Paulina Róża Szadkowska

Application of targeted next-generation sequencing to detect potentially pathogenic alterations in gliomas, circulating tumor DNA and tumor-derived cells

PhD. Thesis completed in the
Laboratory of Molecular Neurobiology
of the Nencki Institute of Experimental Biology
Polish Academy of Sciences

SUPERVISOR:
Prof. Bożena Kamińska-Kaczmarek, PhD., DSc

Warsaw, 2023

Acknowledgments:

I would like to express my sincere gratitude to all those who have supported me throughout my journey:

First and foremost, I would like to express my sincere gratitude to ***Prof. Bożena Kamińska-Kaczmarek*** *for allowing me to grow as a research scientist and for scientific support and guidance throughout the project. For immense help in writing PhD thesis and insightful comments. Also thank you for an opportunity to discover how successful science career can look like.*

I would also like to thank:

Bartosz Wojtaś *who have challenged me to grow and learn. His guidance and wisdom have shaped my perspective and helped me develop the skills I need to succeed in my field.*

Bartłomiej Gielniewski *who have helped me learn how to prepare my first library and how to sequence, which is now one of my favorite things to do at work. His patience and positive attitude is greatly appreciated.*

Adria Jaume Roura-Canalda *who have helped me with bioinformatic analysis and was always a positive person to be around.*

Kamil Wojnicki *that assisted with the statistical analysis and inspired me to join BodyAttack, get in shape and became a true friend.*

Paulina Wiechecka *that showed me how to perform DNA, RNA, and proteins isolations from frozen tissue, how to keep track of samples and patient medical material, and for correcting my polish language mistakes that could occur due to my prolonged stay overseas.*

Beata Kaza *for helping with cell line cultures and keeping the laboratory in order.*

Salwador Cyranowski *for challenging my English skills, helping with manuscript review, and always keeping my iPhone in the best shape.*

I want to thank my ***family*** *for their unwavering love, encouragement, and support which have been the bedrock of my success.*

Iwona Ciechomska *for taking the time to explain to me a lot of biological signaling pathways, transcriptions mechanisms, and apoptosis but mostly for a friendship and support.*

Michał Zawadzki *for giving me an opportunity to test alternative method of blood collection for an alternative liquid biopsy approach and inspiring me to dream big.*

Kasia Poleszak *for teaching me for to isolate nuclear extract and perform EMSA assays, with a lot of patience and great attitude.*

I am grateful to my ***colleagues and collaborators****, who have been instrumental in my research and professional development:* ***Jakub Mieczkowski, Kasia Leszczyńska, Chinchu Jayaprakash, Karolina Stępniak, Marta Maleszewska-Bobińska, Agnieszka Kaczmarczyk, Aleksandra Ellert-Miklaszewska, Anna Lenkiewicz, and other Kamińska Lab Members****. Their contributions and insights have enriched my work and opened new avenues of exploration.*

I also owe a debt of gratitude to my ***friends****, who have always been there to lend a helping hand, offer advice, and provide a listening ear.*

I want to thank all of the ***employees of Nencki Institute*** *for an amazing and friendly work atmosphere, and for an inspiration that is* ***Prof. Agnieszka Dobrzyń.***

In conclusion, I am indebted to all those who have supported me along the way. I could not have achieved my goals without their help, and I am grateful for their contributions to my journey.

The results presented in this dissertation were collected in the frame of the grant GLIOMED STRATEGMED3/307326/6/NCBR/2017 from The National Center of Research and Development, Poland, and the Foundation for Polish Science TEAM-TECH Core Facility project "NGS platform for comprehensive diagnostics and personalized therapy in neuro-oncology". The use of CePT infrastructure financed by the European Union — The European Regional Development Fund within the Operational Programme "Innovative economy" for 2007-2013 is highly appreciated. We thank all the patients for consenting to the use of their biological material for this research.

Table of Contents

Abbreviations

AF – allele frequency

AFM – atomic force microscopy

AFP – alpha-fetoprotein

AKAP – gene encoding for A-kinase anchor proteins

ATRX – gene encoding for the transcription regulator ATP-dependent helicase

BBB – the blood-brain barrier

BBTB – blood-brain-tumor barrier

CA 15-3 – carcinoma antigen 15-3

CA 19-9 – carcinoma antigen 19-9

CA 72-4 – carcinoma antigen 72-4

CA125 – carcinoma antigen 125

CDKN2A – gene coding for the cyclin-dependent kinase inhibitor 2A

CEA – Carcinoembryonic antigens

cfDNA – circulating cell-free DNA

ClinVar – Clinical Genome Resource provided by National Center for Biotechnology Information

CNA – copy number alterations

CNS – central nervous system

COSMIC – Catalog of Somatic Mutations in Cancer

CSF – cerebrospinal fluid

CTC – circulating tumor cells

ctDNA – circulating tumor DNA

ddPCR – droplet digital polymerase chain reaction

DDX10 – gene coding for DEAD-box helicase 10 (ATP-dependent RNA helicase)

dsDNA – double-stranded DNA

DVL1 – gene encoding disheveled segment polarity protein 1

EGFR – gene encoding the epidermal growth factor receptor

EP3000 – gene encoding a E1A Binding Protein P300

ER – estrogen receptor

FANCA – gene coding for Fanconi anemia complementation group

FFPE – Formalin Fixed Paraffin Embedded

GBM – glioblastoma

gDNA – germline white blood cells derived DNA

HER2 – gene encoding human epidermal growth factor receptor 2

ICB – immune checkpoint blockade

IDH – gene encoding for isocitrate dehydrogenase

***IDH1* codon 132 mutation** – mutation in gene IDH1 changing Arg132His

Ivy GAP – Ivy Glioblastoma Atlas Project

KMT2C – gene encoding Histone-Lysine N-Methyltransferase 2C

KMT2D – gene encoding Histone-Lysine N-Methyltransferase 2D

MAF – minor allele frequency

MGMT –gene coding for the O6-methylguanine methyltransferase

MPS – massive parallel sequencing

MRI – magnetic resonance imaging

MTUS2 – gene encoding Microtubule Associated Tumor Suppressor Candidate 2

NF1 – gene encoding the Neurofibromin 1

NGS – next generation sequencing

NMP22 – nuclear matrix protein 22

NSE – neuron-specific enolase

PA – pilocytic astrocytoma

PAI-1 – plasminogen activator inhibitor-1

PCNSL – primary central nervous system lymphoma

PCR – polymerase chain reaction

PDGFRA – gene encoding the platelet-derived growth factor receptor alpha

PgR – progesterone receptor

PKA – protein kinase A

PSA – prostate-specific antigen

PTEN – gene encoding the phosphatase and tensin homolog

qPCR – quantitative polymerase chain reaction

RB1 – retinoblastoma gene

RET – gene encoding a receptor tyrosine kinase (Proto-Oncogene C-Ret)

RT – room temperature

TCGA – the Cancer Genome Atlas

tDNA – tumor-derived DNA

TdT – terminal deoxynucleotidyl transferase

TERT – gene encoding telomerase reverse transcriptase

TMB – tumor mutational burden

TMZ – temozolomide

TP53 – gene encoding the tumor protein 53

UMI – unique molecular identifiers

uPA – urokinase plasminogen activator

VAF – variant allele frequencies

VEGF – gene encoding vascular endothelial growth factor

WGS – whole genome sequencing

WHO – World Health Organization

β-hCG – beta-human chorionic gonadotropin

Summary

Gliomas are primary tumors of the central nervous system. Diagnosis and therapy recommendations are difficult, because of the intertumoral heterogenicity of glial tumors. Current World Health Organization classification of gliomas is based on the pathomorphological and molecular characteristics of the tumor biopsy. Diagnosis is based on pathomorphological features (diffusiveness, proliferation index, a presence of necrosis) and upon specific genetic alterations detected in the tumor, which leads to specific recommendations for the therapy. Next generation sequencing (NGS) is a valuable tool to improve diagnostics of brain tumors.

Traditional tissue biopsy might not present complete mutational spectrum in case of such heterogenic tumors, so alternative methods are being tested to enable more holistic view of each disease. When traditional biopsy or tumor resection are not possible, a liquid biopsy would be of great assistance to clinical practice. Liquid biopsy is a use of bodily fluids to isolate circulating cell free nucleic acids or circulating tumor cells to detect cancer markers for diagnostics, disease monitoring or prognostic. In case of primary brain tumors, cerebrospinal fluid can contain more circulating cell free DNA (cfDNA) or RNA (cfRNA) originating from the tumor, but a lumbar puncture may have side effects, so it is rarely performed on heavily symptomatic primary brain tumor patients.

We sought to evaluate if improvements in cfDNA isolation, library preparation and targeted sequencing would provide reliable information regards genetic alterations in glioblastoma (GBM), most common and deadly primary brain tumor. After analysis of blood derived cfDNA potentially pathogenic variants were detected in 37/84, which based upon the current literature is an improvement from most of the studies.

We employed a target gene panel encompassing 668 cancer-related genes and NGS to a set of diagnostically difficult pediatric glioma tumors. The analysis of DNA isolated from formalin fixed paraffin embedded (FFPE) sections originating from those tumors yielded the whole spectrum of potentially pathogenic mutations, some interesting variants were found, that could be further studied (*MTUS, FANCA, RET*).

Tumor-derived cell cultures are valuable *in vitro* system to study tumorigenesis and screen for therapeutics, however it is not fully known if tumor cells keep their genetic alterations and cultured clones reflect molecular profile of an original tumor. Comparative analysis of somatic mutations present in tumor-derived cell lines and/or original tumors have shown some differences in variant profiles, cell cultures contained more detectable somatic mutations. This can indicate that some somatic variants can be missed in the tissue biopsy, due to its complexity as tumor contains healthy cells, microglia and macrophages that can make background noise decreasing the tumor variants detectability. On the other hand, tumor stem cells can possibly gain mutations during cell culture, as their DNA repair pathways are frequently malfunctioning, and mitosis is maximized by artificial growth factors.

The current classification of gliomas is based upon tumor genotyping. Current diagnostic tests employ molecular analysis of DNA isolated from FFPE or frozen tumor samples. There are many ongoing clinical and research studies improving current diagnostic methods with the aim to create personalized therapy recommendations with use of both blood derived cfDNA and tumor derived cell cultures. Present study demonstrates how tumor derived cell lines and blood derived cfDNA can offer an insight on tumor genetic heterogeneity.

Streszczenie

Glejaki to pierwotne nowotwory ośrodkowego układu nerwowego. Prawidłowa diagnostyka i terapia tych guzów jest utrudniona ze względu na heterogenność guzów i obecność klonów komórek o różnym genotypie. Obecna klasyfikacja glejaków Światowej Organizacji Zdrowia opiera się na patomorfologicznych i molekularnych cechach materiału z biopsji guza. Rozpoznania dokonuje się na podstawie cech patomorfologicznych (dyfuzyjność, wskaźnik proliferacji, obecność martwicy) oraz wykrytych w guzie specyficznych zmian genetycznych, co prowadzi do sformułowania konkretnych zaleceń terapeutycznych. Sekwencjonowanie nowej generacji jest cennym narzędziem usprawniającym diagnostykę guzów mózgu.

Tradycyjna biopsja może nie przedstawiać pełnego spektrum mutacji w przypadku heterogennych guzów, dlatego testowane są alternatywne metody, które umożliwią bardziej całościowe spojrzenie na chorobę. Gdy tradycyjna biopsja lub resekcja nie są możliwe, biopsja płynna byłaby bardzo pomocna w praktyce klinicznej. Biopsja płynna to wykorzystanie płynów ustrojowych do wyizolowania kwasów nukleinowych wolnych od krążących komórek lub krążących komórek nowotworowych w celu wykrycia markerów nowotworowych, które mogą być przydatne w diagnostyce, monitorowaniu choroby lub ustalaniu prognozy. W przypadku pierwotnych guzów mózgu płyn mózgowo-rdzeniowy może zawierać więcej wolnego krążącego DNA (cfDNA) lub RNA (cfRNA) pochodzącego z guza, ale nakłucie lędźwiowe może mieć skutki uboczne, dlatego rzadko wykonuje się je u pacjentów z wyraźnymi objawami pierwotnego guza mózgu.

W naszych badaniach staraliśmy się ocenić, czy ulepszenia w izolacji cfDNA, przygotowaniu biblioteki i ukierunkowanym sekwencjonowaniu dostarczyłyby wiarygodnych informacji dotyczących zmian genetycznych w glejaku wielopostaciowym (GBM), najczęstszym i śmiertelnym pierwotnym guzie mózgu. Po analizie cfDNA pochodzącego z krwi wykryto potencjalnie patogenne warianty w 37/84, co w oparciu o aktualną literaturę stanowi poprawę w stosunku do większości badań.

Zastosowaliśmy panel genów obejmujący 668 genów związanych z rakiem i sekwencjonowanie nowej generacji do ustalenie spektrum mutacji w trudnych diagnostycznie glejakach dziecięcych. Analiza DNA wyizolowanego ze skrawków

utrwalonych w formalinie i zatopionych w parafinie (FFPE) pochodzących z tych guzów ujawniła pełne spektrum potencjalnie patogennych mutacji. Znaleziono kilka interesujących wariantów w genach *MTUS, FANCA, RET*, które powinny być dalej badane.

Hodowle komórkowe pochodzące z guza są cennym systemem *in vitro* do badania powstawania nowotworów i badań przesiewowych pod kątem substancji terapeutycznych. Jednak nie do końca wiadomo, czy komórki nowotworowe zachowują swoje zmiany genetyczne, a klony komórek w hodowli reprezentują pierwotny nowotwór. Analiza porównawcza mutacji somatycznych obecnych w hodowlach wyprowadzonych od pacjentów GBM wykazała pewne różnice w profilach wariantów, a hodowle komórkowe zawierały więcej wykrywalnych mutacji somatycznych. Może to wskazywać, że niektóre warianty somatyczne mogą zostać pominięte w biopsji z guza ze względu na jej różnorodność, ponieważ guz zawiera zdrowe komórki, mikroglej i makrofagi, które mogą powodować szum tła zmniejszający wykrywalność wariantów guza. Z drugiej strony nowotworowe komórki macierzyste mogą prawdopodobnie uzyskać mutacje podczas hodowli komórkowej, ponieważ ich szlaki naprawy DNA często działają nieprawidłowo, a mitoza jest maksymalizowana przez egzogenne czynniki wzrostu. Obecna klasyfikacja glejaków opiera się na genotypowaniu. Stosowane testy diagnostyczne wykorzystują analizę molekularną DNA wyizolowanego z FFPE lub zamrożonych próbek guza. Prowadzonych jest wiele badań klinicznych i podstawowych, których celem jest udoskonalenie obecnych metod diagnostycznych z wykorzystaniem zarówno cfDNA pochodzącego z krwi, jak i hodowli komórek nowotworowych w celu uzyskania zaleceń terapeutycznych w medycynie spersonalizowanej. Przedstawione w rozprawie wyniki badań pokazują, że zarówno linie komórkowe wyprowadzone z guza, jak i cfDNA pochodzące z krwi, dostarczają wiarygodnych informacji na temat heterogenności genetycznej guza, mogą ujawnić więcej potencjalnie patogennych wariantów, które należy wziąć pod uwagę, niż tradycyjna biopsja guza.

1. Introduction

1.1. Uncovering the human genome

Deoxyribonucleic acid (DNA) stores the genetic information of a cell. DNA of a differentiated cell in adults contains a genome, as complete as an embryo, which has been demonstrated by producing a viable offspring a ship Dolly by transfer of an adult single nucleus into an enucleated unfertilized mammalian egg[1]. Changes in DNA induced by aging and lifestyle are still being studied, which so far resulted in the acknowledgment of processes like telomere shortening, epigenetic modifications, copy number variation, and mutations that are induced in a DNA structure and organization over time. Single nucleotide polymorphisms (SNPs) are common, as they variations occur in the human genome at a frequency of one in every 300 bases. That means that on average only 10 million positions out of the human genome (3 billion base-pair) have common variations. Each person's characteristic variations can be used to track inheritance in families and susceptibility to certain diseases, and scientists are working on development of catalogues of SNPs as a tool to use in their efforts to uncover the causes of common illnesses like heart disease, diabetes, or cancer[2–5].

The Human Genome Project launched in 1990 was an international effort to sequence the 3 billion DNA nucleotides in the human genome and is considered by many to be one of the most challenging scientific undertakings of all time. The finished reference sequence produced by the Human Genome Project encompasses about 99% of the gene-containing regions, and it's accuracy reaches 99.99%. The execution of the project led to the achievement of a wide range of other goals, from sequencing the genomes of model organisms to developing new technologies to study whole genomes. Besides uncovering the human genome, the international network of researchers has produced an amazing array of accomplishments that include: an advanced draft of the mouse genome sequence, published in December 2002[6]; an initial draft of the rat genome sequence, produced in January 2002[7]; the identification of more than 3 million human genetic variations, called single nucleotide polymorphisms and full-length complementary DNAs (cDNAs) for over 70 % of known human and mouse genes[8,9].

In 2005 Next Generation Sequencing (NGS) platforms introduced massively parallel, high- throughput sequencing at a fraction of the time. Thanks

to technological developments sequencing cost decreased even by 200,000 times since 2001 as shown in figure I1 (based upon National Human Genome Research Institute publicly available data).

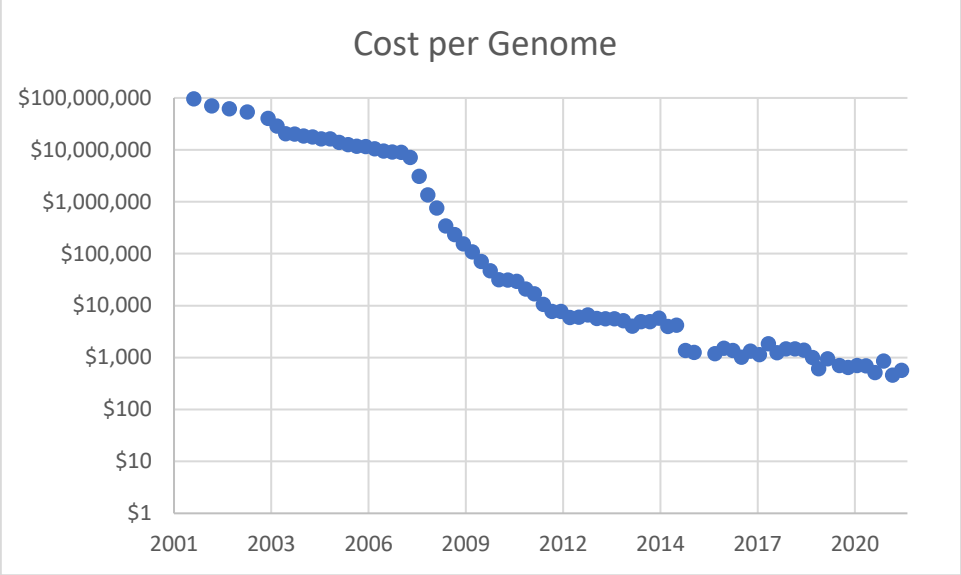


Figure I1. Cost of sequencing per Human Genome in time. Graph generated based upon publicly available data provided in Wetterstrand KA. DNA Sequencing Costs: Data from the NHGRI Genome Sequencing Program (GSP) available at: **www.genome.gov/sequencingcostsdata**. Accessed [20.03.2023].

New whole genome sequencing methods were established: in 2007 ChIP-seq combined chromatin immunoprecipitation and high throughput NGS[10] and allowed for an analysis of chromatin openness and gene regulation mechanisms, and in 2008 massive-scale RNA sequencing methods was developed[11].

The establishment of the genome reference sequence was critical, as it is used by scientists and medical professionals as a standard for comparison of newly generated DNA sequences. Genomic data that was generated from sequencing required better methods of sorting, indexing and visualization. This need yielded in the development of genome browsers such as Ensembl and UCSC Genome Browser[12]. In 2008 genome assembly tools that could reconstruct genomes cost-effectively and in a timely manner emerged (Velvet, ALLPATHS, and SOAPdenovo), leading to the *de novo* assembly of large and high-quality genomes[13]. Over one million SNPs were reported as a common variation

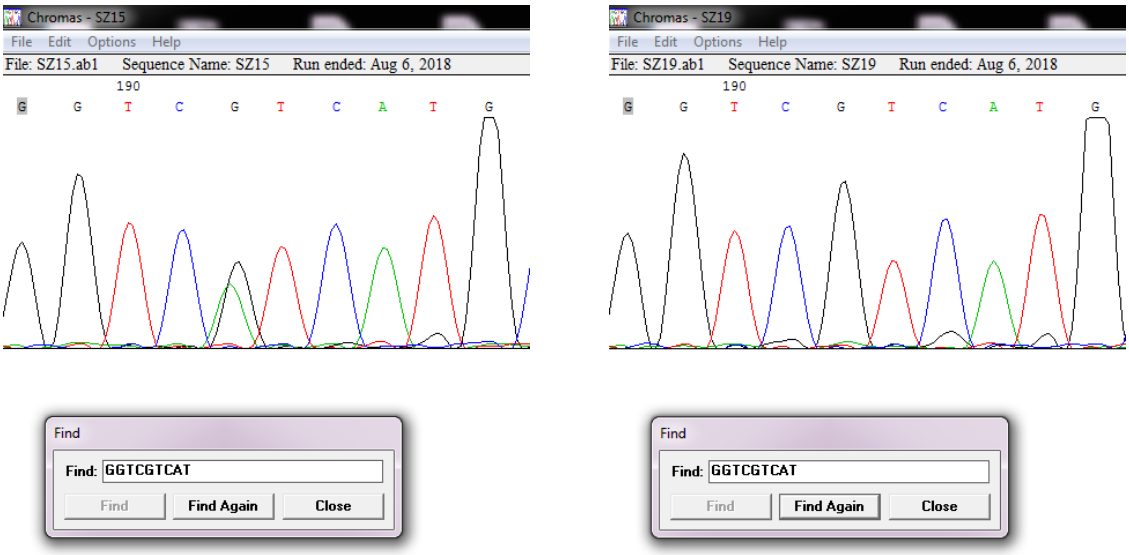
in the human genome[14], the need to catalog common variations in the human genome was the driving force behind the development of the 1000 Genome Project and HapMap. 1000 Genome Projects' aim was to sequence at least one thousand genomes of anonymous participants, from different ethnic groups[15]. By the 2012 project delivered the map of 1092 human genomes[16], and by 2015 the map of 2504 human genomes[17] and became the global reference for human genetic variations[18].

1. 2. Improvements in DNA and RNA Sequencing

The breakthroughs in genome biology and medicine would not be possible without advancements in nucleic acid sequencing technologies. The revolutionary method of sequencing developed by two-time Nobel Laureate Frederick Sanger and his colleagues in 1977 also known as the "chain termination method" or Sanger sequencing[19,20] is still used as a gold standard for DNA sequencing. The method involves chain-terminating dideoxynucleotides (ddNTPs) that are used as inhibitors of chain-elongating DNA polymerases.

Over the years Sanger Sequencing has been modernized and automated into large batch sequencing, but the main idea of the method remained the same. Automated electrophoresis that is used today produces results in a different form, below (figure I2) is an example of the results from Sanger sequencing of isocitrate dehydrogenase (*IDH*) codon 132 (mutation in the *IDH1* resulting in Arg132His).

Sanger sequencing is a parallel method for confirming sequence variants and it can provide a ground against which the NGS assay can be benchmarked and validated.



(**a**) (**b**)

Figure I2. Sanger sequencing results reviewed using Chromas SZ5 software: (**a**) *IDH1*-mutation present; (**b**) no *IDH1*-mutation detected.

1.3. Next-generation sequencing

Next-generation sequencing (NGS) is a massive parallel sequencing platform that enables sequencing of millions up to billions DNA molecules from multiple individuals simultaneously. Most NGS platforms require DNA to be fragmented to a certain length and either amplified or ligated with a custom adapter sequenced, to form a library (a set of fragments representing a genome). During sequencing, these library adapters hybridize to a surface of a glass slide covered with complementary adapters (so-called flow cell), where each library fragment is amplified. Amplification process creates clusters of DNA on the flow cell. Each cluster originates and represents specific single library fragment and acts as an individual sequencing reaction. Massive parallel sequencing technologies differ in sequencing chemistry and engineering configurations of the designed equipment.

NGS can be categorized in 4 main types of sequencing: pyrosequencing, sequencing by ligation, sequencing by synthesis, and ion semiconductor sequencing. Pyrosequencing was first professed in 1993 by Bertil Pettersson, Mathias Uhlen, and Pal Nyren[21]. The method combined solid phase sequencing that uses streptavidin coated magnetic beads that bind to biotinylated DNA, with recombinant DNA polymerase that is lacking a 3' to 5' exonuclease activity

and detection with luminescence of firefly luciferase enzyme[22]. In pyrosequencing each nucleotide incorporation leads to pyrophosphate release which allows monitoring of the reaction. Sequencing by ligation does not incorporate DNA polymerase, but oligonucleotide bases with four fluorescent bases that are ligated to the primer bound sequence and wash out with every step. This method generates short reads. Sequencing by ligation uses a DNA ligase enzyme to identify nucleotides at specific position, rather than DNA polymerase. Ion semiconductor sequencing is a method that detects nucleotides during sequencing by synthesis, but detection is based upon release of hydrogen ions during DNA polymerization process. Other sequencing technologies include PacBio and BGI Technologies. PacBio sequencing technology does not require fragmentation and obtains sequence information through process of the target DNA replication. Nanopore-based DNA sequencing (offered by BGI Technologies) uses measurement
of differences in electrical current signal during a process which involves single strands of DNA traveling through extremely small pores in the membrane.

NGS Illumina technology, that was used in this project, involves sequencing by synthesis, which is based on an addition of a single nucleotide one at a time to sequenced DNA library, washing the unbound nucleotides, imaging, and repeating the process until the full length of the read is obtained. The complete protocol of sequencing includes sample preparation, cluster generation, sequencing, and data analysis. There are many different methods of sample (library) preparation, but all methods include addition of the adaptors to the end fragments of the libraries. Long chains of DNA are fragmentated (enzymatically or mechanically), ends are repaired and through reduced cycle amplifications additional motives are introduced, such as sequencing binding site, indices, and regions complementary to flow cell oligos. Thanks to the indices libraries originating from many different patients can be pooled together, diluted, denatured, and loaded onto a sample loading station on the sequencer. Sequencing takes place on the surface of an oligo covered glass flow cell that is placed in the sequencer. Reagents are loaded into machine, denatured libraries are placed in the sample loading station and run parameters are entered before the process begins.

First part of this process is cluster generation. There are 2 types of oligos present on the flow cells surface. First oligo type is complementary to the libraries

adaptor which allows hybridization of the library to the flow cell surface. Then during polymerase chain reaction (PCR) complementary sequences are synthesized, double stranded molecule is denatured and original template is washed away. Only newly synthesized fragments, complementary to an original template, remain attached to the flow cell. Next, the strand folds over and the adapter region hybridizes to the second type of oligo present at the flow cell, PCR generates complementary strands forming a double stranded bridge. The bridge is denatured, resulting in two single stranded complementary library fragments bound to the flow cell surface. This process (called bridge amplification) is repeated and it occurs simultaneously for millions of clusters resulting in clonal amplification of all of the library fragments. After bridge amplification reverse strands are cleaved and washed off, only the forward strands representing original input library sequence remain tethered to the flow cell and the 3' ends are blocked from priming.

Finally, sequencing phase begins with the extension of the first sequencing primer, after which sequencing by synthesis process can be initiated. During each cycle, only one fluorescently labeled nucleotide can be added to the growing chain, the clusters are excited by a light source, characteristic fluorescent signal is emitted, and images are taken by the machine to register specific signal and its location. This process is called reversible dye termination. Length of the read is defined by the number of cycles. The base call is determined by emission wavelength and signal intensity. All of the identical strands within a given cluster are read simultaneously. Hundreds of millions clusters are sequenced in this massively parallel process. This entire process generates millions to billions of reads, representing all the nucleic acid fragments as shown in figure I3.

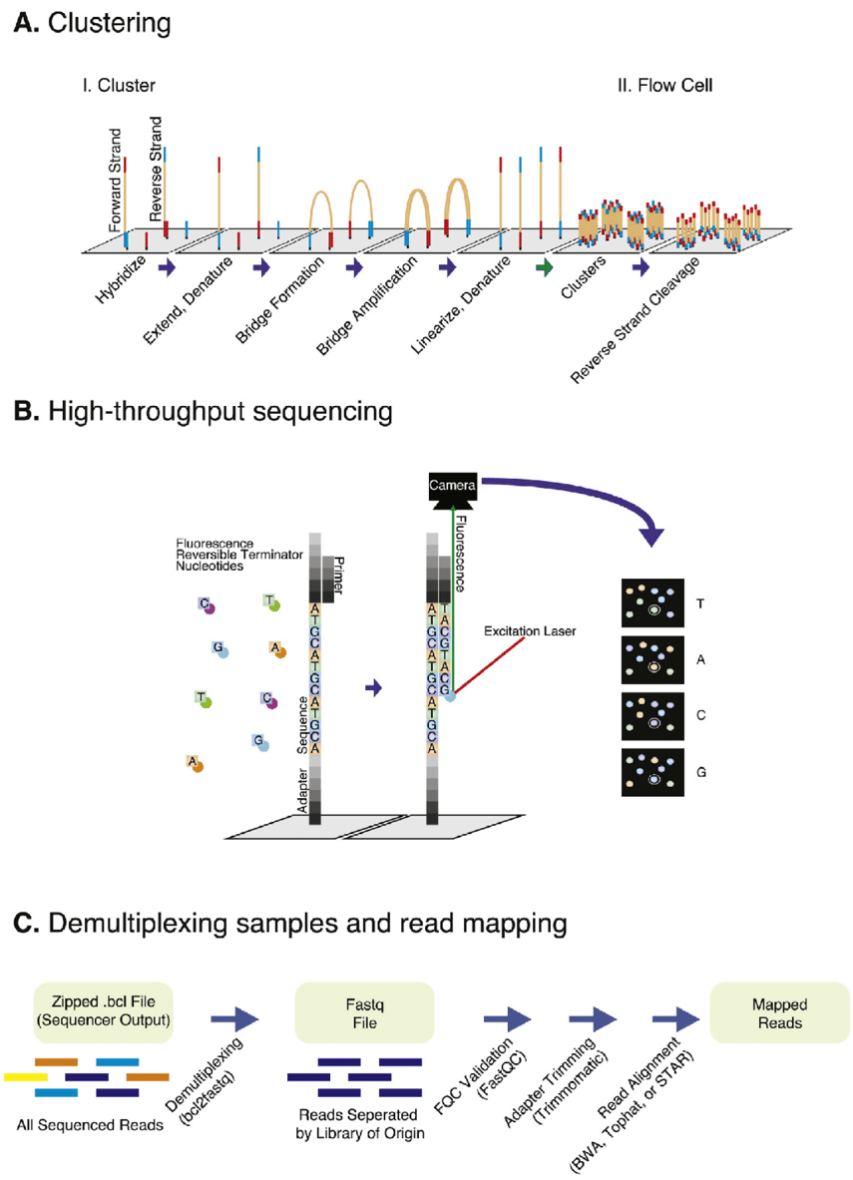


Figure. I3. Illumina sequencing and data processing workflow. Modified from[23].

When the process of sequencing is finished, data that was generated can be acquired in a form of FASTQ files, then the reads need to be aligned to the reference genome, reverse and forward reads are grouped and can be reviewed, analyzed, and compared using for example genome browser. Example below (figure I4) presents review of aligned BAM files using an integrative genome viewer (igv), in the region of Microtubule Associated Tumor Suppressor Candidate 2 gene (*MTUS)*:

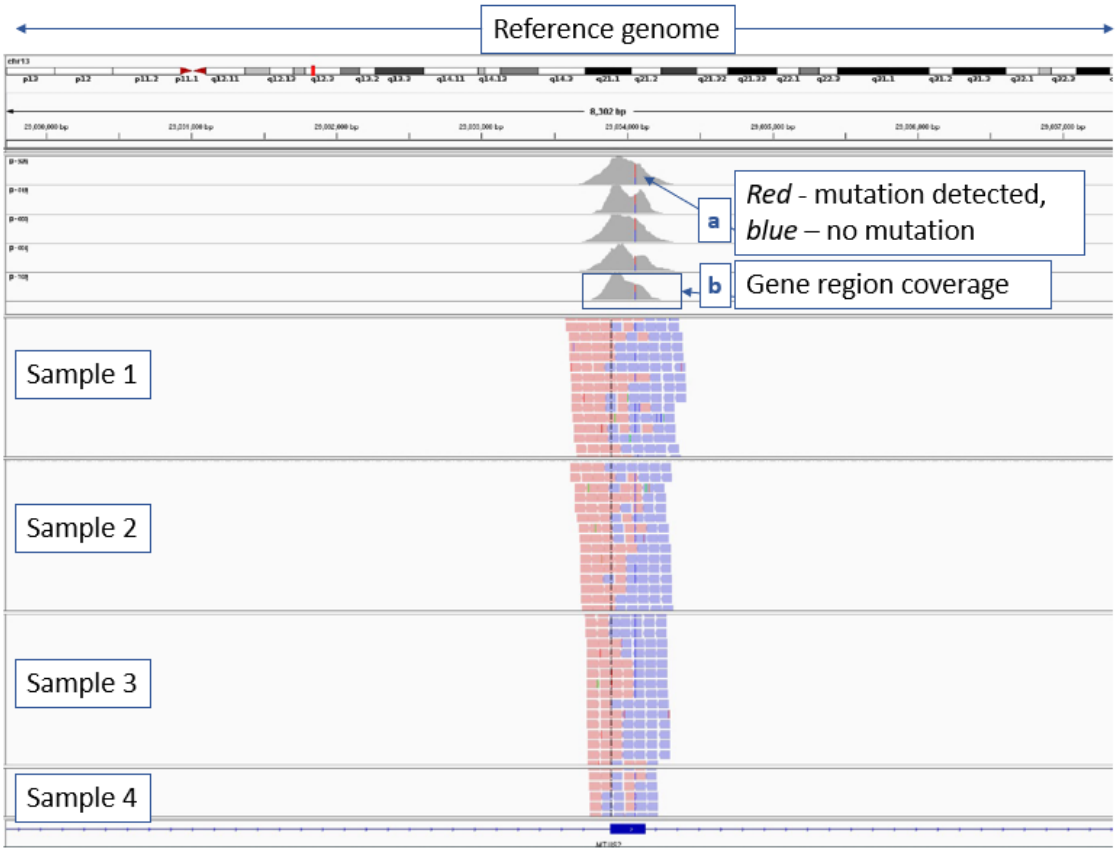


Figure I4. Scheme showing targeted sequencing results via the Integrated Genomic Viewer (IGV). (**a**) Allele frequency – mutation penetration presented in the specific position by a colored line: red – mutated versus blue- wild type consistent with the reference genome; (**b**) Sequencing reads coverage in a specific gene region in distinct samples.

Another technique of DNA sequencing is known as ion semiconductor sequencing, which relies on the detection of hydrogen ions generated during the DNA polymerization process. Hydrogen ions are detected in ion-semiconductor sequencing chips. There are four most commonly known devices used for ion semiconductor sequencing: Ion Personal Genome Machine (PGM), Ion Proton, Ion S5 XL, and Ion S5 system. Disposable chips with a compact array of micro-sized wells are used for performance of massive parallel sequencing reactions in case of PGM technology. Each well contains an individual DNA template and a special ion-sensitive layer that is covering ion sensor that is located below each of the well. This sensors recognize the ions and the change in the solutions pH and if the output voltage is doubled, that means that there are two identical bases on the DNA strand, so the chip reports two identical bases. As there is no optical detection needed, the ION Torrent sequencing technique is very fast.

This technology can allow sequencing of 1–10 Gb with a read length from 150 up to 200 bases, the sequences however are prone to errors (that can range up to 1.7%), which increases in the case of long reads[24].

