Inaugural dissertation

for

obtaining the doctoral degree

of the

Combined Faculty of Mathematics, Engineering and Natural Sciences

of the

Ruprecht - Karls - University

Heidelberg

Presented by

Sarah Benedetto, M.Sc.

born in Biella, Italy

Oral examination: 04 November 2024

Evolutionary dynamics of oligodendrogliomas

Referees: Prof. Dr. Thomas Höfer

Dr. Pei-Chi Wei

A Jaja

Abstract

Genetic changes play crucial roles in shaping the development and progression of tumors, affecting their growth, response to treatment and overall prognosis. This thesis aims at reconstructing the genetic history of oligodendroglioma, a slowly growing yet incurable brain tumor. The dynamics of its ongoing genetic evolution and the resulting progression remain poorly understood. Therefore, in this work I develop a mathematical approach to estimate malignant growth and infer evolutionary trajectories of oligodendrogliomas from whole-genome sequencing data.

I analyse deep whole-genome sequencing data from six primary-relapse oligodendroglioma pairs. This type of data enables quantitative analysis of subclonal evolution during the early stages of the disease and its subsequent progression. By applying a population genetics model to these samples, I identify a common early tumorigenesis pathway characterized by mutations in the *IDH1/2* genes and the loss of chromosomes 1p and 19q, which are defining properties of these tumor entities. Dating the founding cell (most recent common ancestor) of individual patients suggests that oligodendrogliomas originate in childhood and evolve over decades with an increasing growth rate. These dynamics are consistent with rapid oligodendrocyte production in childhood and provides an explanation for the incidence peak of oligodendrogliomas in mid-life. Accelerated growth is linked to the selection of subclones, with or without known driver mutations. Specifically, I find that *TERT* promoter mutations are associated with strong selection, whereas weak subclonal selection is linked to mutations in *CIC*, *NOTCH1*, *ZBTB20* or *ATM*, or could not be explained by a known driver. For individual tumors, I predict the time to recurrence based on the inferred growth rate of the primary tumor. Overall, I show how genome sequencing can provide valuable insights into the dynamics of tumor evolution, allowing for personalized predictions of recurrence based on the genomic analysis of the primary tumor.

In the second part of the thesis, I explore a potential non-genetic factor influencing subclonal selection of oligodendrogliomas: DNA methylation. Specifically, differentially methylated sites between recurrent and primary tumor samples are identified to better understand oligodendrogliomas progression. However, I could not detect evidence for epigenetic subclonal driver events.

Zusammenfassung

Genetische Veränderungen spielen eine entscheidende Rolle bei der Entwicklung und dem Fortschreiten von Tumoren. Sie beeinflussen die Wachstumsgeschwindigkeit des Tumors, das Ansprechen auf Behandlungen und die allgemeine Prognose. Ziel dieser Dissertation ist, die genetische Entwicklung des Oligodendroglioms zu rekonstruieren. Oligondendrogliome sind langsam wachsende, aber unheilbare Hirntumore. Die Dynamik der fortlaufenden genetischen Evolution und des daraus resultierenden Wachstum des Tumors ist noch wenig verstanden. Daher entwickle ich in dieser Arbeit einen mathematischen Ansatz, um Wachstumsparameter zu schätzen und evolutionäre Trajektorien von Oligodendrogliomen anhand von Genomsequenzierungsdaten abzuleiten.
Ich analysiere Genomsequenzierungsdaten von sechs Paaren primärer und rezidivierender Oligodendrogliome. Diese Daten ermöglichen eine quantitative Analyse der subklonalen Evolution in den frühen Stadien der Krankheit und des weiteren Fortschreitens. Anhand eines populationsgenetischen Modells identifiziere ich in allen Proben eine frühe Tumorentstehung, die durch Mutationen in den *IDH1/2*-Genen und den Verlust der Chromosomen 1p und 19q gekennzeichnet ist. Diese intialen Mutationen definieren das Oligodendrogliom. Die Ursprungszelle des Oligodendrioglioms (letzter gemeinsamer Vorfahre aller Zellen des Tumors) wird in allen Patienten auf das Kindesalter datiert. Die Tumore entwickeln sich dann über Jahrzehnte mit einer zunehmenden Wachstumsrate. Diese Dynamik ist konsistent mit einer schnellen Produktion von Oligodendrozyten im Kindesalter und liefert eine Erklärung für die gehäufte Diagnose von Oligodendrogliomen im mittleren Lebensalter. Das Vorkommen selektierter Subklone, mit oder ohne bekannten Treibermutationen, beschleunigt das Tumorwachstum. Insbesondere finde ich, dass Mutationen im *TERT*-Promoter mit einer starken Selektion verbunden sind, während schwache subklonale Selektion mit Mutationen in den *CIC*, *NOTCH1*, *ZBTB20* oder *ATM* Genen verbunden ist, oder nicht durch eine bekannte Treibermutation erklärt werden konnte. Mithilfe des mathematischen Modells schätze ich zudem den voraussichtlichen Zeitpunkt des Wiederkehrens des Tumors nach primärer Behandlung (bei gleichbleibender Wachstumsrate). Insgesamt zeige ich, dass die Genomsequenzierung wertvolle Einblicke in die Dynamik der Tumorevolution liefern kann. Dies ermöglicht personalisierte Vorhersagen des Wiederauftretens des Tumors basierend auf der genomischen Analyse des Primärtumors.

Im zweiten Teil der Dissertation untersuche ich einen potenziellen nicht-genetischen Faktor, der die subklonale Selektion von Oligodendrogliomen beeinflusst: die DNA-Methylierung. Insbesondere werden unterschiedlich methylierte Stellen zwischen rezidivierenden und primären Tumorproben identifiziert, um das Fortschreiten von Oligodendrogliomen besser zu verstehen. Allerdings konnte ich keine Hinweise auf epigenetische Ereignisse finden, die das subklonale Tumorwachstum beschleunigen.

Contents

Chapter 1

Introduction

1.1 Outline of this thesis

This thesis investigates the evolutionary dynamics of six pairs of primary and recurrent oligodendrogliomas by analyzing genetic factors and exploring a potential epigenetic influence on the growth of these tumors.
The first two sections of Chapter 1 provide general background on cancer initiation mechanisms. Specifically, 1.2 offers a general background on mutation accumulation mechanisms and DNA repair, introducing cancer as a mutation-driven disease, whereas 1.3 explains the importance of studying epigenetic changes, despite their relatively understudied nature and incomplete understanding of their contribution to tumorigenesis. Lastly, in 1.4 I provide an introduction to the clinical and biological aspects of oligodendrogliomas.
Chapter 2 focuses on the genetic analysis. I firstly discuss a new computational method to infer tumor heterogeneity from deep whole-genome sequencing data specifically tailored for oligodendrogliomas. Next, I examine clonal evolution in oligodendroglioma, inferring the evolutionary trajectories using the population dynamics model previously described. Furthermore, I uncover tumor growth characteristics, time the tumor's origin and estimate the selective advantage of positively selected subclones, reconstructing the evolutionary history of tumor cell populations from genetic alterations. The results are contextualized within current research in the concluding section.
Chapter 3 focuses on the epigenetic analysis. Specifically, I analyze DNA methylation data from the same cohort of oligodendrogliomas, identifying specific methylome changes associated with cancer progression. I investigate whether any epigenetic drivers is able to explain, the positively selected subclones previously identified in Chapter 2. I could not identify any upmethylated pattern that could drive such faster growth.
In the final Chapter 4, results from both genetic and epigenetic analyses are integrated and put in a broader context to provide a comprehensive understanding about oligoden-

drogliomas' evolution. Moreover, the potential of whole-genome sequencing data to probe evolutionary dynamics in tissues is discussed more generally.
This thesis examines the functional and dynamic aspects of mutations driving the transition from healthy to diseased brain, utilizing deep whole-genome sequencing data and mathematical models to outline cancerous tissue growth dynamics.

1.2 Somatic mutations in cancer cells

Many cancers arise as a result of errors that have occurred in the DNA sequence that transform normal cells into cancerous ones. Mutations arising from these errors can be inherited, build up over time or caused by external factors damaging the genome, such as cigarette smoke, alcohol or ultraviolet radiation from the sun. Over the last decades, advancements in genome sequencing techniques have greatly improved the understanding of cancer genetics.
In this thesis, I will show how a detailed analysis of cancer mutations can reveal the evolutionary dynamics of individual tumors. To set the stage, I will start with a brief discussion on how somatic mutations develop over a lifetime, including a short overview of key historical events, technological advancements and concepts defining cancer evolution. With this foundation, I will examine the role of somatic mutations in the progression from normal to cancer cells. Finally, I will highlight the importance of timing somatic mutational events for understanding tumor evolution.

1.2.1 About somatic mutations and cancer

Brief history

Before the early twentieth century, it was known that the origin of cancer was tightly linked to the unrestricted division of malignant cells and their progeny [49]. However, the reasons behind the malignant transformation, and hence the cancer initiating events, were unclear. In 1914, Theodor Boveri observed unusual chromosomal abnormalities in dividing cancer cells under a microscope [23]. This observation led him to propose that cancers are abnormal clones of cells, caused and characterized by irregularities in hereditary material [23], suggesting that cancers originate from a single cell that bears chromosomal abnormalities. These are then passed on to all descendant of the cell of origin, causing rapid proliferation. Today, it is understood that aneuploidy (a type of genomic instability) is a common genetic feature in solid tumors [53], along with other factors like enabling replicative immortality (where cancer cells maintain their telomeres, allowing indefinite division) and sustaining proliferative signaling (continuous activation of growth factor pathways). These features are among the hallmarks of cancer [47, 48],

which collectively contribute to its development and progression by giving cancer cells the traits needed to survive, proliferate and spread uncontrollably.
This connection between somatic mutations and cancers was one of the earliest established links in cancer research [138]. Over the following decades, numerous studies confirmed and further developed this concept. For example, it was discovered that many chemical agents, such as tobacco smoke, are mutagenic and can lead to cancer [88]. A key breakthrough was the demonstration that introducing genomic DNA from human cancers into normal cells could transform them into cancer cells [74, 30, 29, 97, 133]. This led to the identification of abnormal genes called oncogenes in 1982 [127, 141], which, when mutated, gain functions that drive cancer development [20, 21, 153]. Concurrently, research on hereditary retinoblastoma led to the identification of tumor suppressor genes [69], which typically loose functions by mutations. These findings initiated the still ongoing search for oncogenes driving human cancers.

Cancer as an evolutionary process

In 1976, Peter Nowell used concepts of evolutionary theory to understand cancer formation and development, pioneering the hypothesis of tumor evolution. He proposed that tumors are highly individual, suggesting that each cancer may require a specific treatment approach. Success in therapy could be achieved if no new therapy-resistant clones emerge [113].
Nowell's model proposed that most cancers originate from a single neoplastic cell (exactly like Boveri stated in 1914) and evolve through a selection process that favors somatic alterations promoting the proliferation and survival of the most aggressive clones [113]. Hence, cancer development can be seen as a form of Darwinian evolution [31], involving two key processes: the continuous acquisition of heritable genetic variation in individual somatic cells through mutations, and natural selection acting on the resulting phenotypic diversity [138, 155, 43]. Cancer therapy resistance mechanisms can also be seen as an adaptation of the Darwinian framework: while therapeutic intervention may destroy cancer clones and limit their habitats, it can also unintentionally apply strong selective pressure, leading to the emergence and expansion of resistant variants [45].
Nowadays, the Darwinian framework is widely accepted for understanding the progression of cancer. However, deducing the evolutionary dynamics of individual tumors from, typically, one or a few clinical samples remains a fundamental problem [98, 154, 77]. To address this, several models describing tumor evolution have been proposed. One model describes a linear succession of clonal cell divisions. In this scenario, alterations occur in progenitor cells in a stepwise manner, giving these cells a strong selective advantage that allows them to outcompete earlier clones (see Figure 1.2.1 on the left). As a result, the tumor would consist of clonally identical cells resulting from continuous selective sweeps [32]. However, in 1978, Dexter at al. [34] demonstrated that tumors are composed of

genetically diverse subclones with fundamentally different behaviors, defining tumors as heterogeneous. This supported a model where tumors are formed by multiple subclones simultaneously, each diverging and expanding with varying levels of fitness (or selective advantage) and giving rise to a branching tumor evolution [51, 45] (see Figure 1.2.1 on the right). In Chapter 2, I will show how the branching tumor evolution model will be able to fit my data best.

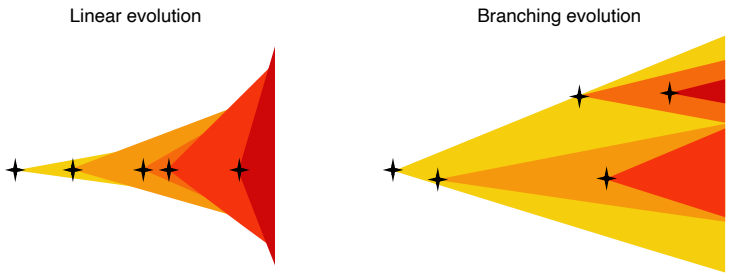


Figure 1.2.1: Linear (left) and branching (right) tumor evolution. Different colors highlight different subclones driven by newly acquired mutations.

Recent advancements of sequencing strategies

Today, genome-wide analysis, through exome or whole-genome sequencing, is well established [157, 96]. In fact, high-throughput DNA sequencing, known as next-generation sequencing, has enabled the sequencing of tens of thousands of primary cancers and matched normal samples across a wide range of cancer types. This has led to projects like The Cancer Genome Atlas (TCGA; `cancergenome.nih.gov`) and the Catalog of Somatic Mutations in Cancer (COSMIC; `cancer.sanger.ac.uk`), which identify common features among different cancer entities.
These advancements have revolutionized the understanding of cancer genetics; for example, previously unknown cancer driver genes were discovered and new mutational signatures have provided deeper insights into cancer development [98]. Moreover, by sequencing each DNA locus several times, the tissue fraction harboring a specific mutation can be estimated from the fraction of reads carrying the mutation [139]. In this way, clonal genetic variants are distinguished from subclonal ones. These efforts have not only advanced the understanding of cancer as a genomic disease, but also provided the data needed to develop tools and resources for rapid detection and analysis of relevant genomic events, guiding individualized treatment options [179, 129, 76]. In this thesis, I will show how analyzing allele frequency patterns can provide even deeper insights into a cancer's evolutionary history.

Passenger and driver mutations

As previously discussed, next-generation sequencing has enabled the identification of numerous somatic mutations in cancer cell genomes [125]. Determining which of these mutations contribute to cancer development is crucial for understanding tumor biology, advancing cancer research and improving targeted therapies.
While most somatic mutations that accumulate in cells are harmless, some can affect genes or regulatory elements, leading to significant consequences such as initiation or progression of cancers [98, 157]. Therefore, each somatic mutation in a cancer cell genome, whatever its structural nature, can be classified based on its impact on cancer development: a driver mutation gives the somatic cell a selective advantage, promoting the growth or survival of the clone originating from the mutated cell, and is thus positively selected during cancer evolution. The remaining and vast majority of mutations, known as passenger mutations, do not confer any growth advantage and have no phenotypic or biological effect on the clone [46, 138]. Large-scale publicly available tumor sequencing projects, developed in recent decades, have provided classifications of variants driving cancer evolution. Examples include Pan-Cancer Analysis of whole-genomes (PCAWG) [1], COSMIC ([144] , `https://cancer.sanger.ac.uk/cmc/home`) and Intogen ([102], `https://www.intogen.org/search`).
Driver genes can be classified into two categories: oncogenes and tumor suppressor genes. They both promote the growth advantage of the cell, but the former ones when activated by a mutation and the latter ones when inactivated by a mutation. Moreover, oncogenes often have recurrent mutations at the same amino acid positions, whereas tumor suppressors typically undergo mutations that truncate the protein throughout their length [157].
Patterns of passenger mutations allow consequences of drivers on clonal evolution to be detected with statistical confidence, which I will exploit in the next chapter.

1.2.2 Mutational processes in cancer

Somatic mutations occur after conception and accumulate with each cell division. These changes can, but do not always, lead to cancer. In a cancer cell's genome, somatic mutations may result from replication errors or DNA damage that is either incorrectly repaired or left unrepaired. These mutations can be either endogenous, originating within the cell, or exogenous, caused by external factors. Endogenous factors include aging, reactive oxygen species, aldehydes, mitotic errors and possibly other unknown mechanisms. Exogenous factors include chemicals, tobacco smoke, ultraviolet (UV) light or radiation exposure [14]. Both endogenous and exogenous mutational processes contribute to various cancers in different ways [98]. The mutations they cause can take various forms: base pair substitutions (single nucleotide variants or SNVs), insertions or deletions (indels) of DNA

segments, DNA rearrangements (also known as structural variants, where DNA is broken and rejoined with a segment from elsewhere in the genome) and copy number losses and gains [98, 26].
The rate of cancer mutations varies significantly [42]. Most cancers carry between 1000 and 20000 somatic point mutations and a few to hundreds indels. The range varies widely due to diverse sequencing depths, but it is known that pediatric tumors and leukemias generally have the fewest mutations, while cancers caused by exogenous mutagens, such as lung cancer (induced by tobacco) or melanomas (induced by UV light), have the highest rates [157, 7, 80].
In addition to genetic variants, the cancer genome also acquires epigenetic changes. These changes alter chromatin structure and gene expression without modifying the DNA sequence and are reflected by, for example, changes in the DNA methylation status [177, 61, 16]. I will provide more into detail about mutations induced by epigenetic factors in Section 1.3.

Signatures of mutational processes in cancer genomes

In the previous section, I discussed how tumor cells (and not only) accumulate, on the genomic level, somatic alterations during cancer development. For certain mutational processes, distinctive patterns of mutations, known as mutational signatures, have been identified. These provide insights into the factors, both endogenous and exogenous, that have caused and influenced cancer development [70], which can be identified by systematically studying mutation spectra [95]. Analysing mutational signatures has shown that some mutagenic processes are active in all tissues, whereas others are tissue- or exposure-specific. For example, it is clear that single base substitution signature (SBS) 1 and SBS5, as well as small insertions and deletions signature (ID) 1 and ID2, are present in all cell types and reflect life-long activity of clock-like mutational processes, which accumulate throughout life at a steady rate and are hence related to aging [5, 7, 110]. However, some cells exhibit additional mutational signatures, that can partly explain the variation in the mutation rate and mutation spectra between tissues. For example, SBS7, found in most skin fibroblasts and melanocytes, is a signature caused by UV light, leading to erroneous repair in pyrimidine dimers [142, 59]. Similarly, SBS2 and SBS13 are associated with activity of endogenous APOBEC cytosine deaminases and have been identified in multiple cell types including lung, breast, colorectal and small intestinal cells [128, 160, 52, 122, 82].
In recent years, mutation signatures have become valuable for cancer diagnosis and prognosis. They can predict therapy response [86, 152, 87] and reflect genomic changes induced by chemotherapy, making them essential tools in cancer research [35]. However, some limitations need to be considered. For instance, a mutational signature can correspond to

multiple types of somatic mutations, complicating the unique decomposition of signatures. Additionally, in recent years, the number of signatures has significantly increased, many of which with unknown origin, and various sub-signatures emerged within a single one (e.g., SBS7 has multiple versions SBS7a, SBS7b, SBS7c and SBS7d). This raises concerns about whether these findings reflect biological phenomena or are merely mathematical artifacts. Therefore, for clinicians, using mutational signatures reliably for clinical stratification is still under discussion [70].
To sum up, mutational signatures can indicate the presence or absence of certain cellular processes in cancer cells. Some of these mutational processes are consistently active throughout the cancer cell's lifetime (clock-like signatures) [5], while others are active only when subject to certain factors and are influenced by the patient's lifestyle [6, 35]. By examining these mutational patterns in cancer genomes, it is possible to uncover the action of many known and unknown mutational processes in cancer [98]. Specifically, I will use signatures to determine if particular mutational processes that occurred during the cancer's evolution could have caused an increase in its mutation accumulation rate.

1.2.3 Progression from normal to cancer cells

Recent research has shown that somatic driver mutations can accumulate in normal tissues, and not just in cancerous ones [126]. This raises important questions about the origins of cancer and how normal tissues transition into tumors.
The transformation from normal to cancer cells is a complex, multi-step process involving a series of genetic and epigenetic mutations that accumulate over a patient's lifetime. Some of these mutations were acquired in the precursors of the first cancer cells (as indicated by Δt_1 in Figure 1.2.2, shown in dark blue). These mutations often result from environmental factors, genetic predispositions or random errors during cell replication, as detailed in Section 1.2.2. Detecting somatic mutations in normal tissues before they develop into cancer is challenging due to the difficulty of identifying low-frequency events in a small number of cells and the limited availability of normal tissue samples from healthy donors. However, targeted sequencing of known cancer driver genes has identified several driver mutations occurring in the skin [100], esophagus [101, 174], colon [116, 166] and brain [41, 63] of healthy individuals.
The remaining somatic mutations in a cancer cell's genome are acquired during the tumor's evolution (Δt_2 in Figure 1.2.2, shown in light blue), making them specific to the tumor's development. Cancer cells within the tumor can further evolve, potentially enhancing their malignancy and capacity to metastasize, thereby spreading cancer to other parts of the body.

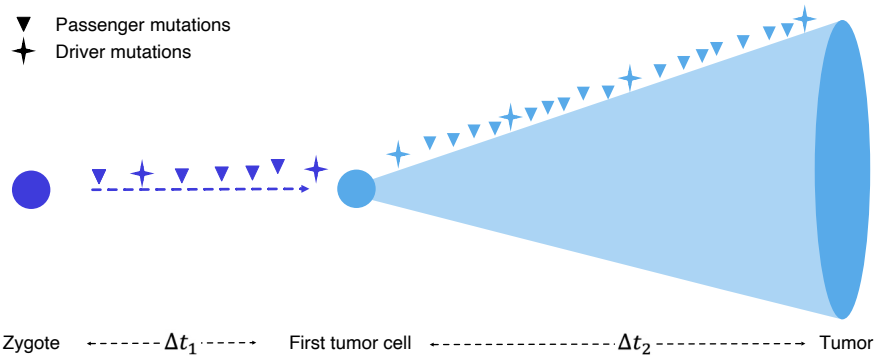


Figure 1.2.2: Scheme showing the progression of a tumor through mutation acquisition. The mutations acquired before the tumor's origin are shown in dark blue, whereas the mutations acquired afterwards are shown in light blue.

To sum up, the catalogue of somatic mutations in a cancer cell is essentially a detailed record of all the mutational processes the cell has undergone throughout the patient's life, from birth to the present. This comprehensive record offers valuable insights into the development of individual tumors. In Chapter 2, I will build on this knowledge to infer quantitative measurements about the tumor's evolutionary dynamics, including timing the tumor's origin. Understanding the tumor's evolutionary history is essential for creating targeted therapies and early detection methods to effectively fight cancer [138, 98].

1.2.4 Mutation timing

In the previous section, I showed how many tumors often transition from benign to malignant lesions through the gradual accumulation of mutations over time. Certain cancers, such as colorectal cancers, esophageal adenocarcinomas and breast carcinomas, follow distinct and well-established progression stages that can be identified using both histological and genetic methods, making early detection more straightforward. Conversely, other cancer types, including brain cancers, do not have a histologically and genetically discernible premalignant phase. One possible reason for this is that catastrophic mutational events, such as telomerase crises, may lead to a rapid accumulation of driver mutations, making it difficult to detect intermediate stages between normal and malignant cells [98]. Accurately timing the acquisition of driver mutations becomes even more crucial, as understanding this progression aids in reconstructing the tumor's evolutionary history and predicting its future course. This task is more straightforward for cancers with known pre-cancerous conditions, as illustrated in the study of colorectal cancer by [54], whereas for tumors without pre-cancerous conditions it is more challenging. Such insights are important for the development of effective therapies and early diagnostic strategies.

Pre-cancerous condition

The term precancerous lesion is referred to areas of a tissue that show certain histological changes that are associated with an increased risk of cancer. In epithelia, these early changes can evolve into adenomas, which are benign growths, where cells present morphological features of cancer but do not yet invade the underlying tissue. After invasion, carcinomas are formed, which are usually associated with stage III malignant cancers. The mostly studied multi-hit model (where the accumulation of each new driver mutation drives a new phase of the cancer's evolution) is the one for colorectal cancer formation and progression. Colorectal cancer's gradual evolution goes through well defined morphological stages (see Figure 1.2.3) and was documented already in the early 90s by Vogelstein and Kinzler [156], who define it as the sequential mutation of genes in a single cell and its progeny.

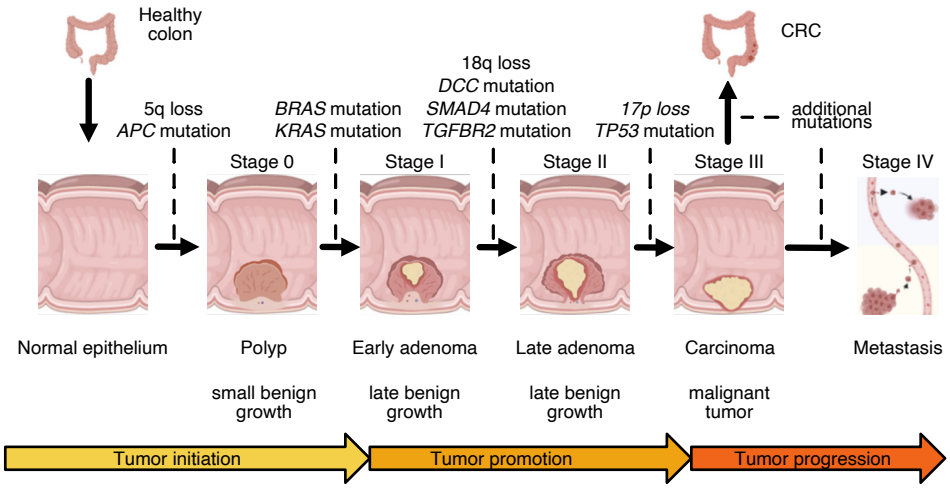


Figure 1.2.3: Colorectal cancer progression. Figure adapted from [156] and [92].

Other examples of pre-cancerous lesions are Barrett's esophagus, which is the pre-malignant condition for oesophageal adenocarcinoma [44], and ductal carcinoma in situ, which represents pre-invasive breast carcinoma [159, 84]. These observations seem consistent with the step-wise model explaining cancer progression, in which a series of events drives successive clonal expansions with progressively more complex phenotypes.

Age

The incidence and the number of mutations in certain tumors are directly correlated with age [145, 108, 78, 17]. Linear regression analysis suggests that more than half of the somatic mutations identified in these tumors arise during the initiating phase, which is when normal cells continuously replenish the tissues [157]. Indeed, recent studies have

confirmed that the burden of somatic mutations in normal tissue cells increases linearly with age [22, 2]. Moreover, deep targeted sequencing of genes from healthy sun-exposed skin, esophagus and brain tissues without tumor diagnoses has revealed positive selection of mutant clones carrying driver mutations [100, 101, 63]. Surprisingly, many of these mutations resemble those commonly found in tumors (including *IDH1* in the brain [41]), indicating a process of positive selection and clonal expansion in normal tissue similar to the one seen in cancer [62, 41]. These findings suggest that somatic mutation-driven clonal expansions may be characteristic of aging in various tissues [57, 95]. However, healthy brains did not show an accumulation of high-frequency variants with age, indicating no age-related increase in detectable oncogenic variants in normal brain tissue [41]. This partly explains why advanced brain tumors generally have fewer mutations than, for instance, colorectal tumors; glial cells in the brain do not replicate like the epithelial cells forming the colon's crypts. Therefore, the initial mutation in brain cancer is likely to occur in a precursor cell with fewer mutations than a colorectal precursor cell.

Telomere length is a measure of biological aging. Telomeres are specific DNA–protein structures at the ends of chromosomes that protect the genome from degradation, recombination, repair and interchromosomal fusion. With each cell division, a portion of telomeric DNA is lost, and when telomere length reaches a critical threshold, the cell undergoes senescence and/or apoptosis. Thus, telomere length serves as a biological clock, inversely correlating with age [151, 58, 132]. Cancer cells often exhibit activation of telomerase reverse transcriptase, which prevents telomere shortening and enables replicative immortality [149, 171, 37].

In Chapter 2, I will show how mutations in the telomerase reverse transcriptase promoter region significantly influence the evolutionary process of the specific cancer entity examined in this thesis, which is *IDH*-mutant and 1p/19q-codeleted oligodendroglioma.

1.3 Epigenetics to explain cancer development

Originally recognized as a genetic disease, cancer is now understood to involve both genetic and epigenetic abnormalities. Mutations are often found in genes that regulate the epigenome and epigenetic changes can lead to gene mutations [16]. While genetic alterations in tumorigenesis have been extensively studied, the role of epigenetic changes is, overall, not equally well understood.
Epigenetics studies how gene expression is heritably regulated without changes to the primary DNA sequence. Disruption of these processes can affect the function of tumor suppressor genes and oncogenes, initiating malignant transformation [61, 103]. Epige-

netic changes, like genetic mutations, can undergo Darwinian natural selection if there is variation within the cell population, stable heritability from mother to daughter cells and resultant phenotypic effects subject to selection [138].
The primary processes responsible for epigenetic regulation include DNA methylation and histone modifications; changes in microRNA expression also contribute [16, 61, 177]. These mechanisms are vital for normal cell development and growth, and their alterations contribute to cancerous phenotypes [16, 61, 177, 25].
In this thesis, I will concentrate on DNA methylation data and analyse data for oligodendrogliomas.

