N. S. 2016. Clinical benefit from resection of recurrent glioblastomas: results of a multicenter study including 503 patients with recurrent glioblastomas undergoing surgical resection. *Neuro Oncol,* 18**,** 96-104.

RIPPAUS, N., F-BRUNS, A., TANNER, G., TAYLOR, C., DROOP, A., BRÜNING-RICHARDSON, A., CARE, M. A., WILKINSON, J., JENKINSON, M. D., BRODBELT, A., CHAKRABARTY, A., ISMAIL, A., SHORT, S. & STEAD, L. F. 2019. JARID2 facilitates transcriptional reprogramming in glioblastoma in response to standard treatment. *biorxiv*.

RIVERA, A. L., PELLOSKI, C. E., GILBERT, M. R., COLMAN, H., DE LA CRUZ, C., SULMAN, E. P., BEKELE, B. N. & ALDAPE, K. D. 2010. MGMT promoter methylation is predictive of response to radiotherapy and prognostic in the absence of adjuvant alkylating chemotherapy for glioblastoma. *Neuro Oncol,* 12**,** 116-21.

ROBBE, P., POPITSCH, N., KNIGHT, S. J. L., ANTONIOU, P., BECQ, J., HE, M., KANAPIN, A., SAMSONOVA, A., VAVOULIS, D. V., ROSS, M. T., KINGSBURY, Z., CABES, M., RAMOS, S. D. C., PAGE, S., DREAU, H., RIDOUT, K., JONES, L. J., TUFF-LACEY, A., HENDERSON, S., MASON, J., BUFFA, F. M., VERRILL, C., MALDONADO-PEREZ, D., ROXANIS, I., COLLANTES, E., BROWNING, L., DHAR, S., DAMATO, S., DAVIES, S., CAULFIELD, M., BENTLEY, D. R., TAYLOR, J. C., TURNBULL, C., SCHUH, A. & PROJECT, G. 2018. Clinical whole-genome sequencing from routine formalin-fixed, paraffin-embedded specimens: pilot study for the 100,000 Genomes Project. *Genet Med,* 20**,** 1196-1205.

ROBERT, M. F., MORIN, S., BEAULIEU, N., GAUTHIER, F., CHUTE, I. C., BARSALOU, A. & MACLEOD, A. R. 2003. DNMT1 is required to maintain CpG methylation and aberrant gene silencing in human cancer cells. *Nat Genet,* 33**,** 61-5.

ROGER BELIZAIRE, S. H. J. K., NAMRATA D. UDESHI, ALEXIS VEDDER, LEI SUN, TANYA SVINKINA, CHRISTINA HARTIGAN, MARIE MCCONKEY, VERONICA KOVALCIK, AMANUEL BIZUAYEHU, CAROLINE STANCLIFT, MONICA SCHENONE, STEVEN A. CARR, ERIC PADRON, AND BENJAMIN L. EBERT 2021. CBL mutations drive PI3K/AKT signaling via increased interaction with LYN and PIK3R1.

ROSENBAUM, D. M., RASMUSSEN, S. G. & KOBILKA, B. K. 2009. The structure and function of G-protein-coupled receptors. *Nature,* 459**,** 356-63.

ROTUNNO, M., BARAJAS, R., CLYNE, M., HOOVER, E., SIMONDS, N. I., LAM, T. K., MECHANIC, L. E., GOLDSTEIN, A. M. & GILLANDERS, E. M. 2020. A Systematic Literature Review of Whole Exome and Genome Sequencing Population Studies of Genetic Susceptibility to Cancer. *Cancer Epidemiol Biomarkers Prev,* 29**,** 1519-1534.

SAH, S., CHEN, L., HOUGHTON, J., KEMPPAINEN, J., MARKO, A. C., ZEIGLER, R. & LATHAM, G. J. 2013. Functional DNA quantification guides accurate next-generation sequencing mutation detection in formalin-fixed, paraffin-embedded tumor biopsies. *Genome Med,* 5**,** 77.

SAKTHIKUMAR, S., ROY, A., HASEEB, L., PETTERSSON, M. E., SUNDSTROM, E., MARINESCU, V. D., LINDBLAD-TOH, K. & FORSBERG-NILSSON, K. 2020. Whole-genome sequencing of glioblastoma reveals enrichment of non-coding constraint mutations in known and novel genes. *Genome Biol,* 21**,** 127.