Both Illumina and Ion Torrent regularly update their platforms by improving sequencing chemistry, nucleotide detection, and throughput. The constant re-assessment of their relative performances is of particular importance as researchers adapt both of these sequencing platforms for use in the clinical setting [24]. For instance, a comparison of the sequences of four microbial genomes produced by MiSeq (Illumina), Pacific Biosciences, and Ion Torrent (PGM) technologies revealed nearly excellent coverage on GC-rich, neutral, and somewhat acceptable coverage on AT-rich genomes. On the Ion Torrent, if a very AT-rich microbial genome was sequenced, significant bias was identified that left around 30% of the genome with no coverage. Ion Torrent data contained slightly more variants than MiSeq, but at the expense of higher false positive rate. Pacific Biosciences data calling required higher coverage depth. Context-specific errors have not been observed when Pacific Biosciences platform was used, contrary to both PGM and MiSeq[25].

1.4. Illumina sequencing platforms

Sequencing platforms from Illumina represent groundbreaking, innovative, high throughput next-generation sequencing systems, that consist of pumps and laminar flow hydraulic systems that assist in performing chemical reactions, vacuum elements for control of flow cell positioning, lasers, and optical systems that allow fluorescent nucleotide sequence registration, temperature control modules that cool the reagents and reaction environment, electrical components (motherboards) and computer control interface that consist of both: Linux and Windows module. This state-of-the-art technological development enables the sequencing of a full genome of 48 people within 44 hours on one machine (Novaseq6000).

The presented PhD project used sequencing on Hiseq1500, the technology that was released in 2012. Hiseq1500 can operate in two modes: Rapid Run and High output. Rapid run can take 16-60 hours (depending on the selected read length that can go up to 2x250 bp) and generate up to 150 million reads per lane with 2 lane flow cell giving a total of 300 million reads per sequencing. The device

used in this project after concentration input optimization could reach in Rapid Mode sequencing run up to 400 million reads per experiment. Hiseq1500 has also an option of High Output mode which involves larger channel 8 lane flow cells and can generate up to 190 million reads per lane in a sequencing run that takes 3-11 days. Hiseq1500 technology includes flat, non-patterned flow cells and 4 fluorescent color signal chemistry and according to product guidelines can generate up to 760 million reads or fragments in an experiment.

New device - Novaseq6000 - was released in 2017 and this modern technology can yield up to 20 billion reads in sequencing that can take up to 48 hours. There are 2 cell line ports that can sequence simultaneously and 4 types of flow cells that can occupy each port. Flow cells are ranging from SP flow cell (2 narrow lanes of this flow cell can generate reads up to 2x250 bp read length and 800million reads) up to S4 flow cell (4 wide lane flow cell that can generate reads up to 2x150 bp read length and 10 billion reads). Technology involves patterned flow cells that have microwells which supposed to decrease problems connected to overcrowding, such as overlap of the signal from 2 different clusters as only microwells are covered with oligos that bind libraries. The new technology involves 2 fluorescent color signal chemistry.

The most modern sequencer Novaseq X from Illumina was released on 29 September 2022. It has the capacity to yield from 1.6 to 52 billion reads and the sequencing takes up to 52 hours. This is the most modern and the highest throughput next-generation sequencing technology that is currently available.

1.5. Alignment of sequencing data

The next step was the analysis of NGS data by the sequence alignment, which consists of comparing 2 (pairwise) or more (multiple) nucleotide sequences against a reference. Illumina obtained reads are usually within the range of 25-250 nucleotide long sequences. Reads can be obtained as “single reads” or “paired reads”, in which case they are representing sequence read from both extremities of the same nucleotide fragment (200-1,000 bp long). Acquired fluorescent signals were translated by an internal Illumina software (CASAVA) into base calls, that are represented in the FASTQ format [26], where each nucleotide is assigned to an ASCII-encoded quality number which is directly corresponding to a PHRED score (Q). PHRED score (Q) is directly translated into the probability p that

the corresponding base call is incorrect by the following equation: $p=10(-Q/10)$ where Q ranges from 0 up to 41, which results in the error rate that ranges from 1 to 0.000079433 at each position. The preprocessing step known as read "trimming" removes low quality regions and involves different algorithms, implementations and tools. Ignoring the low quality base calls is instrumental for any NGS analysis, as it can add unreliable and potentially random false sequences to the dataset, and lead to unrealistic interpretations of data[27].

Upon receiving NGS data from DNA sequencing, there are effectively three major phases of genomic bioinformatic analysis. During the *Alignment* (the first phase) reads are *aligned* or *mapped* to the reference genome, which allows determination of precise location of each base pair in the genome. The second phase (*Variant Calling*) uses those mapped reads to identify genetic variation at different locations in the genome. In the final phase (*Analysis*) the outputs of the first two phases are related to the potential functionality of a gene. This final phase is the most challenging and requires understanding of population genetics to yield proper and correct results.

Identifying genomic variants, including single nucleotide polymorphisms (SNPs) and DNA insertions and deletions (indels), from NGS data is an important part of scientific discovery. A bioinformatics pipeline has been developed to accurately and rapidly identify, and annotate, sequence variants. The pipeline was based on the Broad Institute's recommended for variant discovery analysis: Genome Analysis Toolkit 4 (GATK4) to perform variant calling. After SNPs have been discovered, SnpEff is used to annotate and forecast the impact of variants (https://gatk.broadinstitute.org/hc/en-us). These tools may be modified to work with different technologies, even though they were initially made to process exomes and the entire genomes produced by Illumina sequencing. The GATK has grown to be able to handle genomic data from any organism, with any amount of ploidy, despite initially being intended particularly to research the human genome.

1.6. Alterations in tumor DNA at the origin of tumorigenesis

DNA of the cancer cell contains certain changes that deregulate intracellular processes and drive disease progression. Cancer is driven by few unique features of its genotype leading to specific phenotype characteristics. The whole spectrum of alterations: single nucleotide mutations, deletions, DNA amplifications, genome

rearrangements, gene fusion, copy number changes have been detected in the cancer genomes. These alterations occur in oncogenes, tumor suppressor or regulatory regions that control expression of important genes[27].

The genetic alterations detected in cancer patients could be germline (cancer predisposing, causative- BRCA1/2 for example- present in healthy tissues) or somatic (cancer causative or passenger mutations detectable only in cancer tissue). A reference material is necessary to determine and confirm somatic variants in cancerous tissues, because a great number of DNA alterations are specific to the particular patient hereditary genetic background. Somatic mutations by definition are the ones that occur in cancer cells. Most of the cancer related alterations have been detected thanks to advances in technologies of DNA sequencing, increasing throughput, lowering costs of equipment and reagents, implementation of novel and more efficient methods of data analysis, including artificial intelligence.

It has been reported that besides DNA stored in the genome there are extracellular linear DNA fragments called circulating free DNA (cfDNA) that are 1, 2, 3, or 4 nucleosome size long and are present in all of bodily fluids. cfDNA can be isolated from the whole variety of bodily fluids including: urine, saliva, stool, sputum, pleural fluid, CSF, or peripheral bloods plasma[28,29]. While plasma seems to be most popular source of cfDNA in prenatal paternal or genetic testing, in cancer diagnostics a variety of bodily fluid sources apply have been applied as liquid biopsy. cfDNA extraction from saliva is used in the case of lung, or head and neck squamous carcinoma patients, urine in the case of bladder, prostate, or colorectal cancer patients, and cerebrospinal fluid (CSF) in the case of medulloblastoma, ependyma, glioma or metastatic cancers. In this study cfDNA from blood was analyzed and this issue will be discussed in more details in the liquid biopsy sections 4.7.2 - 4.8.

1.7. Cancer Databases

A reference DNA isolated from the whole blood or other source of healthy tissue is used in the NGS analysis of cancer samples, and it is often necessary to distinguish cancer specific variants. Healthy tissues contain patient specific, common, or irrelevant variants/SNPs that need to be excluded from the analysis. Tumors frequently contain the whole spectrum of mutations, some of them

are pathogenic and can cause a phenotype switch of a cell from normal to malignant. Some mutations that can be only found in cancer genome but not in the reference DNA, could be benign and represent passenger mutations, rather than driver oncogenic variants. Over the years numerous oncogenic, sometimes actionable mutations that have specific therapy recommendations by the World Health Organization (WHO) were discovered and have been registered in databases.

The National Library of Medicine from the National Center for Biotechnology Information (NCBI) is one of publicly available databases that includes a variation viewer with the option of filtering only pathogenic variants (figure I5).

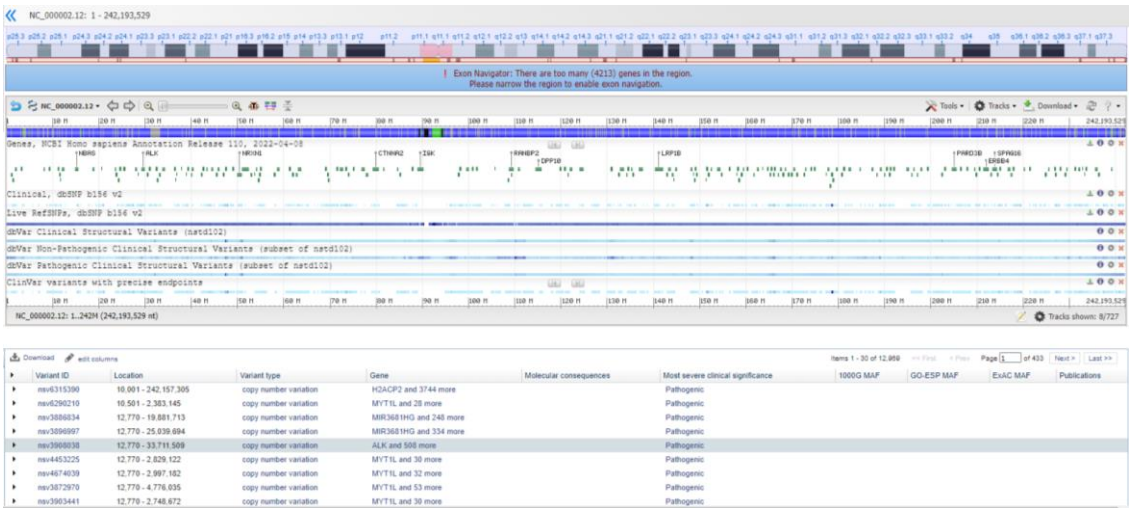


Figure I5. Scheme showing general review of pathogenic variants within the scope of Chr2, registered at NCBI (accesed on 17.2.2023: https://www.ncbi.nlm.nih.gov/variation/view/).

NCBI is a great source of information on clinical significance of mutations, and it classifies variants into 8 different groups: pathogenic, likely pathogenic, uncertain significance, likely benign, benign, conflicting, and not provided. Based upon the reference genome localization or RS ID number (unique label used in most genomic databases to identify the specific SNP) variants can be reviewed as shown in the figure I6 for the *IDH1* gene.

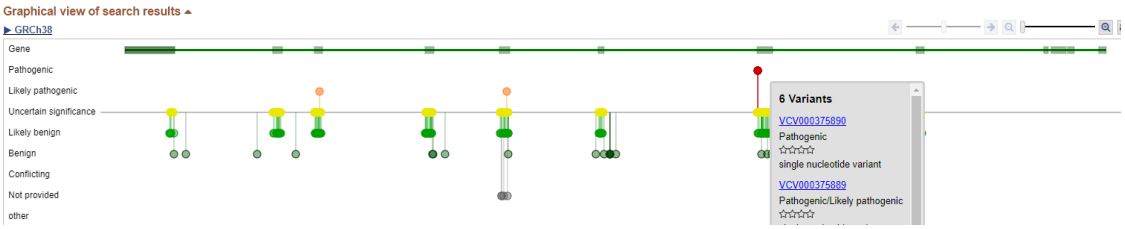


Figure I6. Scheme showing general review of *IDH1* registered variants and their clinical significance clasification provided by NCBI (accesed on 17.2.2023: https://www.ncbi.nlm.nih.gov/clinvar/?gr=1&term=idh1%5Bgene%5D&redir=gene).

Other database that contains a list of registered somatic variants, overall survival information, as well as the frequency of mutation detected for a specific diagnosis is the Genomic Data Commons Data Portal (GDC Data Portal), provided by the National Cancer Institute (accesed at 20.2.2023: https://portal.gdc.cancer.gov/). Data provided by this web tool can be reviewed based upon disease type, primary tumor site, or by the project program: The Cancer Genome Atlas (TCGA), Clinical Proteomic Tumor Analysis Consortium (CPTAC) or Human Cancer Models Initiative (HCMI) for example.

Another commonly used database that contains registered diagnostically relevant variants is the Catalog of Somatic Mutations in Cancer (COSMIC). COSMIC contains data from both scientific literature and the Cancer Genome Project (Sanger Institute)[30,31]. Data provided by the COSMIC website is publicaly available for free to academic researchers. The COSMIC database contains a broad information on specific localization of mutations registered in cancer patients and cancer-derived cell lines. Additionally, actionable variants list can be uploaded, which presents currently ongoing clinical trials and already aproved medication lists for targeted therapy of patients having specific mutations important for a given cancer diagnosis. The example below shows a small fraction of the data that is avaliable under uploaded actionable variants files regarding mutations of the *EGFR* (coding for the epidermal growth factor receptor) along with the proposed treatment of gliomas and large intestine adenocarcinoma (table I1).

Mutation in gene:	DISEASE	STATUS	DRUG COMBINATION	TESTING REQUIRED
EGFR	central nervous system / glioma / astrocytoma Grade IV	Phase 1	Afatinib	
EGFR	central nervous system / glioma / astrocytoma Grade IV	Phase 2	Afatinib + Temozolomide	
EGFR	central nervous system / glioma / astrocytoma Grade IV	Phase 1	BDTX-1535	
EGFR	central nervous system / glioma / astrocytoma Grade IV	Phase 1	CM93	
EGFR	central nervous system / glioma / astrocytoma Grade IV	Phase 2	Dacomitinib	
EGFR	central nervous system / glioma / astrocytoma Grade IV	Phase 3	Depatuxizumab mafodotin + Temozolomide	
EGFR	central nervous system / glioma / astrocytoma Grade IV	Phase 1	ERAS-801	
EGFR	central nervous system / glioma / astrocytoma Grade IV	Phase 2	Erlotinib	
EGFR	central nervous system / glioma / astrocytoma Grade IV	Phase 1	Gamma-retroviral MSGV1 139 scFv EGFRvIII CAR gene-modified T cells	
EGFR	central nervous system / glioma / astrocytoma Grade IV	Phase 2	GC1118	
EGFR	central nervous system / glioma / astrocytoma Grade IV	Phase 1	RO7428731	
EGFR	central nervous system / glioma / astrocytoma Grade IV	Phase 2	Tesevatinib	
EGFR	central nervous system / glioma / astrocytoma Grade IV	Phase 2	Verteporfin	
EGFR	central nervous system / glioma / NS	Phase 1	AMG404 + AMG596	

EGFR	central nervous system / glioma / NS	Phase 2	Cyclophosphamide + EGFRvIII CAR-T cells + Fludarabine + Interleukin-2	
EGFR	central nervous system / glioma / NS	Phase 1	EGFRvIII CAR-T cells	
EGFR	central nervous system / glioma / NS	Phase 2	Irinotecan + Panitumumab	
EGFR	central nervous system / glioma / NS	Unknown	MAb-425	
EGFR	central nervous system / glioma / NS	Phase 2	Poziotinib	
EGFR	central nervous system / NS / NS	Phase 1	EGFR806-CAR T cells	
EGFR	large intestine / carcinoma / adenocarcinoma	Approved FDA	Cetuximab	Required
EGFR	large intestine / carcinoma / adenocarcinoma	Approved FDA	Cetuximab + FOLFIRI	wt KRAS
EGFR	large intestine / carcinoma / adenocarcinoma	Approved FDA	Cetuximab + Irinotecan	wt KRAS
EGFR	large intestine / carcinoma / adenocarcinoma	Approved FDA	FOLFIRI + Ramucirumab	Not required
EGFR	large intestine / carcinoma / adenocarcinoma	Approved FDA	Panitumumab	wt KRAS

Table I1. Scheme showing an example of the COSMIC database actionability table for mutations found in the *EGFR* gene (accesed on 20.2.2023: https://cancer.sanger.ac.uk/cosmic/license).

1.8. Basic facts about gliomas

Gliomas are central nervous system (CNS) tumors that arise from glial progenitors or neural stem cells[32,33]. Malignant gliomas are the most common, primary brain tumor, with glioblastoma (GBM) being the most lethal glioma, with median patient survival time of 15 months; 5 year glioblastoma survival rate is below 7%[34–37]. Complete surgical resection is often not possible as malignant gliomas (WHO grade 3-4) are characterized by highly infiltrative and invasive growth which leads to unavoidable reoccurrence[38]. Glioblastoma treatment includes maximal safe resection, radiotherapy, and chemotherapy with temozolomide (TMZ) and hasn't change for last 17 years[38]. Some advancements in therapy include: the application of tumor-treating electric fields that extend patient survival by 3.9 months as a part of combined therapy with TMZ[39], second line treatment

with bevacizumab[40], virotherapy with oncolytic herpes viruses (RQNestin) that increased tumor infiltrating macrophages, lymphocytes and natural killer cells after viral injection[41,42]. New methods to improve delivery of therapeutics agents through convection-enhanced delivery, blood-brain barrier (BBB) opening, or systemic chemotherapy are being developed[38].

Gliomas are characterized by genomic and cellular heterogeneity. Even within WHO grade 2 or 3 gliomas, there are specific mutations occurring only in certain tumor regions, some mutations are shared by certain areas, while others are common to the whole tumor[43]. The most common and diagnostically relevant are specific mutations in the isocitrate dehydrogenase encoding gene *IDH1* (IDH1 Arg132His) which currently serves to diagnose Glioblastoma as a *IHD* wild type tumor[44] and predicts better survival. The 2021 WHO Classification of Central Nervous System Tumors that included the genetic features categorized 43 different diagnostically classified primary brain tumors according to histopathological features and the mutations that a tumor is carrying. Some tumor types are categorized based upon a single gene alteration, while in case of other tumors diagnosis is based on a set of alterations in multiple genes. Mutations in *IDH1, IDH2, ATRX, TP53, CDKN2A/B* are present in astrocytoma. Mutations in genes coding for Tumor protein 53 (*TP53*) and transcription regulator ATP-dependent helicase (*ATRX*) are strongly associated with astrocytomas, whereas 1p/19q codeletion and telomerase reverse transcriptase gene (*TERT*) mutation are associated with oligodendrogliomas. *EGFR* amplifications, cyclin dependent kinase inhibitor 2A (*CDKN2A)* deletions, phosphatase and tensin homolog *(PTEN)* and neurofibromin 1 (*NF1*) mutations and Chromosome 10 loss are also common[45].

Glioblastoma is a common primary brain tumor, defined by a lack of *IDH* mutations (IDH-wildtype) and frequent mutations within the *TERT* gene promoter, variations in chromosomes 7/10, and alterations in *EGFR.* This classification presents how important for diagnostics and potential anti-tumor therapy is a proper analysis of the DNA mutation spectrum present within the tumor.

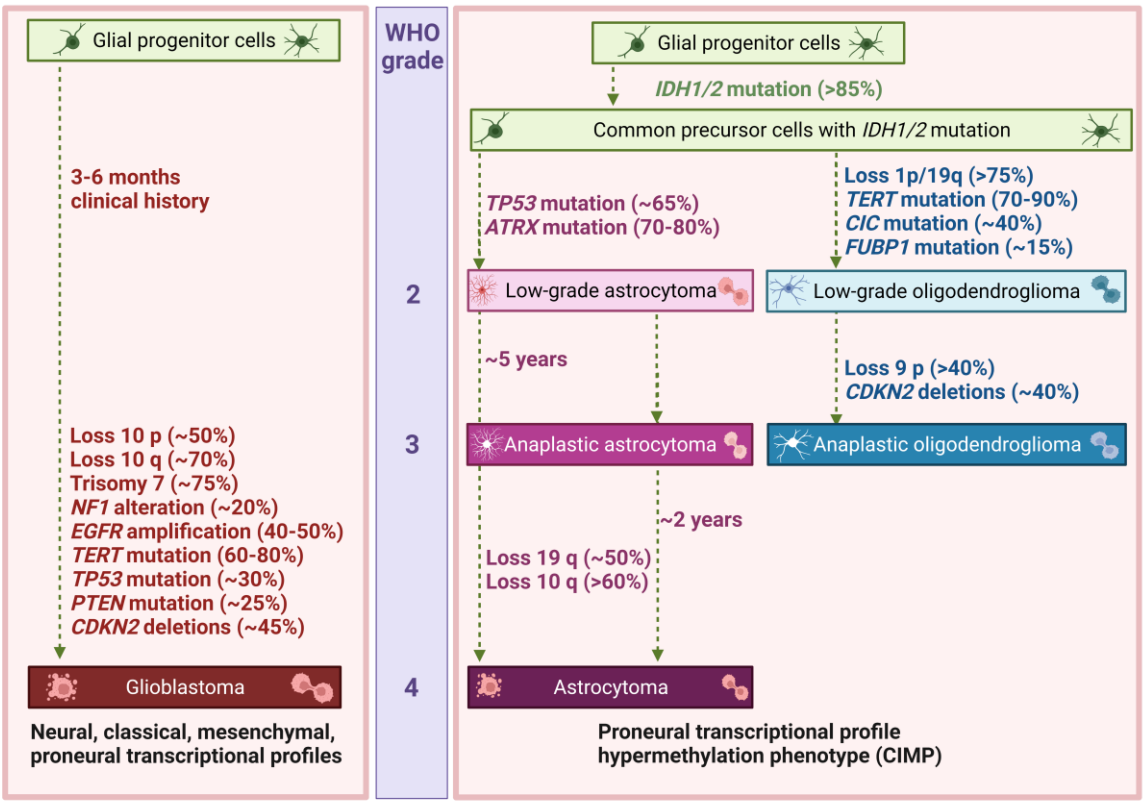


Figure I7. Classification of gliomas based upon genetic alterations (prepared in Biorender adapted from [46]).

As shown in the figure I7, gliomas grade 4 are classified as either *de novo* (primary) glioblastoma or secondary (secondary) grade 4 astrocytomas, which develop from low-grade astrocytomas (WHO grade 2) or by malignant transition from anaplastic astrocytoma (WHO grade III). These two Grade 4 glioma types exhibit distinct genomic changes. *IDH1/2* mutations are frequently seen in WHO grade 2 and 3 gliomas, and in 10% of Grade 4 gliomas. *IDH* mutations are early events that occur before codeletion of chromosomes 1p and 19q in oligodendrogliomas or *TP53* mutation in astrocytomas.

Other important epigenetic modification that can be used as the personalized therapy recommendations is methylation of the *MGMT* gene promotor. *MGMT* encodes the O^6-methylguanine methyltransferase which participates in the DNA damage repair. Individuals with the methylated *MGMT* and subsequent low level of the enzyme had improved survival if treated with an alkylating agent TMZ[45].

Another classification of GBMs (which has a prognostic value) is based upon the bulk RNA sequencing and categorizes gliomas into three different subtypes: mesenchymal (MES), classical (CL), and proneural (PN). CL has predominantly

alterations in *EGFR* (Epidermal growth factor receptor), MES contains mutations in *NF1* (Neurofibromin), whereas the PN subtype includes *PDGFRA* (Platelet-derived growth factor receptor alpha) amplification, *IDH1* and *p53* alterations[47]. This classification is considering both major genetic alterations and transcriptomic signatures which encompasses expression of many genes typical for the microenvironment (immune and stromal genes). GBM patients with the MES signature have the shortest survival among other GBM patients[48].

Many studies have shown that GBMs are very heterogenous[43,49–51]. RNA expression subtypes of gliomas can coexist in one tumor occupying different regions[43,49–51] and a subtype can change over time due to immune response and through the course of therapy[52,53].

1.9. Overview of liquid biopsy

Liquid biopsy is a laboratory test performed on a sample of blood, urine, or other body fluid to look for cancer cells or tumor derived nucleic acids (small pieces of DNA, RNA), proteins, or extra cellular vehicles (EV) released by tumor cells into a person's body fluids. Liquid biopsy allows multiple samples to be taken over time, helping to understand which genetic or molecular changes are appearing in a tumor. A liquid biopsy may be used to detect cancer at an early stage. It may also be used to plan treatment or to find out how well treatment is effective or if cancer recurs.

Circulating DNA was first described by Mandel[54] in 1948. In 1994 a discovery of the mutated DNA in human blood of cancer patients[55] led a field of liquid biopsy in the new direction. For years this method has not been developed well enough to identify the difference between healthy and tumor DNA. The first discovery of fetal cell DNA in mother blood[56] in 1997, clearly presented that circulating DNA originated from different sources can be distinguished, which led to technological developments in this field. The primary goal of medical research and advances in liquid biopsy was to help improve the efficacy of cancer cell DNA detection in the human body at early stages, through a non-invasive test of bodily fluids. Another objective was to provide a fuller spectrum of cancer characteristics from inaccessible tissues and provide the tools that can be used to assess tumor heterogeneity and monitor disease progression[57].

Circulating cell free DNA (cfDNA) are degraded DNA fragments that are released into the blood and all other bodily fluids. A source of cfDNA may vary, as it can originate from normal or malignant cells through different processes: apoptosis, necrosis or secretion[58], as shown in the figure I8.

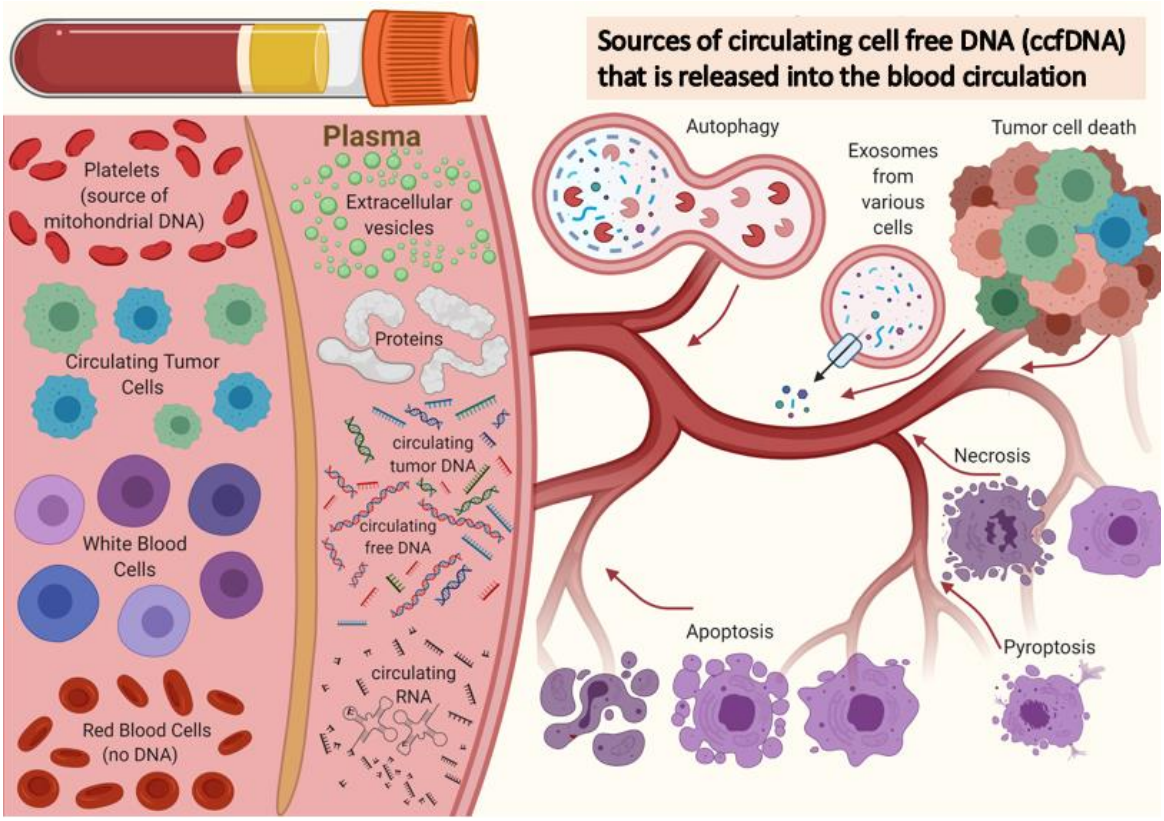


Figure I8. Main sources of circulating cell free DNA (cfDNA) that is being released into the peripheral blood circulation. A scheme prepared using Biorender.

Depending on the mechanism of release, cfDNA can be found in particular structures (microparticles, apoptotic bodies, exosomes) or macromolecular structures (DNA traps, nucleosomes, linked with serum proteins, proteolipid nucleic acid complexes, linked with cell-free membrane parts)[59]. Size of the released fragments of cfDNA can be determined using many different methods: quantitative polymerase chain reaction (qPCR)[59,60], droplet digital PCR (ddPCR)[61], Atomic Force Microscopy (AFM)[62], Massive Parallel Sequencing (MPS)[63], microchip-based capillary electrophoresis[64], agarose gel electrophoresis[65]. Elevated levels of cfDNA are observed in: advanced cancer, sepsis, myocardial infraction, trauma, pregnant woman and transplant graft rejection and all these conditions can influence the size

distribution of released fragments[66–68]. Usually, free cell DNA circulates in fragments ranging between 120-220 bp, with the peak at 167 bp but dimers and trimers of this length also can be found[67]. The size relates to the DNA fragment wrapped around a nucleosome once or multiples thereof. Estimated half-life of cfDNA in circulating blood varies from 2 minutes up to 2 hours[69], this rapid turnover allows a current "snapshot" of the tumor mutational landscape. cfDNA is not only a mixture of DNA fragments originating from both germline and malignant cells, it is also a mixture of nuclear and mitochondrial DNA, which impacts it's structure and stability[59].

Consequently, bodily fluids used in order to extract circulating DNA include blood, saliva, urine, cerebrospinal fluid, semen, and mucus amongst others. While in this dissertation we focus on the circulating free DNA isolated from peripheral blood, there are other types of liquid biopsies that should be mentioned. Circulating tumor cells (CTC) and clusters can be isolated from blood of metastatic cancer patients and assist in indicating a clonal source of currently existing or newly developing metastatic cancer sites. They are extremely rare as it was estimated that there is only one CTC per billion of normal blood cells[70]. Circulating tumor cells have a short half-life in the circulation[71,72] and after that time these cells burst and tumor DNA is released into circulation. Such behavior makes metastatic cancers, that spread through blood circulatory system, the best candidates for ctDNA liquid biopsy taken from blood, as in this case blood can be considered a steady source of tumor DNA. ctDNA isolated from patient blood was shown to be a comprehensive source of mutational signatures from all metastatic sites, as it was shown in the remarkable case study of the gastrointestinal metastatic patient by Parkih and coauthors[73].

On the other hand, liquid biopsy from blood can present an incomplete spectrum of malignant DNA markers, offering a snapshot of oncogenic mutations from angiogenic tumor regions, rather than from hypoxic regions with poor blood circulation. Presence of malignancy in the human body can be indicated by presence of specific markers that can be detected in bloods serum which is used in current medical practices. For example, overexpression (amplification) of *HER2* indicates a necessity of more effective trastuzumab treatment for breast cancer patients[74]. CEA – Carcinoembryonic antigens can be detected in blood and their high concentrations can be used to indicate the presence of certain cancers, i.e. colon cancer. High levels of CEA are also present in patients with ulcerative

colitis, pancreatitis, cirrhosis of the liver[75], or in heavy smokers[76]. Unfortunately, the CEA-based test is not reliable for an early detection screening[77], but it can be used as a monitoring tool during colorectal carcinoma treatment[78]. Other markers that can indicate a cancer presence are: Alpha-fetoprotein (AFP), beta-human chorionic gonadotropin (β-hCG), Neuron-specific enolase (NSE), prostate specific antigen (PSA), carcinoma antigen 15-3 (CA 15-3), carcinoma antigen 19-9 (CA 19-9), carcinoma antigen 125 (CA125), estrogen receptor (ER) and progesterone receptor (PgR), human epidermal growth factor receptor 2 (HER2), Terminal deoxynucleotidyl transferase (TdT), urokinase plasminogen activator (uPA) and plasminogen activator inhibitor-1 (PAI-1), carcinoma antigen 72-4 (CA 72-4), nuclear matrix protein 22 (NMP22) and others. As tests for most of protein markers, hormones, or antigens are less challenging, detection of circulating tumor DNA remains a challenge, as natural characteristic of circulating DNA makes its extraction difficult due to its low concentration in the blood plasma[59].

All other bodily liquids can also be a source of cfDNA or CTC's, depending on the tumor localization. Abnormal cells can be found in the urine of bladder cancer patients, which is utilized in a urine cytology diagnostics[79]. As with most diagnostic non-invasive tests it has a lower detection sensitivity in case of low grade tumors (20-60%), but higher sensitivity of detection of high grade tumors (80%)[80–82]. This accounts for earlier detection of urinary tract cancers through high prevalence of patients and better treatment outcomes. On the other hand, the composition of saliva includes 99% water and 1% protein, which makes it effectively sufficient to conduct diagnostic tests for oral cancer[83]. Tumor cells release cfDNA that can be easily accessed through the salivary glands. In addition, the efficacy of cancer tests using saliva as compared to other forms of liquid biopsy suggests that saliva constitutes a larger component of circulating DNA. In this aspect, circulating DNA detection at an early stage improves outcomes of cancer treatment. As a result, screening methodologies that exhibit high sensitivity are preferred in the scheduling of treatment plans and recurrences in treatment for oral cancer patients[83]. Sputum or mucus are a reliable source of cell-free DNA used for *EGFR* mutation detection in advanced lung adenocarcinoma[84]. On the other hand, seminal cell-free DNA can be used as an assessment tool for prostate cancer[85]. Reported seminal cfDNA amounts per mL of sperm increased in prostate cancer patients

compared to healthy controls even up to 366 times[85]. Finally, cerebrospinal fluid is a good source of circulating tumor DNA in the case of brain tumors[86]. Lumbar puncture cerebrospinal fluid collection procedure can have serious side effects which might include: CNS infection, increased intracranial pressure, or even leptomeningeal tumor spread[53].

1.10. Liquid biopsy in brain tumors diagnostics

Primary brain tumor patients could greatly benefit from a reliable and noninvasive diagnostic method, as brain tumor biopsy carries a lot of risks such as brain swelling, bleeding of the brain, stroke, infection, blood clots, seizures, or even coma. For patients with primary brain tumors two bodily fluid sources of ctDNA were considered: blood plasma and Cerebral Spinal Fluid (CSF). While many studies have shown that blood brain barrier (BBB) in glioma patients is affected by excessive angiogenesis [87] clinical evidence showed a poor BBB permeability in GBMs[88]. BBB is likely one of the main reasons why most studies show that CSF is a better source of ctDNA in case of brain tumors[53,86,89,90]. However CSF collection through lumbar puncture can lead to complications Therefore, a non-invasive blood collection remains a safer alternative to lumbar puncture. Study published in January 2023[91], designed with the goal of comparison of effectivity in detection of ctDNA isolated from plasma, lumbar puncture CSF, and peritumoral CSF using NGS or ddPCR reflected that plasma was not a viable source of ctDNA. It demonstrated that 7/7 samples of plasma were negative for ctDNA signal, as were cellular fractions isolated from CSF. Peritumoral CSF had a pathogenic mutation in gene encoding for an enzyme phosphatidylinositol 3-kinase (*PIK3R1*) that was not detectable in the tumor tissue[91]. *MGMT* promoter methylation was also analyzed using beaming PCR technique and also in this case results from the tumor biopsy were differing from these obtained from cfDNA derived from CSF[91].

There are many ongoing studies that evaluate application of liquid biopsy in the clinical use for glioma patients[92] (mainly in USA, France, Canada, China, Switzerland, and the UK) and these clinical studies use blood, rather than CSF (which is only used in some cases if available for comparison). Studies vary, as some include whole genome sequencing (WGS), whole exome sequencing (WES), RNA-seq, or targeted panel sequencing. Most of the clinical studies

are going to finish in 2024[92] and the main goal is to achieve reliable monitoring method for diagnosed patients.