1.3.1 What is DNA methylation

DNA methylation is a chemical modification where a methyl group is added to the DNA molecule, specifically at the fifth carbon of cytosine residues that are linked to a guanine nucleotide, forming a CpG dinucleotide. This process is catalyzed by DNA methyltransferases (DNMT), resulting in the formation of 5-methylcytosine (5mC) [79, 93]. Certain DNA regions, known as CpG islands, are rich in CpG sites and are typically located at the 5' end of genes, overlapping with gene promoters. CpG sites are also present in gene bodies and other regions; depending on their proximity to CpG islands they are called CpG shores (2 k-bp regions flanking CpG islands), CpG shelves (regions greater than 2 k-bp from CpG shores) and open sea regions (more than 4 k-bp from the nearest CpG islands) [93]. The primary DNA methylation writer enzymes are: DNMT1, which preserves existing methylation patterns after DNA replication, and DNMT3A and DNMT3B, which are de novo enzymes targeting unmethylated CpGs to initiate methylation. DNMT3A and DNMT3B are highly active during embryogenesis and minimally expressed in adult tissues [61].

1.3.2 DNA methylation defines distinct groups of cancers

DNA methylation is a vital epigenetic modification that regulates gene expression and maintains genome stability [61, 119, 177]. This modification can either activate or silence specific genes and varies between cell types, developmental stages and in response to environmental factors and diseases. Abnormal DNA methylation patterns are linked to various diseases, including cancer [60, 119]. Some cancers exhibit unique methylation profiles that define distinct molecular subtypes. Specifically, changes in CpG-island methylation within individual tumors can affect specific loci, potentially defining a unique phenotype [60]. This concept was introduced in 1999 by Toyota et al. as the CpG island methylator phenotype (CIMP) [147]. Initially identified in human colorectal cancer [147], CIMP refers to the cancer-specific hypermethylation of certain genes in specific tumors

[93, 175]. Molecular subtypes based on DNA methylation have been identified in many other cancer types, including acute myeloid leukemia [148], esophageal adenocarcinoma [176], chronic lymphocytic leukemia [114], pediatric brain tumors [27] and gliomas [112, 150], on which I will focus on in the next sections.

G-CIMP status distinguishes low-grade from high-grade gliomas

The primary objective of the methylation classifier of any tumor type is to accurately categorize them based on their methylation profiles. Precise tumor diagnosis is essential for healthcare professionals to predict prognosis and develop personalized treatment plans for patients [119].
Given that the focus of this thesis is on oligodendrogliomas, I will now discuss more in detail the DNA methylation status classification of adult tumors of the central nervous system. In 2010, a glioma CpG island methylator phenotype (G-CIMP) was identified by Noushmehr et al. [112]. This diagnosis was based on an epigenetic biomarker panel comprising seven hypermethylated loci and one hypomethylated locus, validated in silico. Specifics about these eight loci are shown in Table 1.1. A sample is considered G-CIMP high (or G-CIMP+) if at least six genes displayed a combination of *DOCK5* hypomethylation and/or hypermethylation of the other genes in the panel [112, 93].

Gene Name	CpG site ID	Type
DOCK5	cg16849041	Hypomethylated
RHOF	cg09088508	Hypermethylated
FAS1	cg16257983	Hypermethylated
FAS2	cg17120764	Hypermethylated
LGALS3	cg17403875	Hypermethylated
HFE	cg19320816	Hypermethylated
MAL	cg21245652	Hypermethylated
ANKRD43	cg26399201	Hypermethylated

Table 1.1: Table showing sites for determining G-CIMP status in adult gliomas [112].

The G-CIMP status is closely related with somatic mutations in the *IDH1/2* genes [112, 150]. The mechanism behind this association is the following: wild-type *IDH* converts isocitrate into α-ketoglutarate, whereas mutant *IDH* converts α-ketoglutarate into the oncometabolite 2-hydroxyglutarate (2-HG). Hence, 2-HG accumulates in high levels in tumors with *IDH* mutations but is nearly absent in *IDH* wildtype cases [161]. The presence of 2-HG inhibit DNA demethylation [150, 19], leading to widespread DNA and histone hypermethylation, genetic instability and the acquisition of new mutations [60]. Thus, mutations in *IDH1/2* genes define a distinct subset of gliomas with a hypermethylated

G-CIMP phenotype (namely low-grade gliomas), which are associated with better clinical outcomes compared to gliomas without this phenotype (namely high-grade gliomas) [94]. In summary, methylation profiling plays a crucial role in categorizing tumors of the central nervous system into clinically significant subtypes, allowing a more refined diagnosis of brain tumors [65].

Methylation status detects 1p/19q-codeletion

Having established the ability to differentiate low-grade gliomas from high-grade gliomas using the methylation status of specific CpG site, the next step is to further classify low-grade gliomas based on the presence or absence of chromosomes 1p/19q-codeletion. This distinction allows for the differentiation between oligodendrogliomas and astrocytomas. Whole-genome methylation studies have shown that G-CIMP+ gliomas can be divided into subgroups based on their 1p/19q-codeletion status. Specifically, Paul et al. [120] identified 14 CpG sites (shown in Table 1.2) that can allow the distinction between *IDH*-mutant gliomas with 1p/19q-codeletion (oligodendrogliomas) and *IDH*-mutant gliomas with intact 1p/19q loci (astrocytomas).

Gene name	CpG site ID
CD300LB	cg00873351
NA	cg03492827
FLJ37543	cg04437966
FGFR2	cg07250222
TCF7L1	cg07847030
PLCG1	cg07893801
PTPRN2	cg08935418
FGFR2	cg09772154
PRKAG2	cg10363569
NA	cg12210255
MAPKAP1	cg13412754
NA	cg13598010
GPR156	cg19093820
PTPRN2	cg23759393

Table 1.2: Table showing sites for determining 1p/19q-codeletion status [120].

These CpG sites are hypermethylated in oligodendrogliomas compared to diffuse astrocytomas [120]. Additionally, lower methylation levels are associated with poorer survival outcomes. This suggests that the 1p/19q-codeletion induces epigenetic changes and genome-wide hypermethylation, leading to a distinct biological phenotype with better survival rates [146].

1.4 Oligodendrogliomas, *IDH*-mutant and 1p/19q-codeleted

1.4.1 Human brain development

Neuroepithelial cells are the stem cells of the central nervous system and initiate the development of neurons and glial cells. Initially, neuroepithelial cells undergo symmetric division to expand the progenitor pool. They then elongate and transform into radial glial cells. Radial glial cells can produce neurons directly through asymmetric division or indirectly by generating intermediate progenitor cells. Intermediate progenitor cells act as transient amplifying cells, dividing symmetrically multiple times to create clones of neurons [140, 180]. Neurons differentiate and mature as they migrate along radial glial scaffolds of radial glial cells to the cortical plate. This migration occurs in a inside-out manner, with late-born upper-layer neurons migrating past early-born deep-layer neurons, as depicted in Figure 1.4.1 by different shades of green. As neurons approach their destinations, they extend axons and dendrites.

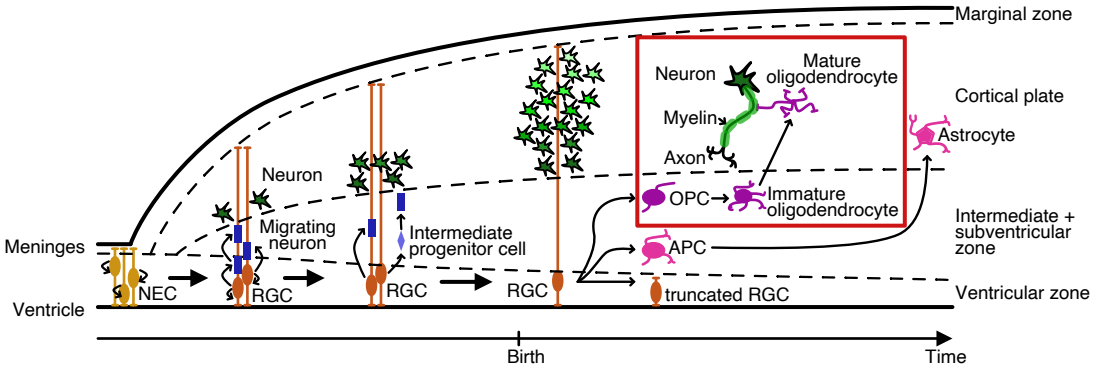


Figure 1.4.1: Neurons (in green) are generated by migration along radial glial cells (orange) and reach the cortical plate. Lighter shades of green indicate later generated neurons. Among non-neuronal cells, astrocytes and oligodendrocytes are generated by radial glial cells (RGCs) post-neurogenesis and mature throughout early postnatal periods. The red box highlights the formation of oligodendrocytes. Figure adapted from [180].

Radial glial cells also give rise to progenitors for glial cells, oligodendrocyte progenitor cells and astrocyte progenitor cells. Oligodendrocytes and astrocytes are produced post-neurogenesis and mature during early postnatal periods, playing roles in modulating neuronal functions, such as regulating synaptic transmission and myelinating neuronal axons [180].

Oligodendrocytes are the myelinating cells of the central nervous system. They facilitate an efficient transmission of electrical signals by wrapping around axons and speeding up impulse transmission [75]. Oligodendrocytes develop from oligodendrocyte progenitor

cells into immature oligodendrocytes through processes of migration, proliferation and differentiation. Eventually, they reach maturity, extending organized branches to myelinate neuronal axons (red box in Figure 1.4.1). Each oligodendrocyte can myelinate multiple neurons, providing support and releasing substances that support neuronal health.

1.4.2 Dynamics of oligodendrocyte generation

Oligodendrogliomas arise from oligodendrocyte progenitors in the brain [121]. The number of these progenitors is highest in infancy and reaches a plateau by around five years of age [172]. After this period, the number of oligodendrocytes remains largely stable throughout life, with an annual turnover rate of 1/300 oligodendrocytes, or about 0.33% per year. This rate, estimated from the dynamics of carbon isotopes formed by nuclear weapon tests, is at least 100-fold lower than the rate measured in mice by genetic fate mapping of oligodendrocyte progenitor cells in transgenic mice (with an oligodendrocyte generation rate of 36.5%–182% per year) [172].
Unlike neurogenesis and astrocytogenesis, oligodendrogenesis begins later and takes longer, as shown in Figure 1.4.2.

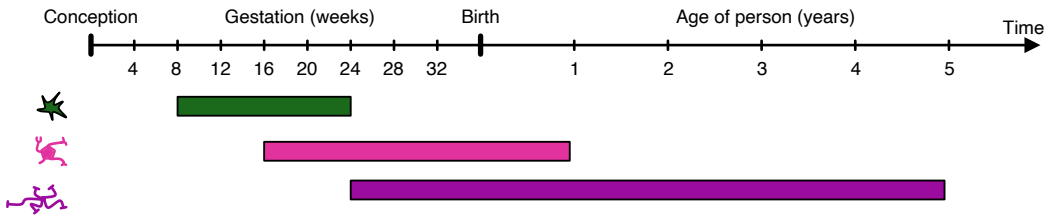


Figure 1.4.2: Timeline of the major events in human brain development: in green neurogenesis, in pink astrocytogenesis and in purple oligodendrogenesis. Figure adapted from [134].

1.4.3 Oligodendroglioma is a relatively recent tumor entity

Over the past decades, the understanding of molecular changes in tumors of the central nervous system has greatly increased. Until the fourth edition of the World Health Organization classification of tumors of the central nervous system published in 2007 [91], these tumors were defined on the basis of their histology only. This meant that they were categorized by their morphological similarities to different potential cells of origin. Starting in 2016, with the updated fourth edition [90], and especially from the 2021 edition [165], the classification underwent a significant advancement: it expanded from purely histological criteria to histological-molecular ones. As a result, the traditional method of diagnosing tumors based on microscopy imaging was improved by introducing molecular

parameters of specific central nervous system tumor entities, increasing the objectivity of the diagnostic process [164]. This aims to enhance diagnostic accuracy and improve patient management by more precisely predicting prognosis and treatment responses.
This shift in classification had a great impact on oligodendrogliomas. Until then, oligodendrogliomas and astrocytomas were grouped together and diagnosed as oligoastrocytoma, a category of diffuse gliomas known for their invasive growth and diagnostic challenges, including high interobserver discordance. By using genetic markers, such as the *IDH1/2* mutations and 1p/19q-codeletion status, these tumors were assigned as either astrocytomas or oligodendrogliomas. As a result, the diagnostic differences between these two cancer entities could be more easily defined and more effective treatment strategies could be developed and tailored to either of the two different cancers. This approach of combining histopathological and molecular features needed a new nomenclature, which would include both the histopathological name and genetic features. Therefore, oligodendrogliomas are now classified as *IDH*-mutant and 1p/19q-codeleted oligodendrogliomas.
The classification of adult tumors of the central nervous system, according to the 2021 edition [165, 89], groups three different categories, namely gliomas, glioneuronal tumors and neuronal tumors. Gliomas are then further characterized into diffuse or more circumscribed, called non-diffuse, gliomas. Adult-type diffuse gliomas are the most common and aggressive malignant neoplasms of the central nervous system and are grouped into distinct subtypes: oligodendrogliomas, *IDH* mutant and whole-arm loss of the long arm of chromosome 1 and the short arm of chromosome 19 (which is shortly called 1p/19q-codeletion) (grades 2 and 3), astrocytoma, *IDH* mutant (grades 2-4) and glioblastoma, *IDH* wild-type (grade 4), as can be seen in Figure 1.4.3.

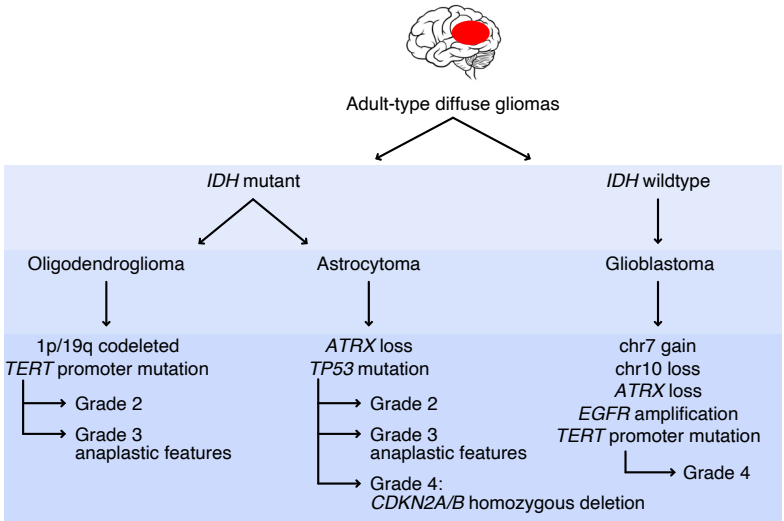


Figure 1.4.3: Classification of adult diffuse gliomas. Figure adapted from [18].

Clinical presentation, diagnosis, prognosis and treatment of oligodendroglioma

Oligodendrogliomas make up around 5% of all malignant central nervous system tumors [146, 118]. They typically peak around midlife, between the fourth and fifth decades of life [146]. Clinically, patients usually present with non-specific symptoms, like headaches, and sometimes with focal neurological deficits or seizures, depending on the tumor's size and location [146, 18].

Accurate classification of these tumors is crucial for diagnosis and treatment planning. The standard diagnostic procedure includes contrast-enhanced magnetic resonance imaging and histopathologic evaluation. However, as I mentioned before, these two procedures alone are insufficient to distinguish oligodendrogliomas from lastrocytomas, necessitating a biopsy to check for specific genetic markers. Oligodendrogliomas are defined by the presence of *IDH* mutations, particularly *IDH1* codon 132 (R132H) and/or the homologous *IDH2* codon 172 (R172K) mutations, along with 1p/19q-codeletion. These key alterations, detectable through next-generation sequencing, occur early in the tumor's development. Oligodendrogliomas are then divided histologically into grades 2 and 3 [18], with grade 3 having a slightly worse prognosis.

Despite their relatively slow growth compared to other brain cancers, oligodendrogliomas are incurable [158]; the primary treatment is surgical resection, though recurrence is inevitable [85, 4]. Postoperative approaches, such as watchful waiting and adjuvant therapy, are considered on a case-by-case basis [65]. Patients under 40 with grade 2 oligodendrogliomas and complete resection can be monitored with serial magnetic resonance imaging, while those with partial resection or older age often require further treatment. For grade 3 oligodendrogliomas, standard therapy includes both radiotherapy and chemotherapy, with several studies (such as the Radiation Therapy Oncology Group 98-02 trial and the European Organization for Research and Treatment of Cancer) showing that adding chemotherapy improves survival [65].

New therapy options are arising [65]. The *IDH* mutations in oligodendrogliomas make them suitable for *IDH*-specific targeted treatments. For example, the NOA16 trial tested an *IDH*-specific vaccine combined with other treatments, demonstrating safety and positive immune responses [124]. The ongoing AMPLIFY-NEOVAC trial is further exploring this vaccine with a checkpoint inhibitor [24]. Additionally, *IDH* inhibitors like enasidenib and ivosidenib, initially used for leukemia, show promise for gliomas [104, 143, 109]. Vorasidenib, another *IDH* inhibitor, has also shown positive results and is currently being tested in the INDIGO trial for primary astrocytoma and oligodendroglioma patients [105].

Chapter 2

Evolutionary trajectories of oligodendrogliomas

In this Chapter, the clonal evolution of *IDH*-mutant and 1p/19q-codeleted oligodendrogliomas is analyzed using whole-genome sequencing data of paired primary and relapsed tumor samples. I will develop a population-dynamic model of mutation acquisition and tumor growth, and will then apply it to the oligodendroglioma data. In this way, I will be able to learn dynamic aspects of the evolution for individual tumors. Finally, the results are discussed in the context of the current state of research.
All data shown in this chapter were collected within the scope of the "SysGlio" consortium, coordinated by Peter Lichter (German Cancer Research Center, Heidelberg).

2.1 Patient and tissue samples collection

Matched tissue samples from the initial surgery, before any tumor-specific treatment, and the second surgery, for recurrent tumor growth, were collected from six patients with *IDH*-mutant and 1p/19q-codeleted oligodendroglioma registered to the German Glioma Network (GGN, www.gliomnetzwerk.de). All patients included in the study provided their written informed consent for participating in the German Glioma Network and for the use of their tissue samples and clinical data for research purposes. The molecular characterization of primary and recurrent oligodendroglioma pairs was performed within the collaborative research project "SysGlio – Systems-based prediction of the biological and clinical behavior of gliomas" consortium, coordinated by Peter Lichter (German Cancer Research Center, Heidelberg) and funded by the German Ministry of Education and Research (BMBF). The study was approved by the institutional review board of the Medical Faculty, Heinrich Heine University, Düsseldorf, Germany (study number 4940). Tumor samples were histologically classified as oligodendrogliomas, *IDH*-mutant and

1p/19q-codeleted by Jörg Felsberg and Guido Reifenberger (Heinrich Heine University, Düsseldorf). According to the 2021 World Health Organization classification of central nervous system tumors [165], four patients were classified as grade 2, while the remaining two were grade 3. Mean patient age was 37 years old, ranging from 24 to 48 years old, and mean time of tumor relapse was 5 years, ranging from 1 to 9 years. Primary samples were retrieved from treatment-naïve tumors after initial surgery. After resection of the primary tumor, one patient received concomitant adjuvant radiation therapy, and all the other patients did not receive any therapy treatment. Table 2.1 provides overall patients' clinical information.

Table 2.1: Patient data including demographics, tumor characteristics and treatment details (RT: radiation therapy).

Patient	Gender	Age at first surgery (years)	WHO grade	Tumor location	Extent of resection	Therapy after 1st surgery	Years between surgeries
P1	F	37	2	Other	Total	RT only	9
P2	F	41	2	Frontal	No data	None	1
P3	M	29	2	Other	Partial	None	2
P4	M	48	3	Temporal	Subtotal	None	4
P5	M	40	2	No data	No data	None	8
P6	F	24	3	Parietal	Total	None	5.5

2.2 Whole genome sequencing

Whole genomes of tumor samples from primary and recurrent resections and matched blood samples, used as germline controls, were sequenced at an average coverage of 150x and 80x, respectively (see Figure 2.2.1). This deep sequencing strategy allows for variant calling at high resolution of allele frequencies.

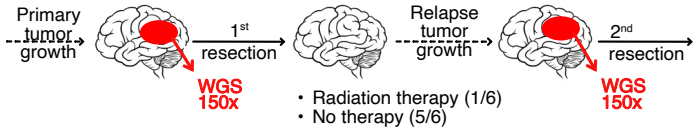


Figure 2.2.1: Sample acquisition and sequencing.

Unfixed deep frozen tissue specimens used for extraction of nucleic acids were histologically evaluated to assure that they contained vital tumor cells with a histologically estimated tumor cell content of 80% or more. Genomic DNA from deep-frozen tissue samples of primary and relapsed oligodendrogliomas was extracted by Bernhard Radlwimmer and Yonghe Wu (both German Cancer Research Center, Heidelberg, with

support from the DKFZ Genomics and Proteomics Core Facility). Sequence alignment (`github.com/DKFZ-ODCF/AlignmentAndQCWorkflows/wiki`), tumor-specific variant calling (with the standard pipeline of DKFZ - DKFZ SNVCalling Workflow (`github.com/DKFZ-ODCF/SNVCallingWorkflow`) and Insertion-Deletion Calling Workflow for Roddy (`github.com/DKFZ-ODCF/indelCallingWorkflow`)) and detection of copy number alterations, tumor cell content estimates (with ACEseq (`github.com/DKFZ-ODCF/ACEseqWorkflow`)) and structural variants (with Sophia (`github.com/DKFZ-ODCF/SophiaWorkflow`)) were performed by Jing Yang and Matthias Schlesner (both German Cancer Research Center, Heidelberg, with support from the DKFZ Omics IT and Data Management Core Facility).

2.3 A model inferring the dynamics of cancer evolution

As introduced in Section 1.2, the development of a tumor is an evolutionary process that begins with a single cell acquiring a mutation that gives it a selective advantage, leading to tumor formation. As the tumor progresses over time and space, different cell populations, or subclones, may appear, expand and diminish due to continuous accumulation of new mutations and selection, resulting in a heterogeneous tumor [45]. Comprehending intratumor heterogeneity by reconstructing subclones is crucial for cancer evolution studies, as it can improve clinical management and can help develop more effective cancer treatments [3]. Hence, in the upcoming sections, I will develop a computational approach to deconvolute the cell population of primary and recurrent oligodendrogliomas into clones and subclones based on the frequencies of accumulated mutations, relying on the variant allele frequency.

2.3.1 The variant allele frequency

To build my inference model I will use deep whole-genome sequencing data. Among other features, this technology provides a table of mutated single nucleotide variants (SNVs) and insertions/deletions (indels) in the genome. Each mutation is mapped to a genomic location, having a certain copy number, and the number of reference reads and variant (mutated) reads are counted. The variant read number measures the independent sequence reads supporting the presence of the variant. The number of reference reads measures the independent reads without the somatic variants. Hence, the sum of variants and reference reads indicates how deeply that genomic locus was sequenced. Sequencing depth varies across the genome, implying that sequencing depth affects the VAF and influences the threshold to set for the analysis (I will discuss this more in detail in the following section). The variant allele frequency (VAF) is the ratio of mutated reads to the total coverage at the

specific locus in the genome

$$\text{VAF} = \frac{\text{\# of mutated reads}}{\text{\# of total reads}} = \frac{\text{\# of mutated reads}}{\text{\# of mutated reads + \# of reference reads}}.$$

This number, between 0 and 1, informs about the proportion of cells carrying a specific mutation and is the key measure for all my analysis. In fact, the VAF is helpful for distinguishing clonal from subclonal mutations within the sequenced tumor data. A high VAF corresponds to a clonal mutation, which was acquired and accumulated in the ancestry of the cell that originated the tumor and hence is present in all tumor cells. In particular, for a disomic locus, a clonal mutation would have VAF ≈ 0.5 in the tumor cell fraction. Lower VAFs indicate a subclonal mutation, which was acquired post tumor origin and hence is present only in a subset of the cells that will form the tumor.

The shape of the variant allele frequency distribution

An effective way to visualize all VAFs is through a histogram, where the abscissa represents the frequency of reads with a specific mutation, while the ordinate shows the number of mutations observed at that frequency. This histogram typically has a distinct shape [167], which I will explain in detail. To be specific, I assume a diploid genome; however, all ideas carry over to other copy numbers.
A variant that is present in every cancer cell is known as a clonal mutation (see Figure 2.3.1 on the left, in dark green). The mutations occur before the tumor's most recent common ancestor (MRCA) and hence I will use them to estimate when the first tumor cell emerged. For genomes without any copy number gains/losses, clonal mutations typically create a peak around a frequency of 0.5 (see Figure 2.3.1 on the right, in dark green), reflecting that such mutations affect one of the two alleles. The spread of this clonal peak is influenced by sequencing depth, as the number of reads for a mutation follows a binomial sampling process. I will explore this aspect further in the next section.

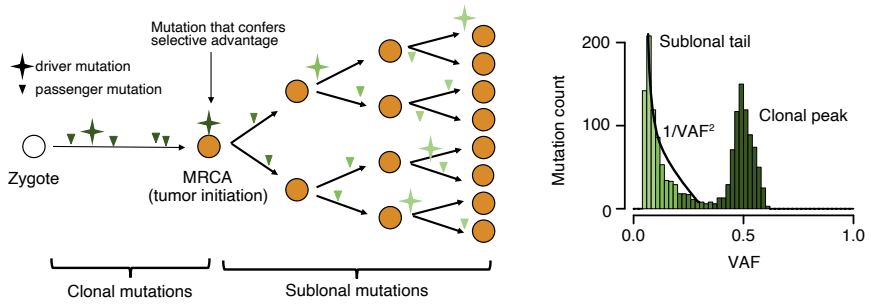


Figure 2.3.1: Mutation accumulation in cancer (on the left) and the corresponding VAF distribution (on the right). Clonal mutations are shown in dark green, whereas subclonal mutations are shown in light green; a lighter shade indicates its later acquisition.

Conversely, variants that arise later in tumor development, and are thus present in only a subset of tumor cells, are referred to as subclonal mutations (shown in Figure 2.3.1 in lighter greens). These are depicted in the histogram at a lower frequency compared to clonal mutations and form the subclonal tail. The subclonal tail follows a $1/\text{VAF}^2$ distribution, as shown in various studies [36, 64, 167].
In summary, the shape of the variant allele frequency histogram contains quantitative information on the tumor's evolution: the clonal peak briefs about mutations acquired prior the tumor's MRCA, while the subclonal peak contains information about intratumoral heterogeneity.

How purity, ploidy and sequencing depth affect the variant allele frequency distribution

Aside from basic cell processes, which I will cover in the next section, other factors influencing the shape of the VAF distribution are sequencing depth, tumor cell content (also known as purity) and changes in the copy number of the tumor's genome. The first two are related to sample acquisition, whereas the latter one is intrinsic to the tumor. I will go more into detail explaining how each of these aspects affect the VAF histogram's shape. Sequencing depth refers to the number of independent reads. As illustrated in Figure 2.3.2 for a diploid genome, lower sequencing depths result in a broader clonal peak, whereas higher ones produce a sharper peak around 0.5. Sequencing depth is the major determinant of the measurement error. Sequencing depth also impacts the ability to detect low-frequency mutations reliably. Specifically, a higher sequencing depth allows for a lower threshold for variant detection, while a lower sequencing depth necessitates a higher threshold. Hence, to accurately interpret the tumor's evolutionary dynamics from the VAF and avoid considering false negatives, it is essential to set an appropriate minimum frequency limit. For my analysis, I have chosen a threshold of 0.05, which is appropriate for the 150x whole-genome sequencing data I will use, as it means that a variant is supported on average by at least seven reads.

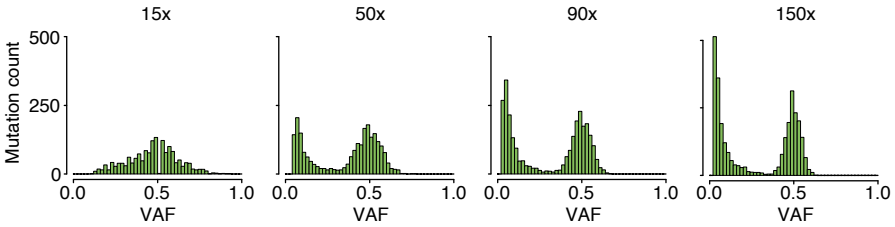


Figure 2.3.2: From left to right: a simulated VAF distribution of variants obtained with a progressively higher sequencing depth, ranging from 15x to 150x.

In bulk whole-genome sequencing, the sequenced sample is formed by a mixture of tumor and normal (non-tumor) cells. The proportion of these cell types defines the tumor cell content (or purity), which ranges from 0 (indicating the absence of tumor cells) to 1 (indicating the presence of exclusively tumor cells). The purity influences the VAF histogram, as the presence of normal cells reduces the frequency of tumor variants. Thus, a lower sample purity results in a shift towards lower VAFs, as shown in Figure 2.3.3.

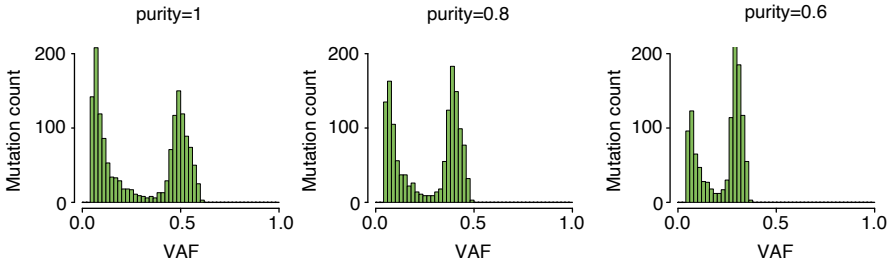


Figure 2.3.3: From left to right, a simulated VAF distribution of variants obtained with a progressively lower purity.

The VAF distribution is also influenced by the copy number of the genomic region the variants are mapped to. Specifically, the higher the copy number, the more clonal peaks will be present. As illustrated in Figure 2.3.4, clonal variants in disomic regions produce a single clonal peak at a frequency of 0.5, corresponding to variants present in one of the two alleles. In contrast, variants in trisomic regions exhibit two distinct peaks: one at a frequency of 1/3, representing variants in one of the three alleles, and another at 2/3, representing variants in two of the three alleles. The latter were present already before the duplication of one of the alleles. The subclonal tail remains evident in both cases.