SCIUSCIO, D., DISERENS, A. C., VAN DOMMELEN, K., MARTINET, D., JONES, G., JANZER, R. C., POLLO, C., HAMOU, M. F., KAINA, B., STUPP, R., LEVIVIER, M. & HEGI, M. E. 2011. Extent and patterns of MGMT promoter methylation in glioblastoma- and respective glioblastoma-derived spheres. *Clin Cancer Res,* 17**,** 255-66.

SEAN DAVIS, P. D., SVEN BILKE, TIM TRICHE JR., MOIZ BOOTWALLA 2025. methylumi: Handle Illumina methylation data. *Bioconductor*.

SESE, B., ENSENYAT-MENDEZ, M., INIGUEZ, S., LLINAS-ARIAS, P. & MARZESE, D. M. 2021. Chromatin insulation dynamics in glioblastoma: challenges and future perspectives of precision oncology. *Clin Epigenetics,* 13**,** 150.

SEVIM, H., PARKINSON, J. F. & MCDONALD, K. L. 2011. Etoposide-mediated glioblastoma cell death: dependent or independent on the expression of its target, topoisomerase II alpha? *J Cancer Res Clin Oncol,* 137**,** 1705-12.

SHEN, R. & SESHAN, V. E. 2016. FACETS: allele-specific copy number and clonal heterogeneity analysis tool for high-throughput DNA sequencing. *Nucleic Acids Res,* 44**,** e131.

SHOMALI, W. & GOTLIB, J. 2018. The new tool "KIT" in advanced systemic mastocytosis. *Hematology Am Soc Hematol Educ Program,* 2018**,** 127-136.

SINGH, D. K., KOLLIPARA, R. K., VEMIREDDY, V., YANG, X. L., SUN, Y., REGMI, N., KLINGLER, S., HATANPAA, K. J., RAISANEN, J., CHO, S. K., SIRASANAGANDLA, S., NANNEPAGA, S., PICCIRILLO, S., MASHIMO, T., WANG, S., HUMPHRIES, C. G., MICKEY, B., MAHER, E. A., ZHENG, H., KIM, R. S., KITTLER, R. & BACHOO, R. M. 2017. Oncogenes Activate an Autonomous Transcriptional Regulatory Circuit That Drives Glioblastoma. *Cell Rep,* 18**,** 961-976.

SINGH, N., MINER, A., HENNIS, L. & MITTAL, S. 2021. Mechanisms of temozolomide resistance in glioblastoma - a comprehensive review. *Cancer Drug Resist,* 4**,** 17-43.

SMITH, M. L., BAGGERLY, K. A., BENGTSSON, H., RITCHIE, M. E. & HANSEN, K. D. 2013. illuminaio: An open source IDAT parsing tool for Illumina microarrays. *F1000Res,* 2**,** 264.

SOTTORIVA, A., SPITERI, I., PICCIRILLO, S. G., TOULOUMIS, A., COLLINS, V. P., MARIONI, J. C., CURTIS, C., WATTS, C. & TAVARE, S. 2013. Intratumor heterogeneity in human glioblastoma reflects cancer evolutionary dynamics. *Proc Natl Acad Sci U S A,* 110**,** 4009-14.

SPITZER, A., JOHNSON, K. C., NOMURA, M., GAROFANO, L., NEHAR-BELAID, D., DARNELL, N. G., GREENWALD, A. C., BUSSEMA, L., OH, Y. T., VARN, F. S., D'ANGELO, F., GRITSCH, S., ANDERSON, K. J., MIGLIOZZI, S., GONZALEZ CASTRO, L. N., CHOWDHURY, T., ROBINE, N., REEVES, C., PARK, J. B., LIPSA, A., HERTEL, F., GOLEBIEWSKA, A., NICLOU, S. P., NUSRAT, L., KELLET, S., DAS, S., MOON, H. E., PAEK, S. H., BIELLE, F., LAURENGE, A., DI STEFANO, A. L., MATHON, B., PICCA, A., SANSON, M., TANAKA, S., SAITO, N., ASHLEY, D. M., KEIR, S. T., LIGON, K. L., HUSE, J. T., YUNG, W. K. A., LASORELLA, A., IAVARONE, A., VERHAAK, R. G. W., TIROSH, I. & SUVA, M. L. 2025. Deciphering the longitudinal trajectories of glioblastoma ecosystems by integrative single-cell genomics. *Nat Genet,* 57**,** 1168-1178.