At the Jonsson Comprehensive Cancer Center, University of California (UCLA), Los Angeles, USA, ongoing phase I trial for recurrent glioblastoma patients combines vaccine and monoclonal antibody and is monitored by liquid biopsy (NCT04201873). British Tessa Jowell BRIAN MATRIX study encompasses 1,000 patients diagnosed with gliomas (WHO grades 2-4) and includes molecular analysis by: epigenomic classification and WES of matched tumor and blood samples (NCT04274283). Different ongoing liquid biopsy trial is taking place in Lyon, France, and includes 150 participants with gliomas (WHO grades 2-4) and monitors only 3 oncogenic markers: *TERT*, *IDH*, and *ATRX* mutations (NCT04931732). Another exciting study will monitor the amounts of cfDNA released into blood circulation of glioblastoma patients before and after BBB disruption, expecting 2-fold increase in cfDNA (NCT05383872). There was a case study with a patient affected with butterfly glioblastoma, in which opening of BBB was followed with intra-arterial delivery of bevacizumab which led to tumor shrinkage as a second line therapy treatment[93]. Opening BBB for both drug delivery and liquid biopsy can assist to current therapeutic approaches.

2. Aims and Goals

Since 2021 when the new diagnostic WHO classification of gliomas was redefined[94], molecular profiling of the tumors had become a necessity. Heterogeneity of gliomas[49,95] indicates limitations to classical tumor biopsy or even histopathological diagnosis. The transcriptomic classification of glioblastoma subtypes at the single-cell level showed neural-progenitor-like (NPC-like), oligodendrocyte-progenitor-like (OPC-like), astrocyte-like (AC-like) and mesenchymal-like (MES-like) signatures confirming a heterogenous nature of gliomas[92]. Tumor evolution during the treatment and therapy[52,96] underlines the importance of molecular profiling during patients care. Single-cell sequencing study that compared tumor-derived explants, cell lines and original tumors showed the retention of intratumor heterogeneity in tumor organoids and clonal evolution of patient-derived cell lines[97].

Liquid biopsies had revolutionized the field of clinical oncology, offering easier tumor sampling, continuous monitoring by repeated sampling, devising personalized therapeutic regimens, and screening for therapeutic resistance. Methods for isolation and analysis of cfDNA for diagnostic sequencing have rapidly evolved over recent years providing abundant details regarding tumor progression, grading, heterogeneity, gene mutations and clonal evolution that pave an avenue for precision medicine. There are numerous controversies in the field of glioma diagnostics regarding a feasibility and usefulness of cfDNA from blood for diagnosis. In this PhD thesis we aimed to improve GBM diagnostics and we addressed several questions including whether cfDNA from serum a viable diagnostically relevant method for GBM detection and monitoring. We assessed if transcriptomic characterization of GBM-derived cells lines can be used to improve GBM diagnostics. We also employed targeted NGS to DNA from FFPE tumor sections to identify potential pathogenic genetic alterations in a set of poorly diagnosed pediatric gliomas.

The specific aims of this study were the following:

1. Collecting the brain tumor patient cohort and matched samples of resected tumors and blood from 100 patients.

2. Optimization of cfDNA isolation, quantification, library preparation and sequencing.
3. Detecting genetic alterations in cfDNA from blood of GBM patients using NGS.
4. Evaluating a spectrum of somatic mutations in tumors and corresponding GBM-patient derived cell cultures maintained at normoxia and hypoxia.
5. Evaluating a spectrum of somatic mutations in poorly characterized pediatric brain tumors.
6. Assessment of the advantages and disadvantages in sequencing and analysis of different types of samples (fresh frozen tumors, cell cultures, FFPE samples).
7. Assessment of the advantages and disadvantages of diagnostic methods based upon molecular analysis of different types of samples (fresh frozen tumors, cell cultures, FFPE samples).

3. Materials and Methods

3.1 Overview of the Project

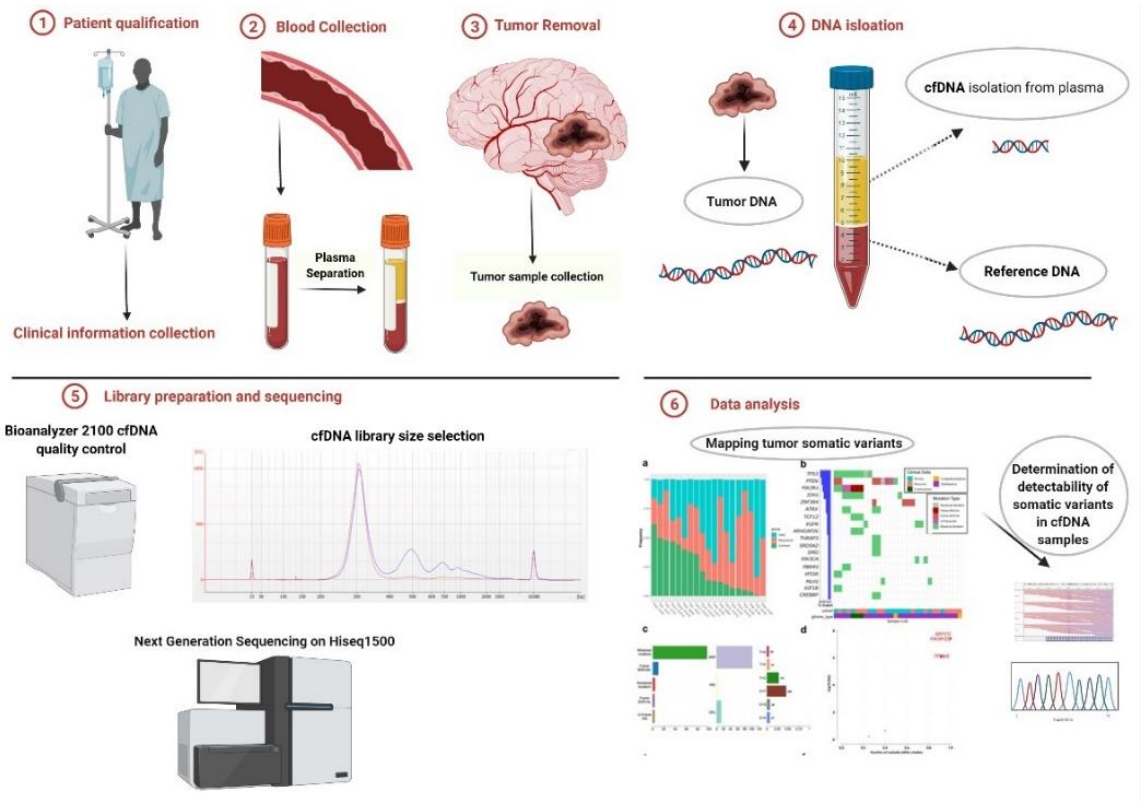


Figure M1. The overview of the liquid biopsy project. Scheme prepared in Biorender.

Patients have been recruited to the surgery clinics. Each patient included in the study gave a written consent for the use of their blood and tumor samples. All the procedures that involved human participants were performed in accordance with the institutional ethical standards and were approved by the ethics committee of the Medical University of Silesia (KNW/0022/KB1/2/I/17). Freshly frozen resected tumors and corresponding blood samples were collected from brain tumor patients. In some cases 10 mL blood samples were collected before and after surgery.

DNA was isolated, libraries were prepared and sequenced then data was analyzed. In the NGS analysis, pre-surgery plasma-derived cfDNA collected from 84 patients was used: 80 patients with WHO G3 and G4 gliomas, 2 patients with primary central nervous system lymphoma (PCNSL), and 2 patients with anaplastic thyroid cancer metastasis and adenocarcinoma lung metastasis.

Copy number alteration targeted sequencing was additionally performed on 4 cfDNA samples that displayed the ctDNA signal in the primary analysis. Overall scope of the liquid biopsy project is described in figure M1.

3.2 Adult Patients Cohort characteristics

The cohort included 126 patients, from which clinical data including age, sex, symptoms, and diagnosis were collected. Majority of the cohort was composed of gliomas: Grade 4 Glioblastoma (n=92), Grade 3 (n=11), Grade 2 (n=15), Grade 1 (n=1), but there were also PCNSL (2), metastatic cancers (4), and brain aneurysm samples (1). The cohort with complete information regarding matched tumor and whole blood reference DNA consisted of 89 pairs. Due to hemolysis some blood samples were not viable for cfDNA isolation, others did not pass quality control. Library preparation of complementary sets of preoperational cfDNA, tDNA and gDNA was successful for a complete cohort of 84.

Complete clinical information for 126 patients is presented in the table M1, below.

Patient ID #	Sex	Age	Grade	Histopathology diagnosis
17	Male	21	1	Astrocytoma pilocytic
2	Male	52	2	Oligoastrocytoma
11	Female	31	2	Diffuse glioma
26	Female	41	2	Diffusive Astrocytoma
27	Male	59	2	Oligodendroglioma
41	Female	60	2	Diffusive Astrocytoma
89	Male	44	2	Oligodendroglioma
92	Male	41	2	Glioma
77	Male	61	2	Glioma
99	Male	66	2	Diffusive Astrocytoma
50	Male	60	2	Diffusive Astrocytoma
57	Female	33	2	Diffuse Astrocytoma
122	Male	77	2	Diffuse Astrocytoma
8	Female	69	2	Oligodendroglioma partim anaplasticum
18	Male	35	2	Diffuse astrocytoma partially anaplasticum
105	Male	41	2	Pleomorphic Xantoastrocytoma
20	Male	66	3	Glial tumor
21	Male	59	3	Oligodendroglioma
23	Male	45	3	Glial tumor
32	Male	37	3	Oligodendroglioma Anaplasticum

44	Female	69	3	Anaplastic Astrocytoma
74	Female	67	3	Anaplastic Astrocytoma
28	Female	55	3	Oligodendroglioma Anaplasticum
109	Female	65	3	Anaplastic Pleomorphic Xantoastrocytoma
114	Male	60	3	Astrocytoma
36	Male	58	3	Oligodendroglioma Anaplasticum
69	Male	63	3	Oligodendroglioma Anaplasticum
87	Female	37	4	Glioblastoma
1	Male	44	4	Glioblastoma
3	Female	61	4	Glioblastoma
4	Male	61	4	Glioblastoma
5	Male	71	4	Glioblastoma
6	Female	65	4	Glioblastoma
10	Female	54	4	Glioblastoma
13	Male	75	4	Glioblastoma
14	Female	46	4	Glioblastoma
22	Female	59	4	Glioblastoma
24	Female	56	4	Glioblastoma
30	Female	59	4	Glioblastoma
31	Female	51	4	Glioblastoma
33	Female	32	4	Glioblastoma
34	Female	64	4	Glioblastoma
40	Male	70	4	Glioblastoma
42	Female	40	4	Glioblastoma
43	Female	72	4	Glioblastoma
55	Female	60	4	Glioblastoma
73	Female	55	4	Glioblastoma
90	Female	81	4	Glioblastoma
91	Male	40	4	Glioblastoma
93	Female	64	4	Glioblastoma
94	Male	67	4	Glioblastoma
95	Male	65	4	Glioblastoma
96	Male	40	4	Glioblastoma
108	Female	58	4	Glioblastoma
112	Male	64	4	Glioblastoma
113	Female	71	4	Glioblastoma
15	Female	61	4	Ganglioglioma
16	Female	67	4	Glioblastoma with Oligodendroglioma
35	Male	64	4	Glioblastoma
60	Female	69	4	Glioblastoma
61	Male	70	4	Glioblastoma
78	Male	64	4	Glioblastoma
79	Female	64	4	Glioblastoma with Oligodendroglioma

80	Female	73	4	Glioblastoma
81	Male	64	4	Glioblastoma
82	Female	59	4	Glioblastoma
100	Male	40	4	Glioblastoma
121	Female	36	4	Giant Cell Glioblastoma
29	Male	56	4	Glioblastoma
46	Male	42	4	Glioblastoma
47	Female	45	4	Glioblastoma
48	Female	56	4	Glioblastoma
52	Male	78	4	Glioblastoma
53	Female	71	4	Glioblastoma
54	Female	49	4	Glioblastoma
56	Female	55	4	Glioblastoma
62	Male	64	4	Glioblastoma
63	Male	78	4	Glioblastoma
75	Female	63	4	Glioblastoma
76	Female	56	4	Glioblastoma
97	Male	55	4	Glioblastoma
88	Male	74	4	Glioblastoma
98	Male	62	4	Glioblastoma
110	Female	63	4	Glioblastoma
111	Female	53	4	Glioblastoma
115	Male	48	4	Glioblastoma
116	Male	41	4	Glioblastoma
117	Male	49	4	Glioblastoma
118	Male	55	4	Glioblastoma with a primitive neuronal component
119	Female	65	4	Glioblastoma
120	Male	59	4	Glioblastoma with a primitive neuronal component
123	Male	60	4	Glioblastoma
7	Male	62	4	Glioblastoma, recurrent
9	Male	73	4	Glioblastoma
12	Male	41	4	Glioblastoma
19	Male	62	4	Glioblastoma
25	Male	61	4	Glioblastoma
37	Male	48	4	Glioblastoma
38	Male	70	4	Glioblastoma
39	Female	74	4	Glioblastoma
64	Female	57	4	Glioblastoma
66	Female	48	4	Glioblastoma
67	Male	68	4	Glioblastoma
68	Female	49	4	Glioblastoma
70	Male	51	4	Glioblastoma
71	Male	68	4	Glioblastoma

72	Female	59	4	Glioblastoma
83	Male	71	4	Glioblastoma
84	Male	57	4	Glioblastoma
85	Male	61	4	Glioblastoma
86	Female	66	4	Giant Cell Glioblastoma
101	Male	70	4	Glioblastoma
102	Female	63	4	Glioblastoma
103	Male	60	4	Glioblastoma
104	Male	63	4	Glioblastoma
106	Male	34	4	Glioblastoma
107	Male	69	4	Glioblastoma
125	Female	71	4	Glioblastoma
126	Male	59	4	Glioblastoma
45	Female	70	meta	Metastasis
51	Female	73	meta	Breast cancer metastasis
59	Male	71	meta	Anaplastic Thyroid Cancer Metastasis
65	Female	56	meta	Adenocarcinoma Lung Metastasis
58	Female	60	PCNSL	Primary Central Nervous System Lymphoma
124	Male	64	PCNSL	Primary Central Nervous System Lymphoma
49	Male	67	Aneurysm	Aneurysm

Table M1. Complete adult patient cohort diagnostic information, with the patient number that refers to a recently published study[98].

3.3 Pediatric Patients Cohort

Formalin Fixed Paraffin Embedded (FFPE) specimens from 35 pediatric tumor patients diagnosed in the Children's Memorial Health Institute were collected. DNA was successfully isolated from 34 samples, libraries were prepared and sequenced. After quality control, samples with PCR duplicates above 88% were removed from further analysis which yielded a complete cohort of 32 analyzed samples (specific PCR duplicates are shown in figure R5 in the results section). Additionally, fresh frozen tumor tissues were collected from 18 juvenile pilocytic astrocytomas. All the procedures that involved human participants were performed in accordance with institutional ethical standards and approved by the Ethics Committee of the Children's Memorial Health Institute, 14/KBE/2012, 27/KBE/2020. Detailed information about the cohort is included in table M2.

Patient ID #	Diagnosis	Grade	Age (years)	Sex	Material received:
1	Glioblastoma	4	-	-	FFPE sections
2	Pilocytic Astrocytoma	1	-	-	FFPE sections
3	Ganglioglioma	1	8	Male	FFPE sections
4	Pilocytic Astrocytoma	1	-	-	FFPE sections
5	Pilocytic Astrocytoma	1	-	-	FFPE sections
6	Pilocytic Astrocytoma	1	-	-	FFPE sections
7	Pilocytic Astrocytoma	1	-	-	FFPE sections
8	Ganglioglioma	1	17	Female	FFPE sections
9	Glioblastoma	4	6	Female	FFPE sections
10	Pediatric Type Oligodendroglioma	2	5	Female	FFPE sections
11	Pleomorphic Xanthoastrocytoma	2	17	Female	FFPE sections
12	Ganglioglioma	1	17	Male	FFPE sections
13	Ganglioglioma	1	12	Female	FFPE sections
14	Glioblastoma	4	6	Female	FFPE sections
15	Anaplastic Astrocytoma	3	11	Male	FFPE sections
16	Dysembryoplastic neuroepithelial tumor	1	17	Female	FFPE sections
17	Ganglioglioma	1	17	Male	FFPE sections
18	Ganglioglioma	1	11	Male	FFPE sections
19	Agiocentric glioma	1	7	Female	FFPE sections
20	Carcinoma plexus choroidei	3	1	Female	FFPE sections
21	Pleomorphic Xanthoastrocytoma	2	16	Male	FFPE sections
22	Pediatric high-grade diffuse glioma	3/4	11	Male	FFPE sections
23	High Grade Glioma	3	1 / 12	Male	FFPE sections
24	Glioblastoma	4	5	Male	FFPE sections
25	Pineoblastoma	4	11	Male	FFPE sections
26	Pineoblastoma	4	11	Male	FFPE sections
27	Carcinoma plexus choroid	3	8 / 12	Female	FFPE sections
28	Glioblastoma	4	6	Female	FFPE sections
29	Diffuse Astrocytoma	2	8	Male	FFPE sections
30	Glioblastoma	4	9	Male	FFPE sections
31	Pediatric Glioblastoma	4	8	Male	FFPE sections
32	Anaplastic Ependymoma	3	4	Male	FFPE sections
33	Ganglioglioma	1	14	Female	FFPE sections
34	Glioblastoma H3K27-mutant	4	8	Male	FFPE sections
35	Embryonal tumors with multilayered rosettes	4	3	Female	FFPE sections
317	Pilocytic Astrocytoma	1	-	-	Frozen tissue
501	Pilocytic Astrocytoma	1	-	-	Frozen tissue
512	Pilocytic Astrocytoma	1	-	-	Frozen tissue
562	Pilocytic Astrocytoma	1	-	-	Frozen tissue
585	Pilocytic Astrocytoma	1	-	-	Frozen tissue
639	Pilocytic Astrocytoma	1	-	-	Frozen tissue

704	Pilocytic Astrocytoma	1	-	-	Frozen tissue
731	Pilocytic Astrocytoma	1	-	-	Frozen tissue
733	Pilocytic Astrocytoma	1	-	-	Frozen tissue
759	Pilocytic Astrocytoma	1	-	-	Frozen tissue
792	Pilocytic Astrocytoma	1	-	-	Frozen tissue
805	Pilocytic Astrocytoma	1	-	-	Frozen tissue
868	Pilocytic Astrocytoma	1	-	-	Frozen tissue
899	Pilocytic Astrocytoma	1	-	-	Frozen tissue
982	Pilocytic Astrocytoma	1	-	-	Frozen tissue
411	Pilocytic Astrocytoma	1	-	-	Frozen tissue
414	Pilocytic Astrocytoma	1	-	-	Frozen tissue
421	Pilocytic Astrocytoma	1	-	-	Frozen tissue

Table M2. Complete diagnostic information regarding the pediatric patients cohort.

3.4 Patient-derived cell cultures

Cell cultures were established in cooperation with Dr Agata Góźdź and the detailed procedure was described in the recent publication[99]. Briefly, resected GBM samples were processed within six hours after resection. Fresh tissue was digested with collagenase IV and DNAse, the culture medium was included 20 ng/mL EGF, 20 ng/mL bFGF, DMEM-F12 with Glutamax, B27 supplement, Penicilin-Streptomycin solution, and 1 IU/mL heparin. Cells were placed onto the cell culture flasks coated with laminin (0.5 µg per cm^2), or to flasks dedicated to suspension cultures to achieve various culture conditions. GBM cells were cultured under two different oxygen conditions, in the incubator at atmospheric (20%) O_2, 5% CO_2, 37°C or in the BioSpherix Xvivo System (Parish, NY, USA) at 5% O_2, 5% CO_2, 37°C. Immediately after displaying an exponential growth rate, cells were harvested for nucleic acid isolation. The set consisting of 8 GBM samples and DNA derived from these cell cultures (three independent passages) was sequenced with the targeted gene panel. Cells were cultured in the form of spheres for most samples, with exception of 2 cell lines: patient ID: #31 and #67, that were cultured as adherent (table M3).

Patient ID #	Diagnosis	Sex	Age
31	Glioblastoma G 4	F	51
67	Glioblastoma G 4	M	68
70	Glioblastoma G 4	M	51
71	Glioblastoma G 4	M	68
73	Glioblastoma G 4	F	55
84	Glioblastoma G 4	M	57
86	Giant Cell Glioblastoma G 4	F	66
83	Glioblastoma G 4	M	71

Table M3. Complete diagnostic information regarding the cohort for which tumor derived cell lines were sequenced and analysed.

3.5 DNA Isolation

3.5.1 DNA isolation from fresh frozen tumor tissue and cell cultures

Tumor DNA (tDNA) was extracted from fresh frozen tumor samples using 1 mL TRIzol reagent (Thermo Fisher Scientific, Waltham, MA, USA). Frozen tissue samples were removed from storage (-80ºC freezer) and placed onto dry ice. Dry, labeled, autoclaved tubes were prepared in the hood. Around 100 mg of frozen tissue was placed in 4ºC 1mL of Trizol reagent and homogenized with Bio-Gen Series PRO2000 PRO Scientific homogenizer. After homogenization samples were left for 5 minutes in room temperature (RT).

Cells cultured at 20% and 5% oxygen were harvested 72 hours after seeding. Cell medium was removed, cells were rinsed with PBS and 0.5 % trypsin was applied; collected cells were then centrifuged 94 *x g* for 5 minutes at RT, then supernatant was removed. Cell pellets were washed with PBS and centrifuged 94 *x g* for 5 minutes at RT. Finally, cells were mechanically disrupted and lysed in 1 mL TRIzol Reagent (Thermo Fisher Scientific), then stored in -80ºC freezer prior to DNA isolation.

TRIzol-cell suspension was equilibrated in RT for 15 minutes. To each sample of homogenized tumors or lysed cells 0.2 mL of chloroform was added, covered tight and shaken vigorously for 15 seconds, which was followed by another 15 minutes incubation at RT. The resulting mix was centrifuged at 12,000 *x g* for 15 minutes (4°C), then upper colorless phase containing RNA was removed. To each sample 0.3 mL of dehydrated ethanol (99.8%) was added and mixed by inversion for 15 seconds, and incubated in RT for 3 minutes. Following the incubation, samples were centrifuged at 10,000 *x g* for 10 minutes at 4°C, which led to a phase

separation. The upper layer supernatant containing the protein fraction was removed. Then pellet was washed with 1 mL of Sodium Citrate Tribasic Dihydrate (0.1 M) and mixed by inversion for 30 seconds, incubated in RT for one the hour, and finally centrifuged at 10 000 g for 10 minutes (4°C). Washes were repeated twice, then DNA pellet was submerged in 1.5 mL 75% ethanol and placed in the fridge for the overnight storage. Next day samples were centrifuged at 7,000 *x g* for 5 minutes (4°C), and supernatant was removed. Samples were left in open vials for 10 minutes to ensure ethanol evaporation. Pellets were suspended in 200 µL of 1 x TE buffer and left in RT for 10 minutes, prior to addition of 6 µL of Proteinase K solution. Finally, samples were placed in the thermo block mixer set up to 56°C and 300 - 400 RPM's for 3 hours, during which time every 20 minutes pellets would be broken apart physically by pipette tip and the solution mixed. After 3 hour incubation 200 µL of 150 mM NaCl 5 mM EDTA was added to each sample and pipette mixed, followed by adding 400 µL of phenol chloroform in isoamyl alcohol and samples were shaken rigorously for 2 minutes. Then samples were centrifuged at 13,000 *x g* for 2 minutes at RT, the top aqueous phase containing DNA was collected to a clean labeled vial. 350 µL of chloroform was added and samples were shaken rigorously for 2 minutes. Then samples were centrifuged at 13,000 *x g* for 2 minutes at RT, top aqueous phase containing DNA was collected to clean labeled vial. 38 µL of 5 M NaCl was added to each sample, mixed by pipetting, then 800 µL of dehydrated ethanol was added, and sample was gently mixed followed by keeping overnight at 4ºC to increase DNA precipitation. In case of large quantities of the starting tumor material DNA would immediately form as a long chain precipitate structure visible with bare eye. Following day samples were centrifuged at 13,000 *x g* for 30 minutes (4°C), then supernatant was removed without disturbing the DNA pellet. Pellet was resuspended in 1 mL of 70% ethanol, then centrifuged at 13 000 *x g* for 10 minutes (4°C), and all ethanol removed without disturbing the DNA pellet. White DNA pellets were left to dry in open tubes until they appeared to be transparent, at which point they were suspended in 50 µL of PCR grade water. At this point DNA pellets were incubated in 4°C for 12-36 hours in order to completely dissolve in water, pipette mixed and in -20ºC for long term storage. Concentration and quality of each sample was measured using Nanodrop (Thermo

Scientific), for quality check parameter A260/280 was used (with optimal value 1,8-2,2).

3.5.2 DNA Isolation from peripheral whole blood

Genomic white blood cell derived DNA (gDNA) was isolated from whole blood samples that were stored and frozen (−20°C) in EDTA-coated tubes prior to isolation. Genomic DNA (gDNA) was isolated using a QIAamp DNA Blood Mini Kit (Qiagen, Hilden, Germany), following the manufacturer's protocol. Concentration and quality of each sample was measured using Nanodrop (Thermo Scientific), for quality check parameter A260/280 was used (with optimal value 1,8-2,2).

3.5.3 DNA Isolation from FFPE tissue sections

DNA was isolated from 34 Formalin Fixed Paraffin Embedded (FFPE) tissue specimens: 3-5 sections were cut from each block at 5 µm setting of microtome and placed in Eppendorf PCR clean 1.5 mL vial. Deparaffinization of each sample was done by adding 1200 µL of Xylene and vortexing. FFPE samples often vary in quality, after deparaffinization (adding xylene) tissue pellets were in different color as shown below:

Figure M3. Photograph of 3 FFPE samples during the depaffinization stage of isolation, each sample had different a colour indicating significant parafinization and quality differences.

Samples were centrifuged in RT at 14,000 *x g* for 5 minutes, the supernatant was removed. Samples were then twice washed in dehydrated ethanol (1200 µL) and centrifuged in RT at 14000 g for 5 minutes, then ethanol was removed. Samples were incubated in 37°C for 10-15 minutes with open lids, so excess of ethanol would evaporate. Then 180 µL of ALT buffer and 20 µL of Proteinase K were added (DNeasy® Blood and Tissue Kit, Qiagen, Hilden, Germany). Samples were incubated at 56°C overnight (12-18 h). Final isolation steps were done according to DNeasy® Blood and Tissue Kit (Qiagen, Hilden, Germany) manufacturer's protocol. Quantification of the resulting DNA was performed using Nanodrop (Thermo Scientific) and Quantus Fluorometer and QuantiFluor Double Stranded DNA System (Promega, Madison, Wisconsin, USA). Quantus measurement was taken into consideration during the library preparation protocol.

3.5.4 Circulating cell-free DNA (cfDNA) isolation

cfDNA was isolated in cooperation with researchers from the Regional Science and Technology Centre, Podzamcze, Poland. Blood for cfDNA isolation was collected in Paxgene tubes (PreAnalytiX, Homberchtikton, Switzerland) and was centrifuged at room temperature (15–25°C) for 15 min at 1,900 × *g*. The obtained plasma was transferred into a 15 mL conical-bottom centrifugation tube without disturbance of the buffy coat. The cellular fraction was centrifuged for 10 min at room temperature (15–25°C) at 1,900 *x g* for further purification. cfDNA was isolated from the obtained plasma using a QIAamp Circulating Nucleic Acid Kit and QIAvac system (Qiagen, Hilden, Germany). The isolation was performed according to the manufacturer's protocol, with appropriate amounts of reagents selected depending on the volume of the input material. The protocol assumes a volume of input material of 1–5 mL of plasma. The obtained cfDNA was stored at −80°C until further processing. Quality and quantity of isolated cfDNA was evaluated using an Agilent Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA). This technology allows measurement of only specific length fragments of double stranded DNA (dsDNA).

3.6 Library Preparation from the acquired material

3.6.1 Large Panel Design (SeqCap EZ Hyper Cap)

In order to capture a wider spectrum of somatic mutations, tumor derived DNA and reference whole blood DNA were sequenced using a broad 664 cancer-related gene panel. SeqCap EZ Custom Enrichment Kit was used - an exome enrichment design which targets the latest genomic annotation GRCh38/hg38. A vast majority of genes (578) were selected from the Roche Nimblegen Cancer Comprehensive Panel (based on NCBI Gene Tests and on Cancer Gene Consensus from Sanger Institute). Additionally, 86 epigenetics-related genes were included (genes coding for histone acetylases/deacetylases, DNA methylases/demethylases histone methylases/demethylases, and chromatin remodeling proteins) based on a literature review[100–102].

3.6.2 Fragmentation, End Repair, A-tailing, Adapter Ligation, PCR Amplification

DNA samples isolated from frozen tumor tissue, cell lines, FFPE tissue sections and whole blood were processed for library preparation using KAPA HyperPlus Kit, according to a SeqCap EZ HyperCap Workflow user's guide (version 2.3). DNA concentration was measured using Quantus Fluorometer with a QuantiFluor ONE Double-Stranded DNA System (Promega, Madison, WI, USA) to quantify an appropriate amount of a starting material for library preparation. For preparation of libraries from FFPE derived DNA 200 ng of material was used, as this type of material is low quality and highly degraded. Fragmentation time was also readjusted for these samples: in some cases enzymatic fragmentation was 30 minutes in others 15 minutes, depending on a sample.

Bioanalyzer measurements were performed after fragmentation in order to ensure proper fragmentation of the libraries (180-220 bp DNA fragments) and readjust fragmentation time if needed.

The rest of the libraries were prepared from 100 ng of starting DNA material (tDNA, gDNA, cell line DNA) and were enzymatically fragmented according to the manufacturers protocol in 37ºC for 20 minutes. DNA fragments were end repaired and A base was added to 3' end. Then indexed adapters were ligated, later double size selection was performed, and libraries were amplified. Total number of PCR cycles was increased in the case of FFPE samples

from 7 (tDNA, gDNA, cell line DNA libraries) to 8 (FFPE DNA libraries). Concentration of resulting libraries was determined by Quantus Fluorometer with QuantiFluor ONE Double Stranded DNA System (Promega, Madison, Wisconsin, USA) and the quality check was done using Agilent Bioanalyzer (Agilent Technologies, Santa Clara, CA, USA).

3.6.3 Hybridization, PCR amplification, Bioanalyzer quality control, Sequencing

Obtained libraries were mixed in equimolar concentrations together to form 1400 ng pool and after COT (Human Cot-1 DNA®, NimbleGen SeqCap EZ Accessory Kit v2, Roche) and complementary adapter oligos (SeqCap Adapter Kit Band hybridization, Roche) were added, each sample was condensed using PCR clean speed vac for 30 minutes at 60°C. Resulting pool was mixed with probes and additional reagents from SeqCap EZ Custom Enrichment Kit, denatured at 95° C for 10 minutes, and incubated in 47° C for at least 17 hours to allow proper probe binding. After an overnight incubation, mixture of pooled libraries was purified using special HyperCap beads (HyperCap Bead Kit, Roche) and later amplified. During that step libraries were enriched in fragments of interest. Quality of obtained libraries was evaluated using Agilent Bioanalyzer with High Sensitivity DNA Kit (Agilent Technologies, Palo Alto, CA, USA). The libraries were run in a rapid-run flow cell and were paired-end sequenced (2 × 76 bp) on a HiSeq 1500 (Illumina, San Diego, CA, USA), as recommended by the manufacturer.

3.7 Library Preparation (cfDNA)

3.7.1 Panel design for targeted sequencing

As cfDNA samples require higher target coverage to detect variants at extremely low allele frequency (a_f), therefore deep sequencing of cfDNA was done using a more narrow 50 gene Sure Select XT HS custom panel (411,483 kbp) which targets the genomic annotation GRCh19/hg19 (Design ID: 3216011). Gene regions were chosen based upon the larger 664 genes panel and the 50 most frequently mutated in the previously analyzed glioma patient cohort of 182 gliomas samples[103] were chosen. Copy number alteration Sure Select XT HS custom probes were covering the same 50 gene regions (Design ID: A3224001).

3.7.2 End Repair, A-tailing, Adapter Ligation, PCR Amplification

The libraries were prepared from the scarce amounts of cfDNA material, which varied between 0.5 and 10 ng depending on the sample, so library preparation would require additional PCR amplification cycles. In order to validate low allele frequency variants, unique molecular identifiers (UMI) technology was used, that was provided by the Sure Select XT HS Target Enrichment System for Illumina Paired-End Multiplexed Sequencing Library (Agilent Technologies, Palo Alto, CA, USA). Libraries were prepared according to the manufacturer's protocol (version C 2 July 2019), with few adjustments. The fragmentation step was skipped for all the cfDNA samples because properly isolated cfDNA has 120–200 bp fragment size, so further fragmentation would result in fragments that are too short. The ends were repaired and a dA-tail was added to the 3′ ends. The next step was the ligation of molecular-barcoded adapters. During this step, unique molecular identifiers were attached to each DNA fragment, labeling each as an original and unique sequence prior to PCR amplification. This step was crucial in allowing a clear verification of false positives during the later stages of the bioinformatic analysis of the sequenced data. The remaining molecular-barcoded adapters were removed by AM Pure Bead purification prior to the PCR-amplification step. Due to the small quantity of cfDNA, 14 PCR cycles were performed, during which SureSelect XT HS Index Primers were added to label each sample.

3.7.3 Right Sided Size Selection

According to some reports[104,105], the enrichment of cfDNA in shorter fragments can improve tDNA detection, so an additional right-sided size-selection step, using AM Pure Beads (Beckman Coulter, Brea, CA, USA), was added to the protocol. After PCR amplification AM Pure Beads were used to purify prepared libraries, and instead of suspending them in 15 µL (as recommended in the protocol version C 2 July 2019), 50 µL of PCR grade water was added. Next samples were vortexed and incubated for 2 minutes prior to placing a strip on a magnetic stand, as described in the following steps of the protocol. Then 50 µL of libraries were collected and placed into a new strip. To 50 µL of libraries, 35 µL of vortexed AM Pure beads were added, samples were vortexed and incubated for 5 minutes in RT. Then, samples were centrifuged (short spin), put on a magnetic

stand and 80 µL of supernatant was collected into a clean strip. Finally, 88 µL of vortexed AM Pure beads were added and samples were vortexed and incubated for 5 minutes in RT. Then samples were centrifuged (short spin), put on a magnetic stand and the supernatant was removed. Ethanol wash was performed twice, by adding 200 µL of freshly prepared 70%, incubating samples for one minute, and removal of the supernatant (all on the magnetic stand). Finally, samples were removed from the magnetic stand, air dried and 53 µL of PCR grade water was added and samples were vortexed. After 2 minutes of incubation, samples went through a short spin, then back on a magnetic stand, and the supernatant containing final libraries after Right-Sided Size Selection was collected and placed into a new tube. The quality of the resulting libraries was evaluated using an Agilent Bioanalyzer with a High-Sensitivity DNA Kit (Agilent Technologies, Palo Alto, CA, USA). The Bioanalyzer plot exemplifying how libraries looked before and after Right-Sided Size Selection is shown in figure M4.

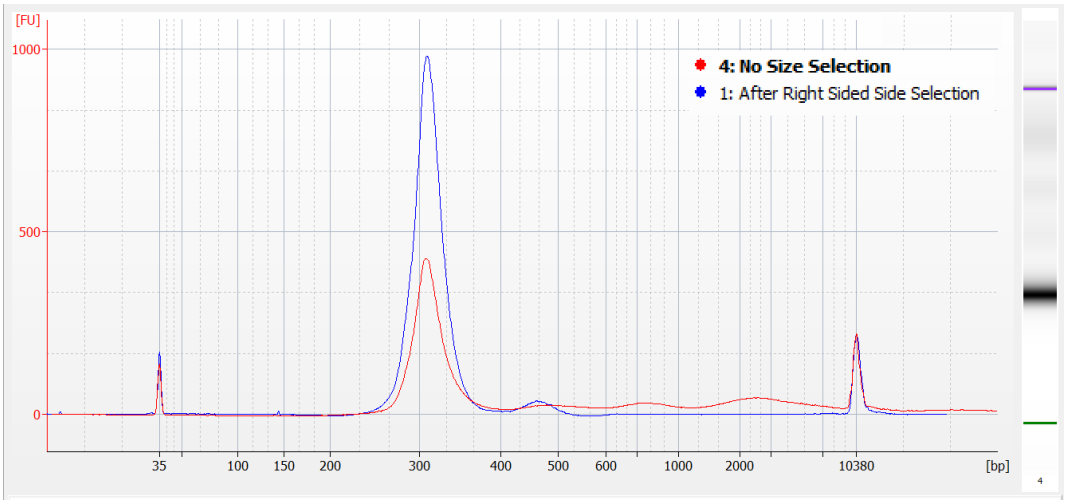


Figure M4. Right Sided Size selection electropherogram prepared using Bioanalyzer shows the reduction in the proportion of long fragments in the final sequenced library sample.