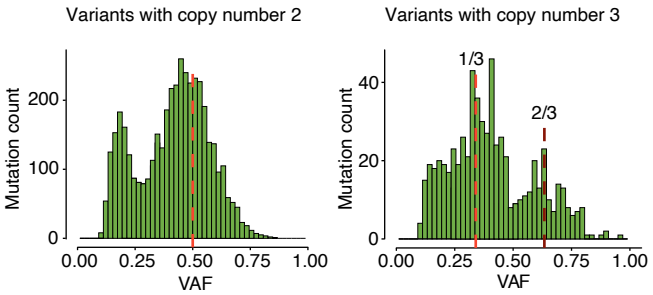


Figure 2.3.4: On the left, a typical VAF of variants mapped to a disomic region, highlighting in red the clonal peak. On the right, a typical VAF of variants mapped to a trisomic region, containing two distinct peaks, shown in different shades of red.

How the mode of evolution influences the variant allele frequency distribution

In addition to copy number, other intrinsic factors related to tumor growth influence the shape of the VAF distribution. The most important ones are the tumor cells' birth rate, death rate and mutation accumulation rate. So far, I discussed the expected shape of the VAF histogram under the assumption of neutral evolution of the tumor, implying that, once the driver mutation occurred, all tumor cells have indistinguishable rates of mutation, proliferation and cell loss (represented by the yellow cells in Figure 2.3.1 and 2.3.5). Now, I will explain how the VAF distribution changes if, at a random time t_{sel}, one of these tumor cells acquires an additional driver mutation that confers a selective advantage to its birth rate with respect to the remainng tumor cells. This driver mutation will enable that cell, and its progeny, to proliferate more rapidly, forming a positively selected subclone (illustrated in dark red on the left in Figure 2.3.5). In the VAF distribution, variants from this subclone will create a distinct peak located at a frequency between the subclonal tail and the clonal peak [169]. The presence of this positively selected subclonal peak sets this VAF histogram apart from the one obtained under neutral evolution (Figure 2.3.1).

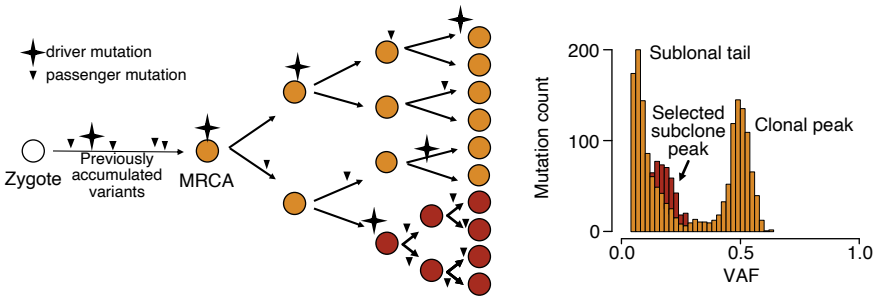


Figure 2.3.5: How the tumor growth dynamics with the presence of a positively selected subclone, which is highlighted by red cells on the left, affect the shape of the VAF histogram with the presence of peak of the selected subclone, on the right.

Both the position and size of the subclonal peak offer quantitative insights into the timing of the subclone's emergence and the dynamics of selection. Specifically, the number of variants in the subclone is formed by the mutations accumulated before its formation and is hence related to the subclone's age t_{sel}. A subclone that arises early in tumor evolution will have fewer variants, as its founder cell had less time to accumulate mutations. Conversely, a subclone that emerges later will exhibit a larger peak. Additionally, the higher the selective advantage, the higher the frequency of the subclone's peak. This is because the selected subclone expands faster and becomes more represented within the tumor quicker.
Thus, the VAF distribution not only provides information about the timing of mutation acquisition, as previously discussed, but also offers insights into the tumor's mode of evolution. In cases where evolution involves ongoing selection, the VAF distribution can

quantitatively reflect the impact of selection: the number of mutations in a subclone indicates its age, while the frequency of these mutations correlates with the selective advantage. Therefore, the VAF histogram is a crucial statistic to infer the tumor's evolutionary history. To do this, developing a mathematical model describing the frequency of somatic variants is essential. In the following sections I will do this, after briefly reviewing existing models for cancer inference from bulk whole-genome sequencing data and explaining why a new model is needed.

2.3.2 Existing models for subclonal deconvolution in cancer

In recent years, various models have been developed for subclonal deconvolution and inference of tumor evolution from whole-genome sequencing data. Examples include BMix, which uses a traditional Bayesian mixture model to analyse variant allele frequencies [28]; VIBER, which employs variational Bayesian inference integrating both VAFs and copy number variants [28]; and MOBSTER, which combines a Bayesian mixture model with a Dirichlet process, focusing primarily on VAFs [28]. Other Bayesian-based models include PyClone [130] and DPClus [111], whereas SciClone [107] uses a Gaussian mixture model. Among these, only MOBSTER can distinguish between a subclonal tail, a selected subclone and a clonal peak. It does so by modeling VAF density with two types of distributions: beta distributions, to capture the clone and subclone, and a Pareto Type-I power law, to represent the neutral tail. Additionally, MOBSTER estimates key parameters related to the tumor's evolutionary dynamics, such as subclone frequency, the number of mutations in the subclone and the selective advantage. However, given that MOBSTER uses a statical approach with a mixture model for subclonal deconvolution, it sometimes fails to detect subclones (see Figure 2.3.6). For example, it does not detect any positively selected subclone in the primary samples of Patients 1, 2 and 5, despite clear evidence of such subclones in at least Patients 1 and 2.
In contrast, the model I developed (which I will explain in detail in Section 2.3.3) dynamically fits evolutionary processes from population genetics and aligns more closely with the approaches described in [73, 71]. My model appears more sensitive to subclonal detection, successfully identifying the subclones that MOBSTER missed (see Figure 2.5.1).
What sets my model apart is its integration of mathematical modeling with the patient's age at the time of the first surgery, allowing the inference, for each individual tumor, of its evolutionary stages in actual years. This innovation provides a timeline for tumor growth and for the development of selected subclones.

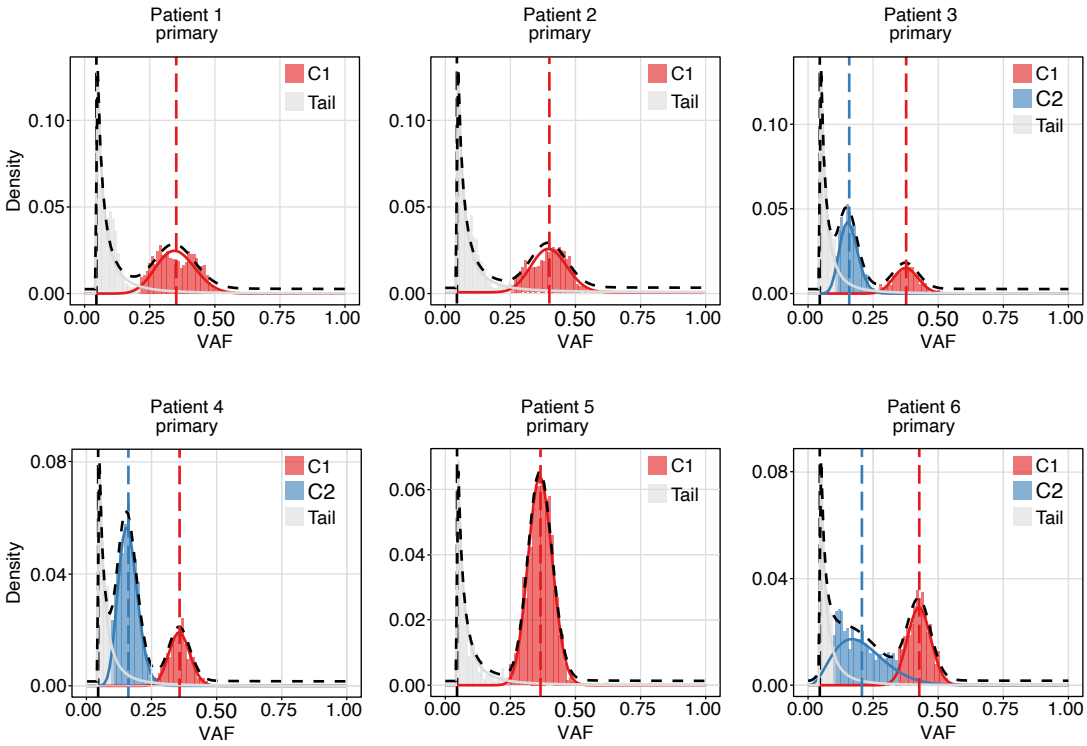


Figure 2.3.6: Results of MOBSTER runs on the six primary samples of oligodendrogliomas. The red cluster C1 corresponds to the clone, whereas the blue one, C2, is the selected subclone. The tail is shown in grey.

2.3.3 Mathematical description of the model

I showed in Section 2.3.1 how the dynamics of cell division, loss and variant acquisition shape the characteristic VAF histogram. The aim of my model is to learn these processes leading to a specific VAF distribution.

Upon tumor initiation, my model distinguishes between two growth scenarios, which I already discussed previously: neutral evolution, where all tumor cells expand at the same rate, and the selection of a fitter subclone, which accelerates tumor growth due to the selective advantage of the subclone cells. In both scenarios, I assume that the tumor originated from a single normal cell that underwent malignant transformation. This cell will be the most recent common ancestor (MRCA) and emerges at time t_{MRCA}. The MRCA cell continues to divide accumulating mutations until the time of the surgery, t_{surgery}, from which the sequenced biopsy was obtained.

In the following pages, I will outline my approach to modeling variant accumulation in tumors, addressing both scenarios. I will begin by presenting the equations that describe tumor cell growth over time in both neutral and positively selected tumors. These equations

will only depend on the birth rate, death rate and selective advantage of the tumor cells. Next, I will incorporate the mutation accumulation rate to simulate VAF distributions. These will then be compared to the observed VAF histograms to estimate the parameters of the model using approximate Bayesian computation.

Modeling the tumor cell population

In the case of neutral evolution, tumor cells divide symmetrically with a net growth rate of $\tau_N = b - d$, where b and d are, respectively, the birth rate (due to division) and the death rate (due to death or differentiation). I assume both of them to be constant over time. In this way, the resulting total cancer cell population at time t, $N_{T,\text{neutral}}(t)$, will grow exponentially with time according to

$$N_{T,\text{neutral}}(t) = e^{\tau_N (t - t_{\text{MRCA}})} \text{ with } t \geq t_{\text{MRCA}}.$$

To incorporate a positively selected subclone into the model, I start by assuming that it arises from a random tumor cell that acquires a driver mutation at time t_{sel}, where $t_{\text{MRCA}} < t_{\text{sel}} < t_{\text{surgery}}$. This driver mutation will confer a selective advantage, which I model as an increased rate of cell division by a factor of $1+s$, signifying a faster exponential growth than the neutral cancer cells due to a net growth rate of $\tau_S = b(1+s) - d$ (see Figure 2.3.7 in red). This implies that the number of cells at time t belonging to the selected subclone $N_S(t)$ grows according to the following equation

$$N_S(t) = e^{\tau_S (t - t_{\text{sel}})} \text{ with } t \geq t_{\text{sel}}.$$

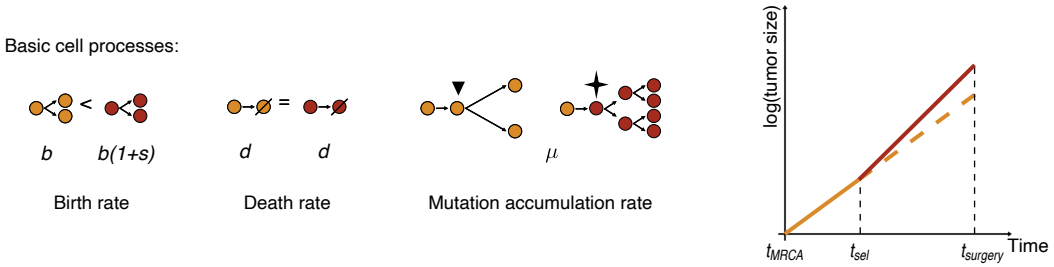


Figure 2.3.7: Tumor cells' rates affect the growth of the tumor. In yellow, tumor's neutral cells, in red, positively selected tumor cells.

In order to get the total amount of cells $N_{T,\text{selection}}(t)$ forming the tumor in the presence of a selected subclone, I will also have to take into account the neutrally growing cells, hence:

$$N_{T,\text{selection}}(t) = N_S(t) + N_N(t),$$

with

$$N_{\mathrm{N}}(t) = \begin{cases} e^{\tau_{\mathrm{N}}(t-t_{\mathrm{MRCA}})}, & \text{if } t_{\mathrm{MRCA}} < t < t_{\mathrm{sel}} \\ e^{\tau_{N}(t-t_{\mathrm{MRCA}})} - e^{\tau_{\mathrm{N}}(t-t_{\mathrm{sel}})}, & \text{if } t \geq t_{\mathrm{sel}}. \end{cases} \tag{2.3.3.1}$$

Modeling the variant allele frequency histogram

Regardless of the evolutionary mode, tumor cells acquire neutral somatic variants at a rate of μ. For simplicity, I assume μ to be sufficiently small to be consistent with Kimura's infinite-sites model, which assumed that the number of nucleotide sites making up the genome is so large while the mutation rate per site is so low that whenever a mutant appears it represents a mutation at a new site [67].
To infer the growth dynamics and potential clonal selection from the VAF distribution in these exponentially growing cell populations, I developed a model to simulate whole-genome sequencing of variants accumulated in a tumor. This model considers tumors starting to grow from a single cell at time t_{MRCA} and reaching a size of 10^9 cells by the time of surgery t_{surgery} [33]. According to [33], a tumor of 10^9 cells is approximately the size of a diagnosable tumor of a few cubic centimeters. This simulation generates a VAF histogram based on the model's parameters, which are specified in Table 2.3.8.

Parameter	Unit	Distribution	Min	Max
d/b	1	Uniform	0	1
μ	SNVs	Uniform	1	200
s	1	Uniform	0	1
t_{sel}	1	Uniform	0	1
n_{clonal}	SNVs	Uniform	0	10000
purity	1	Uniform	0.5	1

Figure 2.3.8: Prior estimates for model fitting.

To determine the expected VAF distribution at time t_{surgery}, a forward branching process is used. According to [115], the expected number of mutations with VAF f in a neutrally evolving tumor is given by

$$S_{\mathrm{neutral}}(f, \mu, t_{\mathrm{surgery}} | b, d) = \int_{t_{\mathrm{MRCA}}}^{t_{\mathrm{surgery}}} \mathbb{P}(i, b, d, t_{\mathrm{surgery}} - t) N_{\mathrm{T,neutral}}(t) b \mu dt,$$

where i denotes the clone size and relates with f according to $i = 2fN_{\mathrm{T,neutral}}$, accounting for a diploid genome; $\mathbb{P}(i, b, d, t_{\mathrm{surgery}} - t)$ accounts for neutral drift during expansion and denotes the probability for a clone to grow to size i from a single cell within $t_{\mathrm{surgery}} - t$.

The probability distribution of this supercritical birth-death process is well known [13]:

$$\mathbb{P}(i,b,d,t') = \begin{cases} x(t') & \text{for } i = 0 \\ (1-x(t'))(1-y(t'))y(t')^{i-1} & \text{for } i \geq 1 \end{cases}$$

where:

$$x(t') = \frac{de^{(b-d)t'} - d}{be^{(b-d)t'} - d}$$

and

$$y(t') = \frac{be^{(b-d)t'} - b}{be^{(b-d)t'} - d}.$$

The second term under the integral, $N_{\text{T,neutral}}(t)b\mu$, describes the number of mutations acquired at time t.
For a tumor evolving with a positively selected subclone, the expected number of mutations with VAF f is given by

$$\begin{aligned} S_{\text{selection}}(f,\mu,t_{\text{surgery}}|b,d) =& S_{\text{I}}(f,\mu,t_{\text{surgery}}|b,d) + S_{\text{II}}(f,\mu,t_{\text{surgery}}|b,d) \\ &+ S_{\text{III}}(f,\mu,t_{\text{surgery}}|b,d). \end{aligned}$$

The terms S_{I}, S_{II} and S_{III} are illustrated in Figure 2.3.9, describing how a mutation (in light blue), acquired either before (subfigure A) or after (subfigure B) the formation of the selected subclone (in red), contributes to the VAF distribution.
Specifically, $S_{\text{I}}(f,\mu,t_{\text{surgery}}|b,d)$ is defined as the sum of all the mutations acquired in the founding tumor clone before the formation of the selected subclone. To quantify this term, I distinguish mutations ending up in the selected subclone (Figure 2.3.9A, nested case, on the left) from mutations that do not (in Figure 2.3.9A, independent case, on the left). The probability that a mutation acquired at time t (in light blue), with $t_{\text{MRCA}} < t < t_{\text{sel}}$, will be in the selected subclone (in red) is

$$\frac{N_{\text{N}}(t_{\text{sel}} - t)}{N_{\text{N}}(t_{\text{sel}})}.$$

In this case, the mutation will be present in the entire selected subclone; otherwise, with a probability of

$$1 - \frac{N_{\text{N}}(t_{\text{sel}} - t)}{N_{\text{N}}(t_{\text{sel}})},$$

the mutation acquired at time t will drift in the founding tumor clone and be absent in the

selected subclone. Thus, S_{I} is defined as:

$$S_{\mathrm{I}}(f,\mu,t_{\mathrm{surgery}}) = \int_{t_{\mathrm{MRCA}}}^{t_{\mathrm{surgery}}} \left[\frac{N_{\mathrm{N}}(t_{\mathrm{sel}}-t)}{N_{\mathrm{N}}(t_{\mathrm{sel}})} \mathbb{P}(i - N_{\mathrm{S}}(t_{\mathrm{surgery}}), b, d, t_{\mathrm{surgery}} - t) + \left(1 - \frac{N_{\mathrm{N}}(t_{\mathrm{sel}}-t)}{N_{\mathrm{N}}(t_{\mathrm{sel}})}\right) \mathbb{P}(i, b, d, t_{\mathrm{surgery}} - t) \right] N_{\mathrm{N}}(t) b\mu \, dt$$

with $t < t_{\mathrm{sel}}$.

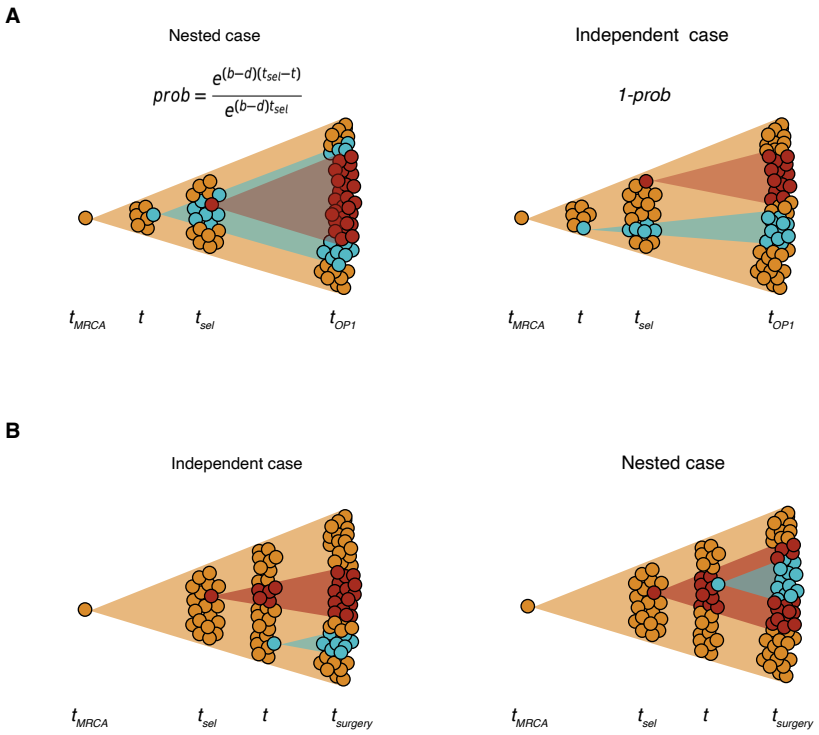


Figure 2.3.9: How a mutation (in light blue) acquired (A) before or (B) after the formation of the selected subclone (in red) contributes to the model.

Otherwise, when considering mutations acquired after the formation of the selected subclone, $t_{\mathrm{sel}} < t$, (Figure 2.3.9B), I distinguish between mutations acquired in the founding tumor clone, which form S_{II} (on the left, the independent case), from those acquired in the selected subclone itself, which form S_{III} (on the right, the nested case). Specifically, in the former case, the blue mutation just arises in the tumor founder clone and hence I employ 2.3.3.1, yielding

$$S_{\mathrm{II}}(f,\mu,t_{\mathrm{surgery}}) = N_{\mathrm{N}}(t_{\mathrm{sel}}) S_{\mathrm{neutral}}(f,\mu,t_{\mathrm{surgery}} - t_{\mathrm{sel}} | b, d).$$

Conversely, S_{III} takes into account the blue mutations acquired within the selected subclone and is thus defined as:

$$S_{\mathrm{III}}(f,\mu,t_{\mathrm{surgery}}|b,d) = S_{\mathrm{neutral}}(f,\mu,t_{\mathrm{surgery}} - t_{\mathrm{sel}}|b(1+s),d).$$

Together, S_{I}, S_{II} and S_{III} define the VAF of mutations acquired in a growing tumor mass with selection of a subclone. The so obtained simulated VAF distribution will then be compared to its measured counterpart by minimizing the distance between the two using approximate Bayesian computation (ABC) (Figure 2.3.10) [68]. In this way I will get a list of values for the parameters of the model. This allows for the inference of the tumor's evolution and growth dynamics retrospectively, which I will describe now more in detail.

Estimating the evolutionary parameters

The time of origin of the selected subclone, t_{sel}, is defined on a scale from 0 (representing t_{MRCA}) to 1 (representing t_{surgery}) and is measured relative to the expansion of the original tumor mass rather than in actual time. Hence, to obtain t_{sel} in actual time (in years), I will calibrate the MRCA's appearance time against the patient's age at diagnosis. From this, I will then derive the actual time estimates for t_{sel}.

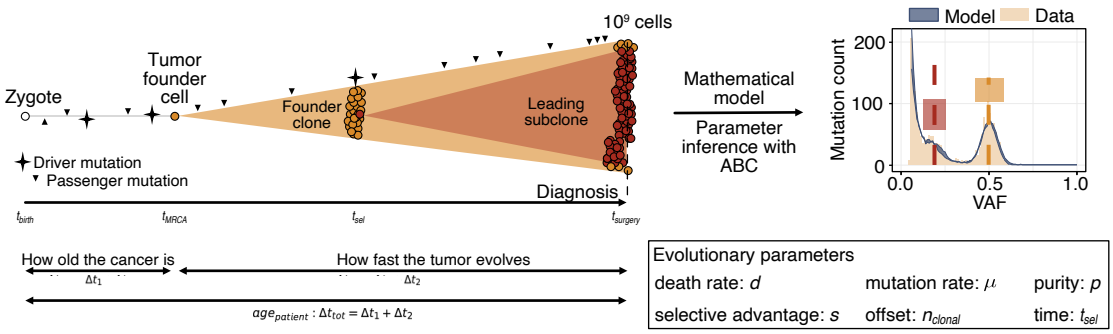


Figure 2.3.10: Description of the mathematical model simulating the VAF from the data. The time estimates and the model's parameters are highlighted in the lower part.

As previously discussed, the growth process of a tumor includes the pre-malignant time span between the patient's birth, t_{birth}, and t_{MRCA}, denoted as Δt_1, and the duration of tumor expansion, denoted as Δt_2, which goes from t_{MRCA} to t_{surgery} (see the lower part of Figure 2.3.10). At t_{MRCA}, the tumor-initiating cell will already contain a certain amount of mutations. These were acquired at a rate of μ, from t_{birth} to the time at which the tumor founder cell arose (as I already discussed in Sections 1.2.3 and 2.3.1). Hence, the tumor cell of origin already contains a set of $\mu b t_{\mathrm{MRCA}}$ somatic variants. Given that the mutation rate is known (as it is one of the inferred model parameters), the number of

pre-malignant generations can be inferred from the number of clonal mutations using the following formula

$$\Delta t_1 = \frac{n_{\text{clonal}}}{b\mu}.$$

Furthermore, the relative loss rate, combined with the final tumor size (assumed to be 10^9 cells), determines the number of generations occurring during tumor growth Δt_2. Assuming a constant cell division rate before and after tumor initiation, Δt_2 is found by solving the following exponential equation, where the last term adjusts for a single cell in the neutral tumor population that mutated into the fitter type:

$$e^{(b-d)\Delta t_2} + e^{(b(1+s)-d)(1-t_{\text{sel}})\Delta t_2} - e^{(b-d)(1-t_{\text{sel}})\Delta t_2} \stackrel{!}{=} 10^9.$$

Note that solving this equation for Δt_2 essentially involves taking the logarithm of the tumor cell number. Hence the estimate of Δt_2 depends on the magnitude of the cell number and not its precise value. Now, given the patient's age at the time of surgery and that same time span relative to the expansion of the tumor mass, $\Delta t = \Delta t_1 + \Delta t_2$, a straightforward ratio can be used to estimate the time (in years) when the first tumor cell originated (t_{MRCA}) and when the advantageous mutation conferring the selective advantage first occurred (t_{sel}) for each patient sample.
Thus, analyzing VAF histograms in conjunction with the age at diagnosis yields evolutionary trajectories for each individual in real-time.

2.4 Genomic landscape in primary and recurrent oligodendrogliomas

To begin my analysis, I compared the mutational burden of small variants (SNVs and indels) and structural variants (SVs, including medium-sized insertions, deletions, inversions, and translocations) between primary and relapse tumor samples. After excluding putative germline mutations by comparing with blood controls, tumor-specific variants were identified. On average, primary tumors had 4003 SNVs (ranging between 2392 and 6051) and 258 indels (ranging between 162 and 356), with 33% to 85% shared with relapse tumors. The overall mutational burden was not significantly higher in relapse tumors (p=0.57, Wilcoxon rank sum test), averaging 4310 SNVs (ranging between 2978 and 8429) and 300 indels (ranging between 186 and 507) (Figure 2.4.1A-B). Moreover, primary tumors had on average 21 SVs (ranging between 8 and 31), while relapse samples had 42 (ranging between 4 and 174) (Figure 2.4.1C), but none targeting known driver genes. Notably, Patient 1, who is the only one who underwent radiation therapy after the initial surgery, showed a significant increase in all variant types in the relapse sample.

Excluding this patient, the mutational burden between primary and relapse samples was comparable and overall stable.
I then analyzed copy number variants (CNVs) and purity estimates results. Purity in all samples was high, ranging from 0.62 to 0.92 (see Figure 2.5.3B), aligning with histology estimates. Furthermore, all tumors had near-diploid genomes, and the copy number profiles between primary and relapse samples of the same patient remained overall stable (see Figure 2.4.1D). Besides the subtype-defining 1p and 19q monosomy, the most common CNVs included the loss of chromosome 14 in four samples, and the loss of chromosome 4/4p, 13, 14 and 18 in three samples (notably, the loss of chromosome 4/4p occurred only in relapse samples). In addition, two samples shared gains of chromosome 7q and 11 and loss of chromosome 15. Non-subtype-defining CNVs were often subclonal and frequently changed after primary resection, indicating ongoing subclonal evolution after tumor initiation (see Section 2.4.1 below).

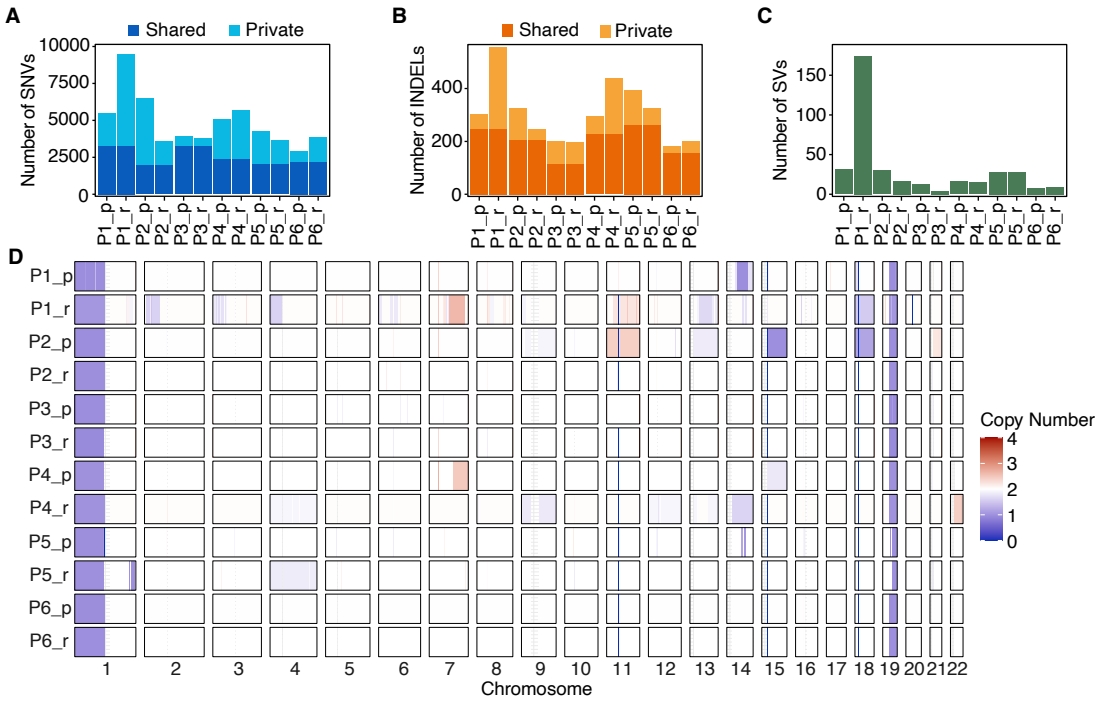


Figure 2.4.1: (A) and (B): Number of SNVs and indels, respectively, for each of the tumor samples. Variants shared between samples of the same patient are shown in a darker colour. (C) Number of SVs for each of the tumor samples. (D) Copy number variation profile for the 12 tumor samples.