STEIERT, T. A., PARRA, G., GUT, M., ARNOLD, N., TROTTA, J. R., TONDA, R., MOUSSY, A., GERBER, Z., ABUJA, P. M., ZATLOUKAL, K., ROCKEN, C., FOLSERAAS, T., GRIMSRUD, M. M., VOGEL, A., GOEPPERT, B., ROESSLER, S., HINZ, S., SCHAFMAYER, C., ROSENSTIEL, P., DELEUZE, J. F., GUT, I. G., FRANKE, A. & FORSTER, M. 2023. A critical spotlight on the paradigms of FFPE-DNA sequencing. *Nucleic Acids Res,* 51**,** 7143-7162.

STEPHAN, G., RAVN-BOESS, N. & PLACANTONAKIS, D. G. 2021. Adhesion G protein-coupled receptors in glioblastoma. *Neurooncol Adv,* 3**,** vdab046.

STROBEL, H., BAISCH, T., FITZEL, R., SCHILBERG, K., SIEGELIN, M. D., KARPEL-MASSLER, G., DEBATIN, K. M. & WESTHOFF, M. A. 2019. Temozolomide and Other Alkylating Agents in Glioblastoma Therapy. *Biomedicines,* 7.

STUPP, R., MASON, W. P., VAN DEN BENT, M. J., WELLER, M., FISHER, B., TAPHOORN, M. J., BELANGER, K., BRANDES, A. A., MAROSI, C., BOGDAHN, U., CURSCHMANN, J., JANZER, R. C., LUDWIN, S. K., GORLIA, T., ALLGEIER, A., LACOMBE, D., CAIRNCROSS, J. G., EISENHAUER, E., MIRIMANOFF, R. O., EUROPEAN ORGANISATION FOR, R., TREATMENT OF

CANCER BRAIN, T., RADIOTHERAPY, G. & NATIONAL CANCER INSTITUTE OF CANADA CLINICAL TRIALS, G. 2005. Radiotherapy plus concomitant and adjuvant temozolomide for glioblastoma. *N Engl J Med,* 352**,** 987-96.

SUCHORSKA, B., WELLER, M., TABATABAI, G., SENFT, C., HAU, P., SABEL, M. C., HERRLINGER, U., KETTER, R., SCHLEGEL, U., MAROSI, C., REIFENBERGER, G., WICK, W., TONN, J. C. & WIRSCHING, H. G. 2016. Complete resection of contrast-enhancing tumor volume is associated with improved survival in recurrent glioblastoma-results from the DIRECTOR trial. *Neuro Oncol,* 18**,** 549-56.

SUN, D., GUO, Y., TANG, P., LI, H. & CHEN, L. 2023. Arf6 as a therapeutic target: Structure, mechanism, and inhibitors. *Acta Pharm Sin B,* 13**,** 4089-4104.

TAMIMI AF, J. M. 2017. Epidemiology and Outcome of Glioblastoma. *Codon Publications; 2017 Sep 27. Chapter 8*.

TANNER, G., BARROW, R., AJAIB, S., AL-JABRI, M., AHMED, N., POLLOCK, S., FINETTI, M., RIPPAUS, N., BRUNS, A. F., SYED, K., POULTER, J. A., MATTHEWS, L., HUGHES, T., WILSON, E., JOHNSON, C., VARN, F. S., BRUNING-RICHARDSON, A., HOGG, C., DROOP, A., GUSNANTO, A., CARE, M. A., CUTILLO, L., WESTHEAD, D. R., SHORT, S. C., JENKINSON, M. D., BRODBELT, A., CHAKRABARTY, A., ISMAIL, A., VERHAAK, R. G. W. & STEAD, L. F. 2024. IDHwt glioblastomas can be stratified by their transcriptional response to standard treatment, with implications for targeted therapy. *Genome Biol,* 25**,** 45.

TANNER, G., WESTHEAD, D. R., DROOP, A. & STEAD, L. F. 2021. Benchmarking pipelines for subclonal deconvolution of bulk tumour sequencing data. *Nat Commun,* 12**,** 6396.