3.7.4 Hybridization capture and sequencing

The hybridization and capture were conducted according to the manufacturer's protocol (version C 2 July 2019). Hybridization was done using a 50 gene panel (Design ID: 3216011) for 84 cfDNA samples and with CNA probes (Design ID: A3224001) for 4 cfDNA samples and 4 corresponding gDNA samples, that had confirmed ctDNA signal by previous NGS 50 genes panel analysis.

The libraries were run on a rapid-run flow cell with the paired-end setting for sequencing (2 × 100 bp) on a HiSeq 1500 (Illumina, San Diego, CA, USA), as recommended by the manufacturer.

3.8 Bioinformatic Analysis

3.8.1 Somatic variants pipeline

DNA from tumor (both frozen and FFPE) specimens, cell cultures, and reference genomic whole blood DNA samples were analyzed as follows. The FASTQ files were obtained from the sequencing raw data, then processed with the trimmomatic program[106] in order to remove low-quality reads and the sequencing adapters. After trimming and filtering, reads were mapped using NextGenMap[107] aligner (http://cibiv.github.io/NextGenMap/)[107] to the human genome (hg38). Duplicates were marked and removed using the Picard tool (https://broadinstitute.github.io/picard/ (accessed on 8 November 2021))[108], then only uniquely mapped and properly oriented reads were used for further analysis. SIFT (Sorting Intolerant From Tolerant) score indicates whether an amino acid substitution will have an impact on protein function. The SIFT scale goes from 0 (deleterious) to 1 (tolerated), with 0 having the most significant predicted impact on the protein function. Strand-supporting-read bias variants were discarded and only filtered variants with SIFT values below 0.05 were selected (0 - 0.05 SIFT value variants are predicted to have a deleterious impact on the protein, source https://ionreporter.thermofisher.com/ionreporter/help/GUID-2097F236-C8A2-4E67-862D-0FB5875979AC.html).

The ProcessSomatic method from VarScan2[109] was used in order to extract high-confidence somatic calls, based upon Fisher's exact test *p*-value and variant allele frequency. The final filtered subset of variants was annotated with usage of Annovar (http://annovar.openbioinformatics.org/en/latest/ (accessed on 8 November 2021))[110], using the latest databases versions (clinvar, cosmic, refGene, avsnp150 and dbnsfp30a). The maftools R library[111] was used to proceed with analysis of the resulting somatic variants. FFPE samples went through quality control after the sequencing and the samples with PCR duplicate rates above 85% were excluded from analysis. The analysis was performed in collaboration with Dr hab Bartosz Wojtaś, Dr Adria Roura, Dr. Sabina Licholai, Dr Tomasz Waller, Dr Tomasz Gubała, and Kacper Żukowski,

3.8.2 cfDNA analysis

Sequencing data from each patient that had a complete set of tDNA, gDNA, and cfDNA was analyzed as described below. tDNA and gDNA samples were processed using a dedicated pipeline based on open-source bioinformatics tools, while the cfDNA samples were analyzed using SureCall (https://www.agilent.com/en/product/next-generation-sequencing/hybridization-based-next-generation-sequencing-ngs/ngs-software/surecall-232880 (accessed on 16 January 2022), which is a dedicated software provided by a Sure Select library-preparation reagent manufacturer. The minimum 10 reads coverage for somatic calls from ctDNA was established. The raw sequencing reads from tDNA and gDNA samples were converted to fastq files with bcl2fastq software (https://emea.support.illumina.com/sequencing/sequencing_software/bcl2fastq-conversion-software.html (which was last accessed on 16 January, 2022)) (Illumina, San Diego, CA, USA). The quality control of obtained reads was performed using the FastQC tool. Raw sequencing reads obtained from cfDNA samples were converted into a fastq format, which enabled the alternative variants to be detected using SureCall software (Agilent Technologies, Palo Alto, CA, USA). Somatic variants detected within the tumors were assigned to the corresponding reads from cfDNA and compared. Variants were annotated and cfDNA reads of sufficient quality were mapped to the hg19 human reference genome, using standard parameters of bwa package with usage of standard parameters. Unfortunately library manufacturer's pipeline was a proprietary software solution, that could not be modified to be compatible with hg38. Next steps included recalibration, de-duplication, and variant-calling with the usage of appropriate tools from the GATK package[112]. Among others Mutect2, BaseRecalibrator, and MarkDuplicates were used. Annotation of vcf files was done using Annovar (http://annovar.openbioinformatics.org/en/latest/ (accessed on 16 January 2022))[110] package with appropriate databases ClinVar, refGene and others. Variants reported in dbSNP (v137) and the 1000 Genomes database were filtered. Mutations, which had at least 10 alternative raw reads in cfDNA and were found also in tDNA (at least 3 reads supporting variant alleles), but not present in gDNA were filtered out and treated as somatic (ctDNA). Potentially pathogenic variants in cfDNA were filtered out as mutations that were present in cfDNA, but not detectable

in a reference gDNA, yet registered in COSMIC or with pathogenic ClinVar clinical significance.

Copy number variations (CNVs) were detected using ADTEx software (http://adtex.sourceforge.net (accessed on 15 February 2022)) with default parameters. CNVs (germline - copy number variations) from each patient were identified using the gDNA sequenced sample and normal human HapMap DNA sample NA18535 (Cornell Institute) for each captured region (the exonic region). Copy number alterations (CNAs) were identified using paired gDNA and cfDNA samples for each exon. Four samples chosen for this analysis were the ones with detectable signal from ctDNA. The data were deposited to European Genome-phenome Archive (EGA http://www.ebi.ac.uk/ega/ (accessed on 25 July 2022)), hosted by the European Bioinformatics Institute (EBI) under accession numbers EGAS00001006451 and EGAD00001009080.

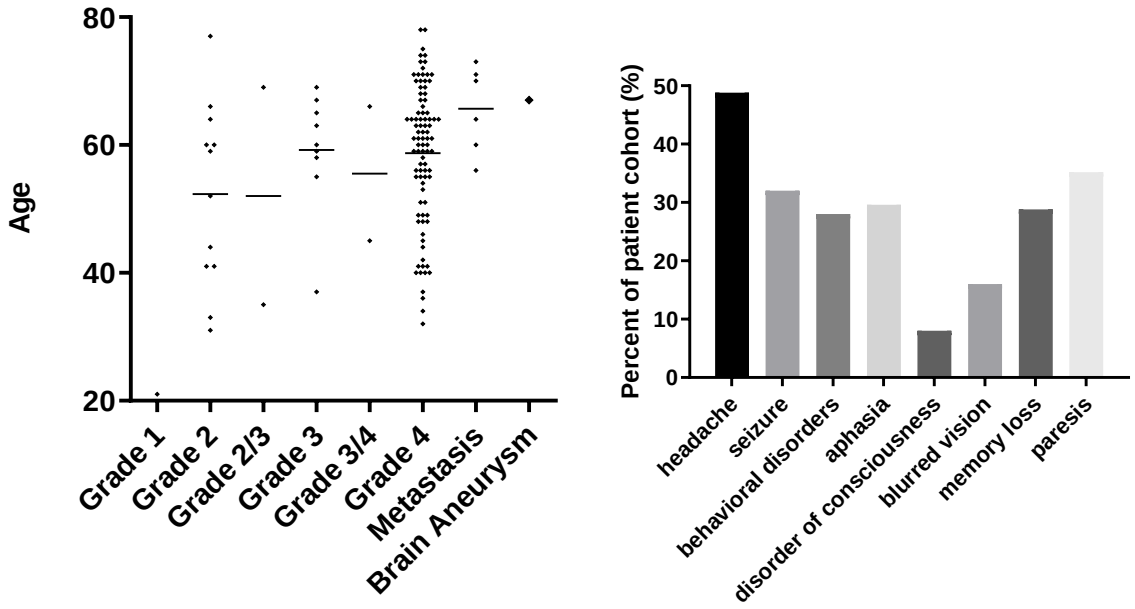
3.9 Statistical Analysis

The statistical significance was calculated using a *t*-test with GraphPad Prism v6 (GraphPad Software, San Diego, CA, USA). *p*-values < 0.05 were considered significant. For comparison of concentration of the pre-surgical to post-surgical cfDNA samples paired t-test was used. To evaluate the difference in early cfDNA isolation (within 24 hours post blood collection) to later cfDNA isolation unpaired t-test was used.

4. Results

4.1 Characteristics of adult glioma patient cohort

Complete cohort clinical data was reviewed. As in a typical primary brain tumor study when tumor samples are processed before a full diagnosis, some patients were misdiagnosed, and the cohort encompassed a brain aneurysm patient and a few metastatic patients (figure R1a). Clinical symptoms and their occurrence are presented in the figure R1b. Headache was the most common symptom occurring in almost 50% of patients, which is connected to inflammation and an intracranial pressure increase due to brain tumor formation. Paresis (weakness or loss of voluntary movement), seizures, aphasia (problems in communication), and behavioral deficits were common symptoms occurring in more than 25% of the patients. Blurred vision and defects of consciousness were much less common, but symptoms are most likely related to the brain region being affected by the tumor growth.

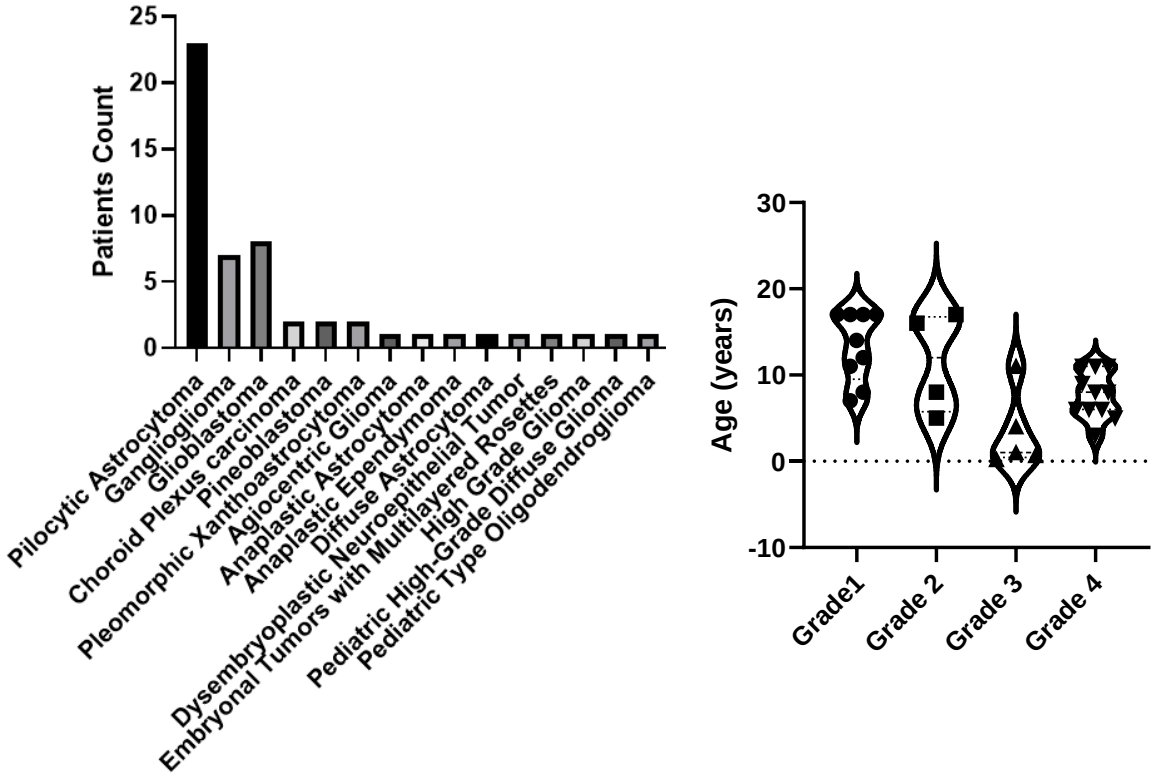


(**a**) (**b**)

Figure R1. Complete clinical data available for the study of the adult patients cohort. (**a**) % of patients at different ages; (**b**) Clinical symptoms occurrence.

4.2 Characteristics of pediatric patients cohort

The most common glioma that occurs in children and adolescents is pilocytic astrocytoma, which is benign and curable if surgically removed[113]. The analyzed pediatric cohort consisted of 23/53 (43%) of patients diagnosed with pilocytic astrocytoma, 8/53 (15%) glioblastoma and 7/53 (13%) had ganglioglioma (figure R2a). The age versus grade in the analyzed cohort was reviewed and the youngest patients were the ones diagnosed with WHO Grade 3 and Grade 4 (figure R2b).



(**a**) (**b**)

Figure R2. Cohort review of the pediatric patients included in this study: (**a**) Number of patients versus diagnosis; (**b**) Age versus grade.

4.3 Determination of quantity of the DNA isolated using TRIzol

TRIzol isolation used in this study is one of the most efficient isolation protocols that can be used to obtain maximal yields of DNA and RNA from the same sample. The amount of the DNA isolated depends on the starting material provided and its potential degradation before it was frozen or placed on dry ice. As different hospitals have different procedures of sample collection, which results in various

yields and cell cultures provide less material than the tissue sample, DNA yield in various samples was determined (figure R3).

Complete DNA yield in most samples was varying from 1-300 micrograms, so the lowest isolated amount (1 µg) would be sufficient for preparation of 10 libraries using the KAPA HyperPlus Kit protocol and for 100 libraries according to the Sure Select XT HS protocol.

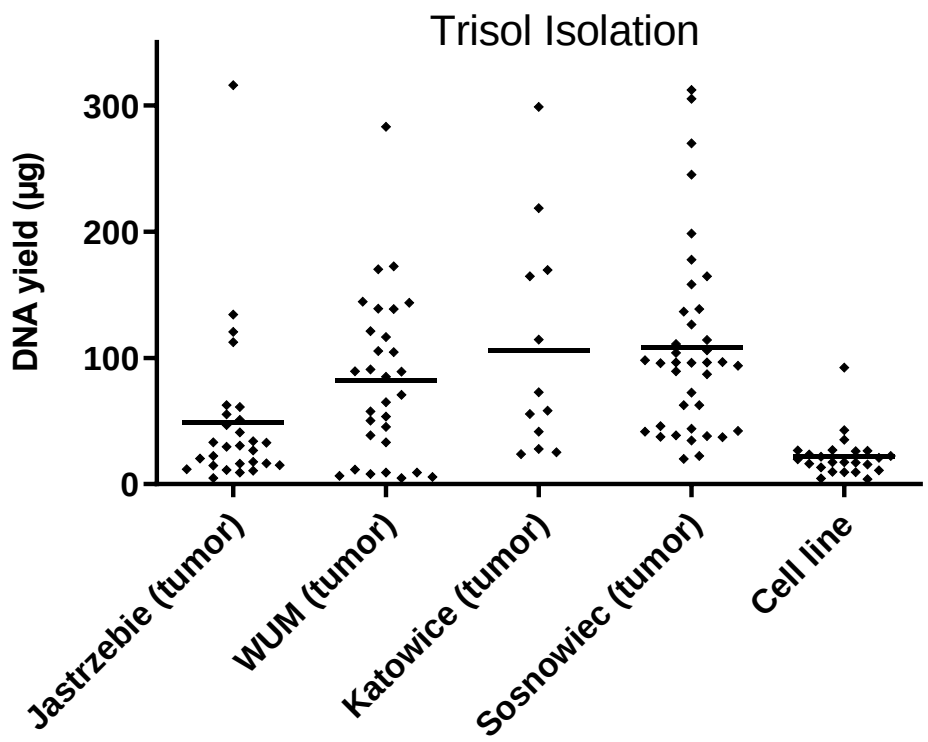


Figure R3. DNA isolation yield depending on the source the fresh, frozen tissue or from the cell lines.

4.4 Performing quality control of FFPE derived samples

Samples isolated from FFPE sections varied between each other and each step: DNA isolation, fragmentation and final library sequencing could bring varying quality results. Nanodrop, Quantus Fluorometer and QuantiFluor Double Stranded DNA System were used to determine DNA amounts. Nanodrop concentration measurements were taken after completion of each DNA isolation and during each stage of library preparation. Unfortunately, this measurement is not as precise as the fluorescent measurement taken by Quantus Fluorometer and QuantiFluor Double Stranded DNA System, therefore library preparation

was based upon the Quantus Fluorometer measurement. Difference in Nanodrop measurements versus the ones taken using Quantus system illustrates the amount of single stranded and degraded DNA fragments present within the sample. These differences are illustrated in the figure R4. Even if there was enough of the material to prepare library, the fragmentation process had varied between samples. Some samples needed to be fragmented for a longer time during library preparation stage and in some cases even after successful fragmentation library preparation would still not yield a sufficient amount and quality of a final library to allow the targeted panel probe hybridization.

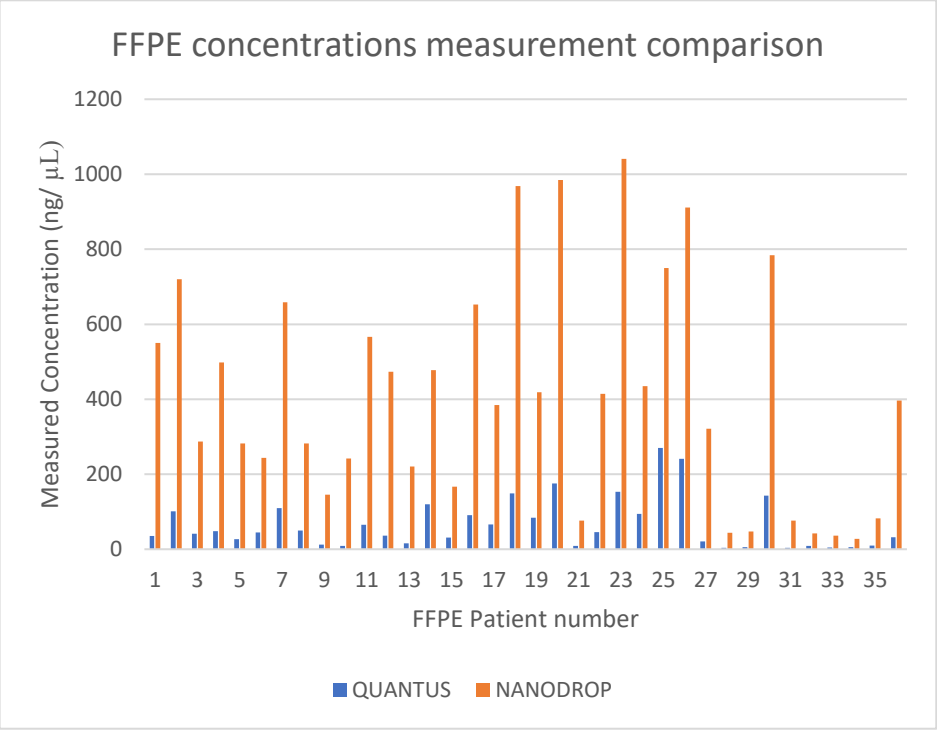


Figure R4. Difference in quantities of the double stranded DNA determined using the Quantus system (blue) and the Nanodrop (orange).

Samples that were hybridized and sequenced still had to go through a quality control, as the substantial amount of reads would contain duplicates which would decrease targeted gene coverage resulting in an insufficient coverage for the conclusive analysis. Samples with PCR duplicates over 85% were excluded from the analysis (figure R5). Final library sequencing range of PCR duplicates in tDNA, gDNA, or cell culture DNA varied between 15-30%, which reflects

the considerable difference in quality between these samples and FFPE sections-derived DNA.

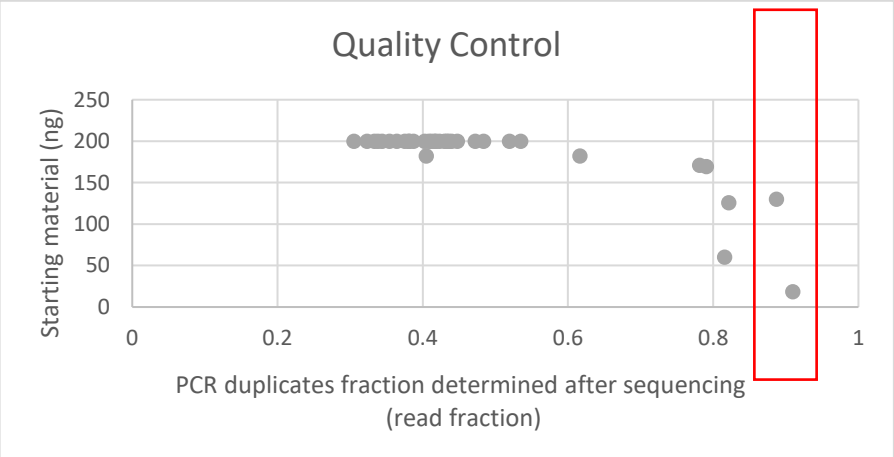


Figure R5. PCR duplicates fraction determined during a quality control stage of the targeted sequencing data of FFPE DNA derived libraries. In the red box 2 samples with around 90% of duplicates, that have been excluded from further analysis.

4.5 Determination of cfDNA that passed quality and quantity control

Isolation of cfDNA is challenging. Recently published study have shown that the isolation of cfDNA from lumbar puncture CSF can yield low amounts of cfDNA, as only in 12.5% of samples had yield higher than 10 ng cfDNA[91]. Herein, in most cases of serum derived cfDNA the isolation yield was low and the amount of cfDNA strongly depended on the storage time prior to serum isolation. The quality and quantity of cfDNA was evaluated using an Agilent Bioanalyzer with a High-Sensitivity DNA Kit. Concentration of cfDNA fragments which were within the size range between 100–500 bp was measured and registered as cfDNA concentrations (figure R6 - blue box), genomic contamination of each sample was also quantified (figure R6 - red box).

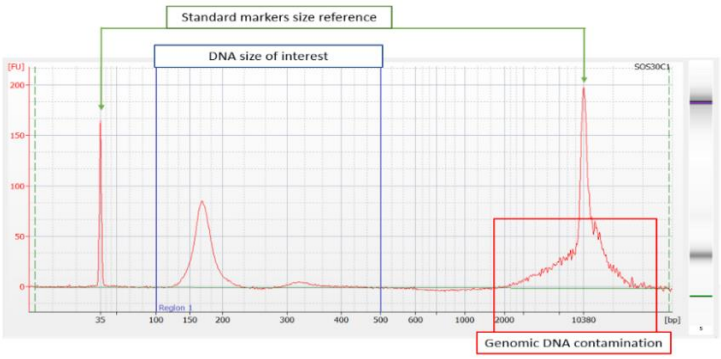
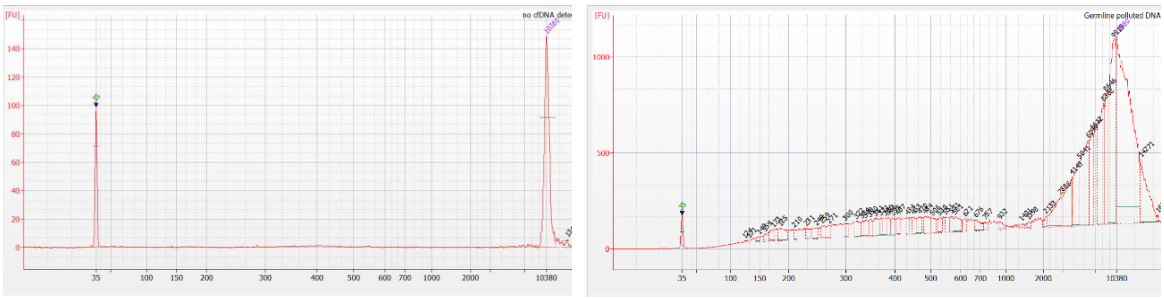


Figure R6. Representative electropherogram generated using Bioanalyzer that presents quantification and quality of the cfDNA sample.

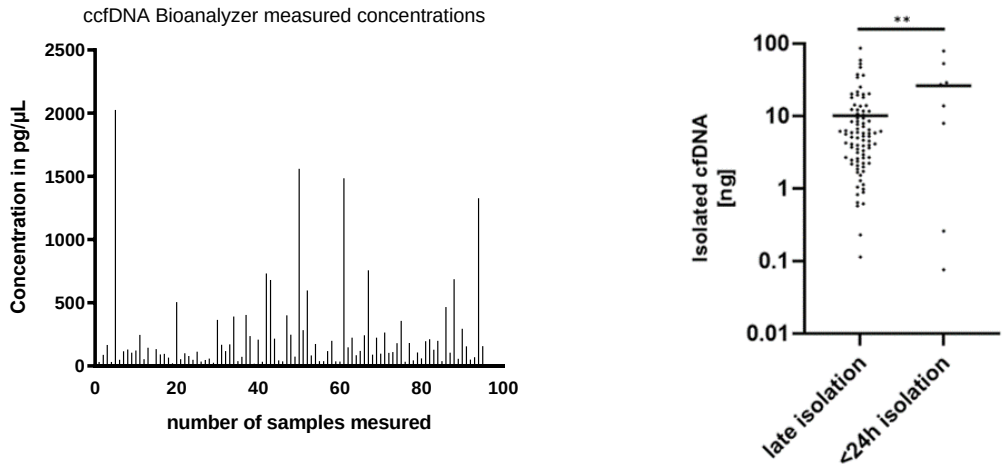
Some samples were excluded from further processing due to a low concentration (figure R7a) or high contamination with long DNA fragments representing most likely the genomic DNA from blood cells (figure R7b). Representative electrograms with examples are shown below.



(**a**) (**b**)

Figure R7. Electropherogram generated using Bioanalyzer measuring cfDNA concentrations: (**a**) no detectable nucleic acid; (**b**) high contamination with long fragments of most likely genomic DNA.

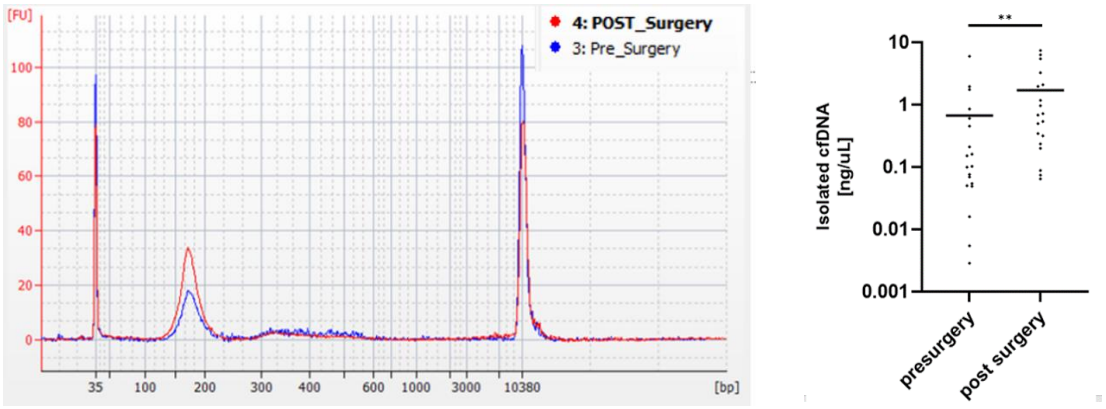
Size of fragments of interest in samples that passed the DNA quality control was between100 bp – 500 bp. Concentration of double-stranded DNA at this specific size range was measured in samples from 95 patients, as illustrated in the figure R8a. Samples of cfDNA that were isolated within 24 hours after the blood collection (n=8) had significantly higher yield when compared to the rest of samples that awaited for isolation for more than 24 hours (n=87), as presented in figure R8b.



(**a**) (**b**)

Figure R8. Quantity of isolated cfDNA in specific samples. (**a**) Concentration of cfDNA measured using Bioanalyzer in samples that passed the quality control; (**b**) Complete isolation yield comparison: amount of cfDNA isolated within 24 hours after bloods collection versus later isolation yield. Statistical significance was calculated with an unpaired t-test.

Concentrations of cfDNA isolated pre- and post-surgery were measured using Bioanalyzer (figure R9a) and compared (figure R9b) (n=19). The comparison of sample quantities showed significantly increased cfDNA levels in blood after surgery as calculated with the paired t-test.



(**a**) (**b**)

Figure R9. Comparison of quantities of cfDNA isolated pre- and post-surgery: (**a**) Bioanalyzer electropherogram showing higher concentration of cfDNA isolated from serum post- surgery; (**b**) Significantly higher cfDNA yield from samples collected post-surgically (significance calculated with a paired t-test; n=19 in each group).

4.6 Identification of genetic alterations in adult patients cohort

4.6.1 Identification of somatic variants in the adult patient cohort

Targeted exome sequencing was used to identify somatic and germline variations in the tDNA samples and gDNA. Variants mutated in tDNA, but not changed in gDNA were labeled as somatic. Figure R10 presents the most frequent somatic mutations found in the cohort, and the most commonly mutated genes were *PTEN*, *TP53*, *EGFR*, *ATRX*, *IDH1*, and *NF1*, which is in line with previous study on the larger cohort from the Polish population[114].

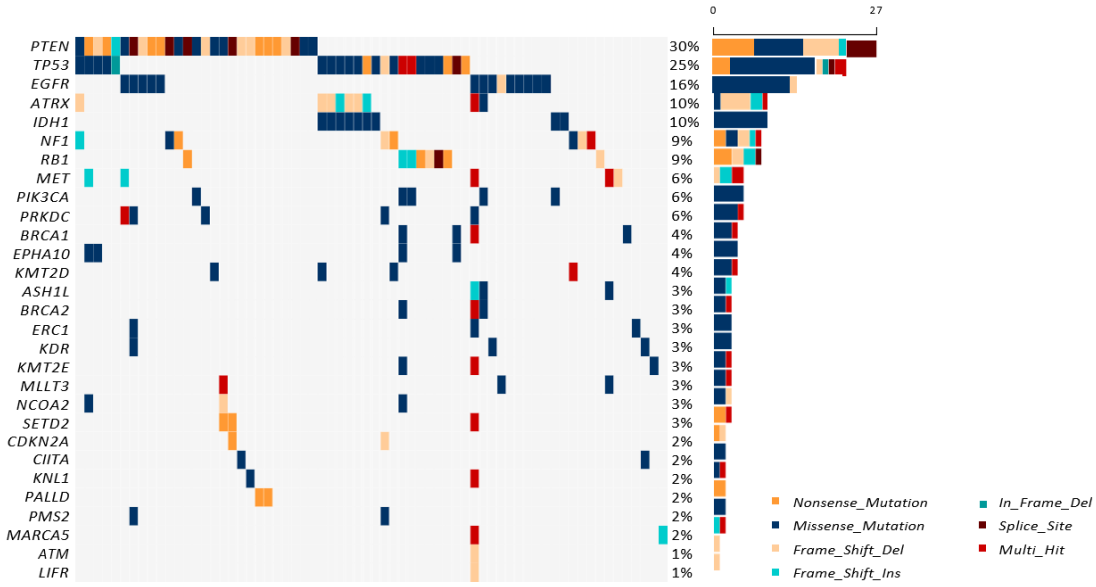


Figure R10. Somatic variants detected in the tumor DNA from adults patients cohort. Color codes for a type of mutations.

OncodriveCLUST algorithm[115] was used to identify oncogenic mutations based on the mutational clustering. The identified several variants enriched at the *TP53* gene (5 clusters), *RECQL4* (1 cluster), *PIK3CA* (2 clusters), and *IDH1* (1 cluster), among many others are depicted in the figure R11.

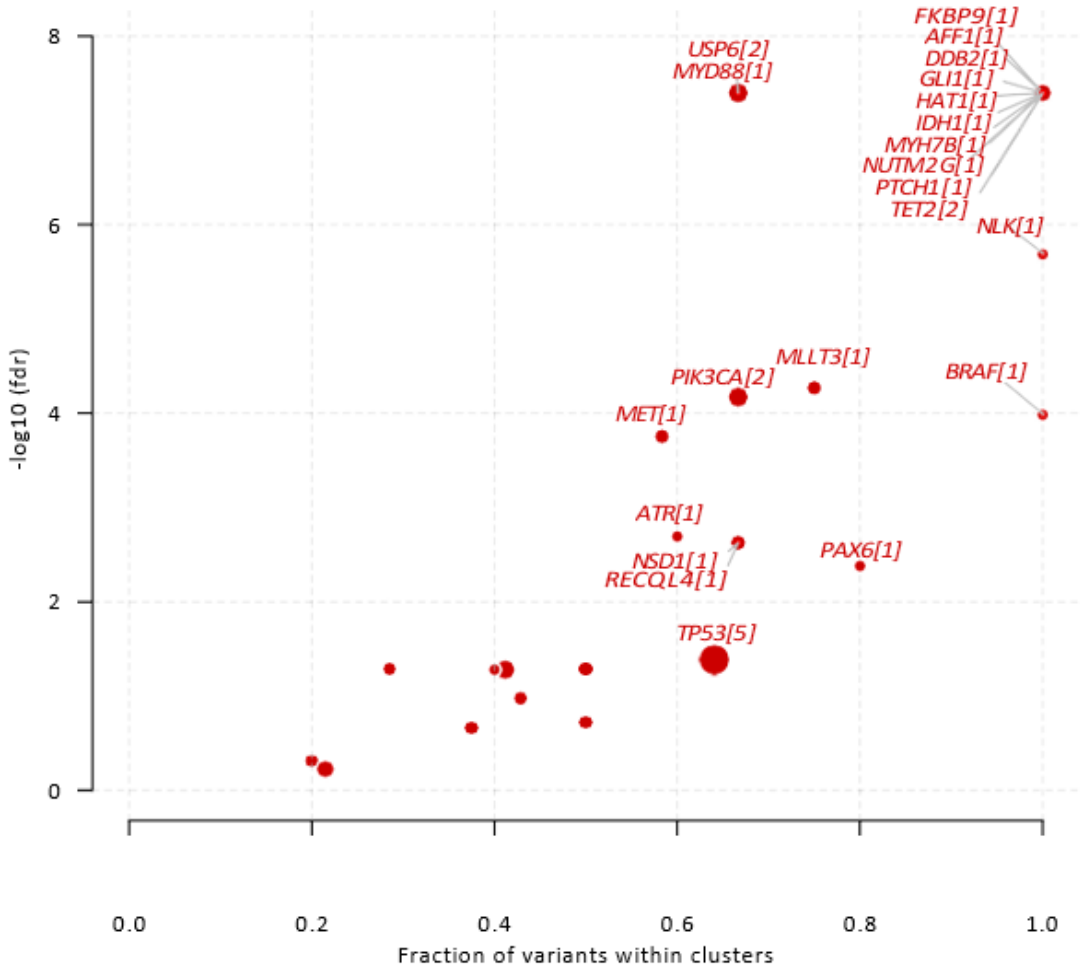


Figure R11. Somatic variants overview: cancer related genes based on mutational clustering;

Allele frequencies of somatic variations were reviewed, as shown in the figure R12, which revealed how mutation penetration varied between the samples and within the examined cohort and could present the estimate of the proportion of cells that carried these mutations. Furthermore, the high-allele frequency of tumor somatic variations facilitates the identification of potential cancer-related genes. We found that, in addition to being frequently mutated, *TP53* and *PTEN* have high variant allele frequencies (VAFs), suggesting the presence of homozygous mutations in some patients (figure R12). Mutations in genes *EGFR, CDKN2A,* and *RB1* exhibited VAFs close to 0.5, which could indicate loss of heterozygosity[96].