While comparing overall mutational burdens provides limited insight into the mutagenic processes involved, deriving mutational signatures can reveal underlying mechanisms of cancer development, as discussed in Section 1.2.2. For instance, the increase in variants

observed in Patient 1 might be attributed to a particular mutational signature, possibly related to therapeutic treatment the patient underwent. To investigate this, I examined the single nucleotide substitution profiles for all primary and relapse samples (Figure 2.4.2). For Patient 1, as well as for all other patients, the profiles remained mostly stable between primary and relapse samples, predominantly featuring C>T and T>C substitutions, similar to patterns seen in normal tissue aging as reported by [99].
Additionally, the analysis of mutational patterns using YAPSA [55] on COSMIC v3 (`cancer.sanger.ac.uk/signatures/sbs/`), revealed that most SNVs were attributed to the clock-like signatures SBS1 and SBS5, except for Patient 6, who had a significant proportion assigned to SBS3, which is strongly associated with defective DNA damage repair (see figure 2.4.3A) [8]. However, when analyzing indel mutation signatures (see Figure 2.4.3B, `cancer.sanger.ac.uk/signatures/id/`), there was no indication of ID6, which is linked to defective DNA damage repair and hence to SBS3. Furthermore, there was no increase in SNV and indel counts for Patient 6 (see Figure 2.4.1). Overall, the mutational profiles were stable, with no evidence of an increased mutation rate due to abnormal mutagenic processes, including therapeutic treatment for Patient 1.
In summary, primary and recurrent oligodendrogliomas had comparable numbers and patterns of somatic mutations, suggesting minimal genetic evolution between the two surgeries, with the exception of the increased mutational burden of Patient 1 upon recurrence.

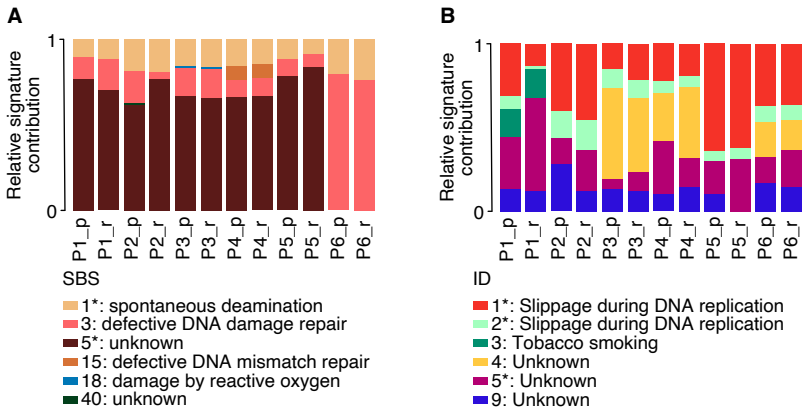


Figure 2.4.3: (A) and (B) Signatures profiles for SNVs and indels, respectively, for the six oligodendroglioma patients.

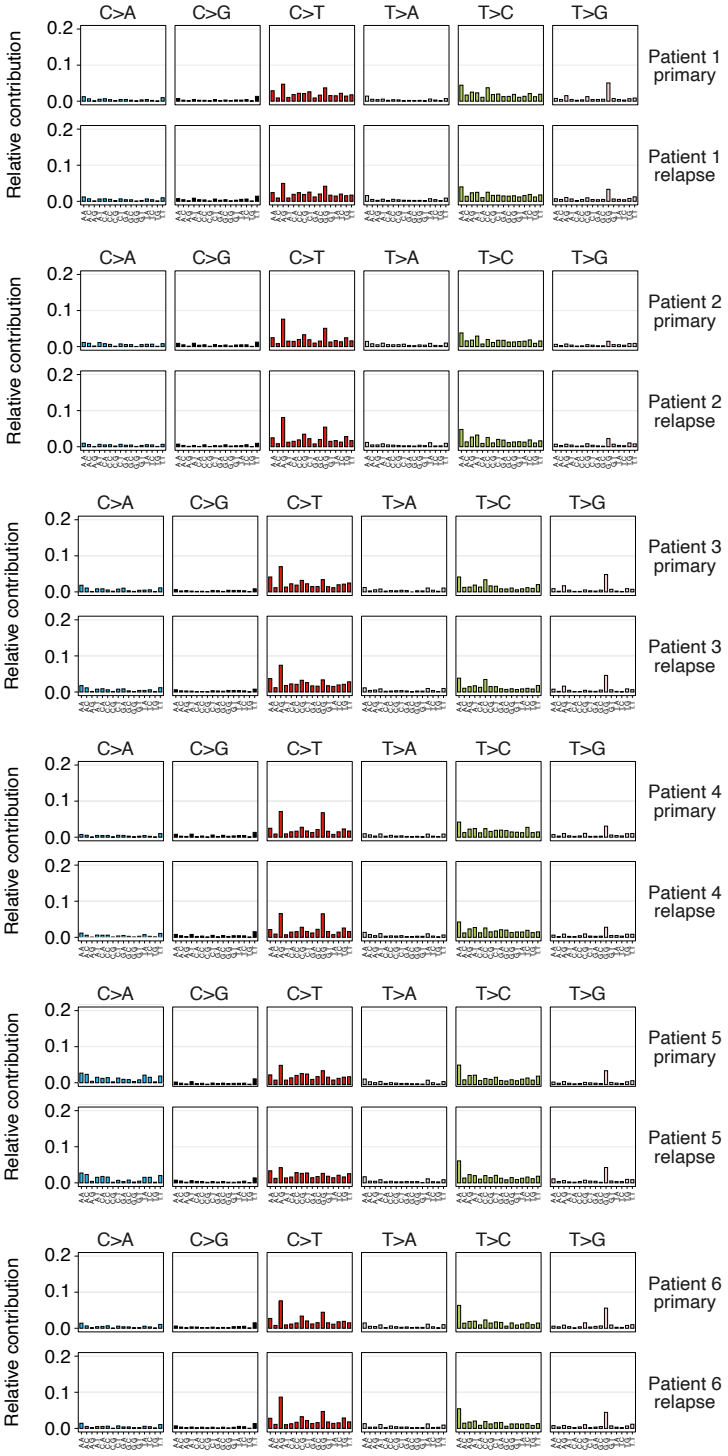


Figure 2.4.2: Single nucleotide substitution profiles in 3 bp context of the 12 tumor samples.

2.4.1 Patterns of driver mutations

As mentioned in Chapter 1.2, most mutations in cancer are passengers, with only a minority driving tumor growth. This raises the question of whether the pattern of driver mutations changes between primary and recurrent oligodendrogliomas, despite their overall similarity in mutational burden and signatures previously analyzed. To address this, I considered a list of driver genes identified for low-grade gliomas from "intOGen" (`intogen.org`) [102], known drivers for gliomas from "COSMIC" (`cancer.sanger.ac.uk/cosmic`) [144] and mutations in the telomerase reverse transcriptase (*TERT*) promoter [66, 15]. *TERT* mutations occur at two mutually exclusive hotspots: C228T (position 1,295,228 on chromosome 5p, a C to T mutation) and C250T (position 1,295,250 on chromosome 5p, with a C to T mutation) [10, 117, 50].

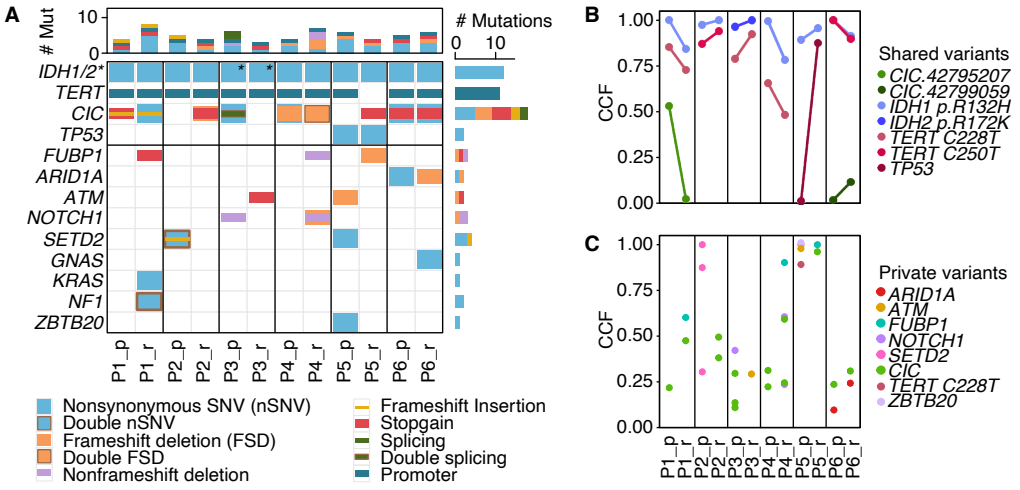


Figure 2.4.4: (A) Upper section, driver mutations that were shared in both samples of a tumor pair; lower section, driver mutations that were found in a single sample of a tumor pair only. *IDH1* mutations are shown without asterisk; *IDH2* mutations are identified with an asterisk. (B) and (C) Cancer cell fraction of driver mutations found in both samples of a tumor pair and in a single sample of a tumor pair, respectively.

On average, each tumor had five small mutations targeting driver genes (Figure 2.4.4A). Besides the subtype-defining mutations in the *IDH* genes (with the few oligodendrogliomas that do not harbor the *IDH1* p.R132H mutation containing the *IDH2* p.R172K mutation) [89], all patients had mutations in the *TERT* promoter. Additionally, mutations in the *CIC* gene was observed in 9 out of 12 samples, with each tumor having two to three distinct mutations in this gene. However, inactivating mutations in the *CIC* gene on chromosome 19q and the *FUBP1* gene on chromosome 1p are often present in oligodendrogliomas: *CIC* functions downstream of growth factor receptor signaling pathways by binding to DNA,

while *FUBP1* is a DNA-binding protein with an unknown role in gliomagenesis [25].
Driver mutations shared between primary and relapse tumors (upper part of Figure 2.4.4A) were mostly found in similar cancer cell fractions in both samples (Figure 2.4.4B). Notably, Patient 1 showed a significant decrease in cancer cell fraction for *CIC* upon recurrence, while Patient 5 exhibited a marked increase in *TP53*. Private driver mutations (lower part of Figure 2.4.4A) appeared in either primary or relapse tumors at varying cancer cell fractions (Figure 2.4.4C). Clonal mutations in the *TERT* promoter, *ATM*, *SETD2* and *ZBTB20* were detected in the primary sample of Patient 5, and in *SETD2* in the primary sample of Patient 2, but these mutations were lost in the corresponding relapse samples. Patient 5, however, increased the frequency of *TP53* and newly acquired *FUBP1* and *CIC* mutations upon recurrence, suggesting clonal replacement between the two surgeries. Additionally, Patient 4 and Patient 6 had different SNVs in *CIC* upon recurrence, and Patient 6 exhibited a frameshift deletion in the SWI/SNF complex gene *ARID1A*, replacing a prior SNV.
In conclusion, the driver's mutations landscape in paired primary and recurrent oligodendroglioma samples confirms a mostly stable mutagenic process during disease progression.

2.4.2 Genetic subclones with and without known driver mutations

To understand subclonal evolution of the oligodendrogliom samples, I analyzed VAF histograms and driver mutations together. As discussed in Section 2.3.1, factors such as sequencing depth, purity and chromosomal changes can influence the VAF of mutations. To account for these factors, I normalized the allele frequencies of variants mapped to chromosomes with copy number gains/losses and adjusted for tumor cell content by calculating the cancer cell fraction (CCF), defined as

$$\mathrm{CCF} = \frac{\mathrm{VAF}}{\mathrm{purity}} \left(\mathrm{purity} * \mathrm{ploidy} + (1 - \mathrm{purity}) * 2\right).$$

Specifically, the term outside the parenthesis is the corrected VAF, which adjusts the frequency of a mutation for the sample's purity estimate. The first term in the parenthesis represents tumor cells with a specific ploidy value, and the second term accounts for diploid non-tumor cells present in the sequenced bulk sample. Computing the CCF allows to plot the distribution of pseudo-heterozygous VAFs, defined as CCF/2, with a minimum threshold of 0.05, to account for sequencing depth. These histograms provide a fair analysis between all variants mapped to different chromosomes and belonging to samples with different purity estimates.
An initial visual inspection revealed substantial peaks at a normalized VAF of around 0.5 in all tumors, formed by clonal variants. Most of these variants remained clonal in the relapse samples, as shown in the second-to-last panels (denoted by 2D) in Figure 2.4.5 and Figure 2.4.6, with few shared variants shifting from the clonal peak in the primary sample

to a lower frequency in the relapse sample, and the other way around.
To rigorously deconvolute subclones, I used VIBER [28] on the 2D pseudo-heterozygous VAF distributions. VIBER, being semi-parametric, attempts to use at most K Binomial clusters. I set a minimum cluster size of 2% of the total number of variants; this explains the presence of the green "NA" cluster in Figure 2.4.5 and Figure 2.4.6. After subclonal decomposition, I highlighted the frequency of the driver mutations, showing their mean frequency and the 95% confidence interval according to a binomial distribution. Different colors indicate distinct clusters to which driver genes are assigned to, according to their pseudo-heterozygous VAF. The 2D VAF scatterplot only includes driver mutations shared between primary and relapse samples. The rightmost plot in the two figures displays how cluster frequencies changed between primary and relapse samples (mean and 95% confidence bound of the mean are shown), along with associated driver mutations to each cluster.
All samples exhibited the presence of 7 to 8 subclones, with many showing changes in VAF between primary and relapse samples. For example, Cluster 7 decreased in frequency in all the relapse samples, whereas Cluster 10 increased. In some cases, distinct subclones could be identified with newly acquired mutations in driver genes. For example, mutations in *NOTCH1* and *ATM* were found in Cluster 6 in Patient 3, while Cluster 8 in Patient 6 featured mutations in *FUBP1*, *CIC* and *TP53*. However, there were exceptions: Cluster 7 in Patient 6 lacked mutations in known driver genes, whereas in other patients the same cluster was associated to mutations in *CIC* and *SETS2*. Moreover, subclones can arise in the absence of selective pressure, either through neutral drift - evidenced by smaller clusters, like Cluster 4, which align with the neutral power-law tail of the VAF histogram - or through the survival of a subset of cells following resection. VIBER decomposes the power-law tail at low frequencies in the histogram into multiple subclones.
Overall, the observed intratumoral heterogeneity, both with and without known driver mutations, raises the question of whether subclonal expansions in oligodendrogliomas are primarily driven by positive selection for genetic drivers or are more influenced by random neutral drift. To answer this question, I will apply my model to the data.

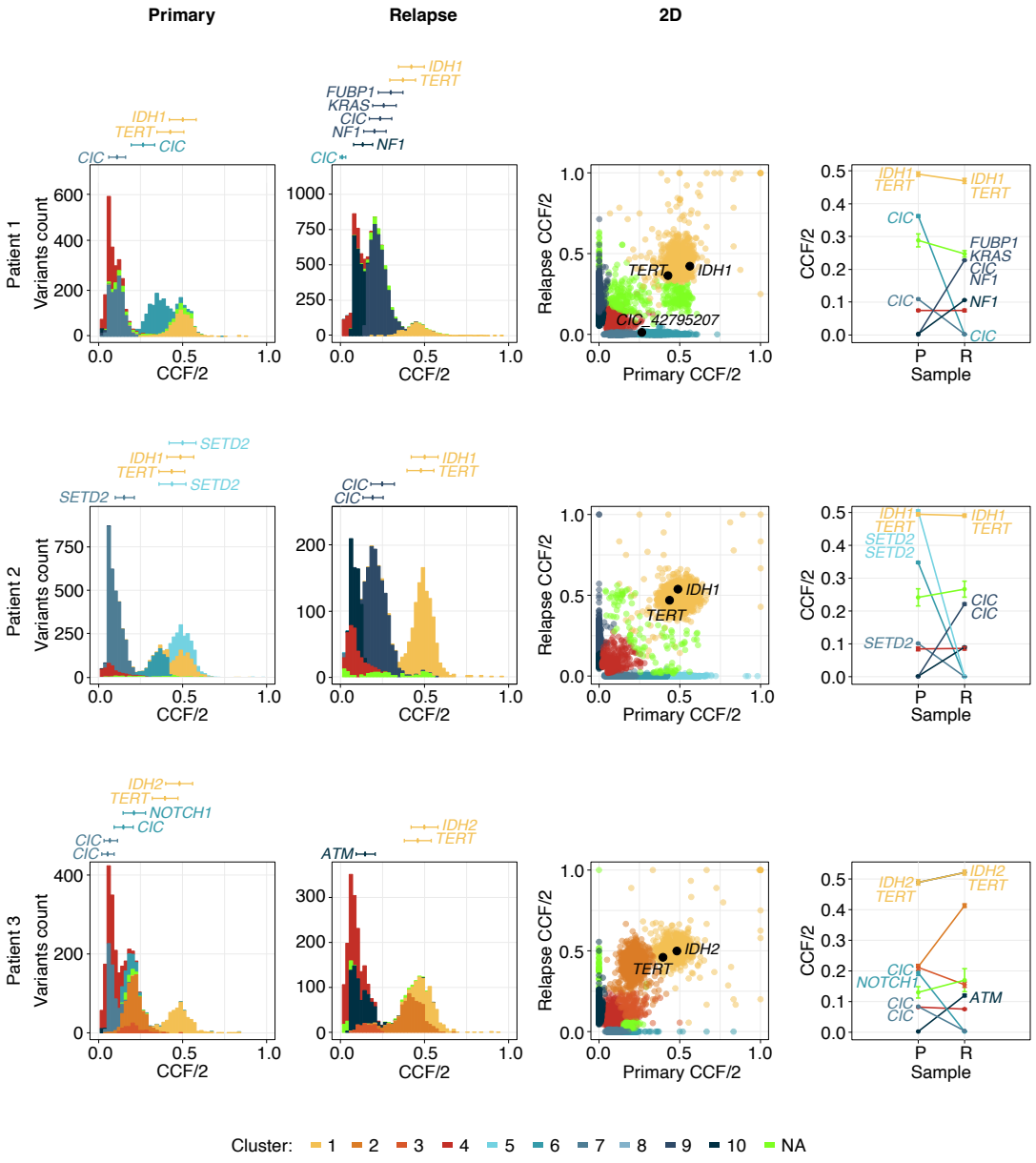


Figure 2.4.5: Pseudo-heterozygous VAF histograms and scatterplot of primary and relapsed tumors of Patients 1, 2 and 3. The plot on the most right shows how the frequencies of each cluster changes between primary and relapse samples.

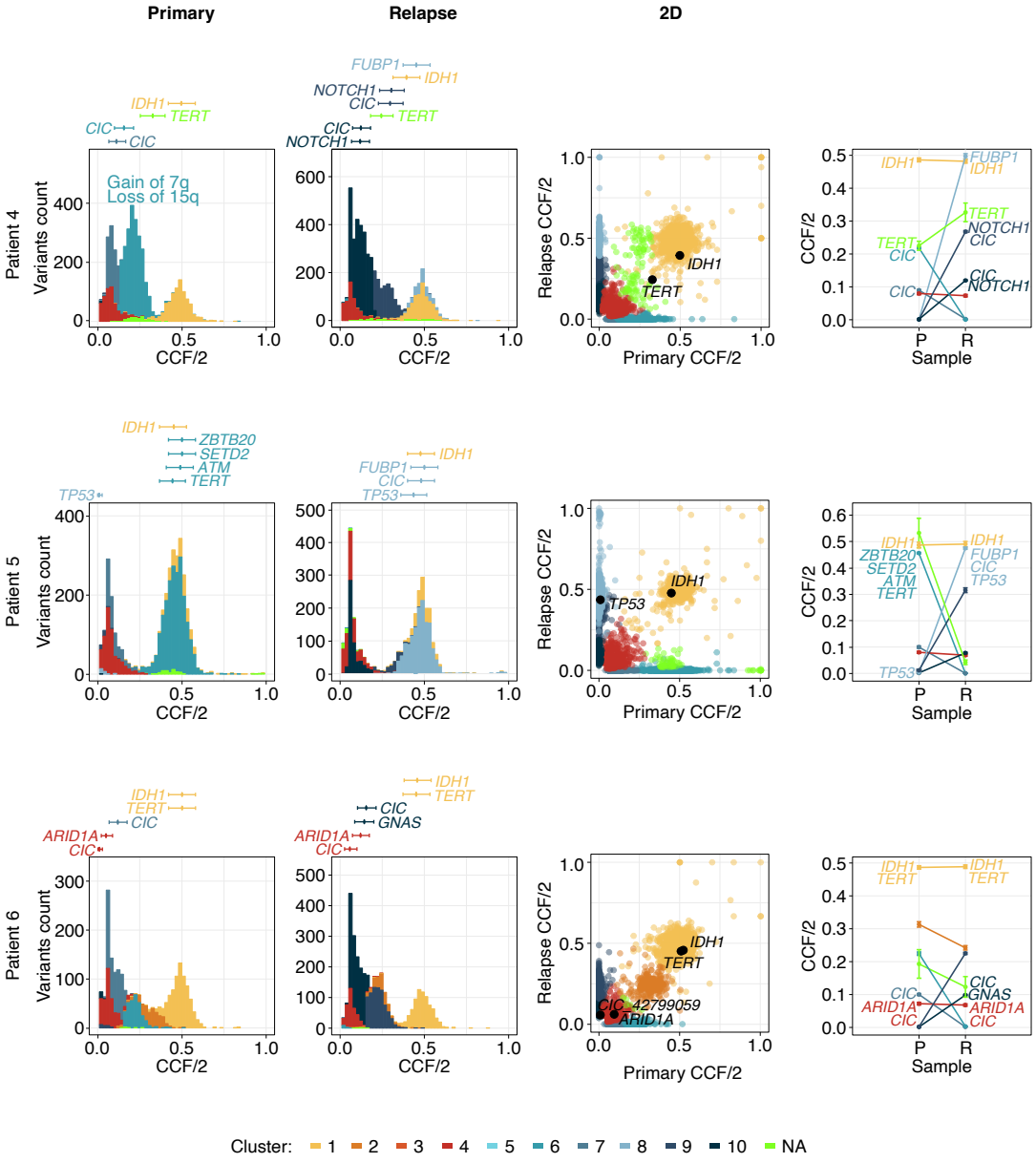


Figure 2.4.6: Pseudo-heterozygous VAF histograms and scatterplot of primary and relapsed tumors of Patients 4, 5 and 6. The plot on the most right shows how the frequencies of each cluster changes between primary and relapse samples.

2.5 Dynamics of tumor growth

As previously shown in Section 2.3.3, it is possible to construct a mathematical model to simulate the VAF distribution. By applying Bayesian inference, it is possible to use the VAF to quantitatively measure the evolutionary dynamics of tumor growth, thereby understanding growth and selection dynamics in primary oligodendrogliomas. Figure 2.5.1 shows the effectiveness of this method, as the model, in blue, fits the data, in pink, reasonably well. Hence, this fit allows to accurately estimate all the parameters of the model, such as selective advantage, growth rate and cancer age.

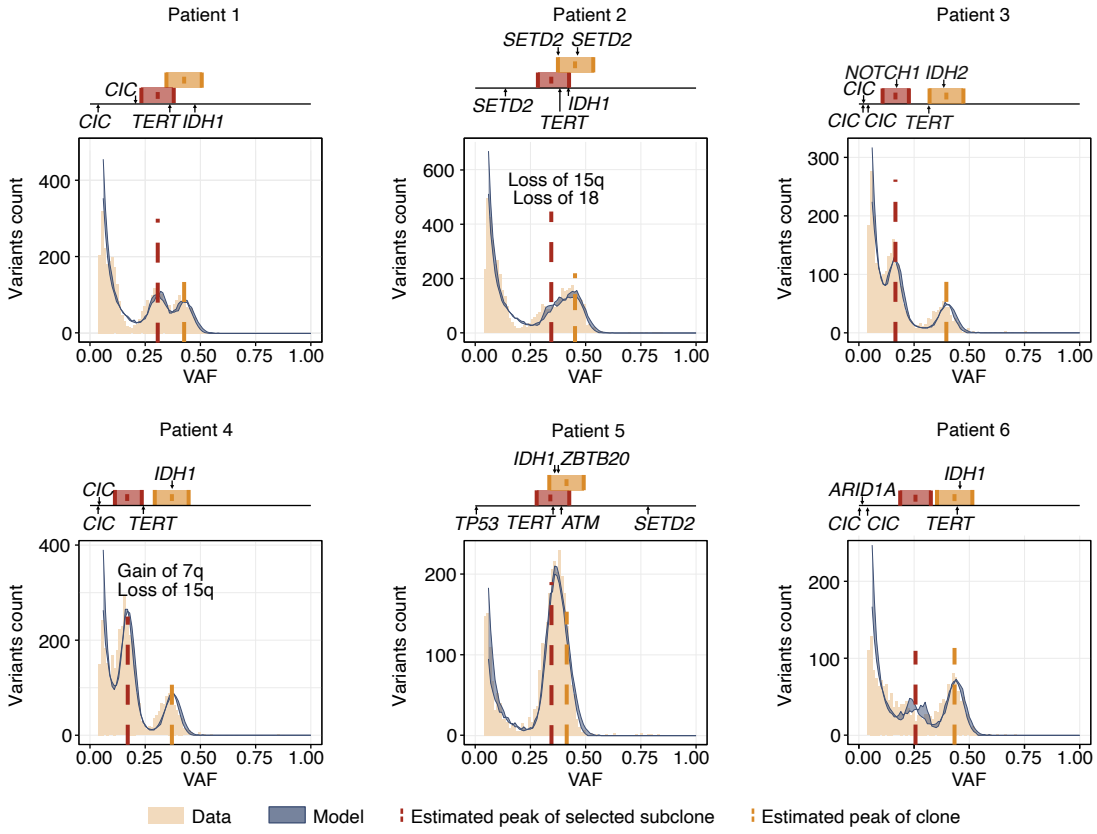


Figure 2.5.1: Model fits the data; on top of the VAF histograms driver mutations are highlighted.

Briefly, the model depends on accurately determining the number of clonal mutations to time the MRCA and using a low VAF threshold for subclonal mutations to resolve the tumor's evolution. Hence, given that the model heavily relies on the number of variants for the inference and quantification of the tumor's parameters, it necessitates data with a relatively high sequencing depth (at least 50x). For example, an attempt to apply the model to publicly available whole-genome sequencing data of ten oligodendrogliomas from the

GLASS consortium [15] showed that, for the majority of samples sequenced at depths ranging from 10x to 25x, the VAF histogram was too poorly defined to allow for a precise analysis. However, the model successfully fitted two datasets sequenced at approximately 55x (Figure 2.5.2).

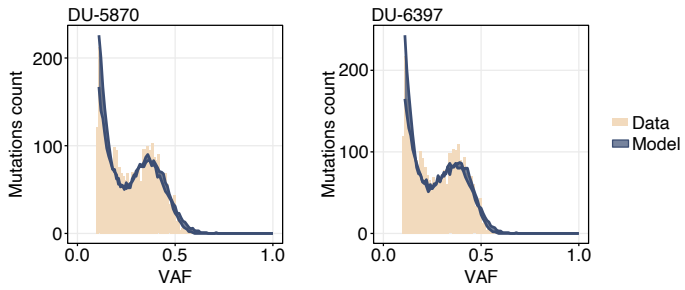


Figure 2.5.2: The model (in blue) fits the VAF histograms of the two samples taken from [15]. On the left Patient TCGA-DU-5870, with sequencing depth of 52x, and on the right Patient TCGA-DU-6397, with sequencing depth of 58x.

2.5.1 Subclonal selection of diverse driver mutations

The first significant finding from the modeling is that in all six primary cases, the model identifies a selected subclone, indicated by the dotted vertical red line in Figure 2.5.1. This suggests the presence of a cell population with a selective advantage over other tumor cells, independent of prior constraints. The size of the selected subclone varies between 42% and 86% (see Figure 2.5.9A). In five of the six cases, driver gene mutations were found at frequencies consistent with the selected clone, indicating that these mutations contributed to the subclonal outgrowth. In the sixth case, no apparent driver was identified for subclone outgrowth. This lack of enrichment for specific drivers in the Patient 6 suggests that oligodendroglioma evolution during tumor progression may involve a variety of oncogenes (maybe even epigenetic, which I will discuss in Chapter 3), indicating that subclonal selection might also be driven by yet unknown factors.

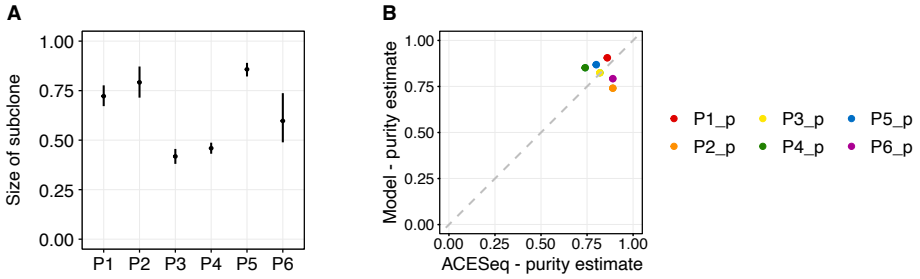


Figure 2.5.3: (A) Mean, with min and max values, of the estimated size of the leading subclone. (B) Comparison of purity estimates from ACESeq and from my model.