TAYLOR, K. R., BARRON, T., HUI, A., SPITZER, A., YALCIN, B., IVEC, A. E., GERAGHTY, A. C., HARTMANN, G. G., ARZT, M., GILLESPIE, S. M., KIM, Y. S., MALEKI JAHAN, S., ZHANG, H., SHAMARDANI, K., SU, M., NI, L., DU, P. P., WOO, P. J., SILVA-TORRES, A., VENKATESH, H. S., MANCUSI, R., PONNUSWAMI, A., MULINYAWE, S., KEOUGH, M. B., CHAU, I., AZIZ-BOSE, R., TIROSH, I., SUVA, M. L. & MONJE, M. 2023. Glioma synapses recruit mechanisms of adaptive plasticity. *Nature*.

TEER, J. K. & MULLIKIN, J. C. 2010. Exome sequencing: the sweet spot before whole genomes. *Hum Mol Genet,* 19**,** R145-51.

TESCHENDORFF, A. E., MARABITA, F., LECHNER, M., BARTLETT, T., TEGNER, J., GOMEZ-CABRERO, D. & BECK, S. 2013. A beta-mixture quantile normalization method for correcting probe design bias in Illumina Infinium 450 k DNA methylation data. *Bioinformatics,* 29**,** 189-196.

THOMAS, A., TANAKA, M., TREPEL, J., REINHOLD, W. C., RAJAPAKSE, V. N. & POMMIER, Y. 2017. Temozolomide in the Era of Precision Medicine. *Cancer Res,* 77**,** 823-826.

TSAI, H. F., C, I. J. & SHEN, A. Q. 2020. Voltage-gated ion channels mediate the electrotaxis of glioblastoma cells in a hybrid PMMA/PDMS microdevice. *APL Bioeng,* 4**,** 036102.

UNO, M., OBA-SHINJO, S. M., CAMARGO, A. A., MOURA, R. P., AGUIAR, P. H., CABRERA, H. N., BEGNAMI, M., ROSEMBERG, S., TEIXEIRA, M. J. & MARIE, S. K. 2011. Correlation of MGMT promoter methylation status with gene and protein expression levels in glioblastoma. *Clinics (Sao Paulo),* 66**,** 1747-55.

VAN DEN BENT, M. J., ERDEM-ERASLAN, L., IDBAIH, A., DE ROOI, J., EILERS, P. H., SPLIET, W. G., DEN DUNNEN, W. F., TIJSSEN, C., WESSELING, P., SILLEVIS SMITT, P. A., KROS, J. M., GORLIA, T. & FRENCH, P. J. 2013. MGMT-STP27 methylation status as predictive marker for response to PCV in anaplastic Oligodendrogliomas and Oligoastrocytomas. A report from EORTC study 26951. *Clin Cancer Res,* 19**,** 5513-22.

VERHAAK, R. G. W., BAFNA, V. & MISCHEL, P. S. 2019. Extrachromosomal oncogene amplification in tumour pathogenesis and evolution. *Nat Rev Cancer,* 19**,** 283-288.

WADDINGTON, C. H. 2012. The epigenotype. 1942. *Int J Epidemiol,* 41**,** 10-3.

WADSWORTH, P. & LEE, W. L. 2013. Microtubule motors: doin' it without dynactin. *Curr Biol,* 23**,** R563-5.

WANG, H., FENG, W., LU, Y., LI, H., XIANG, W., CHEN, Z., HE, M., ZHAO, L., SUN, X., LEI, B., QI, S. & LIU, Y. 2016a. Expression of dynein, cytoplasmic 2, heavy chain 1 (DHC2) associated with glioblastoma cell resistance to temozolomide. *Sci Rep,* 6**,** 28948.

WANG, J., CAZZATO, E., LADEWIG, E., FRATTINI, V., ROSENBLOOM, D. I., 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. 2016b. Clonal evolution of glioblastoma under therapy. *Nat Genet,* 48**,** 768-76.