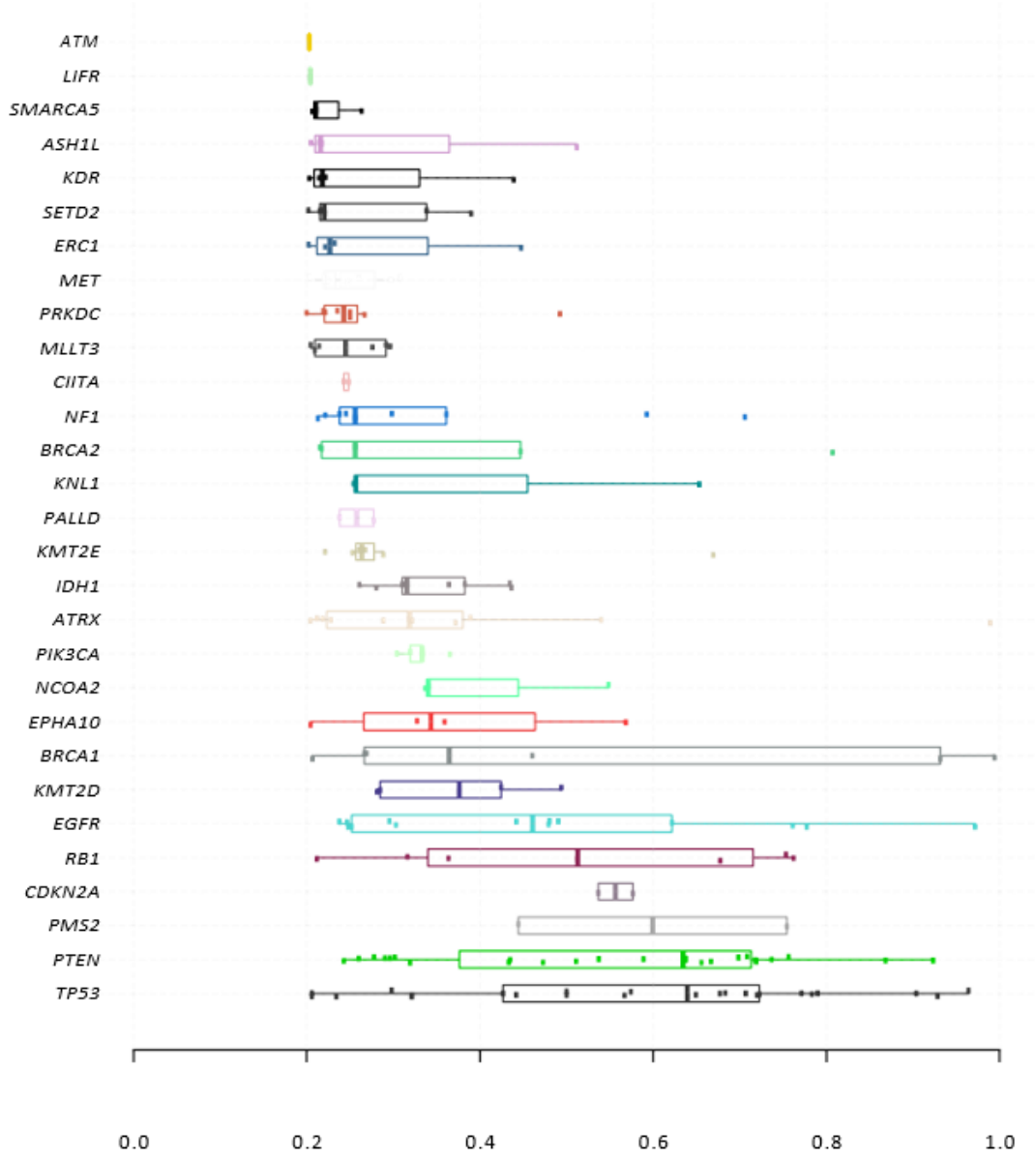


Figure R12. Somatic variants overview: Variant Allele Frequency mutation penetration for specific genes.

4.6.2 Identification of germline variants in the adult patient cohort

Variants found both in a tumor sample and the corresponding gDNA from blood cells were marked as germline variants. Germline variants were reviewed, as possibly cancer predisposing factors. The overview of most common germline mutations the analyzed cohort is shown in the figure R13.

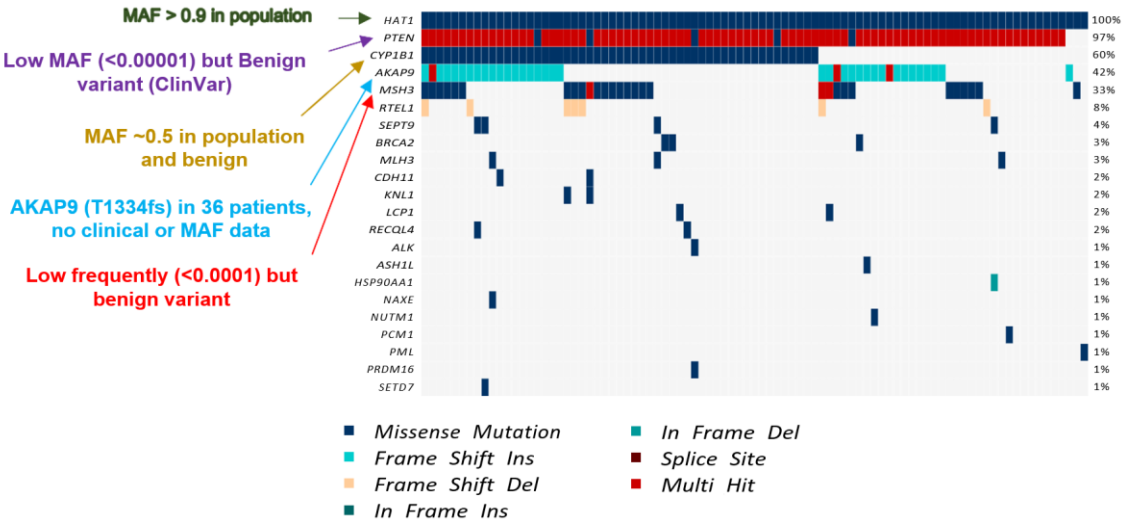


Figure R13. The overview of the most common germline mutations detected in the reference gDNA in the adult patients cohort.

Minor Allele Frequency (MAF) is a parameter that indicates a frequency at which the second most common allele occurs in a given population, which refers to the overall frequency at which specific variant can be detected. If MAF value is below 0.05, we can consider that variant as rare. As shown above, the *HAT1* mutation was present in all of the patients from the current study, but it is a common variant (MAF>0.9). *CYP1B1* presents a similar case, as is common in this cohort (60%) but it is also very common in the general population (MAF=0.5).

The information provided by ClinVar (Clinical Genome Resource provided by National Center for Biotechnology Information) collects the available information concerning genomic variations and how it relates to human health. Specific variants can be reviewed in that database and the registered information about their impact on a phenotype might be described as pathogenic, potentially pathogenic, unknown, or benign. When reviewing germline variants from the results above, *PTEN* and *MSH3* were found to be common in this cohort (97% and 33%), yet they were registered as benign in the ClinVar database. So it is unlikely that they could

be predisposing factors having any impact on tumor development. The overview of germline allele frequency mutation penetration of specific genes is shown in figure R14.

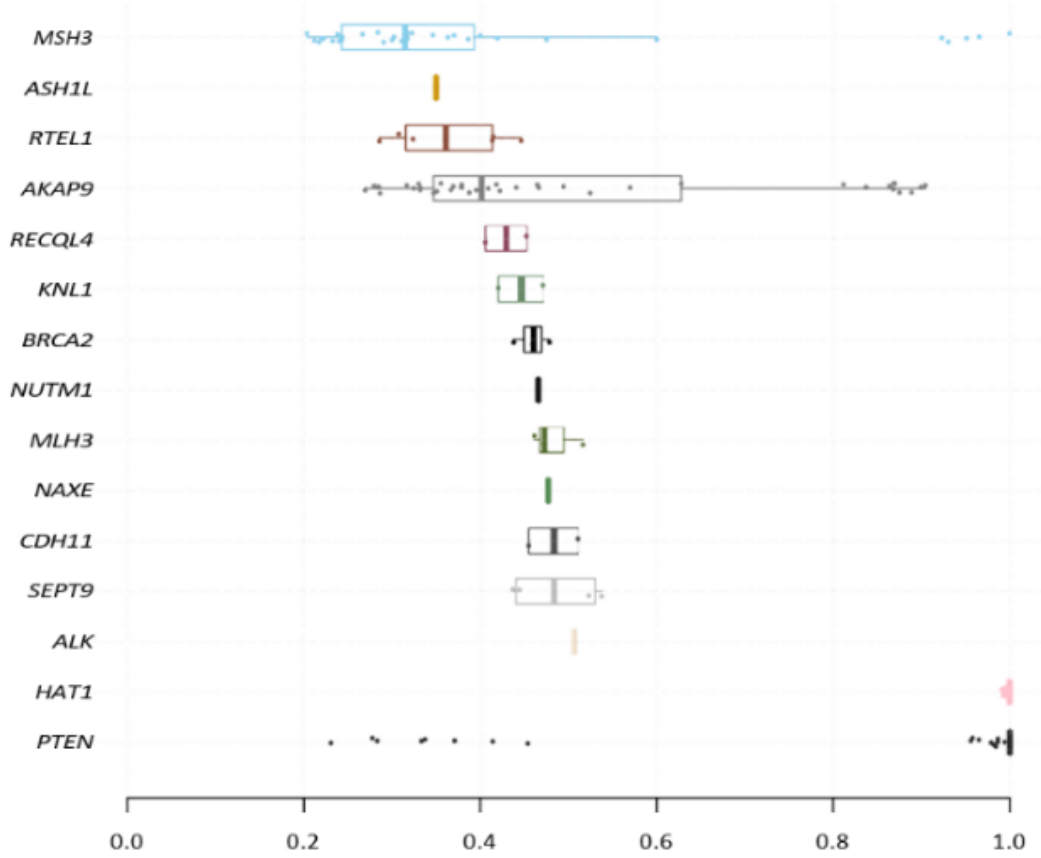


Figure R14. Overview of germline variant Allele Frequency mutation penetration for specific genes.

Finally, the most interesting (novel) germline mutation was detected in the *AKAP9* (T1334fs) gene and it was discovered in 46 % of patients (41 of the 89 patients), as shown in figure R13. As it is shown in figure R14, VAF of germline variants in *AKAP9* vary between the samples, yet an average VAF is close to 0.5 which might imply heterozygous mutations. The *AKAP9* (T1334fs) gene variant is not registered in the ClinVar or has any mutational annotation in MAF (figure R15). This mutation results in a frame shift insertion, which suggests the mutated gene may have a significantly modified structure.

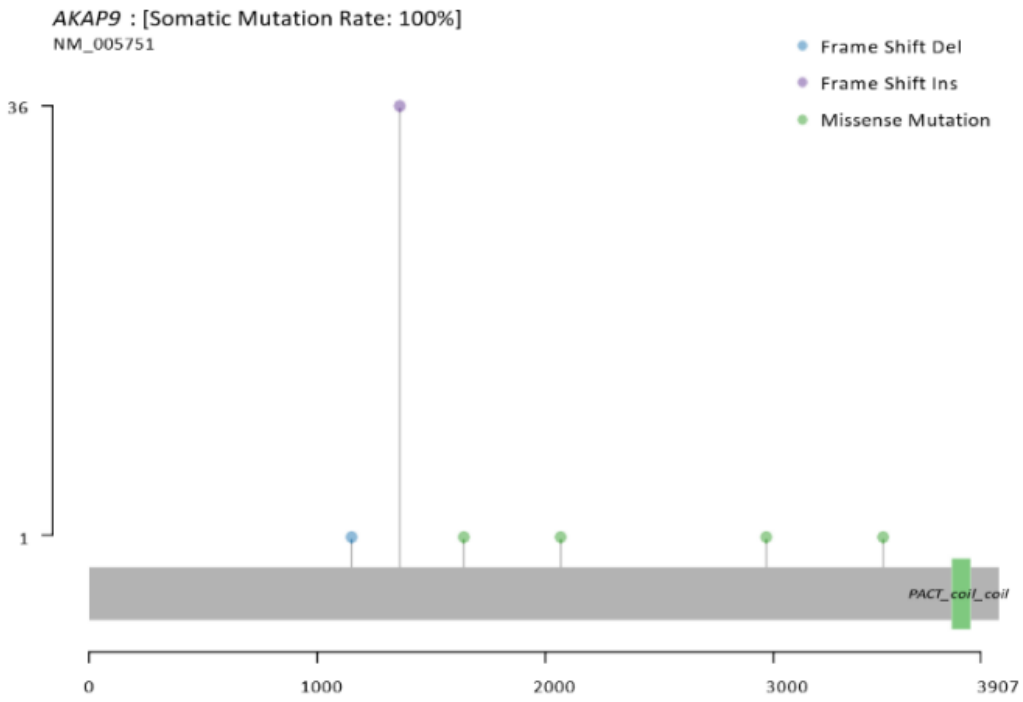


Figure R15. Overview of germline variants and novel *AKAP9* mutations. The *AKAP9* (T1334fs) gene variant present in 36/89 patients, 5 other patients have changes within this gene in different positions.

AKAP9 encodes A-kinase anchoring protein 9. The A-kinase anchor proteins (AKAPs) are a group of structurally diverse proteins that bind to the regulatory subunit of the protein kinase A (PKA), therefore regulating its function. Protein kinases A are the enzymes responsible for phosphorylation, and regulation of many different reaction pathways and enzymes[116].

The AKAP9 gene undergoes alternative splicing, resulting in at least two isoforms that interact with numerous signaling proteins from various signal transduction pathways and localize to the centrosome and the Golgi apparatus. These signaling proteins include type II protein kinase A, serine/threonine kinase protein kinase N, protein phosphatase 1, protein phosphatase 2a, protein kinase C-epsilon and phosphodiesterase 4D3 (accessed on 23.2.2023 https://www.ncbi.nlm.nih.gov/gene/10142#gene-expression).

4.6.3 Pre-surgery versus post-surgery comparison

Only 3 out of post-surgical samples were sequenced and their mutation analysis results were compared to those from the pre-surgical samples. Four variants were found in the presurgical cfDNA (not present in gDNA) that were considered potentially pathogenic. Out of these 4 variants only one in the *SMARCA4* gene was detectable in post-surgical cfDNA. The variant was detectable at 3 times lower allele frequency (AF) in the post-surgical sample, suggesting not a complete surgical tumor removal. Data is shown in the table R1. Results are also presented in graphical form in the figure R16.

Diagnosis	Patient ID #	Gene	Pre-surgery AF	Post-surgery AF	Reference blood AF
Oligoastrocytoma	2	PCM1	0.0118	0	0
Glioblastoma	6	MTOR	0.0213	0	0
Glioblastoma	6	SMARCA4	0.0748	0.0264	0
Glioblastoma	22	PML	0.0101	0	0

Table R1. Comparison of specific variants that have been detected in cfDNA isolated from pre-surgery collected blood versus post-surgery collected blood.

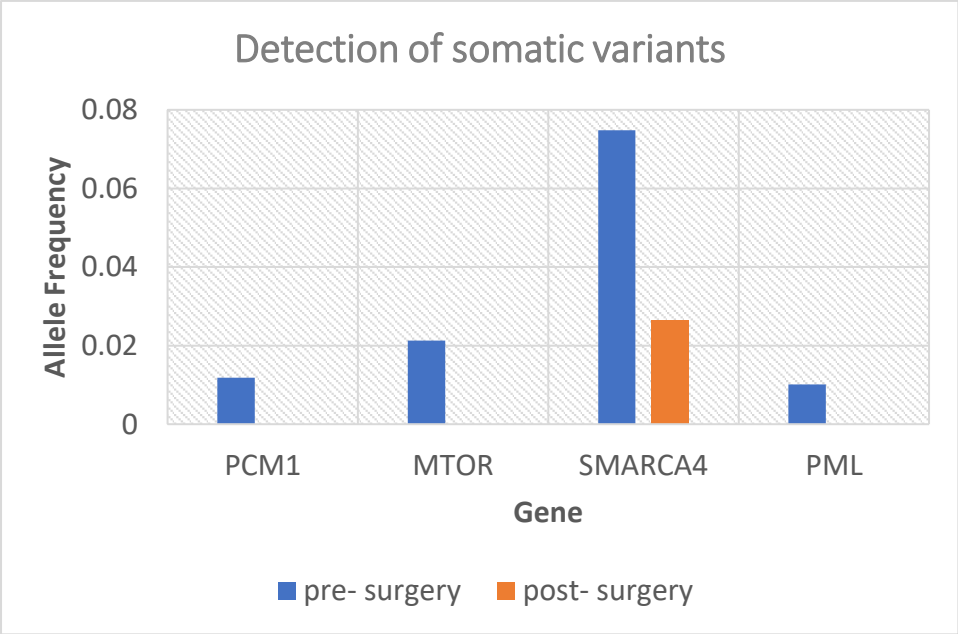


Figure R16. Comparison of the allele frequency of the cfDNA specific signal that was isolated from pre-surgery and post-surgery collected blood.

4.6.4 Identification of tumor mutations in cfDNA

Variants that were found in the tumor DNA and cfDNA, but absent in gDNA, were compared and the allele frequency was reviewed to establish sensitivity of the tumor signal detection. Somatic variants detected in cfDNA are shown below in the figure R17.

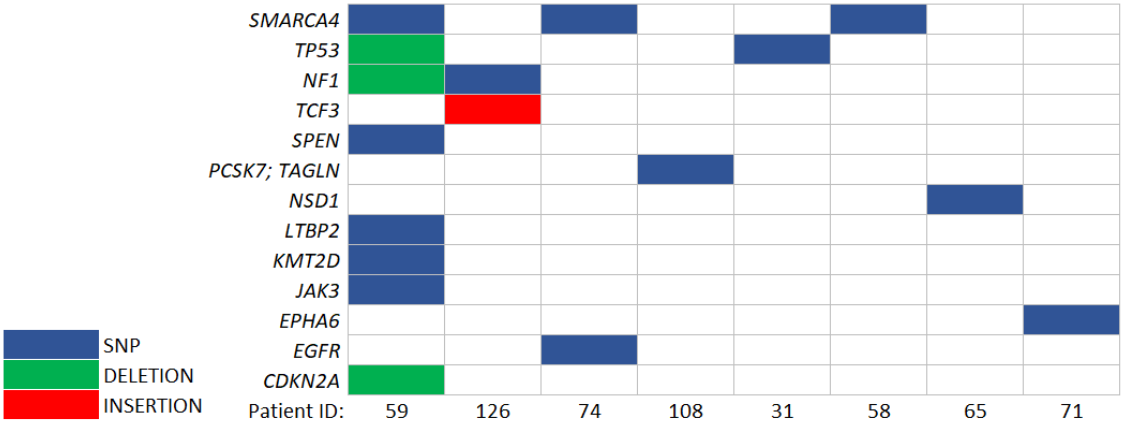


Figure R17. Waterfall Plot: somatic tumor variants detected also in cfDNA.

Patient #31 had in the tumor DNA the detectable mutation in the*TP53* gene that had fully penetrated (97% reads in the tumor were altered), but was barely detectable in the cfDNA (0.9% of reads in cfDNA were altered). Surprisingly, the *SMARCA4* SNP was detected in the PCNSL DNA sample and in cfDNA. Metastatic cancer samples gave a strong somatic variant signal varying from 9% to 24% of altered reads. The detailed overview of the results is shown in table R2.

							gDNA		tumor DNA			cfDNA		
Gene	Chr	Position	ID	Ref	Alt	Diagnostic Information	Reads		Reads		AF	Reads		AF
							All	Alt	All	Alt		All	Alt	
TP53	chr17	7577120	31	C	T	Glioblastoma, Grade 4	106	0	149	144	0.97	1219	11	0.009
SMARCA4	chr19	11170654	58	G	A	Primary Central Nervous System Lymphoma	49	0	55	23	0.42	602	9	0.015
SMARCA4	chr19	11144125	59	C	T	Anaplastic Thyroid Cancer Metastasis	203	0	77	55	0.71	1979	454	0.229
TP53	chr17	7579372		GC	G		324	0	128	76	0.59	1783	435	0.244
SPEN	chr1	16260997		G	T		237	0	141	82	0.58	2308	458	0.198
KMT2D	chr12	49438655		C	G		214	0	153	31	0.2	2238	229	0.102
LTBP2	chr14	75078119		T	G		20	0	15	9	0.6	279	50	0.179
NF1	chr17	29560103		GA	G		176	0	164	42	0.26	2444	227	0.093
CDKN2A	chr9	21971193		GC	G		148	0	95	66	0.69	1199	180	0.15
JAK3	chr19	17952151		G	T		29	0	9	5	0.56	1246	270	0.217
NSD1	chr5	176720936	65	G	C	Adenocarcinoma Lung Metastasis	390	0	418	191	0.46	862	183	0.212
EPHA6	chr3	96728829	71	G	GTT	Glioblastoma, Grade 4	11	0	23	3	0.13	618	14	0.023
SMARCA4	chr19	11144182	74	G	A	Astrocytoma Anaplasticum, Grade 3	43	0	259	76	0.29	1602	17	0.011
EGFR	chr7	55210075		T	G		123	0	3020	1514	0.5	1694	427	0.252
PCSK7; TAGLN	chr11	117076708	108	T	C	Glioblastoma, Grade 4	12	0	67	10	0.15	1539	340	0.221
NF1	chr17	29563087	126	T	G	Glioblastoma, Grade 4	67	0	112	3	0.03	2141	36	0.017
TCF3	chr19	1619749		A	AGGGTG		38	0	73	15	0.21	1281	310	0.242

Table R2. Somatic variants detected in cfDNA.

We checked if the SureSelect XT pipeline analysis software was not responsible for the decrease of sensitivity and low detection of the somatic variants in cfDNA. BAM files (compressed binary SAM files version used to represent sequences aligned up to 128 Mb) generated either with the mutect or Surecall pipeline were reviewed using the Integrative Genomic Viewer (IGV). SNPs occurring in cfDNA, tDNA and gDNA were compared side by side

in the *SMARCA4* (figure R18) and *IDH* genes (figure R19). While the variant in the *SMARCA4* gene was detected in the bioinformatic analysis, the variant in the *IDH1* gene was not. The IGV analysis of cfDNA SNP data has 5 out of a total of 1211 reads that had a well-known substitution in the *IDH1* that is clinically relevant for gliomas. The variant was likely lost during data processing by software quality control. Conclusions from the IGV manual analysis were that: mutect and Surecall pipelines generate data of similar quality, however during the quality control some reads of interest might be lost. Overall tumor variants was detected in cfDNA in 8 out of 84 patients, including 5 out of 80 glioma patients.

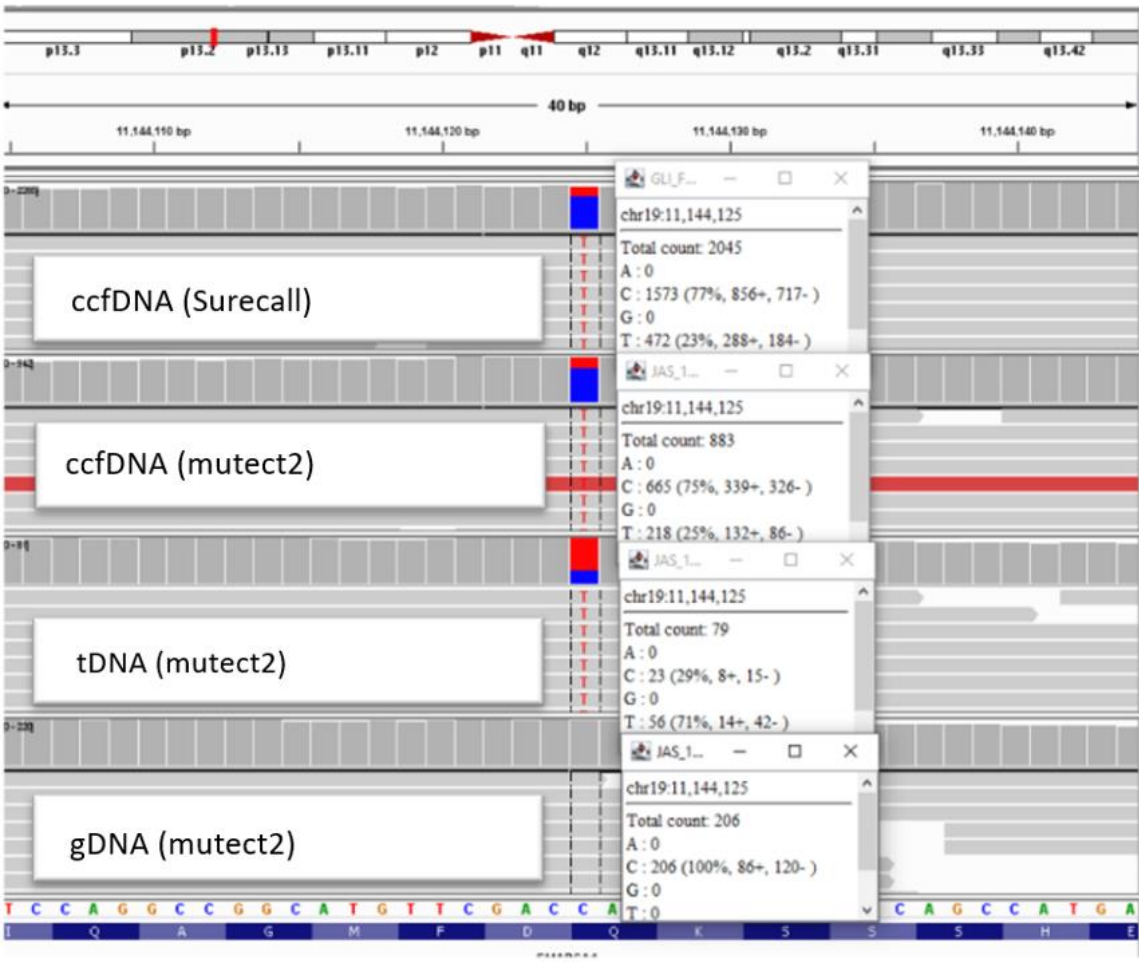


Figure R18. A snapshot of the Integrative Genomic Viewer (IGV) of BAM files in the manual analysis. The *SMARCA4* gene variant is present in cfDNA from both pipelines, high penetration in the tumor sample (red versus blue), no detectable signal in gDNA.

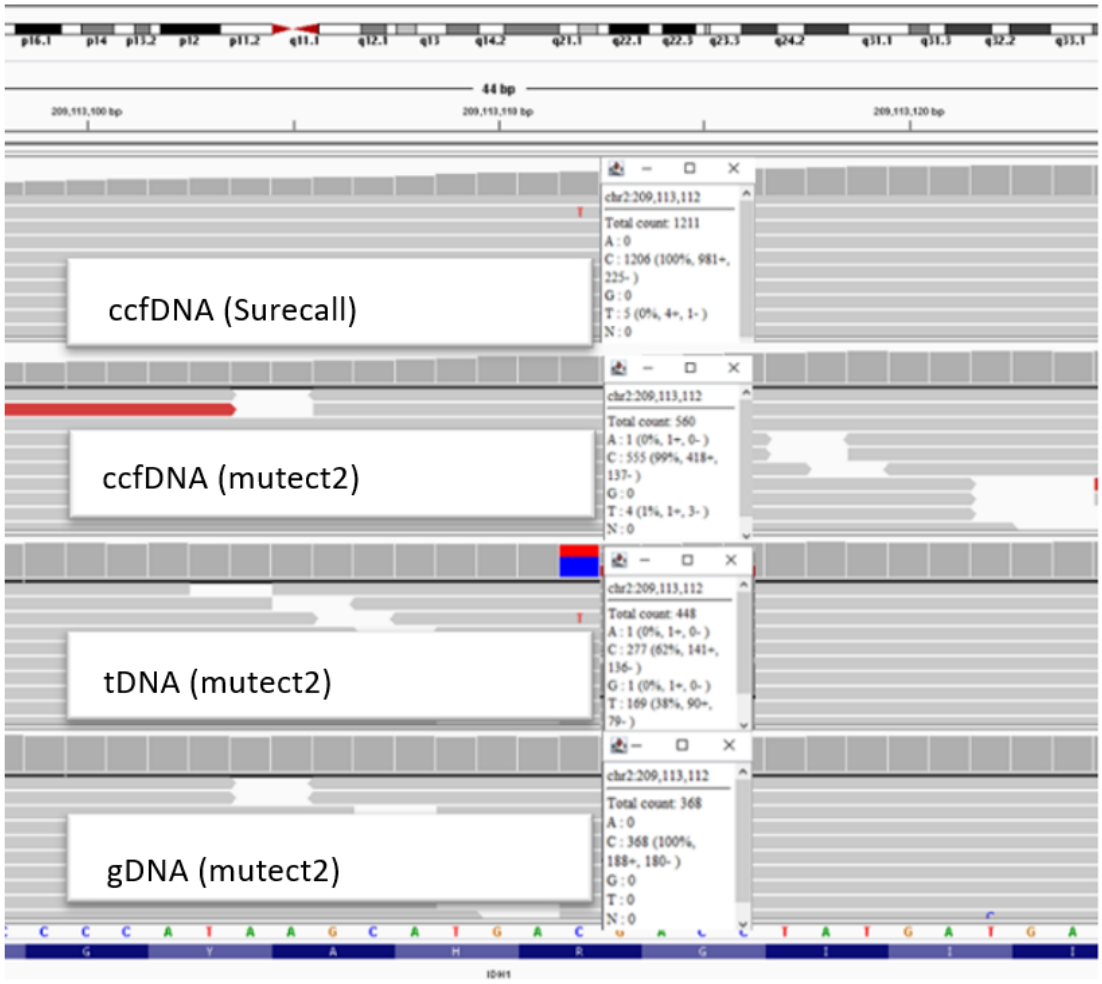


Figure R19. A snapshot of the Integrative Genomic Viewer (IGV) of BAM files in the manual analysis. The *IDH1* point mutation is present in the tumor tissue, barely detectable signal in cfDNA (Surecall 5 reads, mutec 4 reads), no signal in gDNA.

4.6.5 Identification of Copy Number Alteration (CNA) in cfDNA

Copy number alterations (CNA) are somatic alterations (copy number changes that occur only in cancer tissue), which often show a distinctive landscape with synchronous genomic gains and/or losses in gliomas[117]. Recently, a new, interesting method had emerged, that involves CNA analysis with the targeted panel sequencing. We tested its applicability to cfDNA sequencing. Four samples of cfDNA, in which we detected positive ctDNA signals, were chosen. Libraries were prepared from both cfDNA and gDNA using the SureSelect XT library prep kit. Special probes that determine a copy number change with a custom design covering the same gene region as the original SureSelect XT custom panel were used. We found numerous CNAs in cfDNA. The amplifications found in cfDNA of analyzed

samples that were registered in the COSMIC database are reported in the table R3 and visualized as the oncoplot in the figure R20.

Gene	Chrom	Start Position	Chrom Arm	Aberration Size in kb	Patient ID	Diagnosis	COSMIC ID number of cases registered
MTOR	chr1	10861762	p36.22	647.318	31	Glioblastoma, Grade 4	2
KMT2D	chr12	48786986	q13.12	235.435			2
EGFR	chr7	54858413	p11.2	398.301			100+
MUC1	chr1	155081162	q21.3-q22	111.340			1
PRDM16 (UPSTREAM)	chr1	2866654	p36.32	174.452			1
ASH1L (UPSTREAM)	chr1	155093862	q21.3-q22	98.640			1
RECQL4 (UPSTREAM)	chr8	144298772	q24.3	119.645	58	Primary Central Nervous System Lymphoma	4
EGFR	chr7	54821685	p11.2	197.695			100+
PIK3CA	chr3	179061157	q26.32	118.807			12
SARDH (DOWNSTREAM)	chr9	133747060	q34.2	62.409			11
BRAF	chr7	140539751	q34	216.595			20
RECQL4 (DOWNSTREAM)	chr8	144588017	q24.3	175.581	59	Anaplastic Thyroid Cancer Metastasis	4
MTOR (UPSTREAM)	chr1	10861762	p36.22	157.318			2
KMT2D	chr12	48808870	q13.12	213.551			2
EGFR	chr7	54867031	p11.2	537.130			100+
PIK3CA	chr3	178969860	q26.32	230.004			10
RECQL4 (UPSTREAM)	chr8	144325268	q24.3	151.444	65	Adenocarcinoma Lung Metastasis	4
MTOR	chr1	10861762	p36.22	647.318			2
KMT2D (UPSTREAM)	chr12	48786713	q13.12	221.853			2
MUC1	chr1	154944780	q21.3-q22	403.170			1
PRDM16 (UPSTREAM)	chr1	2829228	p36.32	211.878			1
ASH1L	chr1	155093862	q21.3-q22	254.088			1
ARID1A	chr1	26448692	p36.11	562.919			1
TYK2	chr19	10105892	p13.2	511.550			1
CIC (UPSTREAM)	chr19	42023328	q13.2	163.305			2
SPEN (UPSTREAM)	chr1	15618035	p36.21	223.403			1

Table R3. Copy number alterations present in ctDNA; List of genes affected by alterations, detailed position and size of CNAs.

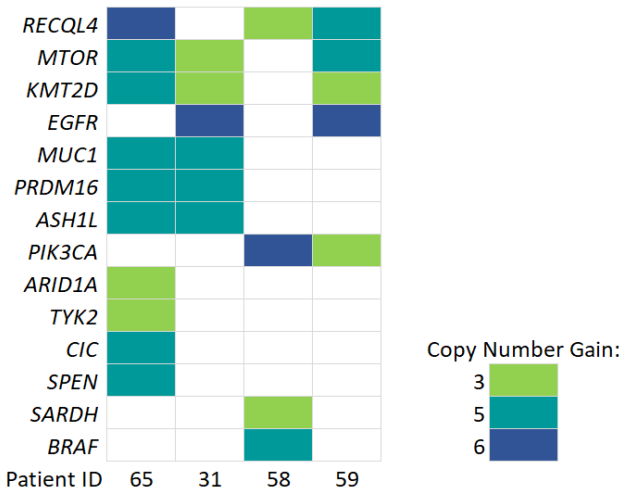


Figure R20. Copy number alterations present in ctDNA; Oncoplot showing overview of detected alterations with the copy gain number.

4.6.6 Detection of potentially pathogenic variants in cfDNA but not in gDNA

ctDNA analysis has often an issue with the false negative results. If the variant wasn't detected it does not guarantee that it wasn't present in original tumor, as we have only analyzed small piece of heterogenic entity and some variants are found only in its specific localizations. Genetic variant databases have improved significantly in recent years, with multiple current data uploads, which by itself can be used as a tool to assist in identification of extremely rare and potentially pathogenic variants (based on TOPMED, EXAC, gnomAD and 1000 genomes data projects). Likely pathogenic, pathogenic, or disease coexisting variants are registered in databases such as COSMIC or ClinVar. GBMs and other malignant gliomas are considered to be genetically heterogenous[49], thus analyzing one tumor fragment is intrinsically limited in capturing a complete mutational spectrum of the tumor. Based on these assumptions, we studied variants that appear only in cfDNA.