I then explored whether the strength of subclonal selection, s, varied among different driver mutations. The results showed significant variation, from mild subclonal selection ($s = 0.11$, potentially associated with a *SETD2* mutation in Patient 2) to rapid expansion ($s = 0.72$, potentially associated with mutations in *ZBTB20*, *SETD2*, *ATM* and a *TERT* promoter mutation in Patient 5), as illustrated in Figure 2.5.4 on the left panel. Interestingly, the largest selective advantages ($s > 0.3$) were inferred in Patient 4 and Patient 5, who exhibited selection for a subclonal *TERT* promoter mutation (in both patients) and/or gain of Chromosome 7q (in Patient 4 only). This suggests that the *TERT* promoter mutation and/or 7q gain may confer a stronger selective advantage compared to other drivers. To further support this observation, I analyzed the normalized ratio of non-synonymous to synonymous substitutions, dN/dS, which is an independent measure of selection based on the assumption that most synonymous mutations are selectively neutral [99, 163, 168, 178]. The results revealed higher dN/dS ratios, indicative of positive selection, in patients with high inferred selective advantages, thus corroborating my inference approach (Figure 2.5.4, panel in the middle).
In sum, both dN/dS ratios and the VAF histograms in primary oligodendrogliomas suggest marked variation in the dynamics of subclonal evolution and strong subclonal selection for *TERT* promoter mutations, which causes an increase in telomerase activity and telomere elongation.

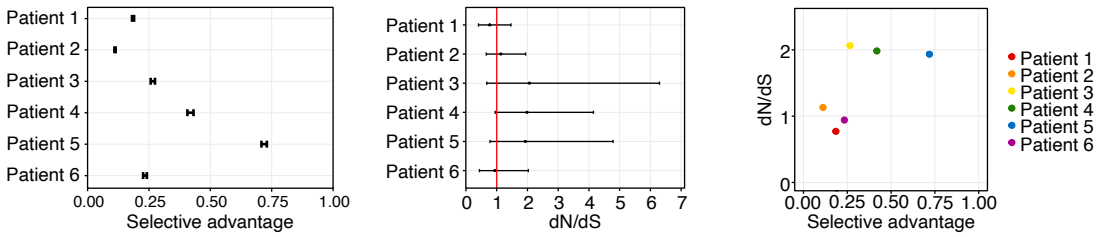


Figure 2.5.4: Estimates of selective advantage, dN/dS ratio of the six primary cases and how they relate to each other.

Importance of *TERT* promoter Telomere length maintenance is crucial for genome stability. In fact, mutations in the *TERT* promoter, which are absent in normal cells, disrupt this stability by enhancing cell proliferation and cancer progression [56]. Specifically, these mutations lead to increased *TERT* expression, necessary for telomerase activation and telomere maintenance, preventing cell death [117]. Such mutations are frequently found in various tumors of the central nervous system [50, 66, 11], including both the most aggressive form of diffuse glioma, astrocytoma, and the least aggressive form, oligodendroglioma, and play a significant role in glioblastoma [117, 73, 137].

Since I have identified a pattern that distinctly separates the six primary cases based on their selective advantage estimates, and given that YAPSA allows for the analysis of mutational signatures using stratified data, I evaluated the relative contribution of signatures from SNVs grouped into clone, subclone or tail signatures (Figure 2.5.5). I initially hypothesized that certain signatures would be predominantly associated with specific clusters. However, the data did not support this, as the relative distribution of signatures was consistent across all three different clusters.

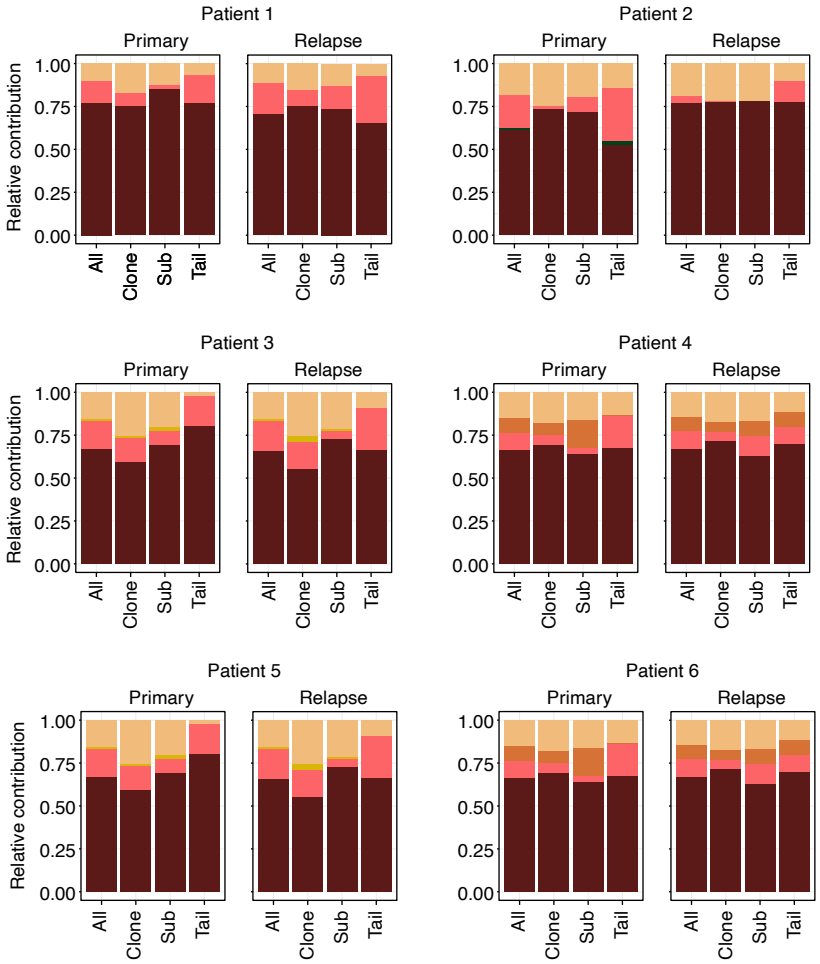


Figure 2.5.5: The relative signature contribution of SNVs to clone, subclone and tail for each tumor sample.

Other aspects that can be derived from the estimated parameters are the tumor growth rates. Due to the inferred intratumoral heterogeneity, the growth rate of the entire tumor mass varies with the relative size of the selected subclone, changing dynamically as the subclone expands. Initially, before the selected subclone emerges, the tumor grows at a

constant rate of $b-d$. As the subclone begins to proliferate, the overall growth rate of the tumor increases, eventually reaching $b(1+s)-d$ when the selected subclone becomes the predominant component of the tumor mass. By the time of primary tumor resection, I estimated that the tumor's annual growth rate increased by a factor of between 1.8 and 3.5, which corresponds to a doubling time of approximately 0.6 to 1.1 years, as shown in Figure 2.5.6.

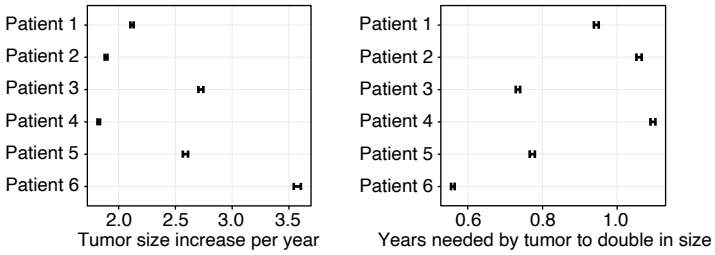


Figure 2.5.6: Estimated tumor size increase per year and doubling time for each patient.

The doubling time is obtained by solving the following equation with respect to Δt (all the other parameters are outputs from the model):

$$2 = e^{(b-d)\Delta t} + e^{(b(1+s)-d)(1-t_{\text{sel}})\Delta t} - e^{(b-d)(1-t_{\text{sel}})\Delta t}.$$

The estimated tumor increase per year is then obtained by a simple ratio $2/\Delta t$, where Δt is the previously derived doubling time.

2.5.2 Decade-long tumor growth

Lastly, I examined the evolutionary history of the tumors in each of the six patients individually. The population-genetics model consistently suggested that tumor initiation occurred in early childhood, between 1 and 9 years of age, averaging at 4.7 years. Given the patients' ages at diagnosis ranging from 24 to 48 years (detailed in Table 2.1), this implies a decade-long tumor growth before diagnosis (Figure 2.5.7). In four of the six patients (Patients 1, 2, 3 and 6), the selected subclone emerged 4.7-8 years after tumor initiation, indicating early-life subclonal selection followed by prolonged expansion. Conversely, in Patient 4 and Patient 5, the selected subclones appeared much later, 18.4 and 20.1 years after tumor initiation, respectively. These subclones also carried a *TERT* promoter mutation, with Patient 4 additionally showing a gain of chromosome 7q and a loss of chromosome 15q.

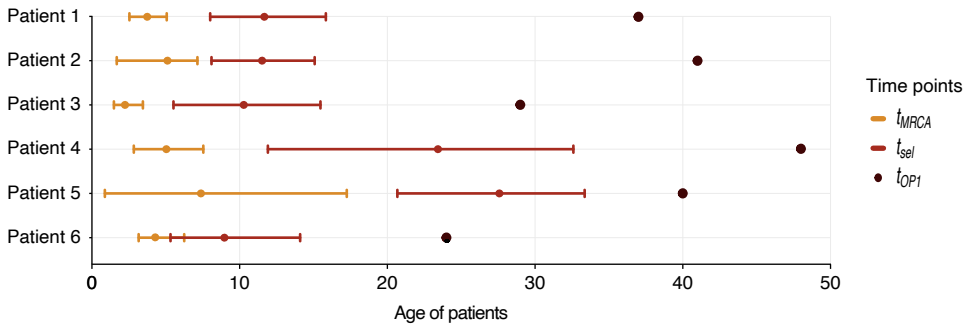


Figure 2.5.7: Time estimates of the six primary cases.

Notably, for the two oligodendroglioma cases from the GLASS study [15] with sufficient sequencing depth (around 55x) and purity (>0.5), the model estimated an early origin (from 0 to 5.3 years, mean: 1.6 years) and a decade-long evolution as well (see Figure 2.5.8B). Moreover, a relatively small selective advantage was inferred (see Figure 2.5.8A).

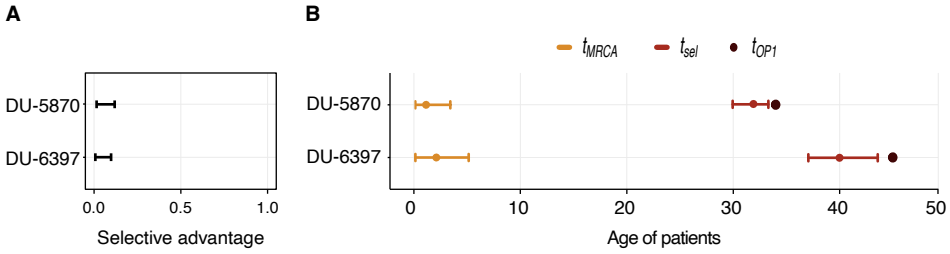


Figure 2.5.8: Selective advantage and time estimates of the two primary cases from [15].

These findings, combined with earlier results discussed in 2.5.1, confirm the model's reliability. For Patients 1, 2, 3 and 6, the subclones formed early, as they have a weaker selective advantage. In contrast, the subclones in Patient 4 and Patient 5, associated with higher selective advantage and *TERT* promoter mutations, emerged later. This implies a significant selective pressure to acquire telomere maintenance mechanisms, enabling subclones with *TERT* promoter mutations to achieve larger sizes despite their later emergence. Conversely, other driver mutations experience less selective pressure and can only reach substantial sizes if they are acquired early in the tumor's development. This suggests that oligodendrogliomas grow inefficiently and face challenges related to cellular senescence. To support this hypothesis I identify the effective mutation rate

$$\mu_{\text{eff}} = \frac{\mu}{(1 - d/b)},$$

defined as the number of variants acquired per effective symmetric self-renewing cell division (the denominator represents the fraction of cell divisions contributing to tumor

growth). The high μ_{eff} values (150 to 434 variants per effective division, Figure 2.5.9 on the left) proves an inefficient tumor growth, as these rates exceed typical estimates of 1 to 10 variants per somatic cell division [12, 72, 81, 106]. However, such a markedly elevated mutation rate is unlikely given the similarity of the mutational profiles in my cohort to normal tissues (Figure 2.4.3). Alternatively, the high effective mutation rate observed in my cohort may report on inefficient expansion of oligodendrogliomas, where only a minority of cell divisions sustain actual tumor growth, in agreement with strong selection for telomere stabilization.
It is also possible to derive the mutation rate of the mutation driving the selected advantage, which is defined as the inverse of the number of divisions until t_{sel}

$$\mu_{driver} = \frac{1}{\int_0^{t_{\text{sel}}} \mu b N_{\text{N}}(t) dt} = \frac{1}{\mu b \int_0^{t_{\text{sel}}} e^{(b-d)t} dt} = \frac{1}{\mu} \frac{b-d}{e^{(b-d)t_{\text{sel}}}}.$$

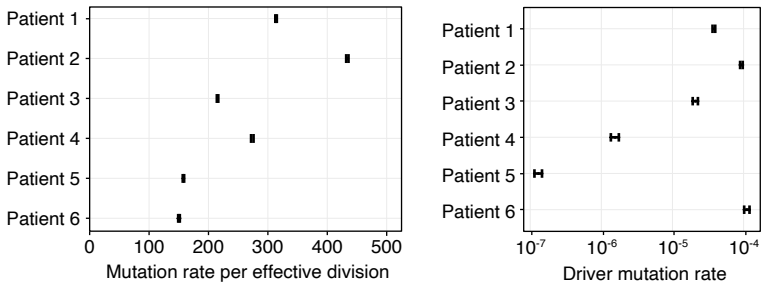


Figure 2.5.9: Mutation rate estimates of the six primary cases.

Taken together, the VAF histograms in primary oligodendrogliomas suggest inefficient tumor growth, driven by strong selection for telomere maintenance and ongoing clonal evolution during tumor progression.

2.6 Growth rate of the primary tumor predicts time to recurrence

Up to this point, my analysis has solely focused on the primary tumor samples, aiming to reconstruct the tumor's evolutionary trajectory up to the day of the first surgery. However, as discussed in Section 1.4.3, recurrence is unfortunately inevitable for oligodendrogliomas. I will use the relapse data to validate the growth dynamics inferred from the primary tumor samples. Specifically, the evolutionary parameters estimated from the primary tumors will be employed to predict the time of tumor regrowth after the initial surgery.

Assuming the primary tumor reached a diagnosable size of 10^9 cells, as described in [33], I consider three possible degrees of resection at t_{surgery}, which the clinicians call partial, subtotal and total. To simulate these resection scenarios, which the clinicians perform based on a case-by-case decision, I associate specific resection percentages to each (as can be seen in Figure 2.6.1). Specifically, for a total resection, I assume that 99-99.99% of the primary tumor cells was removed, indicating that a percentage between 0.01% and 1% of tumor cells were left. For a subtotal incision, about 10% of the primary cells survive, corresponding to 90% resection. Lastly, for a partial resection, around 20% of tumor cells subsist, suggesting the removal of $8 * 10^8$ cells. After the surgery, for each scenario, I estimate the expected time spans for relapse, assuming that the tumor growth dynamics remain unchanged. This implies that the recurrent tumor grows back with the same net growth rate, mutation accumulation rate and selective advantage, ultimately reaching the same size of approximately 10^9 cells. Once I have these three different time span estimates between surgeries, namely $\Delta t_{\text{surgeries, partial}}$, $\Delta t_{\text{surgeries, subtotal}}$ and $\Delta t_{\text{surgeries, total}}$, I can compare them with the known relapse times ($\Delta t_{\text{surgeries, years, known}}$) and derive an appropriate resection percentage for each of the three different resection types. This will allow me to associate a specific percentage of resected cells with partial, subtotal or total surgery.

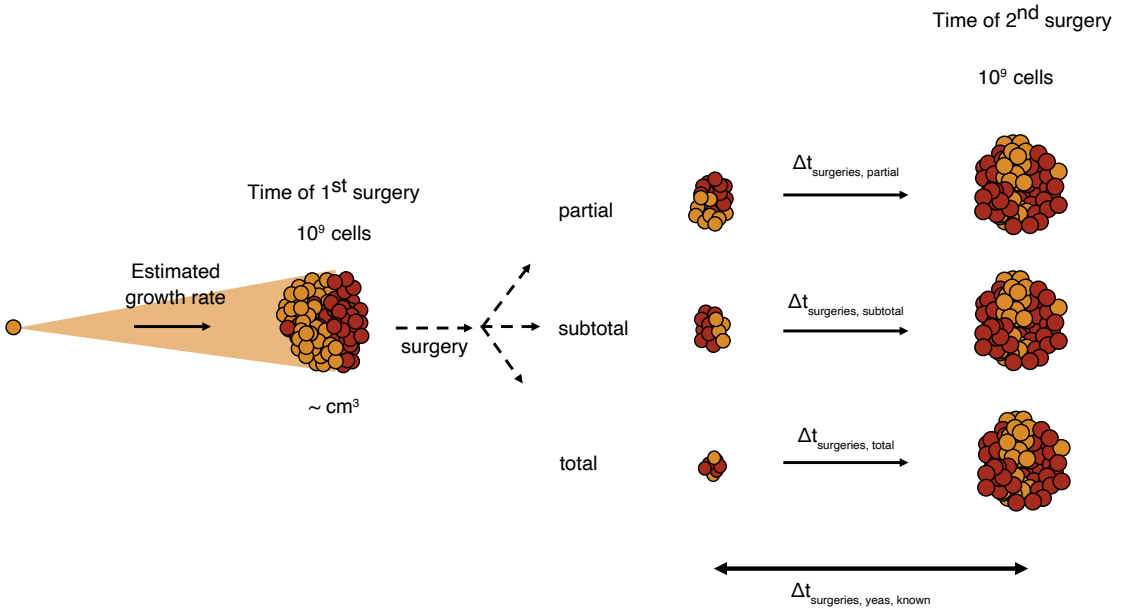
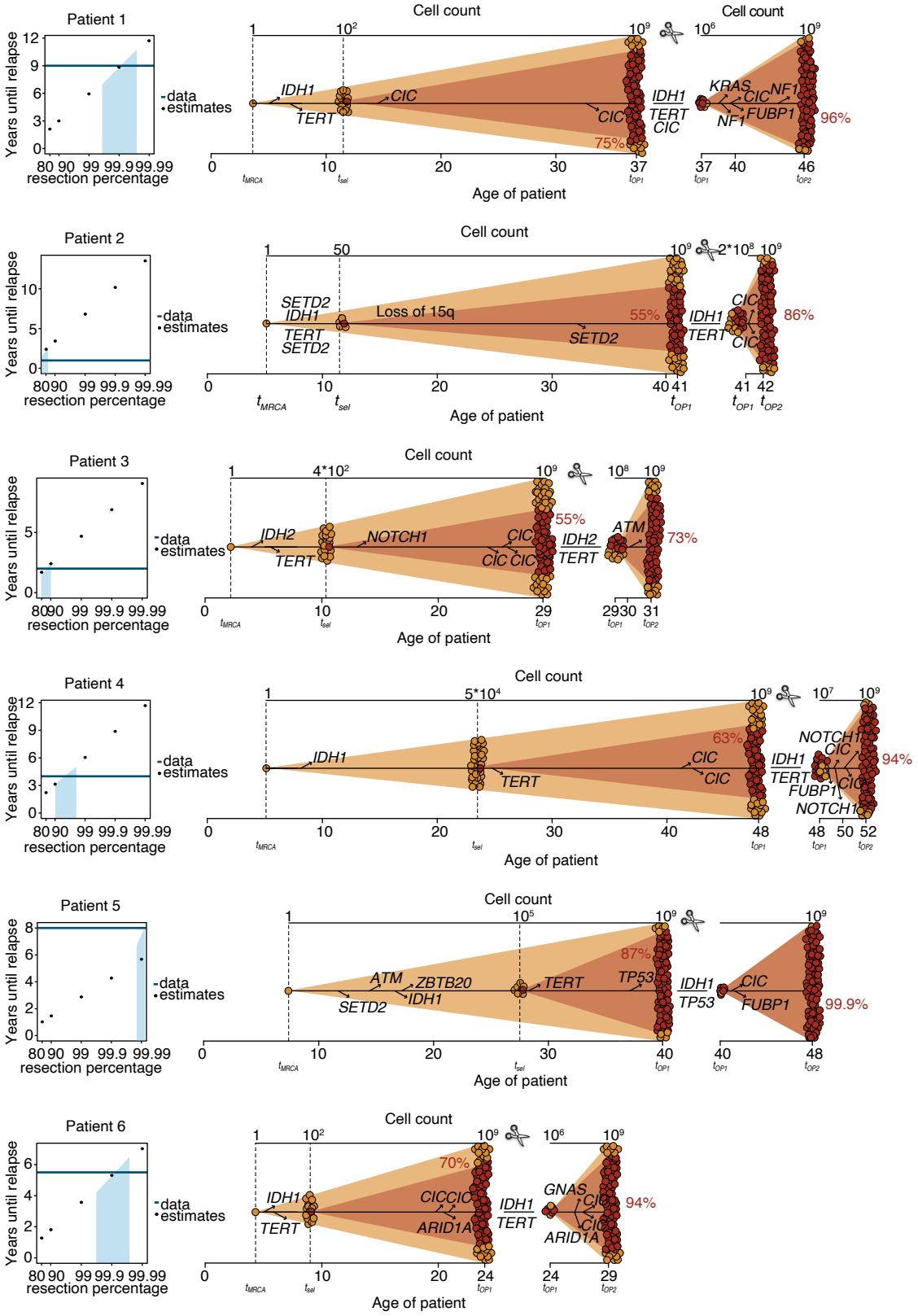


Figure 2.6.1: How I use information from the relapse sample to validate the growth dynamics of the overall tumor.

When comparing the predicted regrowth times with the reported times until tumor progression for different resection types, the cohort fell into two distinct groups (Figure 2.6.2 on the left). The first group, which includes Patients 1, 5 and 6, showed that the reported progression times were best explained by a nearly complete resection of more than 99% of the primary tumor. For Patients 1 and 6, surgical reports confirmed "total resection", aligning with the model's predictions. The second group, consisting of Patients 2, 3 and 4, had reported progression times that were best explained by an incomplete resection of 80-90% of the tumor mass. Surgical reports for Patient 3 and Patient 4 indicated "partial" and "subtotal" resections, respectively, again confirming the model's predictions. Thus, the growth dynamics inferred from primary resections provided reliable predictions for tumor recurrence following both gross and subtotal surgical resections. Remarkably, the dynamic predictions were accurate even though some driver mutations were lost and new ones emerged after the initial surgery and after therapeutic treatment (Figure 2.6.2 on the right). This observation suggests that many subclonal drivers (e.g., *SETD2*, *CIC*) are only weakly selected and do not significantly alter the overall growth dynamics of the tumor. Alternatively, distinct subclones may have similar fitness but occupy different regions, resulting in comparable overall tumor growth dynamics.
In summary, data from matched relapse samples validate the growth dynamics inferred from primary tumor samples. For four patients with available surgical information, the inferred tumor growth rates from the primary tumor accurately predicted the tumor's regrowth after initial surgery.

Figure 2.6.2: (Next page.) Left panels: predicted time till tumor re-growth to 10^9 cells in all six patients, simulating variable extent of tumor resection and using the growth rate estimated from the population genetics model. Surgical reports classified the resection in these tumors as, respectively, total, NA, partial, subtotal, NA and total. Right panels: evolutionary plots illustrating tumor growth (measured in the number of cells), the percentage of cells positively selected at the time of both surgeries and the presence of private and shared driver mutations.


Patient 1
Years until relapse
resection percentage
data
estimates
Cell count
IDH1
TERT
CIC
75%
KRAS
NF1
FUBP1
96%
Age of patient
Patient 2
SETD2
Loss of 15q
55%
86%
Patient 3
IDH2
NOTCH1
ATM
73%
Patient 4
63%
94%
Patient 5
ATM
ZBTB20
SETD2
87%
TP53
99.9%
Patient 6
70%
GNAS
94%

2.7 Discussion

Recent advancements in genomics have provided deeper insights into the genetic factors involved in oligodendroglioma development, revealing complex genetic mosaics defining these tumors [173, 4, 158]. This chapter focuses on broadening the understanding about oligodendrogliomas' genomic evolution by estimating when significant events, such as their origin and the acquisition of mutations that accelerate tumor growth, occur. In addition to outline the tumor's development history, I also aim to predict its future behavior, specifically tumor recurrence, to aid a more accurate treatment planning.

To understand the clonal dynamics driving oligodendroglioma growth and the timing the acquisition of key driver mutations, I developed a population genetics model. This model simulates clonal selection and the accumulation of neutral mutations in an expanding cell mass, based on deep whole-genome sequencing data, as detailed in Section 2.3.3. I applied this model to the primary samples of the six oligodendroglioma patients. My findings indicate that mutations in *IDH1* and *IDH2*, along with 1p/19q codeletions, are clonal events occurring early in tumor evolution, consistent with previous research [162, 25].

Additional alterations in tumor suppressor genes contribute to the tumors' genetic complexity. Notably, *TERT* promoter mutations were found in all primary tumors and were subclonal in two cases. Previous research suggests that in *IDH*-wildtype glioblastomas, *TERT* promoter mutations promote aggressive tumor growth by preventing cancer cell death, rather than serving as founding mutations [73]. My findings suggest a similar role in oligodendrogliomas, where the frequency of *TERT* promoter mutations correlates with tumor aggressiveness. In Section 2.5.1, I discuss the strong selective advantage of *TERT* promoter-mutated subclones, highlighting their role in enhancing cell proliferation [50]. Therefore, *TERT* promoter mutations might serve as both diagnostic and prognostic markers for oligodendrogliomas. Moreover, quantitative modeling of leading subclones in primary tumors revealed positive selection of known driver mutations such as *NOTCH1*, *SETD2*, *CIC*, *ZBTB20* and *ATM*. Interestingly, in Patient 6, clonal selection occurred without detectable variants in known oncogenic drivers. This suggests that oligodendroglioma growth might involve mutations in unidentified driver genes or epigenetic changes, which will be explored in the next chapter.

Unlike cancers like colorectal carcinoma, whose incidence rises steadily with age, oligodendroglioma diagnoses peak between 36 and 40 years and then decline. My birthdating study (Section 2.5), which combines mutation counts, tumor size and patient age with population dynamics models, shows that in all six primary-relapse pairs studied and two additional cases from a previous study [15], the tumor's founding cell originated in childhood, typically before age ten, indicating a path to early tumorigenesis. Interestingly, in the first ten years of life, human oligodendrocyte populations expand and then stabilize, as described in 1.4.2, suggesting that early childhood is a period of vulnerability for develop-

ing oligodendrogliomas.
Lastly, in Section 2.6, I found that the growth dynamics inferred from initially resected samples, together with the extent of surgical resection, could accurately predict the time to progression, despite changes in subclonal composition. These prediction, with further validation and assessment of therapy's impact on the prediction, could be useful improve treatment strategies.
In summary, in this chapter I demonstrated how genomic profiling of temporally separated oligodendroglioma samples can reconstruct the evolutionary history and clonal architectures of individual tumors. By combining deep whole-genome sequencing data with mathematical modeling, I reconstructed a common early tumorigenesis path characterized by 1p/19q chromosomal losses and *IDH1* or *IDH2* mutations [9]. This path, peaking during active glial development, is followed by the acquisition of *TERT* promoter mutations, which enhance telomerase activity and eventually dominate clonally. Understanding that oligodendrogliomas may originate in early childhood and peak in midlife has profound implications for brain tumor biology and early intervention strategies. Investigating these tumors' early origins could lead to novel detection and treatment designs. Given the significance of *TERT* promoter mutations, telomerase-targeted therapy might also be effective, as suggested in [11].

Chapter 3

DNA methylation patterns in oligodendrogliomas

Deregulation of cellular functions in cancer involves both genetic and epigenetic alterations. Thus, there is a need for a comprehensive epigenetic analysis, alongside the genetic one, to fully understand cancer progression. This raises the question of whether epigenetic data, like genetic data (as shown in Chapter 2), can be used to trace the evolutionary pathways of tumors.
To address this question, I will analyze DNA methylation data from the same oligodendrogliomas previously studied, to determine if it can shed light on the events that shape the tumor's evolution. I will specifically investigate whether DNA-methylation factors play a role in the subclonal selection observed in the genetic analysis of oligodendrogliomas, as discussed in Chapter 2. For instance, in Patient 6, no known driver mutation was identified to explain the accelerated growth of the subclone. This suggests that a purely genetic perspective may be inadequate, and incorporating the heterogeneity and evolution of the epigenome could provide new insights.

3.1 Sequencing strategy

The methylation status of the six primary-recurrent paired oligodendroglioma samples was assessed using Illumina 850k arrays by Yonghe Wu (German Cancer Research Center, Heidelberg). The Infinium assay, conducted according to Illumina's standard protocol with the Infinium Human Methylation 850K BeadChip (`illumina.com`), measures DNA methylation levels at CpG sites. This assay uses two types of probes: a methylated (M) one, which binds to methylated DNA regions, and an unmethylated (U) one, which targets unmethylated CpG sites. By comparing the signal intensities of the methylated and unmethylated probes, the assay determines the methylation level at each CpG site,

expressed as the β-value:

$$\beta = \frac{M}{M+U}.$$

The β-value ranges from 0 to 1, where high values indicate high methylation levels and lower values indicate less or no methylation [4].