WANG, Q., HU, B., HU, X., KIM, H., SQUATRITO, M., SCARPACE, L., DECARVALHO, A. C., LYU, S., LI, P., LI, Y., BARTHEL, F., CHO, H. J., LIN, Y. H., SATANI, N., MARTINEZ-LEDESMA, E., ZHENG, S., CHANG, E., SAUVE, C. G., OLAR, A., LAN, Z. D., FINOCCHIARO, G., PHILLIPS, J. J., BERGER, M. S., GABRUSIEWICZ, K. R., WANG, G., ESKILSSON, E., HU, J., MIKKELSEN, T., DEPINHO, R. A., MULLER, F., HEIMBERGER, A. B., SULMAN, E. P., NAM, D. H. & VERHAAK, R. G. W. 2017. Tumor Evolution of Glioma-Intrinsic Gene Expression Subtypes Associates with Immunological Changes in the Microenvironment. *Cancer Cell,* 32**,** 42-56 e6.

WANG, Y., ZHANG, J., LI, W., JIANG, T., QI, S., CHEN, Z., KANG, J., HUO, L., WANG, Y., ZHUGE, Q., GAO, G., WU, Y., FENG, H., ZHAO, G., YANG, X., ZHAO, H., WANG, Y., YANG, H., KANG, D., SU, J., LI, L., JIANG, C., LI, G., QIU, Y., WANG, W., WANG, H., XU, Z., ZHANG, L. & WANG, R. 2021. Guideline conformity to the Stupp regimen in patients with newly diagnosed glioblastoma multiforme in China. *Future Oncol,* 17**,** 4571-4582.

WELLER, M. & LE RHUN, E. 2020. How did lomustine become standard of care in recurrent glioblastoma? *Cancer Treat Rev,* 87**,** 102029.

WELLER, M., STUPP, R., REIFENBERGER, G., BRANDES, A. A., VAN DEN BENT, M. J., WICK, W. & HEGI, M. E. 2010. MGMT promoter methylation in malignant gliomas: ready for personalized medicine? *Nat Rev Neurol,* 6**,** 39-51.

WEMMERT, S., KETTER, R., RAHNENFUHRER, J., BEERENWINKEL, N., STROWITZKI, M., FEIDEN, W., HARTMANN, C., LENGAUER, T., STOCKHAMMER, F., ZANG, K. D., MEESE, E., STEUDEL, W. I., VON DEIMLING, A. & URBSCHAT, S. 2005. Patients with high-grade gliomas harboring deletions of chromosomes 9p and 10q benefit from temozolomide treatment. *Neoplasia,* 7**,** 883-93.

WEN, J., CHEN, W., ZHU, Y. & ZHANG, P. 2021. Clinical features associated with the efficacy of chemotherapy in patients with glioblastoma (GBM): a surveillance, epidemiology, and end results (SEER) analysis. *BMC Cancer,* 21**,** 81.

WEN, P. Y., WELLER, M., LEE, E. Q., ALEXANDER, B. M., BARNHOLTZ-SLOAN, J. S., BARTHEL, F. P., BATCHELOR, T. T., BINDRA, R. S., CHANG, S. M., CHIOCCA, E. A., CLOUGHESY, T. F., DEGROOT, J. F., GALANIS, E., GILBERT, M. R., HEGI, M. E., HORBINSKI, C., HUANG, R. Y., LASSMAN, A. B., LE RHUN, E., LIM, M., MEHTA, M. P., MELLINGHOFF, I. K., MINNITI, G., NATHANSON, D., PLATTEN, M., PREUSSER, M., ROTH, P., SANSON, M., SCHIFF, D., SHORT, S. C., TAPHOORN, M. J. B., TONN, J. C., TSANG, J., VERHAAK, R. G. W., VON DEIMLING, A., WICK, W., ZADEH, G., REARDON, D. A., ALDAPE, K. D. & VAN DEN BENT, M. J. 2020. Glioblastoma in adults: a Society for Neuro-Oncology (SNO) and European Society of Neuro-Oncology (EANO) consensus review on current management and future directions. *Neuro Oncol,* 22**,** 1073-1113.

WILLIAMS, C., PONTÉN, F., MOBERG, C., SÖDERKVIST, P., UHLÉN, M., PONTÉN, J., SITBON, G. & LUNDEBERG, J. 1999. A High Frequency of Sequence Alterations Is Due to Formalin Fixation of Archival Specimens. *The American Journal of Pathology,* 155**,** 1467-1471.

WILSON, A. S., POWER, B. E. & MOLLOY, P. L. 2007. DNA hypomethylation and human diseases. *Biochim Biophys Acta,* 1775**,** 138-62.