First, variants detected in cfDNA, but not in gDNA were filtered and only COSMIC registered variants were included. Most of the selected SNVs were also registered in the ClinVar database as pathogenic or likely pathogenic, some were extremely rare in population (MAF, AF 1000G, gnomAD) as shown in the table R4, presented below. Altogether, we found potentially pathogenic variants (ClinVar) in cfDNA in 25 patients, specifically in 24 glioma patients. These variants were not detected in reference genomic DNA or tumor.

								gDNA (Maftools)			cfDNA (SureCall)		
Gene	**rs ID**	**ID**	**MAF**	**AF 1000G**	**gnomAD**	**ClinVar clinsig**	**Diagnostic Information**	**Reads**		**AF**	**Reads**		**AF**
								All	**Alt**		**All**	**Alt**	
APC	rs752654519	6	-	-	-	pathogenic / likely pathogenic	Glioblastoma, Grade 4	225	0	0	201	5	0.0249
TSC2	rs397515228	11	-	-	-	pathogenic	Diffuse Glioma, Grade 2	204	0	0	303	6	0.0198
APC	rs886039642		-	-	-	pathogenic / likely pathogenic		168	0	0	172	4	0.0233
TSC2	rs45517360	22	-	-	-	pathogenic	Glioblastoma, Grade 4	53	0	0	101	6	0.0594
JAK3	rs145751599	30	0	0.0004	2E-05	uncertain significance	Glioblastoma, Grade 4	212	0	0	200	4	0.02
NF1	rs377662483	31	0	0.0002	2E-05	uncertain significance	Glioblastoma, Grade 4	129	0	0	814	9	0.0111
NF1	rs876657714	33	-	-	-	pathogenic	Glioblastoma, Grade 4	252	0	0	470	4	0.00851
TP53	rs587781589		-	-	-	pathogenic		239	0	0	345	3	0.0087
NSD1	rs587784080		-	-	-	pathogenic		250	0	0	557	5	0.00898
EGFR	rs139236063	34	-	-	4E-06	likely pathogenic	Glioblastoma, Grade 4	120	0	0	2897	39	0.0135
NSD1	rs587784169	50	-	-	-	pathogenic	Diffuse Astrocytoma, Grade 2	154	0	0	428	4	0.00935
NSD1	rs794727176	53	-	-	-	pathogenic	Glioblastoma, Grade 4	239	0	0	335	3	0.00896
NF1	rs746824139	55	-	-	0	pathogenic	Glioblastoma, Grade 4	144	0	0	424	5	0.0118
PTEN	rs190070312	64	-	-	-	pathogenic	Glioblastoma, Grade 4	246	0	0	436	5	0.0115
PTEN	rs121913294	65	-	-	-	likely pathogenic	Adenocarcinoma Lung Metastasis	139	0	0	352	3	0.00852
RECQL4	rs549497811	68	0	0.0002	2E-05	uncertain significance	Glioblastoma, Grade 4	240	0	0	563	9	0.016
BRAF	rs397516894	70	-	-	-	pathogenic	Glioblastoma, Grade 4	228	0	0	372	4	0.0108
NF1	rs376576925		-	-	4E-06	pathogenic		195	0	0	613	6	0.00979
NF1	rs878853884		-	-	-	pathogenic		118	0	0	626	8	0.0128
MTOR	rs587777894	79	-	-	-	pathogenic	Glioblastoma with Oligodendroglioma Component, Grade 4	133	0	0	250	8	0.032
NSD1	rs570278338		-	-	-	pathogenic		65	0	0	132	2	0.0152
PTEN	rs786204927		-	-	-	likely pathogenic		99	0	0	180	7	0.0389

NF1	rs533110479	82	0	0.0002	3E-05	uncertain significance	Glioblastoma, Grade 4	243	0	0	436	5	0.0115
KMT2D	rs886043414	83	-	-	-	pathogenic	Glioblastoma, Grade 4	144	0	0	143	2	0.014
TP53	rs876658483	85	-	-	-	pathogenic	Glioblastoma, Grade 4	198	0	0	347	4	0.0115
TSC2	rs45517179		-	-	-	pathogenic		248	0	0	393	4	0.0102
MED12	rs762659794		0	0.0003	6E-06	uncertain significance		115	0	0	161	4	0.0248
PIK3CA	rs121913279	86	-	-	4E-06	pathogenic FDA recodnised	Giant Cell Glioblastoma, Grade 4	245	0	0	139	2	0.0144
NOTCH1	rs371414501		0	0.0002	2E-05	uncertain significance		178	0	0	165	3	0.0182
SMARCA4	rs563079629		0	0.0002	5E-05	uncertain significance		58	0	0	136	5	0.0368
NF1	rs760703505	87	-	-	8E-06	pathogenic / likely pathogenic	Glioblastoma, Grade 4	241	0	0	485	4	0.00825
PTEN	rs746930141	93	-	-	-	pathogenic	Glioblastoma, Grade 4	70	0	0	199	2	0.0101
APC	rs587779783	99	-	-	-	pathogenic	Diffuse Astrocytoma, Grade 2	250	0	0	643	6	0.00933
NF1	rs199474752	100	-	-	-	likely pathogenic	Glioblastoma, Grade 4	168	0	0	356	3	0.00843
BRAF	rs121913378	105	-	-	-	likely pathogenic	Pleomorphic Xanthoastrocytoma, Grade 2	209	0	0	273	8	0.0293
PTEN	rs587776670	107	-	-	-	pathogenic	Glioblastoma, Grade 4	104	0	0	482	4	0.0083

Table R4. COSMIC registered variants detected in cfDNA, but not in gDNA.

We compared variants present in cfDNA (but not detectable in gDNA) with the identified somatic pathogenic variants from the current study (Results section 1) and from the previous study (n=57)[103]. Likely pathogenic variants detected in cfDNA, confirmed as somatic in the current study, are shown in the table R5.

								gDNA (Maftools)			cfDNA (SureCall)			tumor
Gene	Chr	position	ID	rs ID	Registered in COSMIC	ClinVar clinsig	Diagnostic Information	Reads		AF	Reads		AF	# of patients with somatic variant
								All	Alt		All	Alt		
SMARCA4	19	11144125	59		yes	-	Anaplastic Thyroid Cancer Metastasis	186	0	0	1979	454	0.23	1
PIK3CA	3	178952085	86	rs121913279	-	likely pathogenic	Giant Cell Glioblastoma, Grade 4	245	0	0	139	2	0.01	1
EPHA6	3	97365038	118	rs301948	-	-	Glioblastoma, Grade 4	199	0	0	1328	86	0.06	1
EPHA6	3	97365074	118	rs301949	yes	-	Glioblastoma, Grade 4	179	0	0	1611	103	0.06	1
EGFR	7	55210075	74		yes	-	Astrocytoma, Grade 3	229	0	0	1694	427	0.25	2
EGFR	7	55210075	114		yes	-	Astrocytoma, Grade 3	247	1	0.004	274	3	0.01	2
EGFR	7	55224307	114		yes	likely pathogenic	Astrocytoma, Grade 3	245	0	0	417	9	0.02	1
EGFR	7	55221822	64	rs149840192	yes	-	Glioblastoma, Grade 4	181	3	0.016	703	9	0.01	3

Table R5. Identification of somatic, likely pathogenic variants detected in cfDNA based on comparison with detected somatic variants in this cohort.

Variants detected in cfDNA, confirmed as somatic in the previous study from our group[103], are shown in the R6. One specific benign variant in the *PTEN* (rs12573787) was detected in cfDNA from 10 patients and this is likely a passenger mutation specific to tumor tissue. Likely pathogenic *TP53* and *EGFR* variants were present in 2 additional samples of cfDNA. For example the *EGFR* variant (rs149840192) that was found in cfDNA of the patient 64, was registered in 36 brain tumor cases in the COSMIC database and was confirmed as somatic in 1 patient from our previous study[103] and in 3 patients from the current study.

							gDNA			cfDNA			Tumor DNA
Gene	ID	rs ID	COSMIC (CNS) / Polyphen Pred	GMAF	ClinVar clinsig	Diagnostic Information	Reads		AF	Reads		AF	AF
							All	Alt		All	Alt		
PTEN	2	rs12573787	-/-	0.16	benign	Oligoastrocytoma, Grade 2	60	0	0.00	255	8	0.031	0.6923
	32					Oligodendroglioma Anaplasticum, Grade 3	79	1	0.01	150	62	0.413	
	65					Adenocarcinoma Lung Methastasis	55	1	0.02	61	33	0.541	
	85					Glioblastoma, Grade 4	56	1	0.02	260	140	0.538	
	109					Anaplastic Pleomorphic Xantoastrocytoma, Grade 3	57	0	0.00	168	13	0.077	
	110					Glioblastoma, Grade 4	39	2	0.05	234	144	0.615	
	111					Glioblastoma, Grade 4	107	1	0.01	160	68	0.425	
	117					Glioblastoma, Grade 4	73	1	0.01	299	144	0.482	
	118					Glioblastoma, Grade 4	40	1	0.03	224	114	0.509	
	121					Giant Cell Glioblastoma, Grade 4	52	0	0.00	340	164	0.482	
TP53	90	rs121913343	131/D	-	pathogenic/ likely pathogenic	Glioblastoma, Grade 4	147	3	0.02	1313	27	0.021	0.2619
EGFR	64	rs1057519828	14/D	-	likely pathogenic	Glioblastoma, Grade 4	225	0	0.00	612	22	0.036	0.4502
EGFR	64	rs149840192	36/D	-	likely pathogenic	Glioblastoma, Grade 4	181	3	0.02	703	9	0.013	0.248

Table R6. Comparison of variants detected in cfDNA but not in gDNA and somatic variants from our previous study[103].

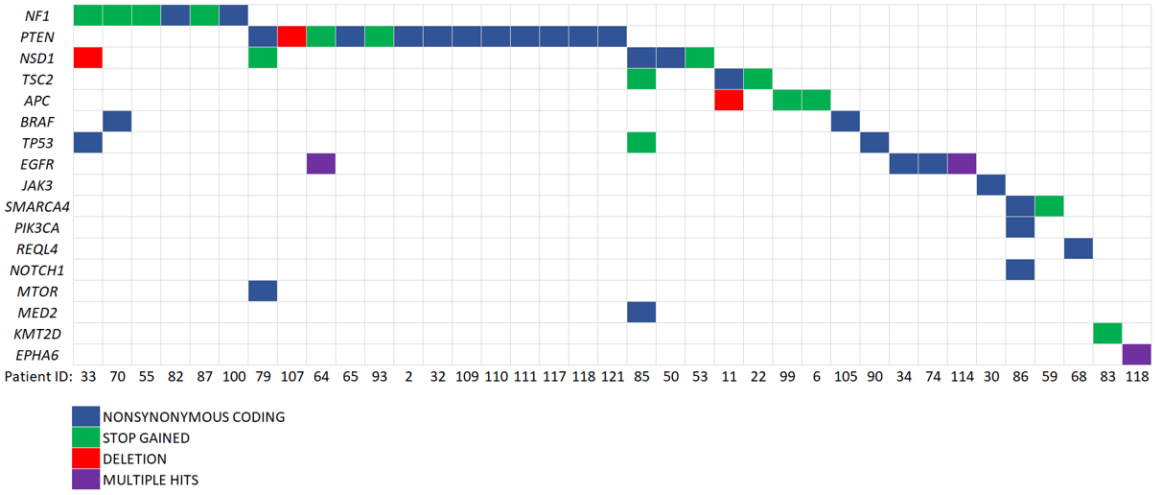


Figure R21. Waterfall Plot: a summary of potentially pathogenic variants detected in cfDNA.

Overall results yielded detection of the potentially pathogenic mutations in cfDNA of 37 out of 84 patients, as presented in the figure R21. Interestingly, some likely pathogenic, somatic variants were identified only in cfDNA which suggests that cfDNA from plasma reflects a different spectrum of genetic alterations then these found in tumors.

4.7 The analysis of cfDNA from plasma collected from a Neck artery – case study

Blood after leaving a brain tumor travels through brain veins and consecutively into superior vena cava, the right anterior of the heart, right ventricle of the heart, lungs, left atrium of the heart, left ventricle of the heart, aorta, tissues, and finally it reaches deep and superficial veins where it can be collected from in a typical blood draw.

We would like to determine if a source of the blood collection has an impact on the detection of circulating tumor DNAs. In the presented case study blood was collected from the superficial vein and neck artery leaving the brain tumor area. cfDNA was isolated from those blood samples and compared to determine if there are any differences in cfDNA between these samples. Overall case study design is pictured in the figure R22.

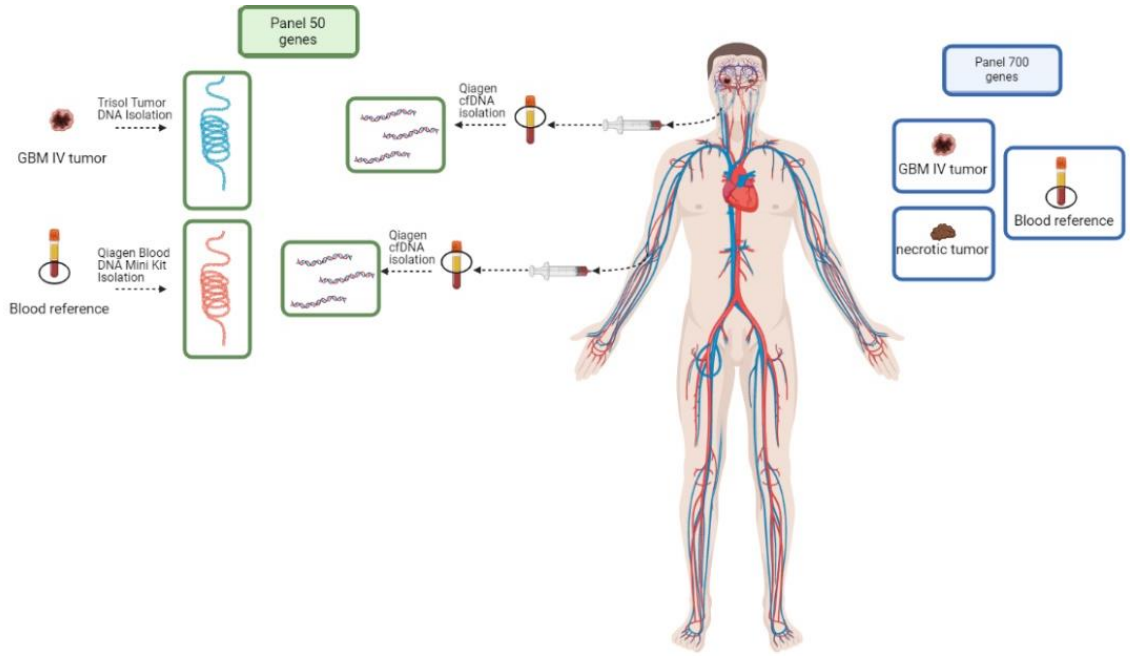


Figure R22. Neck artery liquid biopsy case study design overview.

This case study was done on a single patient, the 59 years old male, diagnosed with the primary brain tumor. Magnetic Resonance Imaging (MRI) before and after surgical resection, and histopathological reports have shown that the patient had 2 brain tumors. First one was initially described as a primary tumor, it had well-defined borders and was later described as a small region of necrotic tissue, not increasing in size, possibly benign. Second tumor was an aggressive and diffusive WHO Grade 4 Glioblastoma. Tumor samples were collected after surgery and DNA was isolated from both tumors and white blood cells from reference blood (gDNA). cfDNA was isolated from blood collected from the neck artery leaving the brain tumor area and from the superficial vein. Large panel sequencing was done tumor DNA and gDNA, so overall somatic variants present in these primary brain tumors could be reviewed in figure R23.

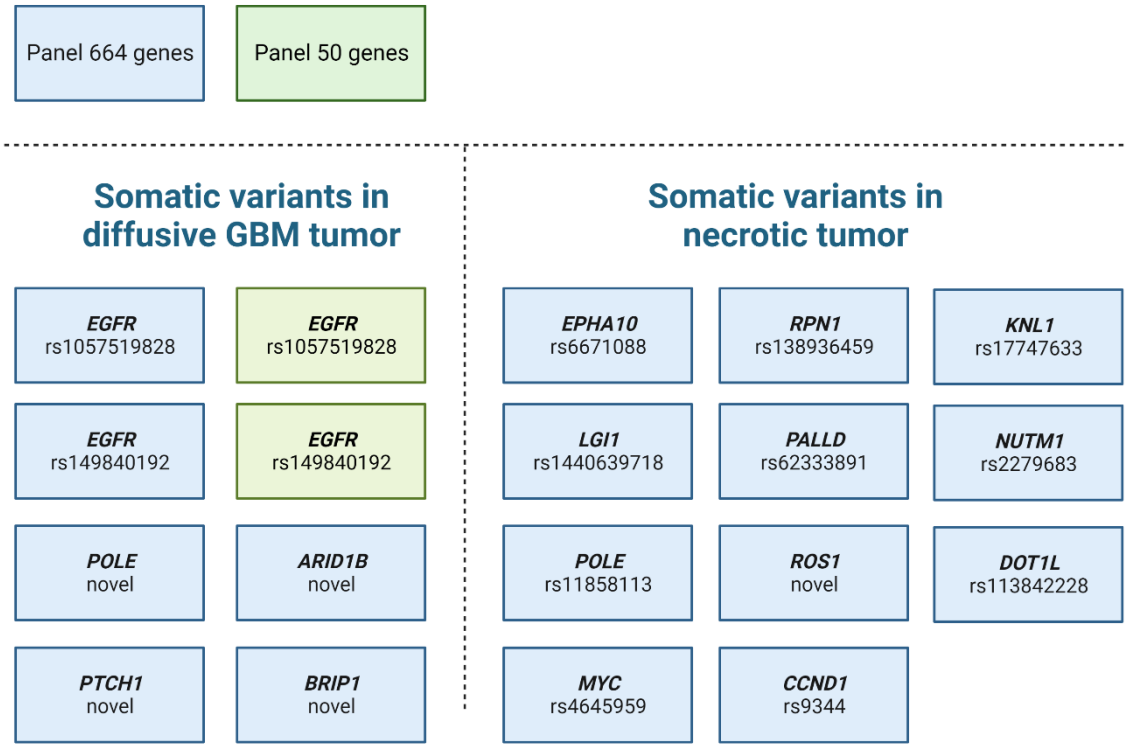


Figure R23. Somatic variants detected in each tumor tissue sample.

Interestingly, both tumors had mutations in distinct genes, which suggests that these tumors occurred independently. As highlighted in green, only few somatic variants detected in the GBM DNA were next detected using the 50 gene panel as unfortunately most of the mutated genes were not included. GBMDNA (t_2DNA), gDNA, ccf_1DNA (superficial vein), and ccf_2DNA(neck artery) were sequenced using the smaller 50 gene panel. Variants that were present in cfDNAs, t_2DNA, but not in gDNA are shown in the table R7.

Gene	Alteration Type	Allele frequency	# of alt reads	Read Depth	ID	COSMIC
SARDH	SNP	0.00995	15	1508	novel	
SARDH	SNP	0.329	5	152	novel	
KMT2C	SNP	0.064	172	2688	rs200559566	registered
KMT2C	SNP	0.0522	98	1879	rs200559566	registered
KMT2C	SNP	0.013	2	154	rs200559566	registered
KMT2C	SNP	0.0155	46	2962	rs202125566	registered
KMT2C	SNP	0.0219	50	2285	rs202125566	registered
KMT2C	SNP	0.0179	3	168	rs202125566	registered
TCF3	Insertion	0.263	425	1616	rs10648013	
TCF3	Insertion	0.242	310	1281	rs10648013	
TCF3	Insertion	0.205	15	73	rs10648013	
NF1	SNP	0.013	36	2774	novel	
NF1	SNP	0.0168	36	2141	novel	
NF1	SNP	0.0268	3	112	novel	
PRDM16	SNP	0.0471	43	912	rs3215936	
PRDM16	SNP	0.0142	36	763	rs3215936	
PRDM16	SNP	0.0159	3	189	rs3215936	

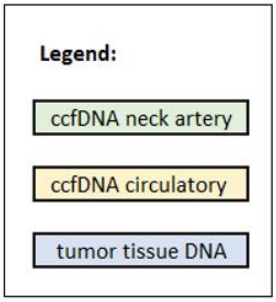


Table R7. Identified somatic variants in tumor DNA and cfDNA isolated from blood collected from different locations.

We found the insertion in the *TCF3* gene encoding the transcription factor E protein, from a family of the helix-loop-helix factors. The AF was very high 20% in the tumor and even higher in cfDNA (24-26%). It is possible that due to tumor heterogeneity this insertion was present in other regions, which would explain a higher signal in cfDNA. The *KMT2C* gene coding for Histone-Lysine N-Methyltransferase 2C (*KMT2C*) had few putative variants that were registered in the COSMIC database as potentially pathogenic and were detected both in tumor and cfDNA, but not in gDNA. This analysis shows that there were potentially pathogenic variants detected in cfDNA of the patient, however method of blood collection has a little impact on the sensitivity.

Somatic mutations in the *DDX10* gene coding for DEAD-Box Helicase 10 (*DDX10)* - an ATP-dependent RNA helicase, were only detected in cfDNA, but not in the tumor. These variants were at slightly higher AF in the circulatory ccf_1DNA, than in ccf_2DNA. The variants found in the *DDX10* gene have not been reported so far in ClinVar or COSMIC; the variant with RS ID: rs375092397 is considered rare according to the MAF dbSNP database (0.004). This data is summarized in table R8.

Gene	Alteration Type	rs ID	Ref Allele	Alt Allele	ID	Allele frequency
DDX10	Insertion	rs375092397	T	TA	rs375092397	0.0301
DDX10	Insertion	rs375092397	T	TA	rs375092397	0.0382
DDX10	Deletion	rs756833840	GTGA	G	rs756833840	0.00642
DDX10	Deletion	rs756833840	GTGA	G	rs756833840	0.00708
DDX10	Deletion	rs761312942	TA	T	rs761312942	0.0632
DDX10	Deletion	rs761312942	TA	T	rs761312942	0.0742

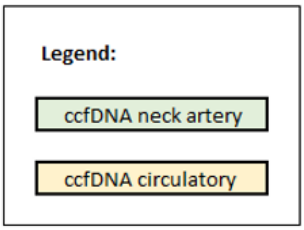


Table R8. Insertions and deletions identified in cfDNA, but not detected in gDNA or tDNA. Comparison of detection levels depending on the collection method of blood for cfDNA isolation.

Loss of heterozygosity in the *EP3000* gene coding for a E1A Binding Protein P300 (*EP3000*) was detected (table R9), as a mutation penetration in gDNA was below 50%, but in tumor tissue a penetration of this variant was 88%. In cfDNA AF is above 0.5, so the loss of heterozygosity is restricted to the tumor and can be detected in cfDNA. This *EP3000* (rs20552) variant is registered as benign in the ClinVar data base.

Gene	RS ID	Allele frequency	HOM/ HET
EP300	rs20552	0.523	HET
EP300	rs20552	0.514	HET
EP300	rs20552	0.877	HOM
EP300	rs20552	0.46	HET

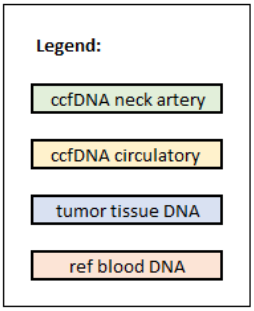


Table R9. Loss of heterozygosity identified in cfDNA. The information on variants identified in cfDNA from the neck artery in green, from the superficial vein in yellow; from tumor in blue and reference blood in pink.

Somatic variants found in the GBM tumor DNA are shown in the table R10. Two point mutations in the *EGFR* gene encoding Epidermal Growth Factor Receptor (*EGFR*) were detected, which is in agreement with the previous analysis that involved larger panel sequencing. Most of the variants were novel and have no RS ID, but from the variants with RS ID only one was categorized as likely pathogenic - the mutation in *EGFR* rs149840192. Point mutations in *IDH1* and *SARDH* were both labeled as of unknown significance in ClinVar.

Gene	Allele frequency	ID
SARDH	0.02	rs552148848
EGFR	0.224	rs1057519828
EGFR	0.279	rs149840192
KMT2C	0.0121	novel
MED12	0.0109	novel
ATRX	0.023	novel
ATRX	0.0225	novel
PIK3CA	0.0105	novel
IDH1	0.0182	rs762792137
DDX10	0.0198	novel

Table R10. Overall results from analysis of tumor DNA in the patient.

4.8 Identification of pathogenic variants in tumors and the corresponding primary cell cultures

We explored if a spectrum of genetic alterations is similar in freshly resected GBMs and primary cell cultures derived from those tumors and kept under normoxia and hypoxia (to better mimic tumor conditions). Targeted sequencing with the 660 genes panel was performed on tDNA, gDNA, DNA from cells kept under 21% O_2 and 5% O_2. We compared clonal variability, rare variants and copy number alterations. This complex analysis was performed on samples from 8 patients. All cell line cultures were maintained in media suitable to keep cancer stem like cells: consisting: DMEM-F12 +B27 +EGF +bFGF. Most of the cell cultures were maintained in the sphere cultures, except two cultures (patient # 31, 67), that grew as adherent cells.

Somatic variants were filtered using gDNA as a reference for all the presented results. Figure R24 shows a spectrum of detected somatic variants detected in DNA from cultured tumor cells.

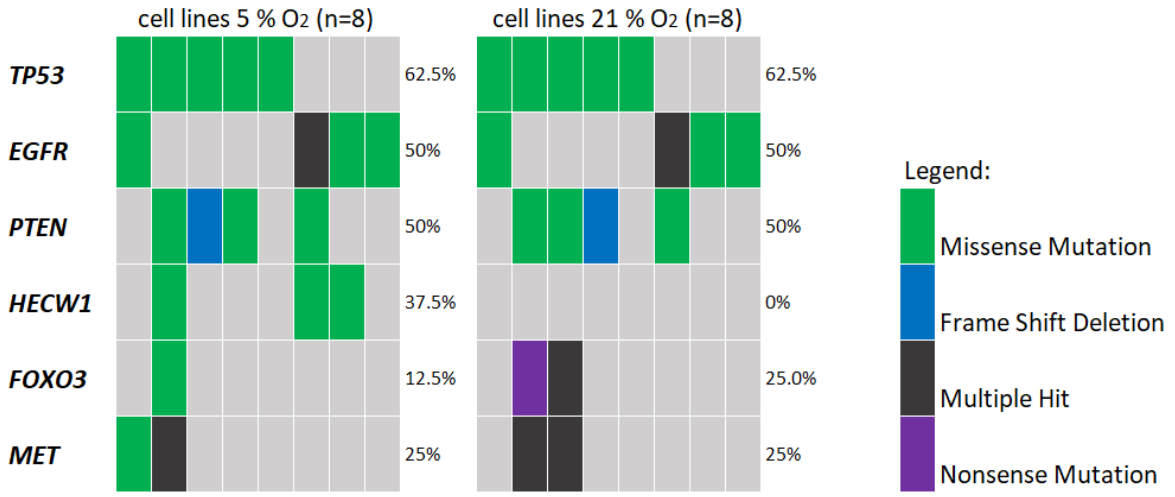


Figure R24. Waterfall plot showing the differences in somatic variants detected in the cell lines derived from the same tumor at 2 different oxygen conditions.

Similar variants in *TP53*, *EGFR*, *PTEN, FOXO3, MET* genes were detected in each cell culture irrespectively of the oxygen conditions. The variant in the *HECW1* codes for a HECT, C2 And WW Domain Containing E3 Ubiquitin Protein Ligase 1 that mediates ubiquitination and degradation of the cell proliferation regulating protein *DVL1* (Dishevelled segment polarity protein 1). The identified mutation in this gene may impact the functionality of the ubiquitin protein ligase E3 and lead to aggregation of *DVL1* which then can subsequently impact cell proliferation. The cells expressing the variant are lost during culturing in normoxia conditions. Comparison of the spectra of somatic variants detected in cell cultures and tumor shows that in most cases the mutations detected in the tumor are kept in cultured cells (figure R25).

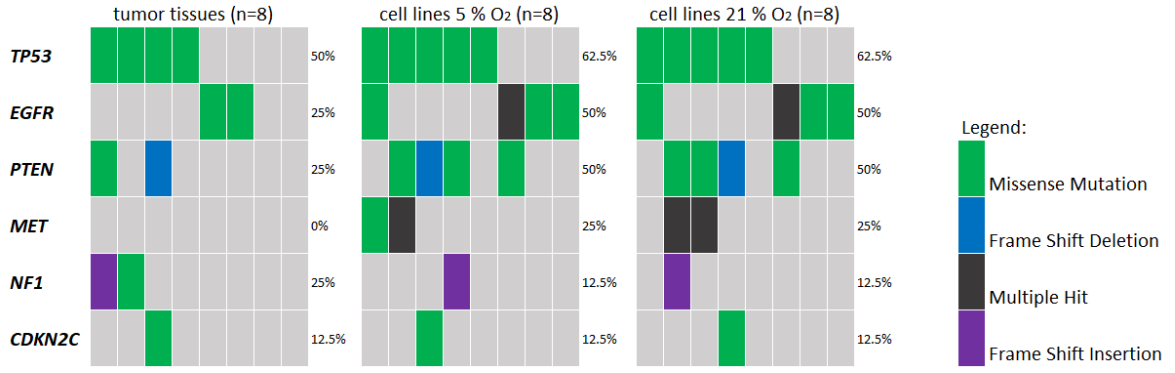


Figure R25. Waterfall plot showing somatic variants detected in the original tumor and tumor-derived cell cultures kept under different oxygen conditions.

The mutation in the *HECW1* that was not detected in cells cultures at 21% O_2 was also not detected in the tumor DNA. This can indicate that this variant appears in cells during clonal selection and accumulates under hypoxic (5% O_2) conditions. The *NF1* mutation detected in the tumor was lost in one of the cell cultures. We found the higher rate of somatic variants in *TP53*, *EGFR*, *PTEN*, *MET* genes in hypoxic cells compared to those in the tumor DNA.

The mutations in the *FOXO3* gene were detected in cells kept under normoxia, but not in the tumors. This suggest that this variant emerges under clonal selection and accumulates during cell culture, while it is hard to find in the tumors. Altogether, the presented data shows that the similar mutational spectrum is detected in the cultured cells irrespectively of oxygen concentration and due to clonal evaluation clones of cells carrying specific mutations accumulate with time in cell cultures.

Whole exome or whole genome sequencing would yield more reliable data for CNA analysis and even that targeted sequencing is not a method of choice for CNA analysis, data was analyzed to compare copy number alterations in specific gene regions. We explored the profiles of CNAs in the same samples to elucidate if patterns of alterations vary between tumors and cell cultures for each patient. The profile of CNAs in tumors and cell cultures is shown in the figure R26. In most cases (except patient TN1) the stronger alteration signal is detectable in cell lines, rather than in tumor. When comparing cell lines and tumors originating from one patient, cultured cell profiles were very similar to each other, but tumor signals differed. That was the case in all but one analyzed sample sets. Only in case of patient T3: normoxia cell cultures and tumor share similar profiles, differing slightly from cell cultured under hypoxia.

Finally, overlap of somatic variants found in the tumors is presented in the figure R27. It reflects differences in somatic variants detected in the original tumor tissue and cell lines derived from the tumor in 2 different oxygen conditions. As shown the co-occurrence is happening less often than finding mutations in genes that are mutually exclusive.

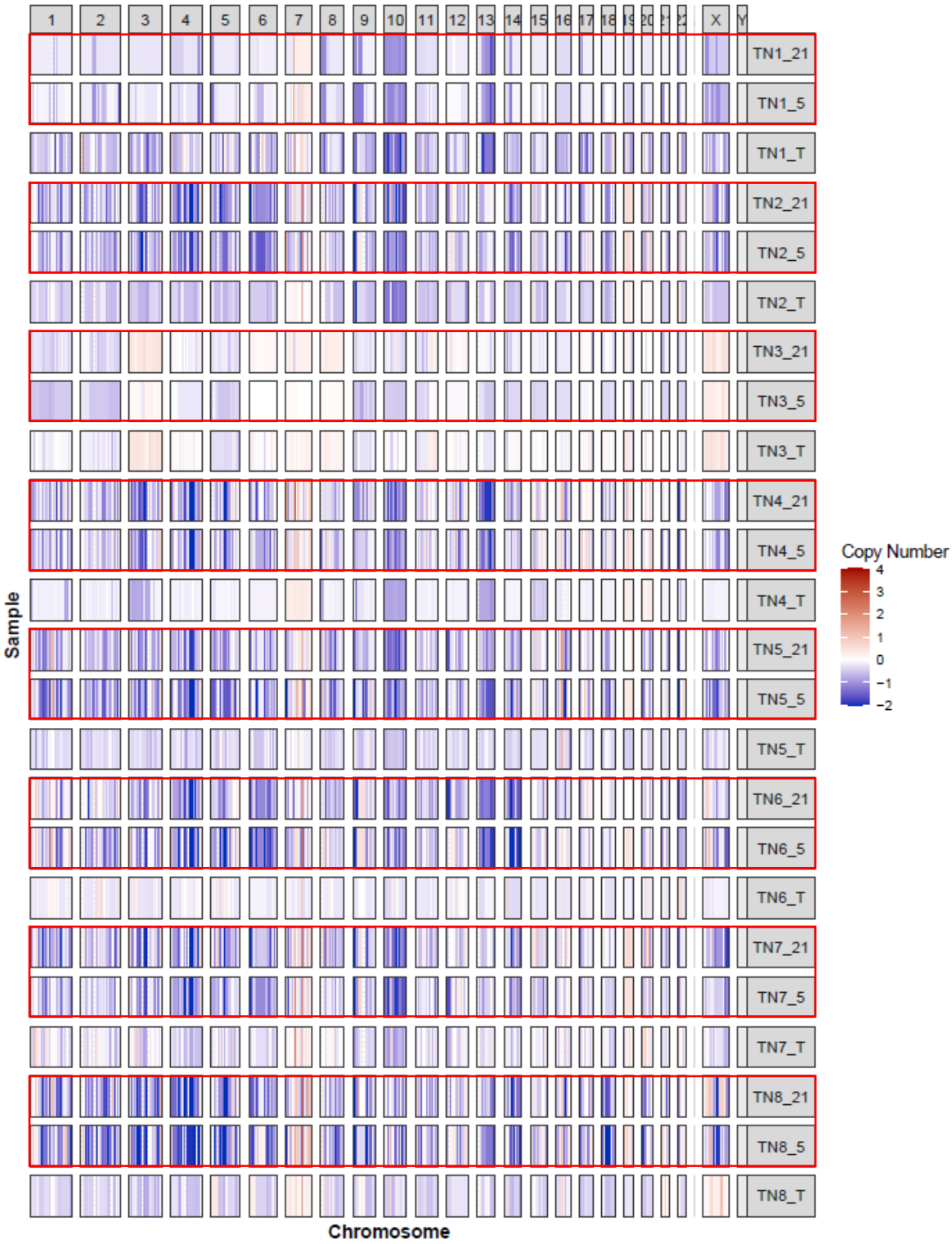


Figure R26. CNA plot showing the differences in somatic variants detected in the original tumor tissue and tumor derived cell cultures under different oxygen conditions (normoxia and hypoxia).

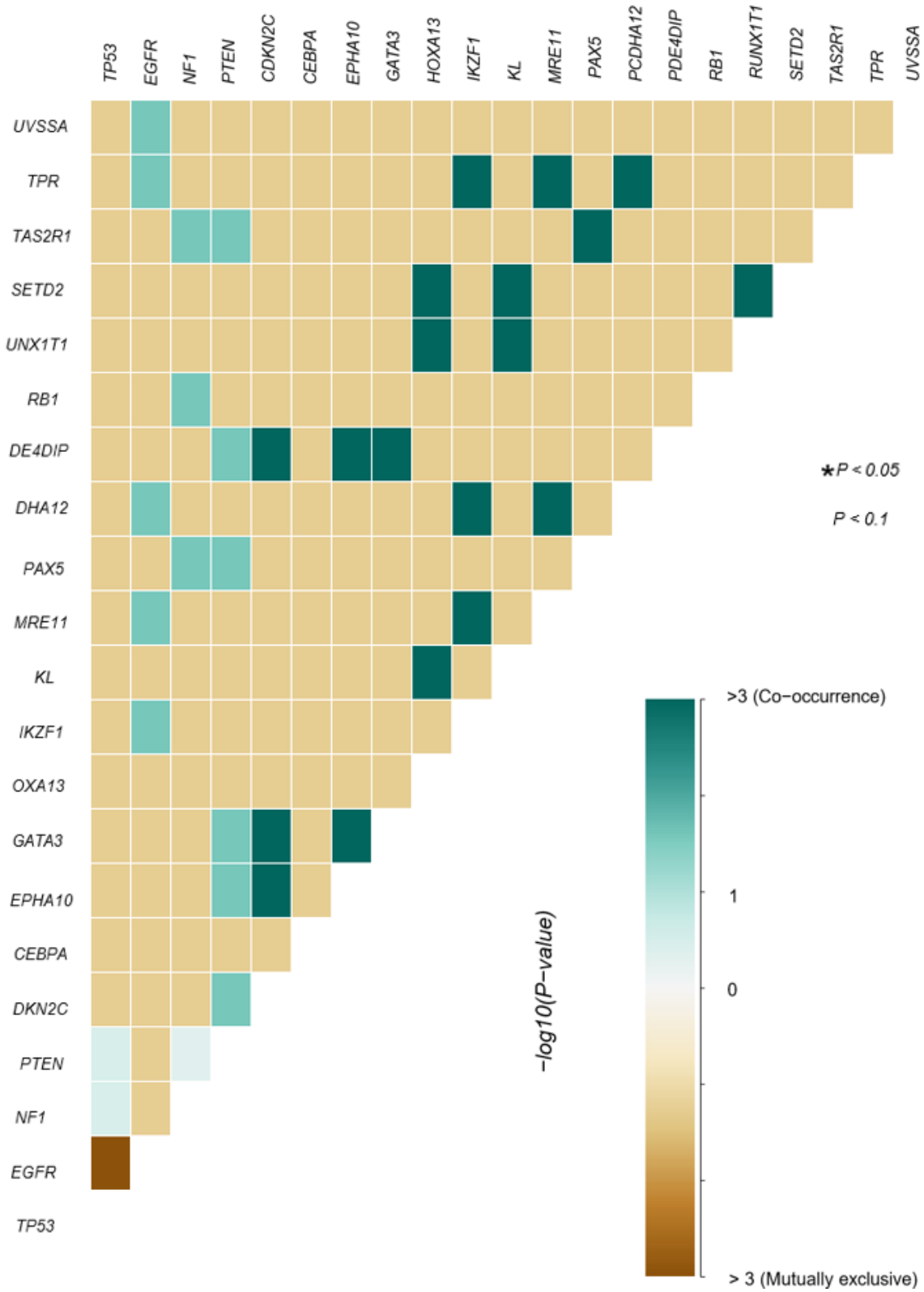


Figure R27. Plot showing somatic variants detected in the original tumor tissue and same tumor derived cell cultures under different oxygen conditions.