3.2 DNA methylation signatures confirm hallmarks of oligodendrogliomas

Genome-wide changes in the methylation status of CpG dinucleotides occur during cancer formation. As mentioned in the previous chapter in Section 1.3.2, some cancers exhibit unique methylomes, such as the CpG Island Methylator Phenotype (CIMP) [147]. The CIMP assumes that the methylation status of specific CpG sites in the promoter regions of various tumor suppressor genes can lead to unchecked cell proliferation and cancer progression. Moreover, the CIMP allows for a clear definition of distinct molecular tumor subtypes. These methylation-based subtypes have been identified in various cancers, including gliomas. Further methylation signatures can distinguish between different subtypes of diffuse adult gliomas, which I will describe in the subsequent sections. Specifically, in Section 3.2.1, I will focus on the primary branching point that distinguishes low-grade gliomas (*IDH*-mutant, shown on the left side of Figure 1.4.3) from high-grade gliomas (*IDH*-wildtype, shown on the right side of Figure 1.4.3), whereas in Section 3.2.2, I will discuss the branching point within *IDH*-mutant gliomas, which separates oligodendrogliomas from astrocytomas based on their 1p/19q-codeletion status.

3.2.1 Glioma CpG island methylator phenotype

As discussed in Section 1.3.2, genetic mutations in the *IDH* genes lead to a specific pattern of early epigenetic alterations known as glioma CIMP (G-CIMP), characterized by extensive remodeling of the DNA methylome [112, 103]. This methylation profile distinguishes low-grade gliomas, such as astrocytomas and oligodendrogliomas, which have a G-CIMP+ phenotype characterized by hypermethylation, from glioblastomas, which have a G-CIMP- phenotype characterized by hypomethylation. When applying the G-CIMP criterion described in 1.3.2 to the 12 tumor samples, I found that all primary samples are classified as G-CIMP+, and their status is maintained at recurrence, as shown in Figure 3.2.1 on the left. This result aligns with expectations, given that all gliomas profiled here are *IDH*-mutant and thus exhibit the characteristic hypermethylation patterns associated with the G-CIMP+ status.

3.2.2 Methylation Classifier identifies 1p/19q-codeletion status

To further classify *IDH*-mutated low-grade gliomas, Paul et al. [120] identified 14 specific CpG positions (shown in Table 1.2) that can distinguish 1p/19q-codeleted from 1p/19q-non-codeleted gliomas. The former ones are oligodendrogliomas, while the latter ones are astrocytomas. This analysis reinforces the classifier's ability to predict tumor types, confirming that the analyzed samples are oligodendrogliomas, as depicted in Figure 3.2.1 on the right.

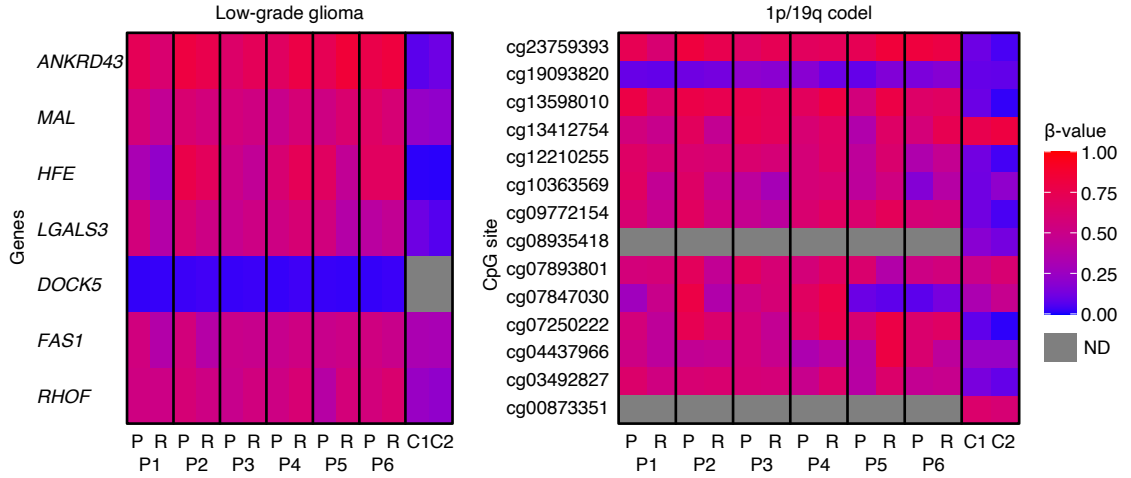


Figure 3.2.1: G-CIMP and 1p/19q-codeleted methylation status confirms that the samples are low grade gliomas with 1p/19q-codeleted, namely oligodendrogliomas. C1 and C2 are two healthy brain tissues downloaded from The Cancer Genome Atlas. ND: no data available.

3.3 β-value distribution

To compare the methylation levels between normal brain tissue and the oligodendroglioma samples, I downloaded methylation data of two healthy brain tissues from The Cancer Genome Atlas portal (`https://tcga-data.nci.nih.gov/`). After filtering out the probes targeting the X and Y chromosomes (as done in [4, 19]), I analysed the difference between methylation levels in healthy brains and oligodendrogliomas.
The standard representation of the methylation levels β is by a density plot. The density plots in Figure 3.3.1 reveal that, irrespective of the sample type (non-tumor or tumor), the majority of CpG sites exhibit either of the two extreme methylation levels. This implies a bimodal distribution of the methylation data, with most β-values clustered near 0 or 1, signifying CpG sites that are either unmethylated in most cells or methylated in most cells. Additionally, in the density plots for the tumor samples, an intermediate peak

at moderate methylation levels is evident, varying in visibility across different samples (the most evident one is in the primary sample of Patient 5). When comparing the DNA methylation profiles between primary and relapse tumor pairs, some patients (such as Patients 2, 3 and 6) show minimal differences, whereas others (Patients 1, 4 and 5) exhibit more substantial variations, implying different paths to recurrence in oligodendrogliomas. The variation between primary and relapse tumors does not correlate with the extent of surgical resection. This is evident in Patient 6, who underwent total resection, yet showed no noticeable change in the β-value distribution. Notably, Patient 1, the only patient to receive concurrent radiation therapy after resection, was also the only patient who exhibited a shift from highly to intermediately methylated sites as the tumor progressed.

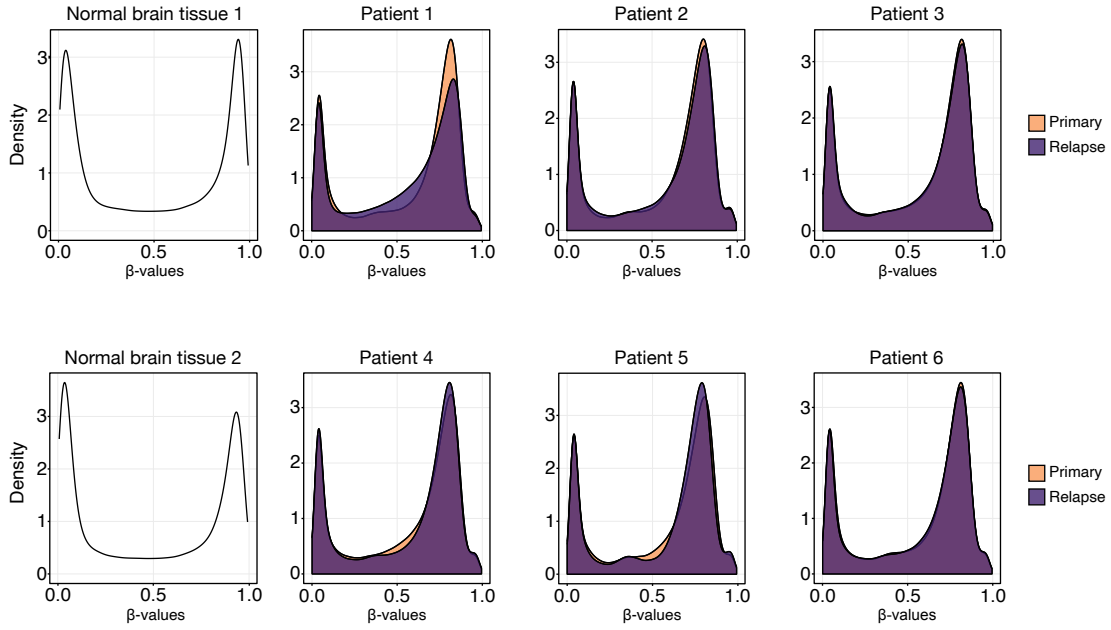


Figure 3.3.1: Density plots representing the distribution of the β-values in normal brain tissue (on the left; in white) and in the 12 oligodendroglioma samples (on the right).

3.3.1 Subclonal selection did not correlate with known hypermethylated tumor suppressor genes

The *CDKN2A/2B* genes on chromosome 9p21 encode the proteins $p16^{INK4a}$/$p14^{ARF}$ and $p15^{INK4b}$, respectively. These tumor suppressor genes are known to be aberrantly expressed in various human cancers [131, 83, 40, 39], including oligodendrogliomas. Specifically, hypermethylation of *CDKN2A/2B* has been identified as a crucial epigenetic mechanism that allows oligodendroglial tumors to evade growth control [170].
I investigated whether I could identify a methylated tumor suppressor in my data that might

help driving positively selected subclones progression. I observed partial hypermethylation of the *CDKN2A/2B* promoter-associated CpG island (chr9:21994101-21995910), as shown in Figure 3.3.3. No clear hypermethylation signal was detected in the *CDKN2A* promoter-associated CpG island (chr9:21974578-21975306), as depicted in Figure 3.3.2.

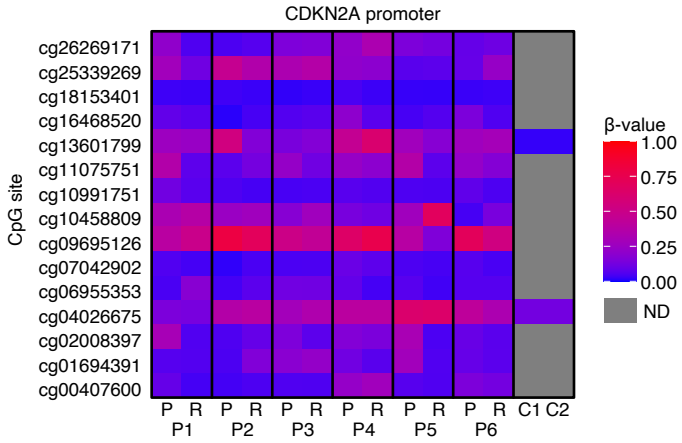


Figure 3.3.2: CDKN2A promoter-associated CpG positions. C1 and C2 are two healthy brain tissues downloaded from The Cancer Genome Atlas. ND: no data available.

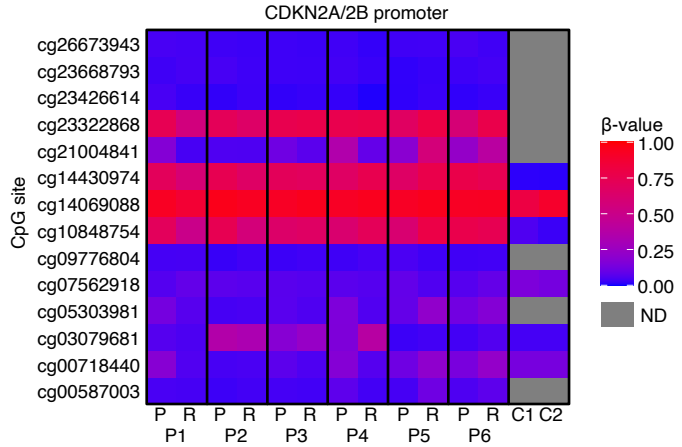


Figure 3.3.3: CDKN2A and CDKN2B promoter-associated CpG positions. C1 and C2 are two healthy brain tissues downloaded from The Cancer Genome Atlas. ND: no data available.

Of note, the available controls from normal brain tissue (taken from The Cancer Genome Atlas) did not cover all investigated CpG sites, which should be improved in the future. Additionally, I did not find any specific CpG site where the β-value correlated with the estimated selective advantage of the subclone. Furthermore, no clear correlation was found

between increased methylation levels of CpG sites in Patient 6, where subclonal selection did not correspond to a known driver, and positive selection.
In conclusion, partial hypermethylation of the *CDKN2A/2B* promoters may contribute to silencing of these genes in all oligodendrogliomas studied here, but does not provide a mechanism for subclonal outgrowth without a known genetic driver.

3.3.2 Comparing methylation levels between samples

To further investigate the differences in methylomes in tumor samples and identify common and specific methylation profiles, I selected the top 1% most variable CpG sites across the six initial and recurrent oligodendrogliomas. Using these sites, I performed two-way unsupervised hierarchical clustering with Euclidean distance and Ward linkage.
Hierarchical clustering grouped initial and recurrent samples from the same patient together (see Figure 3.3.4). This result indicates a higher similarity in DNA methylation profiles between paired samples from individual patients rather than between samples from different patients. Hence, primary and relapse samples exhibit minimal differences, indicating a relatively stable DNA methylation landscape during oligodendroglioma progression, consistent with the findings of [103, 4]. Notably, patients for whom the birth-death model based on genetic data estimated a higher selective advantage for the positively selected subclone (Patients 3, 4 and 5) formed a separate cluster from the others (on the left of Figure 3.3.4).

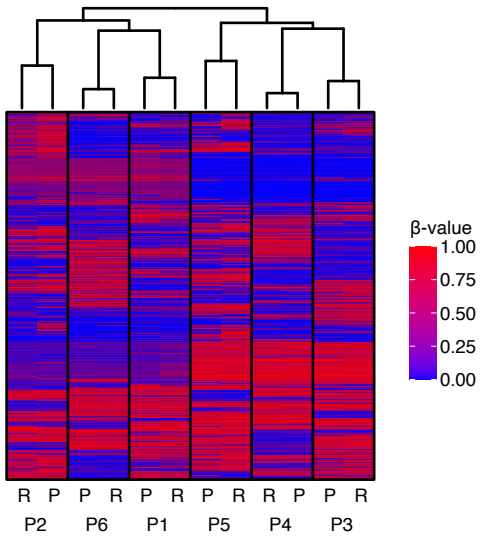


Figure 3.3.4: Unsupervised hierarchical clustering of the top 1% most variable CpG sites.

The same clustering results are also obtained when considering the top 0.5% and 2.5% most methylated CpG sites, as can be seen in Figure 3.3.5 (A) and (B), respectively. This

further supports epigenetic stability of oligodendrogliomas during progression (confirming results in Section 2.4), and points to a specific methylation pattern of tumors with fast growing subclones.

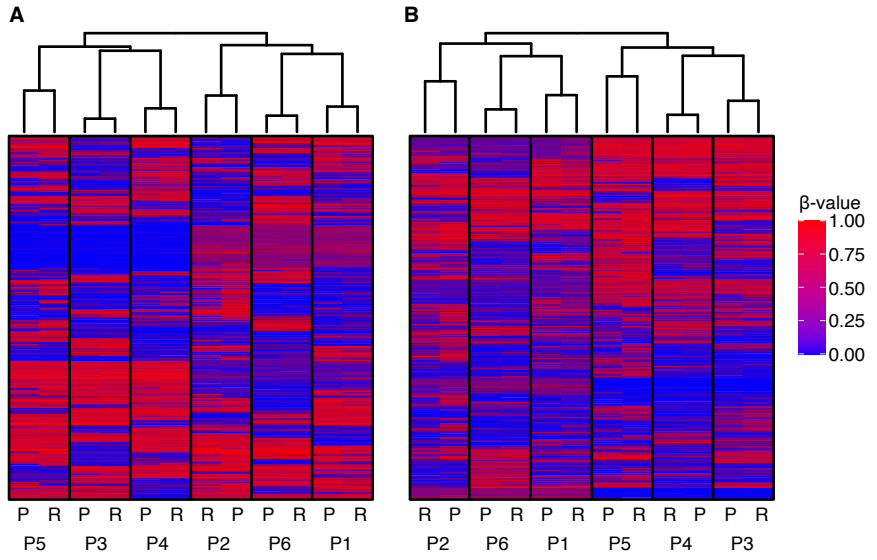


Figure 3.3.5: Unsupervised hierarchical clustering of the (A) top 0.5% and (B) top 2.5% most variable CpG sites.

3.4 Differentially methylated positions

Next I derive changes in methylation level between paired tumor samples of individual patients defined as

$$\Delta\beta = \beta_{\text{recurrent}} - \beta_{\text{primary}}.$$

This approach accounts for age-related DNA methylation differences, as outlined in [103]. Thus, the variations observed between initial and recurrent tumors are, at least in part, due to abnormal changes associated with tumor progression rather than originating from the cells themselves.

To identify CpG sites with significant methylation changes upon recurrence, I focused on the most differentially methylated positions (DMPs), defined as $|\Delta\beta| \geq 0.3$ [38]. I analyzed the number of shared CpG sites between samples of the same patient and identified how many of these sites were differentially methylated (Figure 3.4.1A). The number of shared positions was consistent across patients (mean: 865653). However, only a small fraction of these sites were DMPs (mean: 9120), ranging from 0.01% to 2.67%. I then examined whether, for each patient, the DMPs tended to show increased or decreased methylation levels from primary to recurrence, but no clear pattern emerged (Figure 3.4.1B).

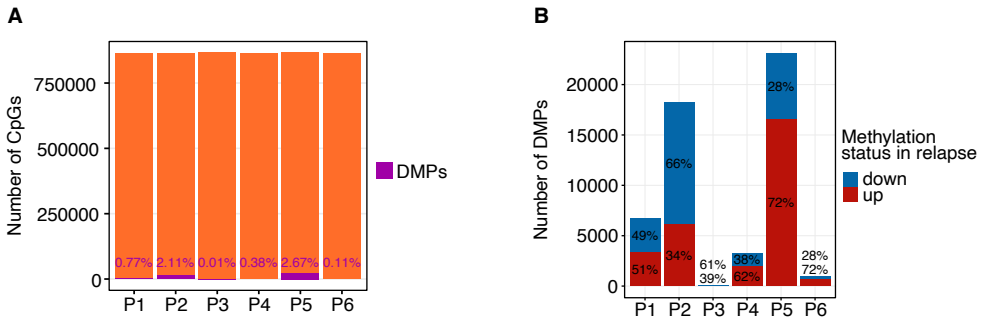


Figure 3.4.1: (A) The total number of CpG sites present in both samples for each patient. Highlighed in purple is the percentage of differentially methylated positions. (B) The number of differentially methylated positions between the primary tumor and recurrence for individual patients is displayed, with colors indicating whether they are down-regulated (in blue) or up-regulated (in red) in the recurrence sample.

Additionally, only a minority of DMPs were shared across patients: 5285 DMPs were present in two out of six patients, 388 DMPs in three out of six patients, 13 DMPs in four out of six patients and only 1 DMP was shared between five out of six patients (Figure 3.4.2A). Furthermore, there was no significant correlation between the number of differentially methylated positions found in a sample and the time until tumor recurrence (Figure 3.4.2B). Notably, no increase in DNA methylation alterations was observed in Patient 1, who received radiation therapy in addition to surgery, suggesting that radiation therapy may not induce focal changes in the methylome, on the contrary to findings by [38].

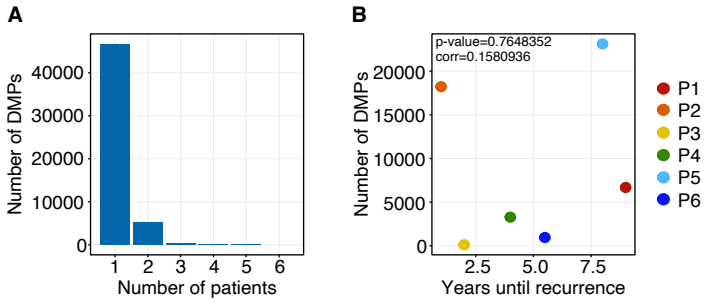


Figure 3.4.2: (A) Only a small number of differentially methylated positions were shared among the majority of patients. (B) The differentially methylated positions showed no correlation with the time between the primary tumor and its recurrence.

For the final analysis, I identified the most different methylated CpG positions using a more relaxed criterion, specifically assuming $|\Delta\beta| > 0.2$ (as done in [103]). I then compared the

number of DMPs between tumor samples classified as grade 2 and grade 3. There was a slightly higher number of hypermethylated DMPs in the grade 3 tumors that relapsed, as shown in Figure 3.4.3. Unfortunately, information on the grade of the relapse samples is not available.

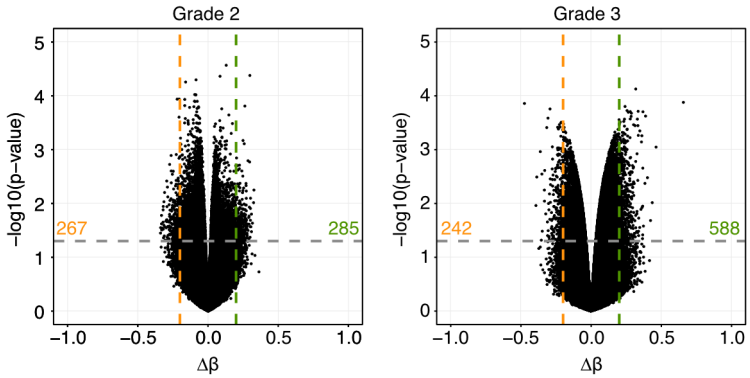


Figure 3.4.3: The average methylation change from initial diagnosis to recurrence at each CpG site was measured in grade 2 patients (left) or grade 3 patients (right). Colored lines represent CpG sites with significant hypomethylation (orange, $\Delta\beta < -0.2$, with total counts provided) or hypermethylation (green, $\Delta\beta > 0.2$, with total counts provided).

Lastly, to connect the genetic analysis from Chapter 2 with the DNA methylation data, I divided the patients into two groups based on their estimated selective advantage: high (Patients 4 and 5) and low (the other four patients). Patients with a low selective advantage had very few common highly differentially methylated sites (fewer than 100 DMPs with $|\Delta\beta| > 0.2$). In contrast, patients with a high selective advantage showed significant methylation changes when the cancer returned, with 1130 CpG sites showing strong hypomethylation ($\Delta\beta < -0.2$) and 3184 CpG sites showing strong hypermethylation ($\Delta\beta > 0.2$), as shown in Figure 3.4.4. When I performed a gene set enrichment analysis on all the differentially methylated positions, I found that they were linked to genes involved in biological processes related to cellular component organization and molecular functions associated with protein and enzyme binding. However, no known genes related to neurobiology were enriched.

Altogether, these results suggest that patients with a higher selective advantage experience a higher amount of differentially methylated changes but no specific gene pathway was enriched.

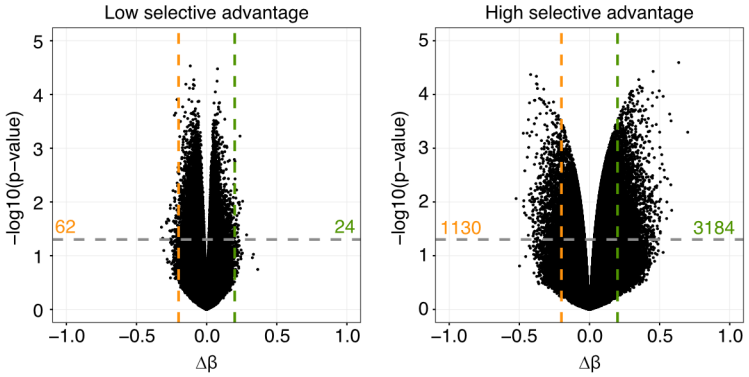


Figure 3.4.4: The average methylation change from initial diagnosis to recurrence at each CpG site was measured in patients estimated to have either a low selective advantage (left) and a high selective advantage (right). Colored lines represent CpG sites with significant hypomethylation (orange, $\Delta\beta < -0.2$, with total counts provided) or hypermethylation (green, $\Delta\beta > 0.2$, with total counts provided).

3.5 Discussion

DNA methylation is an epigenetic mechanism that changes dynamically throughout the human lifespan and aberrant methylation can lead to various diseases, including cancer. In this Chapter, I explored the DNA methylation changes occurring during the progression of adult-type diffuse oligodendroglioma, *IDH*-mutant and 1p/19q-codeleted.

Firstly, I demonstrated the effectiveness of classifying the 12 tumor samples based on specific DNA methylation signatures. The DNA methylation levels at certain CpG sites were able to confirm the diagnoses already made by clinicians and obtained from whole-genome sequencing analysis, affirming that the samples were indeed oligodendrogliomas. Specifically, this methylation analysis confirmed the G-CIMP status in all initial tumors, which was consistently maintained at recurrence. This indicates that these epigenetic changes occur early in tumor development. Considering the strong association between G-CIMP status and the *IDH* gene, and the importance of *IDH* mutations in the formation of oligodendrogliomas as shown in my genetic analysis 2.7, it can be concluded that G-CIMP status is potentially tumor-initiating, a conclusion also supported by [103].

Secondly, hierarchical clustering grouped all the tumor samples by patient identity, regardless of tumor grade, suggesting a higher similarity in DNA methylation profiles between paired samples from individual patients. This result indicates relatively stable DNA methylation over time, supporting the notion of epigenetic stability in oligodendrogliomas at recurrence, consistent with findings from [136, 4].

Lastly, in order to investigate changes in the methylome and identify potential methylation signatures linked to oligodendroglioma recurrence, I analysed the most differentially

methylated CpG positions between paired samples. The results showed that only a few of these positions were common across most patients, confirming findings in [38]. This outcome suggests that there is no common path of DNA methylation changes in recurrence. Furthermore, I did not detect an increase in DMPs across different World Health Organization tumor grades, implying that methylation changes may occur independently of tumor grade and that grade 2 oligodendrogliomas do not differ significantly from grade 3 ones in terms of methylation.
Lastly, the number of detected DMPs was linked to the estimated selective advantage from the computational analysis in Chapter 2. This connection emphasizes the interaction between genetic and epigenetic changes during oligodendroglioma evolution. However, no correlation between the estimated selective advantage and the methylation level in know oligodendroglial tumor suppressor genes could be found. Furthermore, for Patient 6, no potential epigenetic driver was identified that could explain the genetically inferred positive selection of the subclone.

Chapter 4

Conclusions

Resolving the evolutionary history of a tumor is clinically valuable, as prognosis depends on the future course of the evolutionary process and therapeutic response is mostly determined by the evolution of resistant subpopulations. In humans, the details of tumor evolution have remained largely uncharacterized, as longitudinal measurements are impractical (given that patients cannot ethically be biopsied at multiple time points during the progression of the disease) and studies are complicated by intratumoral heterogeneity and between-patient variation. However, recent studies have used mathematical approaches from population genetics to infer the evolutionary history from single time-point samples. This thesis further develops this approach by linking genetic evolution with tumor growth. This approach is possible, because intratumoral heterogeneity, measured in my case by deep whole-genome sequencing, serves a record of mutations that occurred throughout the history of the tumor. Consequently, it becomes feasible to infer the evolutionary parameters that have shaped the tumor's growth and, therefore, to reconstruct its past evolution.

In this thesis, I learn the past dynamics of oligodendroglioma growth, and use the so inferred evolutionary parameters to derive the expected regrowth of the tumor after an initial surgery. Among genetic and chromosomal alterations, *IDH1/2* mutation and 1p/19q-codeletion were present in all tumor cells and are considered to be part of the trunk of the evolutionary tree describing oligodendrogliomas. In addition, the presence of other mutations, first of all the *TERT* promoter mutation, were also frequently observed, appearing to be later events in oligodendroglioma oncogenesis. For the *TERT* promoter mutation especially, I showed how its frequency within the tumor can vary temporally, indicating different expansion rates of the positively inferred selected subclone.

The population genetics model of tumor initiation and progression uniformly dated the origin of oligodendrogliomas in childhood or puberty, when glial development is most active. Thus, the active oligodendrocyte progenitors in early childhood create a time-limited vulnerability for developing this tumor. A confined origin in early life followed

by decade-long tumor growth is consistent with the peak incidence of *IDH*-mutant and 1p/19q-codeleted oligodendrogliomas in midlife. Importantly, this generates a window-of-opportunity for early detection.
Moreover, using the estimates growth rates of primary tumors and the extend of resection at the surgery, I managed to successfully predicting when the recurrent tumor would grow back. I also examined epigenetic data to further analyze the progression of recurrent oligodendrogliomas, assessing whether new epigenetic drivers might challenge the assumption that the growth dynamics of the recurrent tumor mirror those of the primary tumor. No significant epigenetic changes were observed between the first and second surgeries, reinforcing the overall genetic and epigenetic stability of recurrent oligodendrogliomas. Furthermore, I investigated epigenetically driven variants to explain the genetically inferred selective advantage for Patient 6, but I did not find any DNA methylation level pattern that could explain this.
Interestingly, for Patient 1, neither the genetic evolutionary analysis nor the epigenetic progression analysis revealed any significant impact from the administered radiation therapy, aside from an increase in the mutational load.
In summary, this thesis demonstrates that oligodendrogliomas rarely acquire additional driver mutations or genomic/epigenetic aberrations upon recurrence, indicating their genetic and epigenetic stability at recurrence. This stability may contribute to the clinically slow-growing nature of oligodendrogliomas compared to other diffuse gliomas and could inform future tailored therapeutic strategies for this tumor.

Acknowledgements

I thank my Supervisor, Thomas Höfer, for being an exceptional mentor and for proving that when choosing a job you are passionate about, you will never have to work a day in your life. You were understanding and patient with me throughout my learning curve, always offering heartfelt advice from your own experiences, making me feel valued and included. I truly enjoyed our conversations, and it was especially gratifying when I could relate to what you meant from my own journey. Working with you was a pleasure, as I learned something valuable from you every day, both professionally and personally. I hope you enjoyed teaching me about cancer biology as much as I enjoyed offering a few suggestions for your swimming.

I thank Verena for helping me enormously to get significant results during my PhD and for showing me what a strong work ethic looks like. Having you next door at the beginning of this journey was an incredible advantage (dare I saying selective?), which I will never take for granted.