WONG, S. Q., LI, J., TAN, A. Y., VEDURURU, R., PANG, J. M., DO, H., ELLUL, J., DOIG, K., BELL, A., MACARTHUR, G. A., FOX, S. B., THOMAS, D. M., FELLOWES, A., PARISOT, J. P., DOBROVIC, A. & COHORT, C. 2014. Sequence artefacts in a prospective series of formalin-fixed tumours tested for mutations in hotspot regions by massively parallel sequencing. *BMC Med Genomics,* 7**,** 23.

WOODROFFE, R. W., ZANATY, M., SONI, N., MOTT, S. L., HELLAND, L. C., PASHA, A., MALEY, J., DHUNGANA, N., JONES, K. A., MONGA, V. & GREENLEE, J. D. W. 2020. Survival after reoperation for recurrent glioblastoma. *J Clin Neurosci,* 73**,** 118-124.

WU, Q., BERGLUND, A. E. & ETAME, A. B. 2021. The Impact of Epigenetic Modifications on Adaptive Resistance Evolution in Glioblastoma. *Int J Mol Sci,* 22.

XU, L., CHEN, Y., DUTRA-CLARKE, M., MAYAKONDA, A., HAZAWA, M., SAVINOFF, S. E., DOAN, N., SAID, J. W., YONG, W. H., WATKINS, A., YANG, H., DING, L. W., JIANG, Y. Y., TYNER, J. W., CHING, J., KOVALIK, J. P., MADAN, V., CHAN, S. L., MUSCHEN, M., BREUNIG, J. J., LIN, D. C. & KOEFFLER, H. P. 2017. BCL6 promotes glioma and serves as a therapeutic target. *Proc Natl Acad Sci U S A,* 114**,** 3981-3986.

YAMAUCHI, Y., MIURA, Y. & KANAHO, Y. 2017. Machineries regulating the activity of the small GTPase Arf6 in cancer cells are potential targets for developing innovative anti-cancer drugs. *Adv Biol Regul,* 63**,** 115-121.

YANG, M., ZHANG, S., JIANG, R., CHEN, S. & HUANG, M. 2023. Circlehunter: a tool to identify extrachromosomal circular DNA from ATAC-Seq data. *Oncogenesis,* 12**,** 28.

YEHIA, L., NGEOW, J. & ENG, C. 2019. PTEN-opathies: from biological insights to evidence-based precision medicine. *J Clin Invest,* 129**,** 452-464.

ZALCMAN, N., CANELLO, T., OVADIA, H., CHARBIT, H., ZELIKOVITCH, B., MORDECHAI, A., FELLIG, Y., RABANI, S., SHAHAR, T., LOSSOS, A. & LAVON, I. 2018. Androgen receptor: a potential therapeutic target for glioblastoma. *Oncotarget,* 9**,** 19980-19993.

ZHANG, J. J. Y., LEE, K. S., VOISIN, M. R., HERVEY-JUMPER, S. L., BERGER, M. S. & ZADEH, G. 2020. Awake craniotomy for resection of supratentorial glioblastoma: a systematic review and meta-analysis. *Neurooncol Adv,* 2**,** vdaa111.

ZHANG, Y., DUBE, C., GIBERT, M., JR., 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 (Basel),* 10.

ZHAO, Y., YU, L., ZHANG, S., SU, X. & ZHOU, X. 2022. Extrachromosomal circular DNA: Current status and future prospects. *Elife,* 11.

ZHOU, L., GUO, X., CHEN, M., FU, S., ZHOU, J., REN, G., YANG, Z. & FAN, W. 2013. Inhibition of delta-opioid receptors induces brain glioma cell apoptosis through the mitochondrial and protein kinase C pathways. *Oncol Lett,* 6**,** 1351-1357.

ZHUANG, B. C., JUDE, M. S., KONWAR, C., YUSUPOV, N., RYAN, C. P., ENGELBRECHT, H. R., WHITEHEAD, J., HALBERSTAM, A. A., MACISAAC, J. L., DEVER, K., TRAN, T. K., KORINEK, K., ZIMMER, Z., LEE, N. R., MCDADE, T. W., KUZAWA, C. W., HUFFMAN, K. M., BELSKY, D. W., BINDER, E. B., CZAMARA, D., KORTHAUER, K. & KOBOR, M. S. 2025. Accounting for differences between Infinium MethylationEPIC v2 and v1 in DNA methylation-based tools. *Life Sci Alliance,* 8.