4.9 Identification of the mutational landscape in pediatric brain tumors

Despite the progress in identification of the mutational landscape in pediatric brain tumors, molecular diagnosis is not possible for many of them and a search for new mutations continues. In the collaboration with Prof. Wiesława Grajkowska, a neuropathologist for The Children's Memorial Health Institute a collection of poorly diagnosed pediatric tumors has been collected. DNA was isolated from FFPE sections and subjected to targeted sequencing with the 664-genes panel and analysis, however reference gDNA from blood was not available. Potentially pathogenic mutations detected in the analyzed pediatric gliomas are shown below in the figure R28.

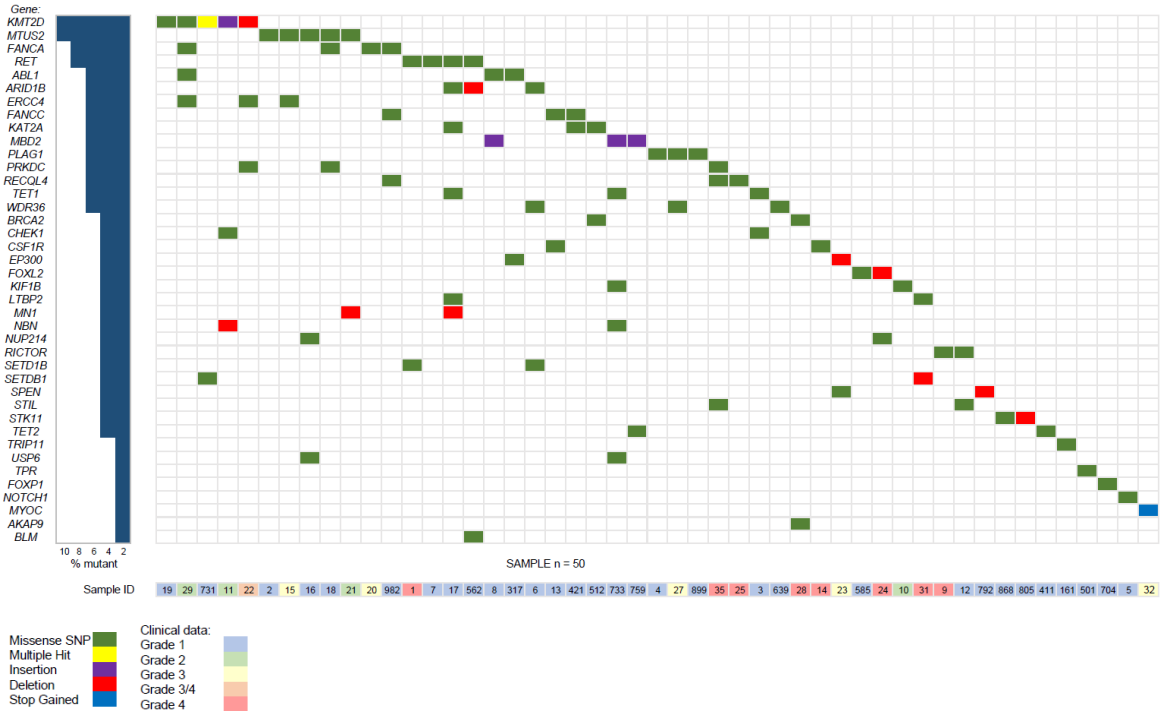


Figure R28. Overview of detected putative mutations in malignant brain tumors. A type of mutations and tumor grades are color coded.

The most commonly mutated genes in the analyzed cohort were *KMT2D* and *MTUS2*, that were altered in 10% of patients. *KMT2D* encodes Histone-Lysine N-Methyltransferase 2D, an epigenetic enzyme that methylates the Lys-4 position of histone H3. KMT2D was implicated in pancreatic carcinogenesis through metabolic reprogramming[118]. All mutations in the *KMT2D* gene are within regions coding for histone lysine N-methyltransferase activity, but in case of each patient, they were in different positions (figure R29).

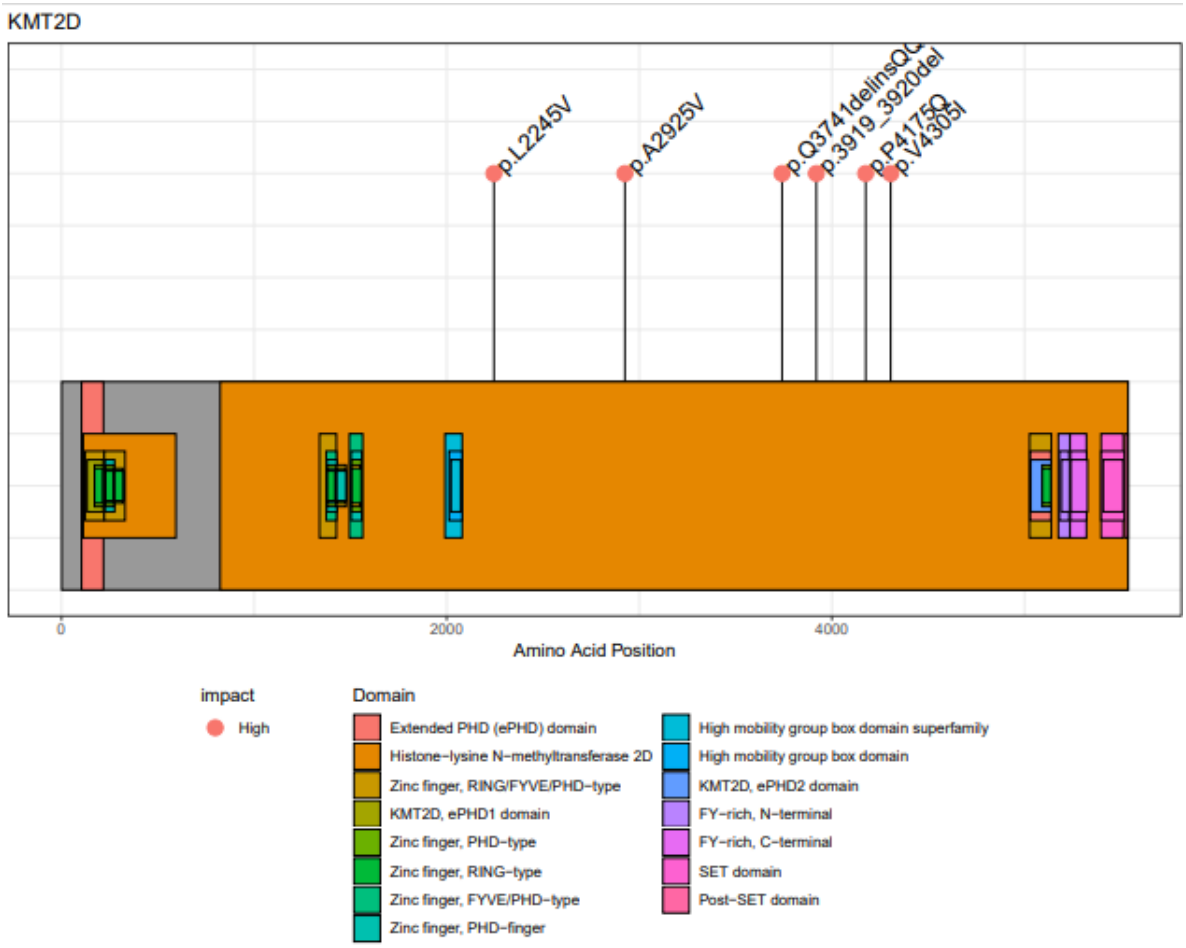


Figure R29. Lollipop plot shows the locations of the detected mutations in the *KMT2D* gene.

Another gene frequently (10%) affected by potentially pathogenic changes in the cohort was *MTUS2* (encodes Microtubule Associated Tumor Suppressor Candidate 2 protein). *MTUS2* is predicted to be involved in microtubule binding and protein homodimerization activity. The detected variant (RS200233522) is not registered in the ClinVar database and the recurrent mutation was found in the exactly same position in all 5 patients, as shown in the figure R30. The recurrent mutation in the *FANCA* gene (coding for Fanconi anemia complementation group) was found in 8% patients. The DNA repair protein FANCA may function in cell cycle checkpoint control or post replication repair. The recurrent mutation in the *FANCA* gene was also found in all 4 cases in the same particular location, as presented in the figure R31. Another potentially pathogenic variant that was detected in 8% of patients was a missense SNP in the *RET* gene. *RET* (Proto-Oncogene C-Ret) is a proto-oncogene, that encodes a receptor tyrosine

kinase for the extracellular signaling molecules, the members of glial cell line derived neurotrophic factor (GDNF) family[119].

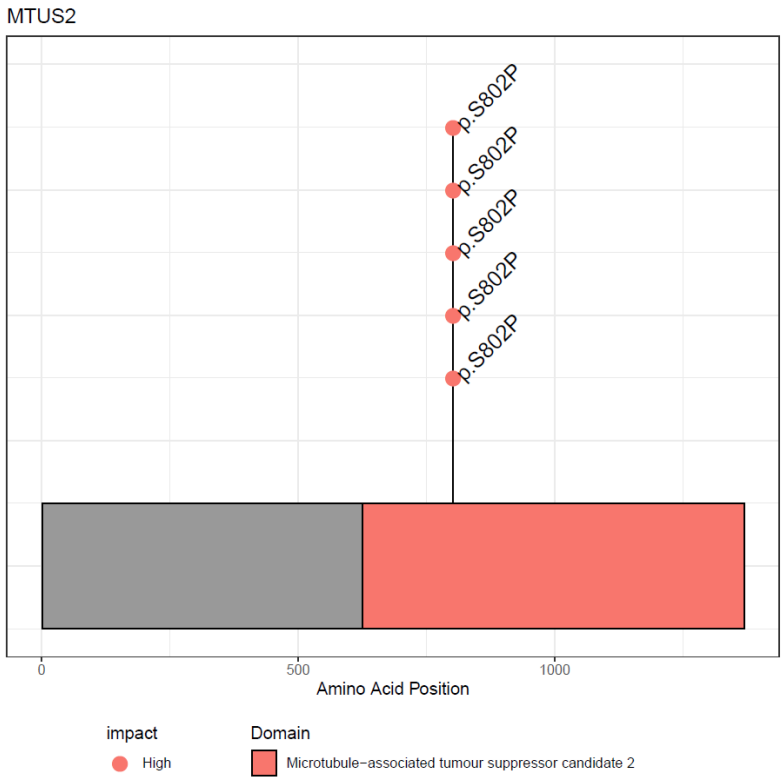


Figure R30. Lollipop shows locations of the detected mutations in the *MTUS2* gene.

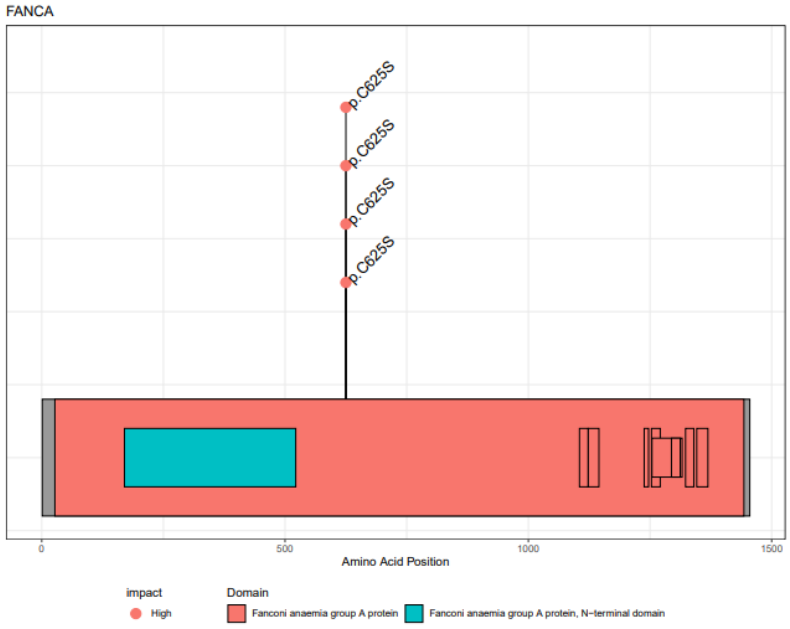


Figure R31. Lollipop shows locations of the detected mutations in the *FANCA* gene.

RET gain-of-function mutations are associated with many types of cancers and specifically the T1038A variant was registered and considered as pathogenic in multiple types of pediatric cancers[120,121]. SNPs concentrate in the region coding for a kinase catalytic domain, as shown in the figure R32. Therefore, the observed variations could change functionality of the important signaling kinase.

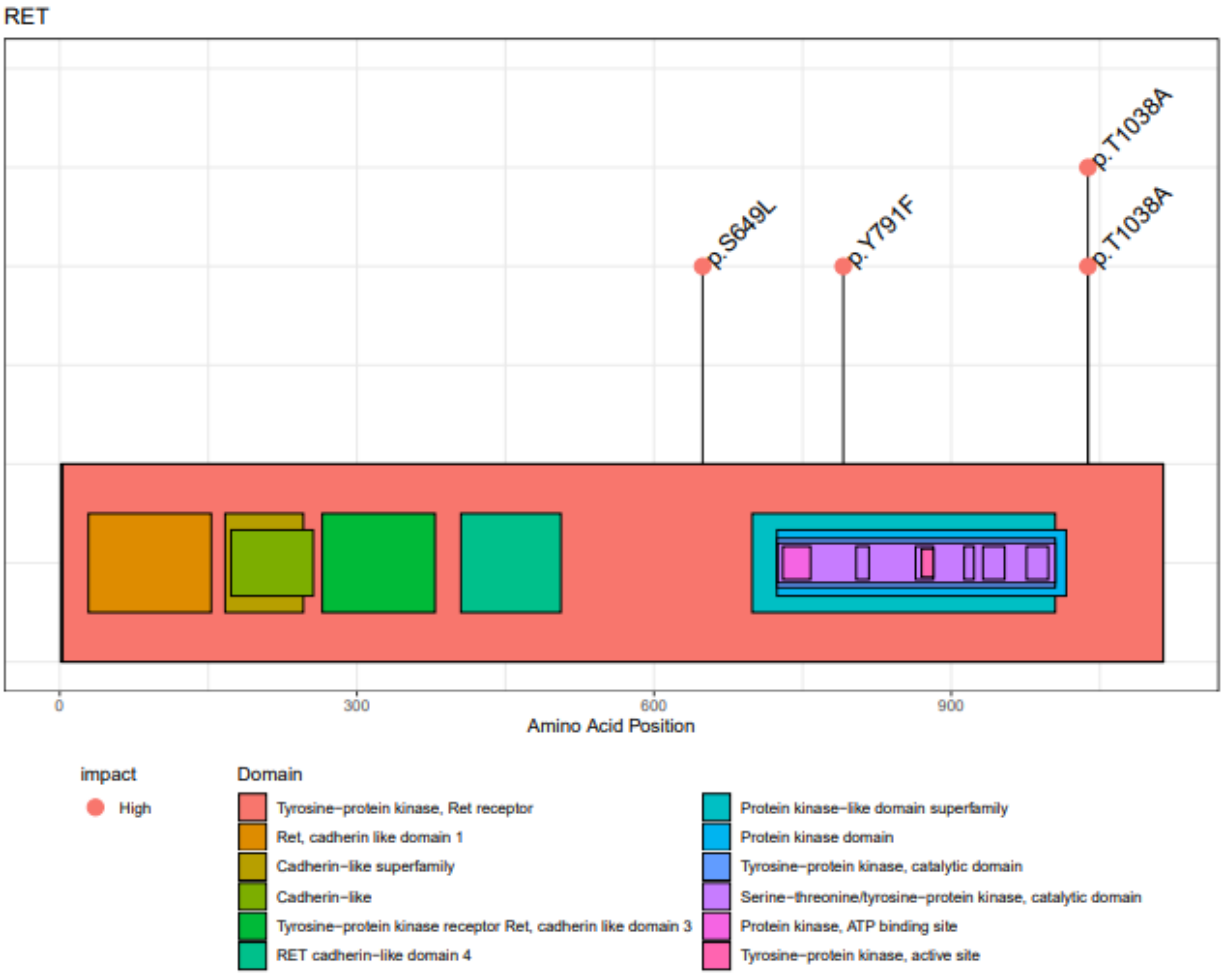


Figure R32. Lollipop shows locations of the detected mutations in the *RET* gene. SNPs concentrate in the region coding for a kinase catalytic domain.

5. Discussion

5.1. Identification of somatic and germline variants in tumors from the adult glioma patients cohort

The current WHO Glioma classification includes the molecular characteristics of the tumor tissue biopsy into diagnosis of brain tumors. Even within a single tumor, genetic alterations, pathophysiology and gene expression exhibit considerable heterogeneity[122]. Nevertheless, most malignant gliomas share similar changes affecting particular cellular signal transduction pathways or cellular activities, some of which present potential for therapeutic intervention[122]. In the current study targeted exome sequencing was used to identify somatic and germline mutations in 94 patients cohort that consisted of 79% of high-grade gliomas and 21% lower grade gliomas. The most commonly mutated genes in our study were *PTEN, TP53, EGFR, ATRX, IDH1* and *NF1*. The identified alterations and their frequencies are similar to previously described cohorts.

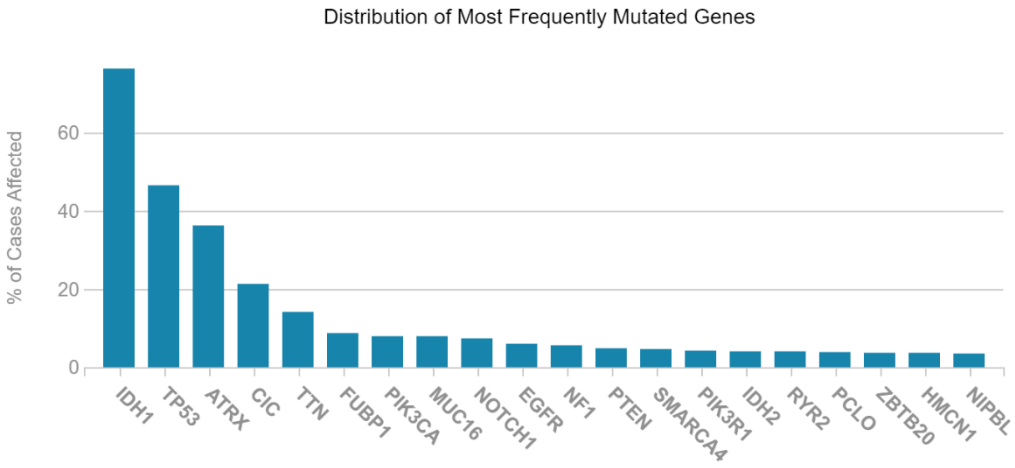


Figure D1. Most common mutations found in Low Grade Gliomas based upon TCGA data from 516 patients analysis (graph prepared using an Exploration Tool from National Cancer Institute GDC Data Portal, accessed on 17.4.2023: https://portal.gdc.cancer.gov/).

According to literature *IDH1* and *IDH2* mutations are present in about 70% of lower-grade gliomas [123] and in our study *IDH1* variant was found in 10%

of samples, but only 21% of our cohort consisted of lower-grade gliomas. The most common variants in lower-grade gliomas that are registered in TCGA study based upon 516 samples, are listed in figure D1.

Overall around 30-40% of gliomas carry *PTEN* mutations[124–126] and mutations in *TP53* are common in majority of high grade gliomas[127–129]. In case of our study mutations in *PTEN* were found in 30% of samples and these detected in *TP53* were present in 25% of analyzed tumors, which is consistent with literature. Figure D2 presents a frequency of the most commonly altered genes in case of 1172 samples of Glioma data, that was deposited to National Cancer Institute.

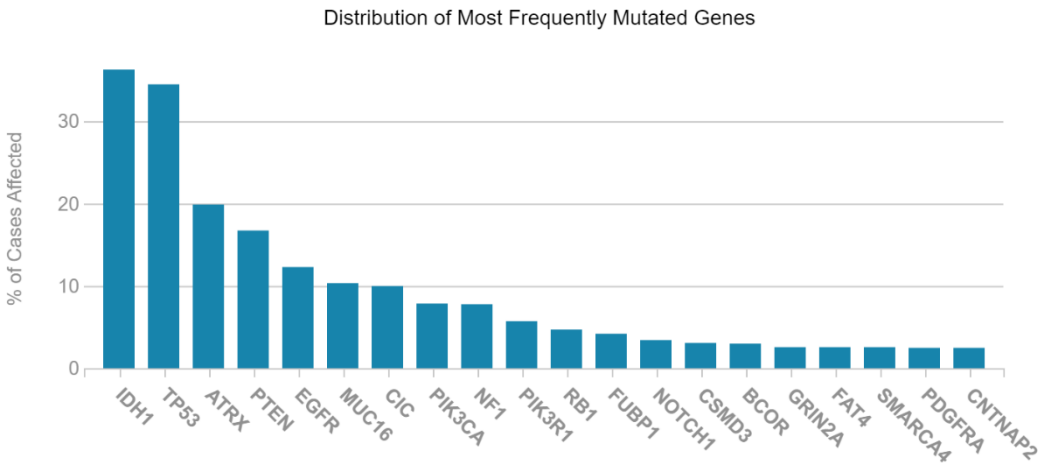


Figure D2. Most common mutations found in Gliomas based upon TCGA cohort data from 1172 patients analysis (graph prepared using an Exploration Tool from National Cancer Institute GDC Data Portal, accessed on 17.4.2023: https://portal.gdc.cancer.gov/).

Gliomas, particularly glioblastomas, have been found to have numerous *EGFR* gene changes, including amplifications, deletions, and single nucleotide polymorphisms (SNPs)[130]. In our study 16% samples contained a somatic variant within the *EGFR* region. *NF1,* that is sometimes called a human glioblastoma suppressor gene, has been reported in gliomas as carrying nonsense mutations, splice mutations, missense changes, and frame shift insertions deletions in about 14% patients[131]. In our cohort variants in this gene were detected in 9% of samples.

Adult patients with WHO grade 2/3 or grade 4 gliomas carry mutations in *ATRX*[132], which can affect the outcome. We reported that 10% of tumor samples had the *ATRX* mutation, which fits with previously published data. The retinoblastoma gene (*RB1)* is a recessive human cancer gene that has a "regulatory" or a "suppressor" function[133], that was found malfunctioning in many serious cancers[133]. *RB1* mutations (nonsense mutations, frame shift insertion, frame shift deletions, splice site mutation) were detectable in 9% of our patients cohort (6% in TCGA Glioblastoma cohort).

After overall review of the mutational landscape, there were no significant differences in a frequency of somatic variants found in analyzed the Polish glioma cohort and global research and database profiles.

Glioma predisposing mutations that occur in *BRCA1*, *NF1*, *P53* were described[134] and our germline variants analysis has shown an additional group of genes affected.

Patients with *PTEN* mutations have higher lifetime risks for a number of malignancies, including melanoma, kidney cancer, and colorectal cancer[135]. In our study the extremely rare (MAF<0.00001) *PTEN* variant was present in 97% of patients, which could suggest that this can be a predisposing factor to Glioma. This particular *PTEN* variant was registered in the ClinVar database as benign and while benign variants are not typically associated with the development of cancer, certain benign variants can be linked to an increased risk of developing cancer. Another very common within our cohort, yet extremely rare in healthy population was a missense mutation within *MSH3* gene (MAF<0.0001), that was present in 33% of analyzed samples.

A novel and interesting point mutation found in the *AKAP9* gene (T1334fs) was not registered in any of current databases and was discovered in over 40% of patients from the cohort. Changes in A-kinase anchor protein 9 (*AKAP9*) expression are linked to or a direct cause of a variety of malignancies, persistent heart failure, and immune system disorders like HIV[136]. Cancer types in which *AKAP* mutations have been identified include pancreatic, ovarian, colorectal, breast cancer, and others[137]. For instance, *AKAP9* mutations were found in endometrial cancers[138]. In this case an endometrioid carcinoma harbored *AKAP9* mutations, which were linked to changes in the PI3K-Akt signaling pathway[138]. Thyroid gland papillary carcinoma, low-grade glioma, NOS,

pancreatic adenocarcinoma, pilocytic astrocytoma, and astrocytoma have the highest prevalence of the *AKAP9-BRAF* Fusion, which is found in 0.07% of AACR GENIE cases[139]. The specific alteration in the *AKAP9* gene (T1334fs) found in this study in 36 patients, is a frame shift insertion and should significantly impact protein structure and therefore it's function. It is a novel mutation, so as it wasn't registered in the international databases, it might suggest it is a specific variant that mainly occurs in the Polish population. This particular variant should be studied further, as *AKAP* proteins have a significant role in cancer development[137].

Over the years, significant progress have been made in finding inherited genetic alterations that can lead to development of primary adult glioma. In general the risk of adult glioma has been conclusively linked to ten distinct hereditary variations in eight chromosomal locations and majority of these variants increase a relative risk of primary adult glioma by 20-40%[140]. There are however more impactful markers like *TP53* variation rs78378222 that gives a two-fold relative risk increase (200%) and the rs557505857 on chromosome 8 that confers a six-fold relative risk increase in *IDH*-mutated astrocytomas and oligodendroglial tumors (600%)[140].

5.2. Assessment of NGS sequencing of cfDNA isolated from GBM patient's blood as a diagnostic tool yet.

With the reduced costs of sequencing, advancements in technology and computational methods, sequencing of exomes, whole genomes, and targeted sequencing are becoming more frequently used as diagnostics tools to assist diagnosis and therapy of glial tumors. Liquid biopsy has emerged as a novel tool that can facilitate tumor diagnosis and offers easy methods for a follow up that includes progress monitoring and further therapy recommendations. When this project has started there were only a couple reports showing the detection of a small set of tumor mutations in cfDNA from GBM patient blood[141]. More success in ctDNA detection had been achieved using CSF as a source of liquid biopsy material[90]. A consensus was that the existing methods do not allow to collect enough cfDNA from blood to make a reliable mutational analysis for primary brain tumor patients[57,141]. We attempted to solve

some technical issues, collect more material and perform further detailed analysis of cfDNA.

In this study we have demonstrated that several improvements in sample isolation, quality and quantity determination and library preparation allowed detection of pathogenic mutations in cfDNA from GBM patients' blood. We have shown that time elapsed after blood collection leading up to plasma separation and cfDNA isolation impacts the final isolation yield. We have implemented size selection protocols to enrich sequenced sample in shorter DNA fragments which should improve sensitivity of ctDNA detection[105,142]. We showed that some of the potentially pathogenic mutations that are detectable in liquid biopsy are not registered in a classical tissue analysis. Overall, we demonstrated that implemented technical improvements in quality control and library preparation allowed for the detection of ctDNA in 8 out of 84 patients, including 5 out of 80 glioma patients. In 32 out of 84 patients, we found potentially pathogenic genetic alterations in cfDNA that were not detectable in tDNA, which can refer to tumor heterogeneity and incomplete tissue sampling. The latter observation is intriguing and deserves more attention to understand a potential source of distinct representation of the mutated genes in tumor and cfDNA. As a main goal of sequencing tumor DNA is for diagnostic purposes, the data presented here show the lack of good representation of tumor mutational landscape in ctDNA. This means that at the current stage sequencing of cfDNA is not sufficient for clinical diagnostics. However, there are some remedies that could improve liquid biopsy based on cfDNA also in GBM patients.

Brain fluid dynamics mainly involves an infiltrate of cerebrospinal fluid and a blood brain barrier (BBB) separate brain parenchyma from blood. It has been reported that BBB in GBM patients is affected by exuberant angiogenesis[87]. As this neovascularization process takes part in the tumor growth and progression, therefore new therapeutic approaches were developed to target vascular endothelial growth factor (*VEGF*) in malignant gliomas[87]. There was a popular view that BBB is uniformly disrupted in all GBMs, however there is an overwhelming clinical evidence that a significant tumor burden could associate with the intact BBB and drugs with poor BBB permeability are not effective in many GBMs[88]. It is likely that opening BBB could improve the ctDNA detection and additionally improve therapeutic effects

of chemotherapy, as was shown in the study in which MRI-guided intra-arterial bevacizumab delivery which led to tumor shrinkage[93]. There is currently ongoing clinical study in USA with a main goal to monitor patients with GBM using blood derived liquid biopsy, while collecting the blood before and after opening the BBB with microbubble resonator (NCT05383872).

Even that cerebrospinal fluid remains a better source of ctDNA than blood in the case of brain tumor patient[90,143], current ongoing clinical studies are focused on using blood rather than CSF[92]. It was established that liquid biopsy from CSF gives less background noise signal ratio, compared to serum[144]. More sensitive assays that can lead to higher detection rates of somatic variants in liquid biopsy for Glioma patients are BEAMing (beads, emulsion, amplification, magnetics) PCR and ddPCR. These techniques can lead to detection of extremely rare genomic signals, but they work only in detection of a specific point mutation. In this case reference material is not needed. Patients with gliomas often have novel mutations within a tumor that can occur in more than 664 different gene regions we tested and can be patient specific. However actionable mutations as mentioned before in section 1.7 can be well described in publicly available databases as specific point mutations, insertions or deletions. Still even if we use the most sensitive methods and most targeted approach the issue of cfDNA isolation and acquiring a sufficient amount of material occurs.

In 2014 the study of blood derived liquid biopsy from glioma patients that employed BEAming PCR based technology detected ctDNA in less than 10% patients[145]. Further technological developments presented more successful results, that detected at least one tumor mutation signal in plasma using targeted sequencing in: 2018 - 51% of patients[146], 2020 - 55% of patients[147]. Truly, as this was a significant increase in sensitivity, having detection of one or more of tumor describing variants in case of only about 50% of patients presents an incomplete diagnostic information, and would not suffice for personalized therapy recommendations or disease monitoring. In 2021 ddPCR assays targeting specific *TETR* promoter mutations would detected ctDNA in plasma with 62.5% sensitivity[148]. Deep sequencing of cfDNA isolated from CSF presented much higher detection rates ranging up to detection of at least one tumor specific variant in 82.5% of cases[149]. In case of routine

diagnostics still there is a room for improvement if reliable clinically relevant method is to be applied[150] .

Study published in Clinical Cancer Research in February 2023 underlined another aspect of CSF versus plasma as a source of liquid biopsy for glioma patients. In this study the amount of isolated cfDNA from two bodily fluids differed significantly, as isolation yield was sufficient to perform ddPCR in case of 40/45 samples isolated from CSF versus 7/42 samples isolated from blood[91], furthermore their ddPCR failed to detect ctDNA in all 7/7 cfDNA samples isolated from blood[91]. The peritumoral CSF was used as a source of ctDNA for NGS with satisfactory results for 3 patients, somatic variants detectable in the liquid biopsy were found with varying AF, sometimes even higher than in the tissue biopsy and had one somatic variant that was not detected in tissue biopsy at all (SNP in gene *PIK3R1*). The data presented in our study shows similar tendencies. Potentially pathogenic variants that were detected in our study, confirmed in the public databases suggest that we were able to detect a signal that was complementary to that of a tissue biopsy, even if it did not represent the full spectrum. Currently published study[91] of peritumoral CSF cfDNA NGS analysis has shown that there is an overlap of pathogenic mutations between the tumor and liquid biopsy, but there were variants in each analyzed case that were only detected in the tumor or only in cfDNA. This suggests that the traditional biopsy and liquid biopsy can complement each other, which was a conclusion of many liquid biopsy studies done on different types of cancers[151,152]. This is consistent with the results of our study[98] where potentially pathogenic variants were detected only in liquid biopsy and some somatic variants were only detectable in tissue biopsy.

As mentioned before, previous comprehensive cancer studies have shown that liquid biopsy in case of glioma patients is much less effective, than in case of other cancers[145,146]. One of the comprehensive studies that tested Guardia Health cfDNA sequencing platform and included over 25000 samples from varying tumor types, concluded that the most detectable tumor signal was in case of Small Cell Lung Cancer (SCLC) patients liquid biopsy with ctDNA signal detectable in 90% of cases and the least detectable signal in the case of Glioblastoma with an average below 50%[146]. The percentage of ctDNA within cfDNA was also estimated with the largest fraction of tumor DNA in the case

of Colorectal cancer (3%), SCLC (2%), and the lowest in case of Gliomas (0.3%)[146].

On the other hand liquid biopsy from blood might represent better overall spectrum of somatic variants present in multiple localizations in the case of metastatic patients[73,153–155]. In our study each liquid biopsy from metastatic patient samples contained the strong tumor signal that was recognized in brain tumor biopsy, which suggests that metastatic cancers are a great candidate for liquid biopsy.

First liquid biopsy to be approved by FDA in early 2000s was the CellSearch® - Circulating Tumor Cell (CTC) isolation technology test that separates and detects cancer cells from peripheral blood sample. It was primarily authorized to monitor adult cancer patients with advanced colorectal, prostate, and breast cancers and it's applications since then extended to monitoring other types of advanced cancers[156,157].

The Cobas® *EGFR* Mutation Test v2 was approved by FDA for Non-Small Cell Lung Cancer monitoring and it includes both: a kit for analysis of blood sample (plasma) or tissue biopsy. This test estimates cfDNA or tDNA samples using PCR screening of the *EGFR* gene in exons 18, 19, 20, and 21 to screen for presence of 42 mutations. It is used to asses relevance of targeted cancer treatment, that can be effective with these alterations[158].

In 2021 two additional tests gained FDA approval for tumor cell DNA detection from blood: The FoundationOne Liquid CDx and Guardant 360® CDx. The FoundationOne Liquid CDx (https://www.foundationmedicine.com/test/foundationone-liquid-cdx) is a test that analyzes blood derived cfDNA and delivers a mutation report, tumor fraction values, microsatellite instability high (MSI-H), and blood tumor mutational burden (bTMB). It includes panel of 324 targetable mutations, rearrangements, microsatellite instability and other markers to guide to FDA approved therapy recommendations for NSCLC, breast, ovarian, and prostate cancers. Guardant360® CDx (https://www.guardantcomplete.com/guardant-portfolio/cdx) is a companion diagnostic FDA approved test that uses targeted NGS of blood derived cfDNA for personalized treatment recommendations for lung, breast and other advanced solid tumors[159].