I thank the entire Höfer group for creating such an inclusive, pleasant and stimulating work environment. I truly enjoyed every single one of our lunch breaks at the CASINO, always starting rigorously at noon, and our group retreats, which allowed us to get to know each other on a more personal level. I thank Ale for our chats in italian and for your quick wit, which lifted my mood. I thank Maurice for being in the office next door the few times I worked late. Your presence made me feel less alone. I thank Tamar for always having some interesting history and arts fact ready to share. I thank Nils for your constructive and very detailed input on work and leisure topics alike, and for making apartment hunting in Heidelberg much easier.

I thank my office mates for sharing so much time with me and making the workdays fly by. I thank Marcel for the great conversations about holiday destinations and food. But I mostly thank Patrick and Franziska for always being there to listen to my concerns, both work and, mostly, life related. I thank Franziska for sharing the room (and the shampoo!) with me at the group retreats and for always being there to pull Patrick's leg with me. And I thank Patrick for laughing with us and for being our office's IT guy, taking everyone's problems at heart. I will never forget the countless "Who wants a tea?" (always black, because I do not like the green one) at around 3 pm, even in boiling hot summer, when we

would end up discussing serious personal matters. You all made this PhD journey so much more enjoyable.
I thank my second Supervisor, Pei-Chi Wei. You showed me the immense value of having a collaborative partner who is approachable, available and just a floor away. I appreciate the support from your group members with experiments and the engaging discussions we shared. Special thanks to Giulia and Lorenzo, who created their own little Italy in TP3 and welcomed me in it. The rich aroma of your coffee spreading through the whole third floor created a home away from home.
I thank Frank Lyko and Angela Goncalves for agreeing to be on my defense committee.
I thank Anna Marciniak-Czochra for being a member of my TAC committee.
I thank my Mom for always, no matter what, being there for me, silently supporting me whenever I needed it. I appreciate the space you gave me to think and come to my own conclusions, without shortcuts, because that is how life is. I thank my Dad for always having advice to share and for putting me first, even during the peak of the pandemic when you wanted to drive to Heidelberg but were stopped every few hundred meters by police checkpoints. And I thank Vera for spending your summer holidays in Heidelberg, even though you live in Pisa, 10 minutes from the seaside, just to spend time with me. Spending time together always reminds me of how important it is to have a sister who is also your best friend.
And lastly, I thank Satish for being the only one who could ease my stress while also causing me immense pain - funny how that works. You were my only confidant, the one who provided a much-needed escape, especially with all the traveling. You taught me not to take all in life so seriously and to see the world from a different perspective. But most importantly, I will always be grateful to you for helping me, even unconsciously, develop empathy and human skills.

Appendix A

List of publications

At the time of this thesis submission, the PhD research has resulted in an article titled "Oligodendrogliomas originate in early childhood, accelerate growth through clonal evolution and have predictable recurrence times" by Sarah Benedetto et al., which has been submitted for publication.
In addition to the work presented in this thesis, I also participated in studies investigating whole-genome sequencing data analysis of various cancer types. While the findings of these studies are not included here, they have resulted in the following publications:

- Milena Simovic, Michiel Bolkestein, Mahmoud Moustafa, John K L Wong, Verena Koerber, **Sarah Benedetto**, Umar Khalid, Hannah Sophia Schreiber, Manfred Jugold, Andrey Korshunov, Daniel H ubschmann, Norman Mack, Stephan Brons, Pei-Chi Wei, Michael O Breckwoldt, Sabine Heiland, Martin Bendszus, Juergen Debus, Thomas Hoefer, Marc Zapatka, Marcel Kool, Stefan M Pfister, Amir Abdollahi, and Aurelie Ernst. Carbon ion radio-therapy eradicates medulloblastomas with chromothripsis in an orthotopic Li-Fraumeni patient-derived mouse model. Neuro-Oncology, 23(12):2028–2041, 12 2021 [135]

- C. Pixberg, M. Zapatka, M. Hlevnjak, **S. Benedetto**, J.P. Suppelna, J. Heil, K. Smetanay, L. Michel, C. Fremd, V. Koerber, M. Ruebsam, L. Buschhorn, S. Heublein, B. Schaefgen, M. Golatta, C. Gomez, A. von Au, M. Wallwiener, S. Wolf, N. Dikow, C. Schaaf, E. Gutjahr, M. Allgaeuer, A. Stenzinger, K. Pfuetze, R. Kirsten, D. Huebschmann, H.-P. Sinn, D. Jaeger, A. Trumpp, R. Schlenk, T. Hoefer, V. Thewes, A. Schneeweiss, and P. Lichter. COGNITION: a prospective precision oncology trial for patients with early breast cancer at high risk following neoadjuvant chemotherapy. ESMO Open, 7(6):100637, 12 2022 [123]

Bibliography

[1] Lauri A. Aaltonen et al. “Pan-cancer analysis of whole genomes”. In: *Nature* 578.7793 (Feb. 2020), pp. 82–93. ISSN: 0028-0836. DOI: 10.1038/s41586-020-1969-6.

[2] Federico Abascal et al. “Somatic mutation landscapes at single-molecule resolution.” In: *Nature* 593.7859 (May 2021), pp. 405–410. ISSN: 1476-4687. DOI: 10.1038/s41586-021-03477-4.

[3] Navid Ahmadinejad, Shayna Troftgruben, Junwen Wang, Pramod B Chandrashekar, Valentin Dinu, Carlo Maley, and Li Liu. “Accurate Identification of Subclones in Tumor Genomes”. In: *Molecular Biology and Evolution* 39.7 (July 2022). ISSN: 0737-4038. DOI: 10.1093/molbev/msac136.

[4] Koki Aihara et al. “Genetic and epigenetic stability of oligodendrogliomas at recurrence”. In: *Acta Neuropathologica Communications* 5.1 (Dec. 2017), p. 18. ISSN: 2051-5960. DOI: 10.1186/s40478-017-0422-z.

[5] Ludmil B Alexandrov, Philip H Jones, David C Wedge, Julian E Sale, Peter J Campbell, Serena Nik-Zainal, and Michael R Stratton. “Clock-like mutational processes in human somatic cells”. In: *Nature Genetics* 47.12 (Dec. 2015), pp. 1402–1407. ISSN: 1061-4036. DOI: 10.1038/ng.3441.

[6] Ludmil B. Alexandrov et al. “Mutational signatures associated with tobacco smoking in human cancer”. In: *Science* 354.6312 (Nov. 2016), pp. 618–622. ISSN: 0036-8075. DOI: 10.1126/science.aag0299.

[7] Ludmil B. Alexandrov et al. “Signatures of mutational processes in human cancer”. In: *Nature* 500.7463 (Aug. 2013), pp. 415–421. ISSN: 0028-0836. DOI: 10.1038/nature12477.

[8] Ludmil B. Alexandrov et al. “The repertoire of mutational signatures in human cancer”. In: *Nature* 578.7793 (Feb. 2020), pp. 94–101. ISSN: 0028-0836. DOI: 10.1038/s41586-020-1943-3.

[9] Christina L. Appin and Daniel J. Brat. "Molecular Pathways in Gliomagenesis and Their Relevance to Neuropathologic Diagnosis". In: *Advances in Anatomic Pathology* 22.1 (Jan. 2015), pp. 50–58. ISSN: 1072-4109. DOI: 10.1097/PAP.0000000000000048.

[10] Hideyuki Arita et al. "A combination of TERT promoter mutation and MGMT methylation status predicts clinically relevant subgroups of newly diagnosed glioblastomas". In: *Acta Neuropathologica Communications* 4.1 (Dec. 2016), p. 79. ISSN: 2051-5960. DOI: 10.1186/s40478-016-0351-2.

[11] Hideyuki Arita et al. "Upregulating mutations in the TERT promoter commonly occur in adult malignant gliomas and are strongly associated with total 1p19q loss". In: *Acta Neuropathologica* 126.2 (Aug. 2013), pp. 267–276. ISSN: 0001-6322. DOI: 10.1007/s00401-013-1141-6.

[12] Taejeong Bae et al. "Different mutational rates and mechanisms in human cells at pregastrulation and neurogenesis". In: *Science* 359.6375 (Feb. 2018), pp. 550–555. ISSN: 0036-8075. DOI: 10.1126/science.aan8690.

[13] Norman T. J Bailey. *The elements of stochastic processes with applications to the natural sciences*. New York, Wiley, 1964.

[14] Jessica L. Barnes, Maria Zubair, Kaarthik John, Miriam C. Poirier, and Francis L. Martin. "Carcinogens and DNA damage". In: *Biochemical Society Transactions* 46.5 (Oct. 2018), pp. 1213–1224. ISSN: 0300-5127. DOI: 10.1042/BST20180519.

[15] Floris P. Barthel et al. "Longitudinal molecular trajectories of diffuse glioma in adults". In: *Nature* 576.7785 (Dec. 2019), pp. 112–120. ISSN: 0028-0836. DOI: 10.1038/s41586-019-1775-1.

[16] Stephen B. Baylin and Peter A. Jones. "Epigenetic Determinants of Cancer". In: *Cold Spring Harbor Perspectives in Biology* 8.9 (Sept. 2016), a019505. ISSN: 1943-0264. DOI: 10.1101/cshperspect.a019505.

[17] Lieze Berben, Giuseppe Floris, Hans Wildiers, and Sigrid Hatse. "Cancer and Aging: Two Tightly Interconnected Biological Processes". In: *Cancers* 13.6 (Mar. 2021), p. 1400. ISSN: 2072-6694. DOI: 10.3390/cancers13061400.

[18] Tamar R. Berger, Patrick Y. Wen, Melanie Lang-Orsini, and Ugonma N. Chukwueke. "World Health Organization 2021 Classification of Central Nervous System Tumors and Implications for Therapy for Adult-Type Gliomas". In: *JAMA Oncology* 8.10 (Oct. 2022), p. 1493. ISSN: 2374-2437. DOI: 10.1001/jamaoncol.2022.2844.

[19] H. Binder et al. "DNA methylation, transcriptome and genetic copy number signatures of diffuse cerebral WHO grade II/III gliomas resolve cancer heterogeneity and development". In: *Acta Neuropathologica Communications* 7.1 (Dec. 2019), p. 59. ISSN: 2051-5960. DOI: 10.1186/s40478-019-0704-8.

[20] Klaus Bister. "Discovery of oncogenes: The advent of molecular cancer research". In: *Proceedings of the National Academy of Sciences* 112.50 (Dec. 2015), pp. 15259–15260. ISSN: 0027-8424. DOI: 10.1073/pnas.1521145112.

[21] Klaus Bister and Hans W. Jansen. "Oncogenes in Retroviruses and Cells: Biochemistry and Molecular Genetics". In: 1986, pp. 99–188. DOI: 10.1016/S0065-230X(08)60199-2.

[22] Francis Blokzijl et al. "Tissue-specific mutation accumulation in human adult stem cells during life". In: *Nature* 538.7624 (Oct. 2016), pp. 260–264. ISSN: 0028-0836. DOI: 10.1038/nature19768.

[23] Theodor Boveri. *Zur Frage der Entstehung Maligner Tumoren*. Gustav Fischer, 1914, pp. 1–64.

[24] Lukas Bunse et al. "AMPLIFY-NEOVAC: a randomized, 3-arm multicenter phase I trial to assess safety, tolerability and immunogenicity of IDH1-vac combined with an immune checkpoint inhibitor targeting programmed death-ligand 1 in isocitrate dehydrogenase 1 mutant gliomas". In: *Neurological Research and Practice* 4.1 (Dec. 2022), p. 20. ISSN: 2524-3489. DOI: 10.1186/s42466-022-00184-x.

[25] Daniel P Cahill, David N Louis, and John Gregory Cairncross. "Molecular background of oligodendroglioma: 1p/19q, IDH, TERT, CIC and FUBP1". In: *CNS Oncology* 4.5 (Oct. 2015), pp. 287–294. ISSN: 2045-0907. DOI: 10.2217/cns.15.32.

[26] Vincent L. Cannataro, Jeffrey D. Mandell, and Jeffrey P. Townsend. "Attribution of Cancer Origins to Endogenous, Exogenous, and Preventable Mutational Processes". In: *Molecular Biology and Evolution* 39.5 (May 2022). ISSN: 0737-4038. DOI: 10.1093/molbev/msac084.

[27] David Capper et al. "DNA methylation-based classification of central nervous system tumours." In: *Nature* 555.7697 (Mar. 2018), pp. 469–474. ISSN: 1476-4687. DOI: 10.1038/nature26000.

[28] Giulio Caravagna et al. "Subclonal reconstruction of tumors by using machine learning and population genetics". In: *Nature Genetics* 52.9 (Sept. 2020), pp. 898–907. ISSN: 1061-4036. DOI: 10.1038/s41588-020-0675-5.

[29] Geoffrey M. Cooper and Paul E. Neiman. "Transforming genes of neoplasms induced by avian lymphoid leukosis viruses". In: *Nature* 287.5783 (Oct. 1980), pp. 656–659. ISSN: 0028-0836. DOI: 10.1038/287656a0.

[30] Geoffrey M. Cooper, Sharon Okenquist, and Lauren Silverman. "Transforming activity of DNA of chemically transformed and normal cells". In: *Nature* 284.5755 (Apr. 1980), pp. 418–421. ISSN: 0028-0836. DOI: 10.1038/284418a0.

[31] Charles Darwin. *The Origin of Species*. 1859. ISBN: 9781108005487. DOI: 10.1017/CBO9780511694295.

[32] Alexander Davis, Ruli Gao, and Nicholas Navin. "Tumor evolution: Linear, branching, neutral or punctuated?" In: *Biochimica et Biophysica Acta (BBA) - Reviews on Cancer* 1867.2 (Apr. 2017), pp. 151–161. ISSN: 0304419X. DOI: 10.1016/j.bbcan.2017.01.003.

[33] Ugo Del Monte. "Does the cell number 10
9 still really fit one gram of tumor tissue?" In: *Cell Cycle* 8.3 (Feb. 2009), pp. 505–506. ISSN: 1538-4101. DOI: 10.4161/cc.8.3.7608.

[34] D L Dexter, H M Kowalski, B A Blazar, Z Fligiel, R Vogel, and G H Heppner. "Heterogeneity of tumor cells from a single mouse mammary tumor." In: *Cancer research* 38.10 (Oct. 1978), pp. 3174–81. ISSN: 0008-5472.

[35] Rodrigo D. Drummond, Alexandre Defelicibus, Mathilde Meyenberg, Renan Valieris, Emmanuel Dias-Neto, Rafael A. Rosales, and Israel Tojal da Silva. "Relating mutational signature exposures to clinical data in cancers via signeR 2.0". In: *BMC Bioinformatics* 24.1 (Nov. 2023), p. 439. ISSN: 1471-2105. DOI: 10.1186/s12859-023-05550-3.

[36] Rick Durrett. "Population genetics of neutral mutations in exponentially growing cancer cell populations". In: *The Annals of Applied Probability* 23.1 (Feb. 2013). ISSN: 1050-5164. DOI: 10.1214/11-AAP824.

[37] Hueng-Chuen Fan, Fung-Wei Chang, Jeng-Dau Tsai, Kao-Min Lin, Chuan-Mu Chen, Shinn-Zong Lin, Ching-Ann Liu, and Horng-Jyh Harn. "Telomeres and Cancer". In: *Life* 11.12 (Dec. 2021), p. 1405. ISSN: 2075-1729. DOI: 10.3390/life11121405.

[38] Sandra Ferreyra Vega, Thomas Olsson Bontell, Teresia Kling, Asgeir Store Jakola, and Helena Carén. "Longitudinal DNA methylation analysis of adult-type IDH-mutant gliomas". In: *Acta Neuropathologica Communications* 11.1 (Feb. 2023), p. 23. ISSN: 2051-5960. DOI: 10.1186/s40478-023-01520-1.

[39] Andrea R Florl, Knut H Franke, Dieter Niederacher, Claus-Dieter Gerharz, Hans-Helge Seifert, and Wolfgang A Schulz. "DNA Methylation and the Mechanisms of CDKN2A Inactivation in Transitional Cell Carcinoma of the Urinary Bladder". In: *Laboratory Investigation* 80.10 (Oct. 2000), pp. 1513–1522. ISSN: 00236837. DOI: 10.1038/labinvest.3780161.

[40] William D. Foulkes, Tamar Y. Flanders, Pamela M. Pollock, and Nicholas K. Hayward. "The CDKN2A (p16) Gene and Human Cancer". In: *Molecular Medicine* 3.1 (Jan. 1997), pp. 5–20. ISSN: 1076-1551. DOI: 10.1007/BF03401664.

[41] Javier Ganz et al. "Rates and Patterns of Clonal Oncogenic Mutations in the Normal Human Brain". In: *Cancer Discovery* 12.1 (Jan. 2022), pp. 172–185. ISSN: 2159-8274. DOI: 10.1158/2159-8290.CD-21-0245.

[42] Levi A. Garraway and Eric S. Lander. "Lessons from the Cancer Genome". In: *Cell* 153.1 (Mar. 2013), pp. 17–37. ISSN: 00928674. DOI: 10.1016/j.cell.2013.03.002.

[43] Robert A Gatenby. "Commentary: Carcinogenesis as Darwinian evolution? Do the math!" In: *International Journal of Epidemiology* 35.5 (Oct. 2006), pp. 1165–1167. ISSN: 1464-3685. DOI: 10.1093/ije/dyl192.

[44] David Graham, Gideon Lipman, Vinay Sehgal, and Laurence B Lovat. "Monitoring the premalignant potential of Barrett's oesophagus'". In: *Frontline Gastroenterology* 7.4 (Oct. 2016), pp. 316–322. ISSN: 2041-4137. DOI: 10.1136/flgastro-2016-100712.

[45] Mel Greaves and Carlo C. Maley. "Clonal evolution in cancer". In: *Nature* 481.7381 (Jan. 2012), pp. 306–313. ISSN: 0028-0836. DOI: 10.1038/nature10762.

[46] Christopher Greenman et al. "Patterns of somatic mutation in human cancer genomes". In: *Nature* 446.7132 (Mar. 2007), pp. 153–158. ISSN: 0028-0836. DOI: 10.1038/nature05610.

[47] Douglas Hanahan and Robert A Weinberg. "The Hallmarks of Cancer". In: *Cell* 100.1 (Jan. 2000), pp. 57–70. ISSN: 00928674. DOI: 10.1016/S0092-8674(00)81683-9.

[48] Douglas Hanahan and Robert A. Weinberg. "Hallmarks of Cancer: The Next Generation". In: *Cell* 144.5 (Mar. 2011), pp. 646–674. ISSN: 00928674. DOI: 10.1016/j.cell.2011.02.013.

[49] David Hansemann. "Ueber asymmetrische Zelltheilung in Epithelkrebsen und deren biologische Bedeutung". In: *Archiv für Pathologische Anatomie und Physiologie und für Klinische Medicin* 119.2 (Feb. 1890), pp. 299–326. ISSN: 0945-6317. DOI: 10.1007/BF01882039.

[50] Tsimur Hasanau, Eduard Pisarev, Olga Kisil, Naosuke Nonoguchi, Florence Le Calvez-Kelm, and Maria Zvereva. "Detection of TERT Promoter Mutations as a Prognostic Biomarker in Gliomas: Methodology, Prospects, and Advances". In: *Biomedicines* 10.3 (Mar. 2022), p. 728. ISSN: 2227-9059. DOI: 10.3390/biomedicines10030728.

[51] G H Heppner. "Tumor heterogeneity." In: *Cancer research* 44.6 (June 1984), pp. 2259–65. ISSN: 0008-5472.

[52] Guus R. M. van den Heuvel, Leonie I. Kroeze, Marjolijn J. L. Ligtenberg, Katrien Grünberg, Erik A. M. Jansen, Daniel von Rhein, Richarda M. de Voer, and Michel M. van den Heuvel. "Mutational signature analysis in non-small cell lung cancer patients with a high tumor mutational burden". In: *Respiratory Research* 22.1 (Dec. 2021), p. 302. ISSN: 1465-993X. DOI: 10.1186/s12931-021-01871-0.

[53] Andrew J Holland and Don W Cleveland. "Losing balance: the origin and impact of aneuploidy in cancer". In: *EMBO reports* 13.6 (June 2012), pp. 501–514. ISSN: 1469-221X. DOI: 10.1038/embor.2012.55.

[54] Md. Sanower Hossain et al. "Colorectal Cancer: A Review of Carcinogenesis, Global Epidemiology, Current Challenges, Risk Factors, Preventive and Treatment Strategies". In: *Cancers* 14.7 (Mar. 2022), p. 1732. ISSN: 2072-6694. DOI: 10.3390/cancers14071732.

[55] Daniel Hübschmann et al. "Analysis of mutational signatures with yet another package for signature analysis". In: *Genes, Chromosomes and Cancer* 60.5 (May 2021), pp. 314–331. ISSN: 1045-2257. DOI: 10.1002/gcc.22918.

[56] Mohammad A. Jafri, Shakeel A. Ansari, Mohammed H. Alqahtani, and Jerry W. Shay. "Roles of telomeres and telomerase in cancer, and advances in telomerase-targeted therapies". In: *Genome Medicine* 8.1 (Dec. 2016), p. 69. ISSN: 1756-994X. DOI: 10.1186/s13073-016-0324-x.

[57] Siddhartha Jaiswal and Benjamin L. Ebert. "Clonal hematopoiesis in human aging and disease". In: *Science* 366.6465 (Nov. 2019). ISSN: 0036-8075. DOI: 10.1126/science.aan4673.

[58] Hong Jiang et al. "Proteins induced by telomere dysfunction and DNA damage represent biomarkers of human aging and disease". In: *Proceedings of the National Academy of Sciences* 105.32 (Aug. 2008), pp. 11299–11304. ISSN: 0027-8424. DOI: 10.1073/pnas.0801457105.

[59] Seung-Gi Jin, Dean Pettinga, Jennifer Johnson, Peipei Li, and Gerd P. Pfeifer. "The major mechanism of melanoma mutations is based on deamination of cytosine in pyrimidine dimers as determined by circle damage sequencing". In: *Science Advances* 7.31 (July 2021). ISSN: 2375-2548. DOI: 10.1126/sciadv.abi6508.

[60] Peter A. Jones and Stephen B. Baylin. "The Epigenomics of Cancer". In: *Cell* 128.4 (Feb. 2007), pp. 683–692. ISSN: 00928674. DOI: 10.1016/j.cell.2007.01.029.

[61] R Kanwal and S Gupta. "Epigenetic modifications in cancer". In: *Clinical Genetics* 81.4 (Apr. 2012), pp. 303–311. ISSN: 0009-9163. DOI: 10.1111/j.1399-0004.2011.01809.x.

[62] Scott R. Kennedy, Yuezheng Zhang, and Rosa Ana Risques. "Cancer-Associated Mutations but No Cancer: Insights into the Early Steps of Carcinogenesis and Implications for Early Cancer Detection". In: *Trends in Cancer* 5.9 (Sept. 2019), pp. 531–540. ISSN: 24058033. DOI: 10.1016/j.trecan.2019.07.007.

[63] Michael J. Keogh et al. "High prevalence of focal and multi-focal somatic genetic variants in the human brain". In: *Nature Communications* 9.1 (Oct. 2018), p. 4257. ISSN: 2041-1723. DOI: 10.1038/s41467-018-06331-w.

[64] David A. Kessler and Herbert Levine. "Large population solution of the stochastic Luria–Delbrück evolution model". In: *Proceedings of the National Academy of Sciences* 110.29 (July 2013), pp. 11682–11687. ISSN: 0027-8424. DOI: 10.1073/pnas.1309667110.

[65] Tobias Kessler, Jakob Ito, Wolfgang Wick, and Antje Wick. "Conventional and emerging treatments of astrocytomas and oligodendrogliomas". In: *Journal of Neuro-Oncology* 162.3 (May 2023), pp. 471–478. ISSN: 0167-594X. DOI: 10.1007/s11060-022-04216-z.

[66] Patrick J. Killela et al. "TERT promoter mutations occur frequently in gliomas and a subset of tumors derived from cells with low rates of self-renewal". In: *Proceedings of the National Academy of Sciences* 110.15 (Apr. 2013), pp. 6021–6026. ISSN: 0027-8424. DOI: 10.1073/pnas.1303607110.

[67] Motoo Kimura. "The Neutral Theory of Molecular Evolution". In: *Scientific American* 241.5 (Nov. 1979), pp. 98–126. ISSN: 0036-8733. DOI: 10.1038/scientificamerican1179-98.

[68] Emmanuel Klinger, Dennis Rickert, and Jan Hasenauer. "pyABC: distributed, likelihood-free inference". In: *Bioinformatics* 34.20 (Oct. 2018), pp. 3591–3593. ISSN: 1367-4803. DOI: 10.1093/bioinformatics/bty361.

[69] Alfred G. Knudson. "Mutation and Cancer: Statistical Study of Retinoblastoma". In: *Proceedings of the National Academy of Sciences* 68.4 (Apr. 1971), pp. 820–823. ISSN: 0027-8424. DOI: 10.1073/pnas.68.4.820.

[70] Gene Koh, Andrea Degasperi, Xueqing Zou, Sophie Momen, and Serena Nik-Zainal. "Mutational signatures: emerging concepts, caveats and clinical applications". In: *Nature Reviews Cancer* 21.10 (Oct. 2021), pp. 619–637. ISSN: 1474-175X. DOI: 10.1038/s41568-021-00377-7.

[71] Verena Körber, Sabine A. Stainczyk, Roma Kurilov, Kai-Oliver Henrich, Barbara Hero, Benedikt Brors, Frank Westermann, and Thomas Höfer. "Neuroblastoma arises in early fetal development and its evolutionary duration predicts outcome". In: *Nature Genetics* 55.4 (Apr. 2023), pp. 619–630. ISSN: 1061-4036. DOI: 10.1038/s41588-023-01332-y.

[72] Verena Körber et al. "Detecting and quantifying clonal selection in somatic stem cells". In: *bioRxiv* (Jan. 2023), p. 2021.12.15.472780. DOI: 10.1101/2021.12.15.472780. URL: http://biorxiv.org/content/early/2023/10/10/2021.12.15.472780.abstract.

[73] Verena Körber et al. "Evolutionary Trajectories of IDHWT Glioblastomas Reveal a Common Path of Early Tumorigenesis Instigated Years ahead of Initial Diagnosis". In: *Cancer Cell* 35.4 (Apr. 2019), pp. 692–704. ISSN: 15356108. DOI: 10.1016/j.ccell.2019.02.007.

[74] T G Krontiris and G M Cooper. "Transforming activity of human tumor DNAs." In: *Proceedings of the National Academy of Sciences* 78.2 (Feb. 1981), pp. 1181–1184. ISSN: 0027-8424. DOI: 10.1073/pnas.78.2.1181.

[75] Sarah Kuhn, Laura Gritti, Daniel Crooks, and Yvonne Dombrowski. "Oligodendrocytes in Development, Myelin Generation and Beyond". In: *Cells* 8.11 (Nov. 2019), p. 1424. ISSN: 2073-4409. DOI: 10.3390/cells8111424.

[76] Ioannis D. Kyrochristos, Demosthenes E. Ziogas, Anna Goussia, Georgios K. Glantzounis, and Dimitrios H. Roukos. "Bulk and Single-Cell Next-Generation Sequencing: Individualizing Treatment for Colorectal Cancer". In: *Cancers* 11.11 (Nov. 2019), p. 1809. ISSN: 2072-6694. DOI: 10.3390/cancers11111809.

[77] Lukáš Lacina, Matúš Čoma, Barbora Dvořánková, Ondřej Kodet, Nikola Melegová, Peter Gál, and Karel Smetana. "Evolution of Cancer Progression in the Context of Darwinism". In: *Anticancer Research* 39.1 (Jan. 2019), pp. 1–16. ISSN: 0250-7005. DOI: 10.21873/anticanres.13074.

[78] Ezio Laconi, Fabio Marongiu, and James DeGregori. "Cancer as a disease of old age: changing mutational and microenvironmental landscapes". In: *British Journal of Cancer* 122.7 (Mar. 2020), pp. 943–952. ISSN: 0007-0920. DOI: 10.1038/s41416-019-0721-1.

[79] Cristina M Lanata, Sharon A Chung, and Lindsey A Criswell. "DNA methylation 101: what is important to know about DNA methylation and its role in SLE risk and disease heterogeneity". In: *Lupus Science & Medicine* 5.1 (July 2018), e000285. ISSN: 2053-8790. DOI: 10.1136/lupus-2018-000285.

[80] Michael S. Lawrence et al. "Mutational heterogeneity in cancer and the search for new cancer-associated genes". In: *Nature* 499.7457 (July 2013), pp. 214–218. ISSN: 0028-0836. DOI: 10.1038/nature12213.

[81] Henry Lee-Six et al. "Population dynamics of normal human blood inferred from somatic mutations". In: *Nature* 561.7724 (Sept. 2018), pp. 473–478. ISSN: 0028-0836. DOI: 10.1038/s41586-018-0497-0.

[82] Henry Lee-Six et al. "The landscape of somatic mutation in normal colorectal epithelial cells". In: *Nature* 574.7779 (Oct. 2019), pp. 532–537. ISSN: 0028-0836. DOI: 10.1038/s41586-019-1672-7.

[83] Chunjie Li et al. "The functional role of inherited CDKN2A variants in childhood acute lymphoblastic leukemia". In: *Pharmacogenetics and Genomics* 32.2 (Feb. 2022), pp. 43–50. ISSN: 1744-6872. DOI: 10.1097/FPC.0000000000000451.

[84] Esther H. Lips et al. "Genomic analysis defines clonal relationships of ductal carcinoma in situ and recurrent invasive breast cancer". In: *Nature Genetics* 54.6 (June 2022), pp. 850–860. ISSN: 1061-4036. DOI: 10.1038/s41588-022-01082-3.

[85] Guanzheng Liu et al. "Genomic alterations of oligodendrogliomas at distant recurrence". In: *Cancer Medicine* 12.16 (Aug. 2023), pp. 17171–17183. ISSN: 2045-7634. DOI: 10.1002/cam4.6327.