CHAPTER 5 – DISCUSSION

Glioblastoma (GBM) stands as an exceptionally aggressive and challenging malignancy, primarily due to its difficult brain location for biopsy sampling and its rapid recurrence despite current therapies. This necessitates a profound understanding of its molecular drivers and evolutionary dynamics. While the field rapidly leveraged cutting-edge technologies from bulk genomic and transcriptomic studies to single-cell and spatial omics to try and unearth the pathobiology underlying gliomagenesis, comprehensive longitudinal analyses of matched primary and recurrent tumours remain scarce, despite potentially yielding the best chance to decipher mechanisms behind GBM's treatment resistance and short patient survival. My PhD research aimed to bridge this gap by developing robust bioinformatic pipelines and applying them to longitudinal GBM cohorts, thereby illuminating critical molecular alterations, identifying treatment-driven evolutionary paths, and proposing therapeutic and diagnostic strategies.

A fundamental challenge in molecular oncology, particularly with clinically derived FFPE samples, is ensuring sequencing data fidelity. Recognising that high-quality biological insights stem from high-quality data, a core component of my thesis involved meticulously optimising whole-exome (WES) and whole-genome sequencing (WGS) analysis pipelines. Initially, technical artefacts like FFPE-induced C>T transitions and errors from overlapping paired-end reads inflated variant calls and compromised variant allele frequency (VAF) estimates crucial for clonal tracking. My work successfully addressed these issues by utilising ClipBam to correct for overlapping reads and other methods, which significantly improved data reliability. This brought WES variant counts in line with established GBM cohorts like TCGA, enabling accurate tumour mutational burden (TMB) estimation and robust identification of key GBM mutations (e.g., *EGFR, PTEN, TP53, RB1*). While WGS data remained more challenging due to FFPE constraints and lower depth, it proved invaluable for reliable copy number aberration (CNA) detection, identifying characteristic GBM gains on chromosome 7 and losses on chromosomes 10 and 9. This optimised pipeline, a significant success of my work, laid the essential groundwork for confident downstream analyses, especially for resolving clonal dynamics and tracking genetic alterations over time. To further elevate data quality, future efforts should focus on higher sequencing depths for WGS and adopting shorter paired-end read lengths (e.g., 75 bp) to minimise problematic overlaps inherent to fragmented FFPE DNA. Establishing a panel of normals (PON) from non-malignant brain tissue would also significantly reduce false positives. Moreover, the modular design of the pipeline, adaptable to platforms like Nextflow, can serve as a blueprint for developing robust workflows for other omics data, such as RNA sequencing and methylation analysis.

With this validated pipeline established, my research shifted to deciphering the molecular dynamics of GBM progression, identifying both drivers of treatment resistance and vulnerabilities associated with therapeutic sensitivity by analysing longitudinal changes in variant profiles. A central and specific discovery of this work was the identification of ERBB signalling pathways as a critical axis undergoing differential selection pressure in GBM evolution. My focused investigation revealed that while *ERBB2* was selected for during tumour progression, *ERBB4* was selected against. This differential selection, involving specific variants in *EGFR*, *PTEN*, and *PIK3CA*, intricately modulates ERBB receptor dimerisation, downstream signalling, and cellular fate. Notably, even in the absence of precise cancer cell fraction (CCF) estimates and after carefully excluding variants located in copy number variable regions, the differential selection within the ERBB signalling pathway remained a statistically significant and robust finding. This highlights its intrinsic biological importance and identifies it as a prime candidate for future targeted therapeutic and mechanistic studies. The profound finding of differential selection within the ERBB signalling pathway warrants experimental validation in laboratory settings (e.g., *in vitro* and *in vivo* models) to confirm the functional impact of these selection pressures and specific variants. Beyond ERBB, a key future goal is to enhance pathway analysis specificity. Although we used public gene sets in this thesis, our planned effort to perform gene set enrichment analysis with custom, GBM-specific gene sets developed by our GliomaGenomics group via PathScore was unfortunately halted. This remains an important recommendation for future work, pending feedback from the PathScore developer on enabling GBM-specific gene sets.

Furthermore, investigating the role of extrachromosomal DNA (ecDNA) is vital to understanding its contribution to rapid tumour evolution and the emergence of drug resistance. A deeper understanding of ecDNA's mechanisms could pave the way for novel therapeutic strategies specifically targeting these unstable genetic elements, thereby overcoming a major hurdle in cancer treatment.

My thesis also ventured into the epigenetic landscape of GBM, specifically examining the methylation status of the MGMT promoter in paired primary and recurrent glioblastoma samples. While MGMT methylation is a well-established prognostic and predictive biomarker, my assessment, using the MGMT-STP27 classifier, primarily served to compare findings to existing literature and investigate its association with treatment-driven tumour progression behaviours. My analysis revealed that MGMT methylation status was largely stable between primary and recurrent tumours across both discovery and validation cohorts (approximately 86-89% consistent). However, a small but clinically relevant subset of patients exhibited methylation switching, predominantly a loss of methylation at

recurrence, implying an acquired resistance mechanism. Crucially, my work specifically checked the association between MGMT status changes and the "Up-responder" or "Down-responder" subtypes. Interestingly, these changes occurred at similarly low frequencies in both groups, and chi-square testing revealed no significant differences in switching rates between them, suggesting that MGMT changes are not uniquely enriched in either subtype but rather reflect broader tumour-intrinsic or treatment-related factors.

A considerable challenge in this chapter was the overall small number of samples, especially when trying to do more detailed analyses, such as identifying reliable differentially methylated regions (DMRs) at the probe or regional level. This issue became even more noticeable when I divided the samples into "Up-responder" and "Down-responder" groups, as it further reduced the number of cases in each subgroup. This limitation affected the confidence and power of the results beyond the MGMT analysis. While this is a well-known problem in GBM research, it is difficult to overcome due to limitations in what tissue can be collected and preserved during surgery. Clinical tissue banks are extremely valuable, but diagnostic priorities and variability in sample quality constrain them.

To make the most of available samples and improve statistical power, one practical direction is to develop a methylation-based stratification method. The current classification method of the Up and Down-responder subtypes used in this thesis is derived from RNA-seq data. However, not all archival or FFPE-derived samples have matched RNA-seq available. Creating a methylation-based classifier would allow researchers to assign subtype labels to samples that cannot be classified through transcriptomics, enabling the inclusion of larger methylation cohorts in downstream analyses. This approach would also help harmonise datasets from different sources and increase the power to detect robust epigenetic signatures linked to treatment response.

In addition, moving beyond promoter regions and investigating enhancer methylation and transcription factor motif enrichment will be key to uncovering new regulatory mechanisms involved in tumour progression and resistance. These enhancer-based changes are increasingly recognised as important in GBM biology, and focusing on them could reveal patterns not captured through conventional promoter analysis.

Another important direction is the use of experimental model systems designed to reflect the molecular features seen in patient tumours. The findings in this thesis, such as the selective pressure on ERBB genes and the behaviour of MGMT methylation, offer a strong foundation for guiding model development. Once validated, these models would allow researchers to study treatment resistance

and progression in a controlled environment, as well as explore how different molecular profiles affect response to therapy. Importantly, they also provide the opportunity to perform multi-omics analyses — combining genome, methylome, and transcriptome data from the same sample. Although integrating these layers is technically challenging, the tools are improving, and this kind of analysis holds great promise for refining patient stratification and identifying therapeutic targets.

Ultimately, my PhD research has provided a comprehensive look into the molecular intricacies of glioblastoma, from fundamental data quality challenges to complex pathway dynamics and epigenetic shifts. By developing robust analytical tools and applying them to precious longitudinal GBM samples, I have identified genetic and epigenetic alterations that may influence therapeutic response. The findings not only clarify mechanisms of resistance and sensitivity but also propose biological targets and strategies for patient stratification. The journey ahead in GBM research requires a concerted, multi-faceted approach. Building on the foundation laid by my thesis, future efforts should prioritise integrating multi-modal omics data with functional validation, leveraging cutting-edge single-cell and spatial technologies to dissect tumour heterogeneity, and expanding longitudinal cohort studies to capture the full spectrum of tumour evolution. Moreover, the application of AI and machine learning will be instrumental in synthesising these vast datasets for predictive modelling and biomarker discovery. Ultimately, by continually pushing the boundaries of genomic and epigenomic analysis, coupled with rigorous experimental validation, we, researchers, can pave the way for a new era of precision oncology for glioblastoma patients, offering more effective, tailored treatments and improved survival outcomes.