[86] Manjiao Liu, Sijian Xia, Xu Zhang, Bei Zhang, Linlin Yan, Meijia Yang, Yong Ren, Hao Guo, and Jie Zhao. "Development and validation of a blood-based genomic mutation signature to predict the clinical outcomes of atezolizumab therapy in NSCLC". In: *Lung Cancer* 170 (Aug. 2022), pp. 148–155. ISSN: 01695002. DOI: 10.1016/j.lungcan.2022.06.016.

[87] Zhichao Liu, Guo Lin, Zeping Yan, Linduo Li, Xingchen Wu, Jingrong Shi, Jianxing He, Lei Zhao, Hengrui Liang, and Wei Wang. "Predictive mutation signature of immunotherapy benefits in NSCLC based on machine learning algorithms". In: *Frontiers in Immunology* 13 (Sept. 2022). ISSN: 1664-3224. DOI: 10.3389/fimmu.2022.989275.

[88] Lawrence A. Loeb and Curtis C. Harris. "Advances in Chemical Carcinogenesis: A Historical Review and Prospective". In: *Cancer Research* 68.17 (Sept. 2008), pp. 6863–6872. ISSN: 0008-5472. DOI: 10.1158/0008-5472.CAN-08-2852.

[89] David N Louis et al. "The 2021 WHO Classification of Tumors of the Central Nervous System: a summary". In: *Neuro-Oncology* 23.8 (Aug. 2021), pp. 1231–1251. ISSN: 1522-8517. DOI: 10.1093/neuonc/noab106.

[90] DN Louis, H Ohgaki, OD Wiestler, and WK Cavenee. *World Health Organization Classification of Tumours of the Central Nervous System*. 4th updated edition. Lyon: International Agency for Research on Cancer, 2016.

[91] DN Louis, H Ohgaki, OD Wiestler, and WK Cavenee. *World Health Organization Histological Classification of Tumours of the Central Nervous System*. 4th edition. Lyon: International Agency for Research on Cancer, 2007.

[92] Asma Ismail Mahmod, Shatha Khaled Haif, Ayah Kamal, Israa A. Al-ataby, and Wamidh H. Talib. "Chemoprevention effect of the Mediterranean diet on colorectal cancer: Current studies and future prospects". In: *Frontiers in Nutrition* 9 (Aug. 2022). ISSN: 2296-861X. DOI: 10.3389/fnut.2022.924192.

[93] Tathiane M Malta, Camila F de Souza, Thais S Sabedot, Tiago C Silva, Maritza S Mosella, Steven N Kalkanis, James Snyder, Ana Valeria B Castro, and Houtan Noushmehr. "Glioma CpG island methylator phenotype (G-CIMP): biological and clinical implications". In: *Neuro-Oncology* 20.5 (Apr. 2018), pp. 608–620. ISSN: 1522-8517. DOI: 10.1093/neuonc/nox183.

[94] Tathiane M. Malta et al. "The Epigenetic Evolution of Glioma Is Determined by the IDH1 Mutation Status and Treatment Regimen". In: *Cancer Research* 84.5 (Mar. 2024), pp. 741–756. ISSN: 0008-5472. DOI: 10.1158/0008-5472.CAN-23-2093.

[95] Freek Manders, Ruben van Boxtel, and Sjors Middelkamp. "The Dynamics of Somatic Mutagenesis During Life in Humans". In: *Frontiers in Aging* 2 (Dec. 2021). ISSN: 2673-6217. DOI: 10.3389/fragi.2021.802407.

[96] Elaine R. Mardis. "The Impact of Next-Generation Sequencing on Cancer Genomics: From Discovery to Clinic". In: *Cold Spring Harbor Perspectives in Medicine* 9.9 (Sept. 2019), a036269. ISSN: 2157-1422. DOI: 10.1101/cshperspect.a036269.

[97] C. J. Marshall, L. M. Franks, and A. W. Carbonell. "Markers of Neoplastic Transformation in Epithelial Cell Lines Derived From Human Carcinomas". In: *JNCI: Journal of the National Cancer Institute* 58.6 (June 1977), pp. 1743–1751. ISSN: 1460-2105. DOI: 10.1093/jnci/58.6.1743.

[98] Iñigo Martincorena and Peter J. Campbell. "Somatic mutation in cancer and normal cells". In: *Science* 349.6255 (Sept. 2015), pp. 1483–1489. ISSN: 0036-8075. DOI: 10.1126/science.aab4082.

[99] Iñigo Martincorena, Keiran M. Raine, Moritz Gerstung, Kevin J. Dawson, Kerstin Haase, Peter Van Loo, Helen Davies, Michael R. Stratton, and Peter J. Campbell. "Universal Patterns of Selection in Cancer and Somatic Tissues". In: *Cell* 171.5 (Nov. 2017), pp. 1029–1041. ISSN: 00928674. DOI: 10.1016/j.cell.2017.09.042.

[100] Iñigo Martincorena et al. "High burden and pervasive positive selection of somatic mutations in normal human skin". In: *Science* 348.6237 (May 2015), pp. 880–886. ISSN: 0036-8075. DOI: 10.1126/science.aaa6806.

[101] Iñigo Martincorena et al. "Somatic mutant clones colonize the human esophagus with age". In: *Science* 362.6417 (Nov. 2018), pp. 911–917. ISSN: 0036-8075. DOI: 10.1126/science.aau3879.

[102] Francisco Martínez-Jiménez et al. "A compendium of mutational cancer driver genes". In: *Nature Reviews Cancer* 20.10 (Oct. 2020), pp. 555–572. ISSN: 1474-175X. DOI: 10.1038/s41568-020-0290-x.

[103] Tali Mazor et al. "DNA Methylation and Somatic Mutations Converge on the Cell Cycle and Define Similar Evolutionary Histories in Brain Tumors". In: *Cancer Cell* 28.3 (Sept. 2015), pp. 307–317. ISSN: 15356108. DOI: 10.1016/j.ccell.2015.07.012.

[104] Ingo K. Mellinghoff et al. "Vorasidenib and ivosidenib in IDH1-mutant low-grade glioma: a randomized, perioperative phase 1 trial". In: *Nature Medicine* 29.3 (Mar. 2023), pp. 615–622. ISSN: 1078-8956. DOI: 10.1038/s41591-022-02141-2.

[105] Ingo K. Mellinghoff et al. "Vorasidenib in IDH1- or IDH2-Mutant Low-Grade Glioma". In: *New England Journal of Medicine* 389.7 (Aug. 2023), pp. 589–601. ISSN: 0028-4793. DOI: 10.1056/NEJMoa2304194.

[106] Brandon Milholland, Xiao Dong, Lei Zhang, Xiaoxiao Hao, Yousin Suh, and Jan Vijg. "Differences between germline and somatic mutation rates in humans and mice". In: *Nature Communications* 8.1 (May 2017), p. 15183. ISSN: 2041-1723. DOI: 10.1038/ncomms15183.

[107] Christopher A. Miller et al. "SciClone: Inferring Clonal Architecture and Tracking the Spatial and Temporal Patterns of Tumor Evolution". In: *PLoS Computational Biology* 10.8 (Aug. 2014), e1003665. ISSN: 1553-7358. DOI: 10.1371/journal.pcbi.1003665.

[108] Léa Montégut, Carlos López-Otín, and Guido Kroemer. "Aging and cancer". In: *Molecular Cancer* 23.1 (May 2024), p. 106. ISSN: 1476-4598. DOI: 10.1186/s12943-024-02020-z.

[109] Poornimaa Murali and Ramanathan Karuppasamy. "Exploring the potential of nutraceutical to combat gliomas: focus on mIDH2 protein". In: *Frontiers in Physics* 12 (Feb. 2024). ISSN: 2296-424X. DOI: 10.3389/fphy.2024.1345834.

[110] Serena Nik-Zainal et al. "Mutational Processes Molding the Genomes of 21 Breast Cancers". In: *Cell* 149.5 (May 2012), pp. 979–993. ISSN: 00928674. DOI: 10.1016/j.cell.2012.04.024.

[111] Serena Nik-Zainal et al. "The Life History of 21 Breast Cancers". In: *Cell* 149.5 (May 2012), pp. 994–1007. ISSN: 00928674. DOI: 10.1016/j.cell.2012.04.023.

[112] Houtan Noushmehr et al. "Identification of a CpG Island Methylator Phenotype that Defines a Distinct Subgroup of Glioma". In: *Cancer Cell* 17.5 (May 2010), pp. 510–522. ISSN: 15356108. DOI: 10.1016/j.ccr.2010.03.017.

[113] Peter C. Nowell. "The Clonal Evolution of Tumor Cell Populations". In: *Science* 194.4260 (Oct. 1976), pp. 23–28. ISSN: 0036-8075. DOI: 10.1126/science.959840.

[114] Christopher C Oakes et al. "DNA methylation dynamics during B cell maturation underlie a continuum of disease phenotypes in chronic lymphocytic leukemia." In: *Nature genetics* 48.3 (Mar. 2016), pp. 253–64. ISSN: 1546-1718. DOI: 10.1038/ng.3488.

[115] Hisashi Ohtsuki and Hideki Innan. "Forward and backward evolutionary processes and allele frequency spectrum in a cancer cell population". In: *Theoretical Population Biology* 117 (Oct. 2017), pp. 43–50. ISSN: 00405809. DOI: 10.1016/j.tpb.2017.08.006.

[116] Sigurgeir Olafsson et al. “Somatic Evolution in Non-neoplastic IBD-Affected Colon”. In: *Cell* 182.3 (Aug. 2020), pp. 672–684. ISSN: 00928674. DOI: 10.1016/j.cell.2020.06.036.

[117] Nathalie Olympios, Vianney Gilard, Florent Marguet, Florian Clatot, Frédéric Di Fiore, and Maxime Fontanilles. “TERT Promoter Alterations in Glioblastoma: A Systematic Review”. In: *Cancers* 13.5 (Mar. 2021), p. 1147. ISSN: 2072-6694. DOI: 10.3390/cancers13051147.

[118] Quinn T Ostrom, Gino Cioffi, Kristin Waite, Carol Kruchko, and Jill S Barnholtz-Sloan. “CBTRUS Statistical Report: Primary Brain and Other Central Nervous System Tumors Diagnosed in the United States in 2014–2018”. In: *Neuro-Oncology* 23.Supplement_3 (Oct. 2021), pp. iii1–iii105. ISSN: 1522-8517. DOI: 10.1093/neuonc/noab200.

[119] Jin Woo Park, Kwanghoon Lee, Eric Eunshik Kim, Seong-Ik Kim, and Sung-Hye Park. “Brain Tumor Classification by Methylation Profile”. In: *Journal of Korean Medical Science* 38.43 (2023). ISSN: 1011-8934. DOI: 10.3346/jkms.2023.38.e356.

[120] Yashna Paul, Baisakhi Mondal, Vikas Patil, and Kumaravel Somasundaram. “DNA methylation signatures for 2016 WHO classification subtypes of diffuse gliomas”. In: *Clinical Epigenetics* 9.1 (Dec. 2017), p. 32. ISSN: 1868-7075. DOI: 10.1186/s13148-017-0331-9.

[121] Anders I. Persson et al. “Non-Stem Cell Origin for Oligodendroglioma”. In: *Cancer Cell* 18.6 (Dec. 2010), pp. 669–682. ISSN: 15356108. DOI: 10.1016/j.ccr.2010.10.033.

[122] Mia Petljak et al. “Characterizing Mutational Signatures in Human Cancer Cell Lines Reveals Episodic APOBEC Mutagenesis”. In: *Cell* 176.6 (Mar. 2019), pp. 1282–1294. ISSN: 00928674. DOI: 10.1016/j.cell.2019.02.012.

[123] C. Pixberg et al. “COGNITION: a prospective precision oncology trial for patients with early breast cancer at high risk following neoadjuvant chemotherapy”. In: *ESMO Open* 7.6 (Dec. 2022), p. 100637. ISSN: 20597029. DOI: 10.1016/j.esmoop.2022.100637.

[124] Michael Platten et al. “A vaccine targeting mutant IDH1 in newly diagnosed glioma”. In: *Nature* 592.7854 (Apr. 2021), pp. 463–468. ISSN: 0028-0836. DOI: 10.1038/s41586-021-03363-z.

[125] Julia R. Pon and Marco A. Marra. “Driver and Passenger Mutations in Cancer”. In: *Annual Review of Pathology: Mechanisms of Disease* 10.1 (Jan. 2015), pp. 25–50. ISSN: 1553-4006. DOI: 10.1146/annurev-pathol-012414-040312.

[126] Zahraa Rahal, Paul Scheet, and Humam Kadara. "Somatic Mutations in Normal Tissues: Calm before the Storm." In: *Cancer discovery* 14.4 (Apr. 2024), pp. 605–609. ISSN: 2159-8290. DOI: 10.1158/2159-8290.CD-23-1508.

[127] E. Premkumar Reddy, Roberta K. Reynolds, Eugenio Santos, and Mariano Barbacid. "A point mutation is responsible for the acquisition of transforming properties by the T24 human bladder carcinoma oncogene". In: *Nature* 300.5888 (Nov. 1982), pp. 149–152. ISSN: 0028-0836. DOI: 10.1038/300149a0.

[128] Pieter A. Roelofs, John W.M. Martens, Reuben S. Harris, and Paul N. Span. "Clinical Implications of APOBEC3-Mediated Mutagenesis in Breast Cancer". In: *Clinical Cancer Research* 29.9 (May 2023), pp. 1658–1669. ISSN: 1078-0432. DOI: 10.1158/1078-0432.CCR-22-2861.

[129] Paul Roepman et al. "Clinical Validation of Whole Genome Sequencing for Cancer Diagnostics". In: *The Journal of Molecular Diagnostics* 23.7 (July 2021), pp. 816–833. ISSN: 15251578. DOI: 10.1016/j.jmoldx.2021.04.011.

[130] Andrew Roth, Jaswinder Khattra, Damian Yap, Adrian Wan, Emma Laks, Justina Biele, Gavin Ha, Samuel Aparicio, Alexandre Bouchard-Côté, and Sohrab P Shah. "PyClone: statistical inference of clonal population structure in cancer". In: *Nature Methods* 11.4 (Apr. 2014), pp. 396–398. ISSN: 1548-7091. DOI: 10.1038/nmeth.2883.

[131] R Schneider-Stock, H Walter, C Haeckel, K Radig, J Rys, and A Roessner. "Gene alterations at the CDKN2A (p16/MTS1) locus in soft tissue tumors." In: *International Journal of Oncology* (Aug. 1998). ISSN: 1019-6439. DOI: 10.3892/ijo.13.2.325.

[132] Masood A Shammas. "Telomeres, lifestyle, cancer, and aging". In: *Current Opinion in Clinical Nutrition and Metabolic Care* 14.1 (Jan. 2011), pp. 28–34. ISSN: 1363-1950. DOI: 10.1097/MCO.0b013e32834121b1.

[133] Chiaho Shih, L. C. Padhy, Mark Murray, and Robert A. Weinberg. "Transforming genes of carcinomas and neuroblastomas introduced into mouse fibroblasts". In: *Nature* 290.5803 (Mar. 1981), pp. 261–264. ISSN: 0028-0836. DOI: 10.1038/290261a0.

[134] John C. Silbereis, Sirisha Pochareddy, Ying Zhu, Mingfeng Li, and Nenad Sestan. "The Cellular and Molecular Landscapes of the Developing Human Central Nervous System". In: *Neuron* 89.2 (Jan. 2016), pp. 248–268. ISSN: 08966273. DOI: 10.1016/j.neuron.2015.12.008.

[135] Milena Simovic et al. “Carbon ion radiotherapy eradicates medulloblastomas with chromothripsis in an orthotopic Li-Fraumeni patient-derived mouse model”. In: *Neuro-Oncology* 23.12 (Dec. 2021), pp. 2028–2041. ISSN: 1522-8517. DOI: 10.1093/neuonc/noab127.

[136] Camila Ferreira de Souza et al. “A Distinct DNA Methylation Shift in a Subset of Glioma CpG Island Methylator Phenotypes during Tumor Recurrence”. In: *Cell Reports* 23.2 (Apr. 2018), pp. 637–651. ISSN: 22111247. DOI: 10.1016/j.celrep.2018.03.107.

[137] Lucy F. Stead and Roel G.W. Verhaak. “Doomed from the TERT? A Two-Stage Model of Tumorigenesis in IDH-Wild-Type Glioblastoma”. In: *Cancer Cell* 35.4 (Apr. 2019), pp. 542–544. ISSN: 15356108. DOI: 10.1016/j.ccell.2019.03.009.

[138] Michael R. Stratton, Peter J. Campbell, and P. Andrew Futreal. “The cancer genome”. In: *Nature* 458.7239 (Apr. 2009), pp. 719–724. ISSN: 0028-0836. DOI: 10.1038/nature07943.

[139] Samuel P Strom. “Current practices and guidelines for clinical next-generation sequencing oncology testing.” In: *Cancer biology & medicine* 13.1 (Mar. 2016), pp. 3–11. ISSN: 2095-3941. DOI: 10.28092/j.issn.2095-3941.2016.0004.

[140] Lakshmi Subramanian, Maria Elisa Calcagnotto, and Mercedes F. Paredes. “Cortical Malformations: Lessons in Human Brain Development”. In: *Frontiers in Cellular Neuroscience* 13 (Jan. 2020). ISSN: 1662-5102. DOI: 10.3389/fncel.2019.00576.

[141] Clifford J. Tabin, Scott M. Bradley, Cornelia I. Bargmann, Robert A. Weinberg, Alex G. Papageorge, Edward M. Scolnick, Ravi Dhar, Douglas R. Lowy, and Esther H. Chang. “Mechanism of activation of a human oncogene”. In: *Nature* 300.5888 (Nov. 1982), pp. 143–149. ISSN: 0028-0836. DOI: 10.1038/300143a0.

[142] Jessica Tang et al. “The genomic landscapes of individual melanocytes from human skin”. In: *Nature* 586.7830 (Oct. 2020), pp. 600–605. ISSN: 0028-0836. DOI: 10.1038/s41586-020-2785-8.

[143] “Targeted glioma therapy makes strides”. In: *Nature Biotechnology* 41.7 (July 2023), pp. 884–884. ISSN: 1087-0156. DOI: 10.1038/s41587-023-01869-7.

[144] John G Tate et al. “COSMIC: the Catalogue Of Somatic Mutations In Cancer”. In: *Nucleic Acids Research* 47.D1 (Jan. 2019), pp. D941–D947. ISSN: 0305-1048. DOI: 10.1093/nar/gky1015.

[145] Cristian Tomasetti, Bert Vogelstein, and Giovanni Parmigiani. "Half or more of the somatic mutations in cancers of self-renewing tissues originate prior to tumor initiation." In: *Proceedings of the National Academy of Sciences of the United States of America* 110.6 (Feb. 2013), pp. 1999–2004. ISSN: 1091-6490. DOI: 10.1073/pnas.1221068110.

[146] Craig A. Tork and Christopher Atkinson. *Oligodendroglioma*. 2024.

[147] Minoru Toyota, Nita Ahuja, Mutsumi Ohe-Toyota, James G. Herman, Stephen B. Baylin, and Jean-Pierre J. Issa. "CpG island methylator phenotype in colorectal cancer". In: *Proceedings of the National Academy of Sciences* 96.15 (July 1999), pp. 8681–8686. ISSN: 0027-8424. DOI: 10.1073/pnas.96.15.8681.

[148] Minoru Toyota, Kenneth J. Kopecky, Mutsumi-Ohe Toyota, Kam-Wing Jair, Cheryl L. Willman, and Jean-Pierre J. Issa. "Methylation profiling in acute myeloid leukemia". In: *Blood* 97.9 (May 2001), pp. 2823–2829. ISSN: 1528-0020. DOI: 10.1182/blood.V97.9.2823.

[149] Aristidis Tsatsakis et al. "Role of telomere length in human carcinogenesis (Review)". In: *International Journal of Oncology* 63.1 (May 2023), p. 78. ISSN: 1019-6439. DOI: 10.3892/ijo.2023.5526.

[150] Sevin Turcan et al. "IDH1 mutation is sufficient to establish the glioma hypermethylator phenotype". In: *Nature* 483.7390 (Mar. 2012), pp. 479–483. ISSN: 0028-0836. DOI: 10.1038/nature10866.

[151] Alexander Vaiserman and Dmytro Krasnienkov. "Telomere Length as a Marker of Biological Age: State-of-the-Art, Open Issues, and Future Perspectives". In: *Frontiers in Genetics* 11 (Jan. 2021). ISSN: 1664-8021. DOI: 10.3389/fgene.2020.630186.

[152] Arne Van Hoeck, Niels H. Tjoonk, Ruben van Boxtel, and Edwin Cuppen. "Portrait of a cancer: mutational signature analyses for cancer diagnostics". In: *BMC Cancer* 19.1 (Dec. 2019), p. 457. ISSN: 1471-2407. DOI: 10.1186/s12885-019-5677-2.

[153] Harold E. Varmus. "Retroviruses and Oncogenes I". In: *Bioscience Reports* 10.5 (Oct. 1990), pp. 413–430. ISSN: 0144-8463. DOI: 10.1007/BF01152288.

[154] Roberto Vendramin, Kevin Litchfield, and Charles Swanton. "Cancer evolution: Darwin and beyond". In: *The EMBO Journal* 40.18 (Sept. 2021). ISSN: 0261-4189. DOI: 10.15252/embj.2021108389.

[155] T L Vincent and R A Gatenby. "An evolutionary model for initiation, promotion, and progression in carcinogenesis." In: *International journal of oncology* 32.4 (Apr. 2008), pp. 729–37. ISSN: 1019-6439.

[156] Bert Vogelstein and Kenneth W. Kinzler. "The multistep nature of cancer". In: *Trends in Genetics* 9.4 (Apr. 1993), pp. 138–141. ISSN: 01689525. DOI: 10.1016/0168-9525(93)90209-Z.

[157] Bert Vogelstein, Nickolas Papadopoulos, Victor E. Velculescu, Shibin Zhou, Luis A. Diaz, and Kenneth W. Kinzler. "Cancer Genome Landscapes". In: *Science* 339.6127 (Mar. 2013), pp. 1546–1558. ISSN: 0036-8075. DOI: 10.1126/science.1235122.

[158] Josephine Volovetz, Defne Bayik, and Justin D. Lathia. "Origin and development of oligodendroglioma". In: *Oligodendroglioma*. Elsevier, 2019, pp. 79–87. DOI: 10.1016/B978-0-12-813158-9.00007-4.

[159] Jing Wang, Baizhou Li, Meng Luo, Jia Huang, Kun Zhang, Shu Zheng, Suzhan Zhang, and Jiaojiao Zhou. "Progression from ductal carcinoma in situ to invasive breast cancer: molecular features and clinical significance". In: *Signal Transduction and Targeted Therapy* 9.1 (Apr. 2024), p. 83. ISSN: 2059-3635. DOI: 10.1038/s41392-024-01779-3.

[160] Yichen Wang et al. "APOBEC mutagenesis is a common process in normal human small intestine". In: *Nature Genetics* 55.2 (Feb. 2023), pp. 246–254. ISSN: 1061-4036. DOI: 10.1038/s41588-022-01296-5.

[161] P S Ward, J R Cross, C Lu, O Weigert, O Abel-Wahab, R L Levine, D M Weinstock, K A Sharp, and C B Thompson. "Identification of additional IDH mutations associated with oncometabolite R()-2-hydroxyglutarate production". In: *Oncogene* 31.19 (May 2012), pp. 2491–2498. ISSN: 0950-9232. DOI: 10.1038/onc.2011.416.

[162] Takuya Watanabe, Sumihito Nobusawa, Paul Kleihues, and Hiroko Ohgaki. "IDH1 Mutations Are Early Events in the Development of Astrocytomas and Oligodendrogliomas". In: *The American Journal of Pathology* 174.4 (Apr. 2009), pp. 1149–1153. ISSN: 00029440. DOI: 10.2353/ajpath.2009.080958.

[163] Donate Weghorn and Shamil Sunyaev. "Bayesian inference of negative and positive selection in human cancers". In: *Nature Genetics* 49.12 (Dec. 2017), pp. 1785–1788. ISSN: 1061-4036. DOI: 10.1038/ng.3987.

[164] P. Wesseling and D. Capper. "WHO 2016 Classification of gliomas". In: *Neuropathology and Applied Neurobiology* 44.2 (Feb. 2018), pp. 139–150. ISSN: 0305-1846. DOI: 10.1111/nan.12432.

[165] WHO Classification of Tumours Editorial Board. *World Health Organization Classification of Tumours of the Central Nervous System*. 5th edition. Lyon: International Agency for Research on Cancer, 2021.

[166] Neshika Wijewardhane, Lisa Dressler, and Francesca D. Ciccarelli. "Normal Somatic Mutations in Cancer Transformation". In: *Cancer Cell* 39.2 (Feb. 2021), pp. 125–129. ISSN: 15356108. DOI: 10.1016/j.ccell.2020.11.002.

[167] Marc J Williams, Benjamin Werner, Chris P Barnes, Trevor A Graham, and Andrea Sottoriva. "Identification of neutral tumor evolution across cancer types". In: *Nature Genetics* 48.3 (Mar. 2016), pp. 238–244. ISSN: 1061-4036. DOI: 10.1038/ng.3489.

[168] Marc J Williams, Luis Zapata, Benjamin Werner, Chris P Barnes, Andrea Sottoriva, and Trevor A Graham. "Measuring the distribution of fitness effects in somatic evolution by combining clonal dynamics with dN/dS ratios". In: *eLife* 9 (Mar. 2020). ISSN: 2050-084X. DOI: 10.7554/eLife.48714.

[169] Marc J. Williams, Benjamin Werner, Timon Heide, Christina Curtis, Chris P. Barnes, Andrea Sottoriva, and Trevor A. Graham. "Quantification of subclonal selection in cancer from bulk sequencing data". In: *Nature Genetics* 50.6 (June 2018), pp. 895–903. ISSN: 1061-4036. DOI: 10.1038/s41588-018-0128-6.

[170] Marietta Wolter, Julia Reifenberger, Britta Blaschke, Koichi Ichimura, Esther E. Schmidt, V. Peter Collins, and Guido Reifenberger. "Oligodendroglial Tumors Frequently Demonstrate Hypermethylation of the CDKN2A (MTS1, p16 ^{INK4a}), p14 ^{ARF} , and CDKN2B (MTS2 , p15 ^{INK4b}) Tumor Suppressor Genes". In: *Journal of Neuropathology & Experimental Neurology* 60.12 (Dec. 2001), pp. 1170–1180. ISSN: 0022-3069. DOI: 10.1093/jnen/60.12.1170.

[171] X. Wu, C. I. Amos, Y. Zhu, H. Zhao, B. H. Grossman, J. W. Shay, S. Luo, W. K. Hong, and M. R. Spitz. "Telomere Dysfunction: A Potential Cancer Predisposition Factor". In: *JNCI Journal of the National Cancer Institute* 95.16 (Aug. 2003), pp. 1211–1218. ISSN: 0027-8874. DOI: 10.1093/jnci/djg011.

[172] Maggie S.Y. Yeung et al. "Dynamics of Oligodendrocyte Generation and Myelination in the Human Brain". In: *Cell* 159.4 (Nov. 2014), pp. 766–774. ISSN: 00928674. DOI: 10.1016/j.cell.2014.10.011.

[173] Stephen Yip et al. "Concurrent CIC mutations, IDH mutations, and 1p/19q loss distinguish oligodendrogliomas from other cancers". In: *The Journal of Pathology* 226.1 (Jan. 2012), pp. 7–16. ISSN: 0022-3417. DOI: 10.1002/path.2995.

[174] Akira Yokoyama et al. "Age-related remodelling of oesophageal epithelia by mutated cancer drivers". In: *Nature* 565.7739 (Jan. 2019), pp. 312–317. ISSN: 0028-0836. DOI: 10.1038/s41586-018-0811-x.

[175] Ming Yu, William D. Hazelton, Georg E. Luebeck, and William M. Grady. "Epigenetic Aging: More Than Just a Clock When It Comes to Cancer". In: *Cancer Research* 80.3 (Feb. 2020), pp. 367–374. ISSN: 0008-5472. DOI: 10.1158/0008-5472.CAN-19-0924.

[176] Ming Yu et al. "Subtypes of Barrett's oesophagus and oesophageal adenocarcinoma based on genome-wide methylation analysis". In: *Gut* 68.3 (Mar. 2019), pp. 389–399. ISSN: 0017-5749. DOI: 10.1136/gutjnl-2017-314544.

[177] Xinyang Yu, Hao Zhao, Ruiqi Wang, Yingyin Chen, Xumei Ouyang, Wenting Li, Yihao Sun, and Anghui Peng. "Cancer epigenetics: from laboratory studies and clinical trials to precision medicine". In: *Cell Death Discovery* 10.1 (Jan. 2024), p. 28. ISSN: 2058-7716. DOI: 10.1038/s41420-024-01803-z.

[178] Luis Zapata, Oriol Pich, Luis Serrano, Fyodor A. Kondrashov, Stephan Ossowski, and Martin H. Schaefer. "Negative selection in tumor genome evolution acts on essential cellular functions and the immunopeptidome". In: *Genome Biology* 19.1 (Dec. 2018), p. 67. ISSN: 1474-760X. DOI: 10.1186/s13059-018-1434-0.

[179] Eric Y. Zhao, Martin Jones, and Steven J.M. Jones. "Whole-Genome Sequencing in Cancer". In: *Cold Spring Harbor Perspectives in Medicine* 9.3 (Mar. 2019), a034579. ISSN: 2157-1422. DOI: 10.1101/cshperspect.a034579.

[180] Yi Zhou, Hongjun Song, and Guo-li Ming. "Genetics of human brain development". In: *Nature Reviews Genetics* 25.1 (Jan. 2024), pp. 26–45. ISSN: 1471-0056. DOI: 10.1038/s41576-023-00626-5.

Disclaimer:
The text of this thesis is original and was fully written by Sarah Benedetto. It has been proofread using ChatGPT.