Currently there are no FDA-approved liquid biopsy diagnostic tests for gliomas, but there are ongoing clinical trials evaluating potential of this method for monitoring and diagnostics. The interventional phase I NCT04201873 trial that will conclude in 2024 evaluates the gene expression signature of RNA isolated from peripheral blood before and after treatment using nanoString RNAseq. Matching tumor and blood targetable mutations are evaluated in whole genome classification in the NCT04274283 observational study that will concluded in 2025. Three specific oncogenic pathways (*IDH, TERT, ATRX*) are investigated as a potential diagnostic genes and liquid biopsy using CicTeloDIAG is explored for monitoring WHO grade 2-4 glioma, pancreatic, lung and colon cancers (NCT04931732). SensiScreen glioma study (NCT04539431) that includes 220 glioma participants focuses on validation of a cheaper and more sensitive PCR platform for liquid biopsy (from both blood and CSF) comparison to tumor DNA. Another study (NCT05383872) with a purely diagnostic purpose registers increase of cfDNA in liquid biopsy after BBBD (blood – brain barrier disruption) with Microbubble resonator.

The clinical study that includes 20 GBM patients, but it's mainly focused on prostate cancer patients (NCT05281731) involves new technological development called sonobiopsy -that combines high-intensity focused ultrasound with ultrasound imaging, and it is used to generate thermal ablation in cancer tissue. This ongoing interventional clinical study uses deep sequencing to analyze cfDNA from blood after the sonobiopsy procedure. Exploratory study number NCT05133154 includes 30 low-grade gliomas, 10 high grade gliomas, and 10 control patients is focused on liquid biopsy search of CTCs (circulating tumor cells), TEP (tumor educated platelets), and specific markers (*IDH, 1p19q, ATRX*) for diagnosis and monitoring. Finally, deep learning MRI (Magnetic Resonance Imaging) liquid biopsy method is evaluated for grading and molecular subtype prediction in the observational study (NCT05536024) on 500 glioma patients

At the beginning of liquid biopsy, a concept was to establish a simple blood test for early diagnosis of non-symptomatic cancers. It was supposed to be the tool of early detection protecting patients from late diagnosis and cancer-related death. Today's research has shown that liquid biopsy can be used as complementary method to modern diagnostics during advanced

stages of cancer monitoring. Currently liquid biopsy is used to predict metastatic lesions (CellSearch®), or as a complementary or only source for personalized therapy recommendations (The FoundationOne Liquid CDx and Guardant 360® CDx). Study published in JAMA Oncology in 2019 had shown that using both tissue and liquid biopsy can significantly improve ability to detect clinically relevant mutations in the case of stage 4 NSCLC[160], which indicates that these methods are complementary and each test can reveal different actionable mutations present in cancer.

Even though there are clinically approved liquid biopsy technologies for some advanced cancers that are effective in tracking ctDNA, each method can yield false negative results and if possible tissue biopsy is recommended, which is a main reason why FDA approved liquid biopsies are defined as assisting diagnostic methods. There is a perspective that currently ongoing clinical studies and scientific research might yield clinically approved methods for a liquid biopsy for glioma patients.

5.3. Preservation of the GBM mutational spectra in cells cultured in normoxia and hypoxia.

Low-passage, serum-free, patient-derived GBM cell cultures have proven to be useful in predicting the responses to drugs[161]. Cell lines derived from primary tumors can be used for drug screening to assist in the recommendation of the best course of personalized therapy. Recent molecular profiles of hundreds of cell lines from The Cancer Cell Line Encyclopedia and thousands of tumor samples from the Cancer Genome Atlas allowed comparison of cell lines and tumors and showed frequent changes consistent with clonal evolution under culture conditions. The large-scale studies show accumulation of mutational signatures associated with DNA replication and repair with passaging of cells[156].

Our study shows that the overall mutation spectrum in tumors was generally mimicked in the cell cultures. However, we found several mutations in cells that were not detected in tumors. This might suggest that sampling and sequencing of tumors might show a narrow view and for example did not detect mutations in rare tumor stem cells that can thrive in cell cultures and recapitulate a tumor with a different genetic set up. Variants specific to stem

cell cultured clones could thrive and dominate in cell cultures but might be overlooked in the analysis of bulk tumor. This suggests that stem cell's molecular profile (high proliferating potential and resilience) can be overlooked in the traditional biopsy, because there is a small percentage of them within the bulk tumor. This can implicate the importance of the incorporation of tumor derived cells and tumor genomic analysis in grasping a complete spectrum of glioma complexity.

On the other hand, cancer stem cells that are in high proliferation states can exhibit tendency to acquire new morphological, chromosomal, and mutational changes during the culturing process[162]. Accumulation of mutations has been observed in non-cancer derived stem cells *in vitro* cell cultures[163]. Cancer stem cells due to their intrinsic instability and frequent inability to repair properly DNA damage can accumulate and acquire mutations during mitosis, that is highly induced by synthetic growth factors in cell culture environment. This is why to verify that a cell line is a valid model for a cancer research project, a precise genetic characterization is necessary sometimes even repeated during different time points of the project.

Cell lines derived from many types of tumors have been studied and analyzed to model molecular mechanisms of disease. Many aberrations that are carried by tumor derived cells, can assist in pharmacogenomics studies[164]. It is clear that cells grown in culture are devoid of interactions with stromal cells. The tumor microenvironment is critical for tumor development and progression, therefore drug testing should be carried out in tumor explants or *in vivo* models. Tumor explants exhibit similar transcriptional heterogeneity as bulk tumors, whereas glioma sphere cell cultures represent more uniform profiles of a more uniform transcriptional state[97]. Some studies claim glioma sphere cultures carry much higher number of distinct mutations than the matched tumors[97]. Our results were in line with these currently published[97] findings, which suggest that sphere cell cultures can provide a good model for personalized therapy recommendations (6/8 cell lines in our study were derived in the form of sphere cultures). Whereas tumor explants seem to be a potentially viable model for mimicking cancer microenvironment.

Drug response assays that use tumor-derived cell cultures are also known as "*ex vivo* drug sensitivity assays" or "tumor organoid assays". The main idea

is to establish cell culture from a patient's tumor and to expose cell culture to various anti-cancer drugs to determine which is the most effective against tumor cells. There are FDA-approved drug response assays: OncoType DX Breast Cancer Assay, for example, which is used to predict likelihood of recurrence and chemotherapy benefits.

There are no FDA-approved cell culture assays for GBM yet, there are however studies that establish patient-derived GBM organoids with a goal of setting up drug screening assays and personalized treatment recommendations - the Ivy Glioblastoma Atlas Project (Ivy GAP) for example. The Ivy GAP is a research venture that started in 2015 as a collaborative effort with a goal of creation of comprehensive atlas of molecular features of GBM and has a goal of development of new models for drug discovery and testing based upon organoid cell culture. Another research project that works on development of tumor organoid assays is Avatar project, it's main focus is to test various drugs and their combinations to determine the most effective treatment for each patient.

Although tumor organoid assays have a great potential, additional research is required to ascertain their therapeutic utility and efficacy, because they are not currently frequently employed in clinical practice. These assays do, however, hold significant promise for enhancing our knowledge of GBM and other malignancies and creating more efficient and individualized treatments.

5.5. Mutational spectrum of pediatric brain tumors reveals several targetable candidate genes

Using FFPE blocks for DNA targeted sequencing can provide access to large specimen collections and allows re-analyses of old samples from the hospital archives. Unfortunately, DNA from this type of archival material can vary in quality, is frequently fragmented and is definitely lower quality than fresh tissue DNA. Quality between samples can vary and even successful DNA isolation and library preparation can still yield NGS data with too many PCR cycle duplicates to pass quality control and provide results. Despite those limitations, we successfully isolated DNA from the cohort of poorly diagnosed pediatric brain tumors and performed targeted sequencing. As in this case a reference blood genomic DNA was not available, we took

the advantage of the existing databases to identify variants that do not occur in general population, therefore are likely somatic variants. We found several candidate variants in genes coding for important intracellular proteins.

The most commonly mutated genes in the analyzed cohort were *KMT2D* and *MTUS2*, that were altered in 10% of patients. *KMT2D* encodes an epigenetic enzyme that methylates the Lys-4 position of histone H3. All mutations in the *KMT2D* gene are within regions coding histone lysine N-methyltransferase activity but in each patient they were in different positions. *KMT2D* was implicated in pancreatic carcinogenesis through metabolic reprogramming[165]. There are several inhibitors that have been reported to target KMT2D, including MI-2[166], which is a small molecule inhibitor that selectively binds to the catalytic SET domain of KMT2D and inhibits its methyltransferase activity. Another inhibitor of KMT2D is UNC1999[167], which inhibits the catalytic activity of KMT2D and its paralog KMT2C (also known as MLL3). In addition to these small molecule inhibitors, there are also RNA interference (RNAi) approaches that can be used to selectively knock down KMT2D expression in cells[168]. These approaches use small interfering RNAs (siRNAs) or short hairpin RNAs (shRNAs) to target and degrade *KMT2D* mRNA. It should be noted that the development and use of KMT2D inhibitors is still an active area of research, and their efficacy and safety in treating diseases is still being investigated.

The recurrent mutation in the *MTUS2* gene occurs in the exactly same position in 5 patients. The detected variant (RS200233522) is not registered in ClinVar, which shows MTUS2 is predicted to be involved in microtubule binding and a protein homodimerization activity.

The recurrent mutation was found in the *FANCA* gene (coding for Fanconi anemia complementation group). The DNA repair protein FANCA may function in cell cycle checkpoint control or post replication repair. It is a part of the complex of proteins called Fanconi anemia (FA) pathway, which is responsible for repairing DNA damage that occurs during DNA replication, therefore it's role is to maintain genome stability. The FA pathway is activated in response to a type of DNA damage called interstrand crosslinks (ICLs), which occur when the two strands of DNA become covalently linked[169]. ICLs are highly toxic to cells and can block DNA replication and transcription, leading to cell death or genetic

mutations if left unrepaired. FANCA is one of the key components of the FA pathway, and it plays a critical role in the recognition and repair of ICLs. It works by binding to the damaged DNA and recruiting other proteins to form a complex that repairs the damage[169]. In addition to its role in DNA repair, FANCA has also been implicated in the cell cycle checkpoint control, specifically, it is thought to be involved in the G2/M checkpoint, which ensures that DNA damage is repaired before the cell enters mitosis[170]. In the absence of FANCA, cells may be more prone to entering mitosis with unrepaired DNA damage, which can lead to genetic instability and cell death. Overall, FANCA plays a critical role in maintaining genome stability and preventing the development of cancer and other genetic diseases [170].

The potentially pathogenic variant was detected in the *RET* gene in 8% of patients. RET (rearranged during transfection) is a receptor tyrosine kinase for the extracellular signaling molecules, the members of glial cell line derived neurotrophic factor (GDNF) family[119]. *RET* gain-of-function mutations are associated with many types of cancers and specifically the T1038A variant was registered and considered as pathogenic in multiple types of pediatric cancers[120,121]. The RET signaling pathway has been implicated in the development and progression of gliomas[171–173].

RET is activated by the binding of its ligands (GDNF family of ligands - GFLs, including GDNF, neurturin, artemin, and persephin.

Studies have shown that RET is overexpressed in a subset of gliomas, particularly in low-grade gliomas and glioblastomas[171–174]. Activation of the RET pathway in glioma cells has been shown to promote cell proliferation, survival, migration, and invasion[173]. Inhibition of RET signaling using small molecule inhibitors or RNAi approaches has been shown to suppress cell growth and induce apoptosis[175]. One potential therapeutic strategy for targeting RET in gliomas is the use of selective RET inhibitors (Alectinib for example). Several small molecule RET inhibitors are currently in clinical trials for the treatment of various types of cancers[176], including gliomas. Another potential approach is the use of RNAi-based therapies to target RET expression in glioma cells[177]. These approaches have shown promise in preclinical studies and hold potential for future clinical development.

Currently ongoing clinical trial (NCT05009992) at National Institute of Neurological Disorders and Stroke (NINDS) includes participants of ages from 2 - 39 years old and represents an adaptive platform for children and adults with Diffuse Gliomas and WHO Grade 3 Gliomas that includes combinational therapy with Paxalisib the Inhibitor of the PI3K/AKT/mTOR pathway. Trial is in a phase 2 and it is accepting participants (**https://clinicaltrials.gov/ct2/show/NCT05009992**). Paxalisib is already FDA approved for treatment of an aggressive and very rare childhood brain cancer (teratoid or atypical rhabdoid tumors).

6. Summary and conclusions

In this PhD thesis I presented a set of experiments exploring a potential of targeted NGS sequencing in uncovering the mutational landscape of malignant gliomas in adults, in circulating tumor DNA, patient derived cells cultures and FFPE material from poorly characterized pediatric tumors.

These specific aims of the study were achieved:

1) Collection and characterization of the brain tumor patient cohort and matched samples of resected tumors and blood from 100 patients.
2) Protocols of cfDNA isolation, quantification, library preparation and sequencing were established and improved resulting in high quality of sequencing results.
3) Tumor confirmed variants were detected in 8 out of 84 cfDNA samples.
4) Potentially pathogenic genetic alterations were detected in 37 out of 84 cfDNA samples.
5) A spectrum of potentially pathogenic genetic alterations that were detected in cfDNA isolated from blood of GBM patients was not detectable in tumor tissue, which shows non-overlapping spectrum of liquid (cfDNA) and tumor biopsy.
6) Analysis of somatic mutations in tumors and corresponding GBM-patient derived cell cultures maintained at normoxia and hypoxia showed retaining of mutational spectrum.
7) A spectrum of somatic mutations in poorly characterized pediatric brain tumors was characterized and some new mutations were discovered.

Altogether, the presented data shows a great potential of targeted sequencing in improving diagnosis and selecting the best therapy for glioma patients, for which current treatment is not effective. Despite some disadvantages liquid biopsy holds a promise of complementing current tumor biopsy-based diagnostic. Patient derived cell cultures even if maintained under normoxia retained the spectrum of mutations and allow prescreening of potential therapeutics.

References:

1. Wilmut I, Schnieke AE, McWhir J, Kind AJ, Campbell KHS. Viable offspring derived from fetal and adult mammalian cells. *Nature*. 1997;385(6619):810-813. doi:10.1038/385810a0
2. Chanock S. Candidate genes and single nucleotide polymorphisms (SNPs) in the study of human disease. *Dis Markers*. 2001;17:89-98.
3. Suh Y, Vijg J. SNP discovery in associating genetic variation with human disease phenotypes. *Mutation Research/Fundamental and Molecular Mechanisms of Mutagenesis*. 2005;573(1):41-53. doi:https://doi.org/10.1016/j.mrfmmm.2005.01.005
4. Bhagwat M. Searching NCBI's dbSNP database. *Curr Protoc Bioinformatics*. 2010;Chapter 1:Unit-1.19. doi:10.1002/0471250953.bi0119s32
5. Gorlov IP, Gorlova OY, Frazier ML, Spitz MR, Amos CI. Evolutionary evidence of the effect of rare variants on disease etiology. *Clin Genet*. 2011;79(3):199-206. doi:https://doi.org/10.1111/j.1399-0004.2010.01535.x
6. Chinwalla AT, Cook LL, Delehaunty KD, et al. Initial sequencing and comparative analysis of the mouse genome. *Nature*. 2002;420(6915):520-562. doi:10.1038/nature01262
7. Twigger S, Lu J, Shimoyama M, et al. Rat Genome Database (RGD): mapping disease onto the genome. *Nucleic Acids Res*. 2002;30(1):125-128. doi:10.1093/nar/30.1.125
8. Lander ES, Linton LM, Birren B, et al. Initial sequencing and analysis of the human genome. *Nature*. 2001;409(6822):860-921. doi:10.1038/35057062
9. Collins FS, Green ED, Guttmacher AE, Guyer MS, Institute on behalf of the USNHGR. A vision for the future of genomics research. *Nature*. 2003;422(6934):835-847. doi:10.1038/nature01626
10. Mardis ER. ChIP-seq: welcome to the new frontier. *Nat Methods*. 2007;4(8):613-614. doi:10.1038/nmeth0807-613
11. Cloonan N, Forrest ARR, Kolle G, et al. Stem cell transcriptome profiling via massive-scale mRNA sequencing. *Nat Methods*. 2008;5(7):613-619. doi:10.1038/nmeth.1223
12. Packer A. The clickable genome. *Nat Rev Genet*. 2007;8(1):S17-S17. doi:10.1038/nrg2243
13. Nagarajan N, Pop M. Sequence assembly demystified. *Nat Rev Genet*. 2013;14(3):157-167. doi:10.1038/nrg3367
14. Altshuler D, Donnelly P, Consortium TIH. A haplotype map of the human genome. *Nature*. 2005;437(7063):1299-1320. doi:10.1038/nature04226
15. Consortium 1000 Genomes Project, Abecasis GR, Altshuler D, et al. A map of human genome variation from population-scale sequencing. *Nature*. 2010;467(7319):1061-1073. doi:10.1038/nature09534
16. Consortium 1000 Genomes Project, Abecasis GR, Auton A, et al. An integrated map of genetic variation from 1,092 human genomes. *Nature*. 2012;491(7422):56-65. doi:10.1038/nature11632
17. Sudmant PH, Rausch T, Gardner EJ, et al. An integrated map of structural variation in 2,504 human genomes. *Nature*. 2015;526(7571):75-81. doi:10.1038/nature15394

18. Consortium 1000 Genomes Project, Auton A, Brooks LD, et al. A global reference for human genetic variation. *Nature*. 2015;526(7571):68-74. doi:10.1038/nature15393
19. Sanger F, Air GM, Barrell BG, et al. Nucleotide sequence of bacteriophage φX174 DNA. *Nature*. 1977;265(5596):687-695. doi:10.1038/265687a0
20. Sanger F, Nicklen S, Coulson AR. DNA sequencing with chain-terminating inhibitors. *Proceedings of the National Academy of Sciences*. 1977;74(12):5463-5467. doi:10.1073/pnas.74.12.5463
21. Nyren P, Pettersson B, Uhlen M. Solid Phase DNA Minisequencing by an Enzymatic Luminometric Inorganic Pyrophosphate Detection Assay. *Anal Biochem*. 1993;208(1):171-175. doi:https://doi.org/10.1006/abio.1993.1024
22. Nyrén P, Lundin A. Enzymatic method for continuous monitoring of inorganic pyrophosphate synthesis. *Anal Biochem*. 1985;151(2):504-509. doi:https://doi.org/10.1016/0003-2697(85)90211-8
23. Chaitankar V, Karakülah G, Ratnapriya R, Giuste FO, Brooks MJ, Swaroop A. Next generation sequencing technology and genomewide data analysis: Perspectives for retinal research. *Prog Retin Eye Res*. 2016;55:1-31. doi:https://doi.org/10.1016/j.preteyeres.2016.06.001
24. Lahens NF, Ricciotti E, Smirnova O, et al. A comparison of Illumina and Ion Torrent sequencing platforms in the context of differential gene expression. *BMC Genomics*. 2017;18(1):602. doi:10.1186/s12864-017-4011-0
25. Quail MA, Smith M, Coupland P, et al. A tale of three next generation sequencing platforms: comparison of Ion Torrent, Pacific Biosciences and Illumina MiSeq sequencers. *BMC Genomics*. 2012;13(1):341. doi:10.1186/1471-2164-13-341
26. Cock PJA, Fields CJ, Goto N, Heuer ML, Rice PM. The Sanger FASTQ file format for sequences with quality scores, and the Solexa/Illumina FASTQ variants. *Nucleic Acids Res*. 2010;38(6):1767-1771. doi:10.1093/nar/gkp1137
27. Del Fabbro C, Scalabrin S, Morgante M, Giorgi FM. An Extensive Evaluation of Read Trimming Effects on Illumina NGS Data Analysis. *PLoS One*. 2013;8(12):e85024-. https://doi.org/10.1371/journal.pone.0085024
28. Kirsch C, Weickmann S, Schmidt B, Fleischhacker M. An Improved Method for the Isolation of Free-Circulating Plasma DNA and Cell-Free DNA from Other Body Fluids. *Ann N Y Acad Sci*. 2008;1137(1):135-139. doi:https://doi.org/10.1196/annals.1448.035
29. Leest P van der, Boonstra PA, Elst A ter, et al. Comparison of Circulating Cell-Free DNA Extraction Methods for Downstream Analysis in Cancer Patients. *Cancers (Basel)*. 2020;12(5):1222. doi:10.3390/cancers12051222
30. Forbes S, Clements J, Dawson E, et al. COSMIC 2005. *Br J Cancer*. 2006;94(2):318-322. doi:10.1038/sj.bjc.6602928
31. Forbes SA, Bhamra G, Bamford S, et al. The Catalogue of Somatic Mutations in Cancer (COSMIC). *Curr Protoc Hum Genet*. 2008;57(1):10.11.1-10.11.26. doi:https://doi.org/10.1002/0471142905.hg1011s57
32. Canoll P, Goldman JE. The interface between glial progenitors and gliomas. *Acta Neuropathol*. 2008;116(5):465-477. doi:10.1007/s00401-008-0432-9
33. Sanai N, Alvarez-Buylla A, Berger MS. Neural Stem Cells and the Origin of Gliomas. *New England Journal of Medicine*. 2005;353(8):811-822. doi:10.1056/NEJMra043666
34. Liu Z, Li H, He L, et al. Discovery of Small-Molecule Inhibitors of the HSP90-Calcineurin-NFAT Pathway against Glioblastoma. *Cell Chem Biol*. 2019;26(3):352-365.e7. doi:10.1016/j.chembiol.2018.11.009

35. Preusser M, Lim M, Hafler DA, Reardon DA, Sampson JH. Prospects of immune checkpoint modulators in the treatment of glioblastoma. *Nat Rev Neurol*. 2015;11(9):504-514. doi:10.1038/nrneurol.2015.139
36. Zhou Y, Wu W, Bi H, Yang D, Zhang C. Glioblastoma precision therapy: From the bench to the clinic. *Cancer Lett*. 2020;475:79-91. doi:https://doi.org/10.1016/j.canlet.2020.01.027
37. Grauwet K, Chiocca EA. Glioma and microglia, a double entendre. *Nat Immunol*. 2016;17:1240-1242.
38. Shah AH, Heiss JD. Neurosurgical Clinical Trials for Glioblastoma: Current and Future Directions. *Brain Sci*. 2022;12(6). doi:10.3390/brainsci12060787
39. citations-20221205T134554.
40. Reardon DA, Brandes AA, Omuro A, et al. Effect of Nivolumab vs Bevacizumab in Patients With Recurrent Glioblastoma: The CheckMate 143 Phase 3 Randomized Clinical Trial. *JAMA Oncol*. 2020;6(7):1003-1010. doi:10.1001/jamaoncol.2020.1024
41. Friedman GK, Johnston JM, Bag AK, et al. Oncolytic HSV-1 G207 Immunovirotherapy for Pediatric High-Grade Gliomas. *New England Journal of Medicine*. 2021;384(17):1613-1622. doi:10.1056/NEJMoa2024947
42. Bernstock JD, Hoffman SE, Chen JA, et al. The Current Landscape of Oncolytic Herpes Simplex Viruses as Novel Therapies for Brain Malignancies. *Viruses*. 2021;13(6). doi:10.3390/v13061158
43. Suzuki H, Aoki K, Chiba K, et al. Mutational landscape and clonal architecture in grade II and III gliomas. *Nat Genet*. 2015;47(5):458-468. doi:10.1038/ng.3273
44. Osborn AG, Louis DN, Poussaint TY, Linscott LL, Salzman KL. The 2021 World Health Organization Classification of Tumors of the Central Nervous System: What Neuroradiologists Need to Know. *American Journal of Neuroradiology*. Published online June 16, 2022. doi:10.3174/ajnr.A7462
45. Hegi ME, Stupp R. In search of molecular markers of glioma in elderly patients. *Nat Rev Neurol*. 2013;9(8):424-425. doi:10.1038/nrneurol.2013.127
46. Taal W, Bromberg JEC, van den Bent MJ. Chemotherapy in glioma. *CNS Oncol*. 2015;4(3):179-192. doi:10.2217/cns.15.2
47. Verhaak RGW, Hoadley KA, Purdom E, et al. Integrated Genomic Analysis Identifies Clinically Relevant Subtypes of Glioblastoma Characterized by Abnormalities in PDGFRA, IDH1, EGFR, and NF1. *Cancer Cell*. 2010;17(1):98-110. doi:https://doi.org/10.1016/j.ccr.2009.12.020
48. Zhang Z, Chen J, Huo X, et al. Identification of a mesenchymal-related signature associated with clinical prognosis in glioma. *Aging*. 2021;13(9):12431-12455. doi:10.18632/aging.202886
49. Becker AP, Sells BE, Haque SJ, Chakravarti A. Tumor Heterogeneity in Glioblastomas: From Light Microscopy to Molecular Pathology. *Cancers (Basel)*. 2021;13(4). doi:10.3390/cancers13040761
50. Sottoriva A, Spiteri I, Piccirillo SGM, et al. Intratumor heterogeneity in human glioblastoma reflects cancer evolutionary dynamics. *Proceedings of the National Academy of Sciences*. 2013;110(10):4009-4014. doi:10.1073/pnas.1219747110
51. Parker NR, Khong P, Parkinson JF, Howell VM, Wheeler HR. Molecular Heterogeneity in Glioblastoma: Potential Clinical Implications. *Front Oncol*. 2015;5. doi:10.3389/fonc.2015.00055
52. Wang Q, Hu B, Hu X, et al. Tumor Evolution of Glioma-Intrinsic Gene Expression Subtypes Associates with Immunological Changes in the Microenvironment. *Cancer Cell*. 2017;32(1):42-56.e6. doi:https://doi.org/10.1016/j.ccell.2017.06.003

53. Miller AM, Shah RH, Pentsova EI, et al. Tracking tumour evolution in glioma through liquid biopsies of cerebrospinal fluid. *Nature*. 2019;565(7741):654-658. doi:10.1038/s41586-019-0882-3
54. Mandel P. Les acides nucleiques du plasma sanguin chez 1 homme. C R Seances Soc. Biol. Fil. *C R Seances Soc Biol Fil*. 1948;142:241-243.
55. Sorenson GD, Pribish DM, Valone FH, Memoli VA, Bzik DJ, Yao SL. Soluble normal and mutated DNA sequences from single-copy genes in human blood. *Cancer Epidemiology, Biomarkers & Prevention*. 1994;3(1):67-71.
56. Lo YM, Corbetta N, Chamberlain PF, et al. Presence of fetal DNA in maternal plasma and serum. *Lancet*. 1997;350(9076):485-487. doi:10.1016/s0140-6736(97)02174-0
57. Gorgannezhad L, Umer M, Islam MN, Nguyen NT, Shiddiky M. Circulating tumor DNA and liquid biopsy: Opportunities, challenges, and recent advances in detection technologies. *Lab Chip*. 2018;18. doi:10.1039/C8LC00100F
58. Rostami A, Lambie M, Yu CW, Stambolic V, Waldron JN, Bratman S v. Senescence, Necrosis, and Apoptosis Govern Circulating Cell-free DNA Release Kinetics. *Cell Rep*. 2020;31(13):107830. doi:https://doi.org/10.1016/j.celrep.2020.107830
59. Thierry AR, el Messaoudi S, Gahan PB, Anker P, Stroun M. Origins, structures, and functions of circulating DNA in oncology. *Cancer and Metastasis Reviews*. 2016;35(3):347-376. doi:10.1007/s10555-016-9629-x
60. Jiang WW, Zahurak M, Goldenberg D, et al. Increased plasma DNA integrity index in head and neck cancer patients. *Int J Cancer*. 2006;119(11):2673-2676. doi:https://doi.org/10.1002/ijc.22250
61. Alcaide M, Cheung M, Hillman J, et al. Evaluating the quantity, quality and size distribution of cell-free DNA by multiplex droplet digital PCR. *Sci Rep*. 2020;10(1):12564. doi:10.1038/s41598-020-69432-x
62. Mouliere F, el Messaoudi S, Pang D, Dritschilo A, Thierry AR. Multi-marker analysis of circulating cell-free DNA toward personalized medicine for colorectal cancer. *Mol Oncol*. 2014;8(5):927-941. doi:10.1016/j.molonc.2014.02.005
63. Jiang P, Chan CWM, Chan KCA, et al. Lengthening and shortening of plasma DNA in hepatocellular carcinoma patients. *Proc Natl Acad Sci U S A*. 2015;112(11):E1317-25. doi:10.1073/pnas.1500076112
64. Yu SCY, Chan KCA, Zheng YWL, et al. Size-based molecular diagnostics using plasma DNA for noninvasive prenatal testing. *Proc Natl Acad Sci U S A*. 2014;111(23):8583-8588. doi:10.1073/pnas.1406103111
65. Martignetti J, Camacho-Vanegas O, Priedigkeit N, et al. Personalized Ovarian Cancer Disease Surveillance and Detection of Candidate Therapeutic Drug Target in Circulating Tumor DNA. *Neoplasia*. 2014;16:97-103. doi:10.1593/neo.131900
66. Dwivedi DJ, Toltl LJ, Swystun LL, et al. Prognostic utility and characterization of cell-free DNA in patients with severe sepsis. *Crit Care*. 2012;16(4):R151. doi:10.1186/cc11466
67. Shi J, Zhang R, Li J, Zhang R. Size profile of cell-free DNA: A beacon guiding the practice and innovation of clinical testing. *Theranostics*. 2020;10:4737-4748. doi:10.7150/thno.42565
68. Chan KCA, Leung SF, Yeung SW, Chan ATC, Lo YMD. Persistent Aberrations in Circulating DNA Integrity after Radiotherapy Are Associated with Poor Prognosis in Nasopharyngeal Carcinoma Patients. *Clinical Cancer Research*. 2008;14(13):4141. doi:10.1158/1078-0432.CCR-08-0182

69. Kustanovich A, Schwartz R, Peretz T, Grinshpun A. Life and death of circulating cell-free DNA. *Cancer Biol Ther*. 2019;20(8):1057-1067. doi:10.1080/15384047.2019.1598759
70. Yu M, Stott S, Toner M, Maheswaran S, Haber DA. Circulating tumor cells: approaches to isolation and characterization. *Journal of Cell Biology*. 2011;192(3):373-382. doi:10.1083/jcb.201010021
71. Meng S, Tripathy D, Frenkel EP, et al. Circulating Tumor Cells in Patients with Breast Cancer Dormancy. *Clinical Cancer Research*. 2004;10(24):8152. doi:10.1158/1078-0432.CCR-04-1110
72. Krog BL, Henry MD. Biomechanics of the Circulating Tumor Cell Microenvironment. *Adv Exp Med Biol*. 2018;1092:209-233. doi:10.1007/978-3-319-95294-9_11
73. Parikh AR, Leshchiner I, Elagina L, et al. Liquid versus tissue biopsy for detecting acquired resistance and tumor heterogeneity in gastrointestinal cancers. *Nat Med*. 2019;25(9):1415-1421. doi:10.1038/s41591-019-0561-9
74. Iqbal N, Iqbal N. Human Epidermal Growth Factor Receptor 2 (HER2) in Cancers: Overexpression and Therapeutic Implications. *Mol Biol Int*. 2014;2014:852748. doi:10.1155/2014/852748
75. Maestranzi S, Przemioslo R, Mitchell H, Sherwood RA. The Effect of Benign and Malignant Liver Disease on the Tumour Markers CA19-9 and CEA. *Ann Clin Biochem*. 1998;35(1):99-103. doi:10.1177/000456329803500113
76. Sajid K, Parveen R, Sabih D, et al. Carcinoembryonic antigen (CEA) levels in hookah smokers, cigarette smokers and non-smokers. *J Pak Med Assoc*. 2007;57:595-599.
77. Duffy MJ, van Dalen A, Haglund C, et al. Clinical utility of biochemical markers in colorectal cancer: European Group on Tumour Markers (EGTM) guidelines. *Eur J Cancer*. 2003;39(6):718-727. doi:https://doi.org/10.1016/S0959-8049(02)00811-0
78. Duffy M. Carcinoembryonic Antigen as a Marker for Colorectal Cancer: Is It Clinically Useful? *Clin Chem*. 2001;47:624-630. doi:10.1093/clinchem/47.4.624
79. Villicana P, Whiting B, Goodison S, Rosser CJ. Urine-based assays for the detection of bladder cancer. *Biomark Med*. 2009;3(3):265. doi:10.2217/bmm.09.23
80. Mireille G, Yves F, Francois M, et al. Diagnostic Accuracy of Urinary Cytology, and Deoxyribonucleic Acid Flow Cytometry and Cytology on Bladder Washings During Followup for Bladder Tumors. *Journal of Urology*. 1997;157(5):1660-1664. doi:10.1016/S0022-5347(01)64827-4
81. Rife CC, Farrow GM, Utz DC. Urine cytology of transitional cell neoplasms. *Urol Clin North Am*. 1979;6(3):599-612. http://europepmc.org/abstract/MED/505675
82. Wiener HG, Vooijs GP, van't Hof-Grootenboer B. Accuracy of urinary cytology in the diagnosis of primary and recurrent bladder cancer. *Acta Cytol*. 1993;37(2):163-169. http://europepmc.org/abstract/MED/8465635
83. Krishna Prasad RB, Sharma A, Babu HM. An insight into salivary markers in oral cancer. *Dent Res J (Isfahan)*. 2013;10(3):287-295. https://pubmed.ncbi.nlm.nih.gov/24019794
84. Wang Z, Zhang L, Li L, et al. Sputum Cell-Free DNA: Valued Surrogate Sample for Detection of EGFR Mutation in Patients with Advanced Lung Adenocarcinoma. *The Journal of Molecular Diagnostics*. 2020;22(7):934-942. doi:https://doi.org/10.1016/j.jmoldx.2020.04.208

85. Ponti G, Maccaferri M, Mandrioli M, et al. Seminal cell-free DNA assessment as a novel prostate cancer biomarkers. *Pathology & Oncology Research*. 2018;24:941-945. doi:10.1007/12253-018-0416-6
86. Escudero L, Martínez-Ricarte F, Seoane J. ctDNA-Based Liquid Biopsy of Cerebrospinal Fluid in Brain Cancer. *Cancers (Basel)*. 2021;13(9):1989. doi:10.3390/cancers13091989
87. Fischer I, Gagner JP, Law M, Newcomb EW, Zagzag D. Angiogenesis in Gliomas: Biology and Molecular Pathophysiology. *Brain Pathology*. 2005;15(4):297-310. doi:https://doi.org/10.1111/j.1750-3639.2005.tb00115.x
88. Sarkaria JN, Hu LS, Parney IF, et al. Is the blood–brain barrier really disrupted in all glioblastomas? A critical assessment of existing clinical data. *Neuro Oncol*. 2018;20(2):184-191. doi:10.1093/neuonc/nox175
89. Pentsova EI, Shah RH, Tang J, et al. Evaluating Cancer of the Central Nervous System Through Next-Generation Sequencing of Cerebrospinal Fluid. *Journal of Clinical Oncology*. 2016;34(20):2404-2415. doi:10.1200/JCO.2016.66.6487
90. Mattos-Arruda L de, Mayor R, Ng CKY, et al. Cerebrospinal fluid-derived circulating tumour DNA better represents the genomic alterations of brain tumours than plasma. *Nat Commun*. 2015;6(1):8839. doi:10.1038/ncomms9839
91. Orzan F, De Bacco F, Lazzarini E, et al. Liquid Biopsy of Cerebrospinal Fluid Enables Selective Profiling of Glioma Molecular Subtypes at First Clinical Presentation. *Clinical Cancer Research*. Published online February 15, 2023:OF1-OF15. doi:10.1158/1078-0432.CCR-22-2903
92. Eibl RH, Schneemann M. Liquid biopsy and glioblastoma. *Explor Target Antitumor Ther*. 2023;4(1):28-41. doi:10.37349/etat.2023.00121
93. Zawadzki M, Walecki J, Kostkiewicz B, Kostyra K, Walczak P, Janowski M. Follow-up of intra-arterial delivery of bevacizumab for treatment of butterfly glioblastoma in patient with first-in-human, real-time MRI-guided intra-arterial neurointervention. *J Neurointerv Surg*. 2021;13(11):1037. doi:10.1136/neurintsurg-2021-017900
94. Louis DN, Perry A, Wesseling P, et al. The 2021 WHO Classification of Tumors of the Central Nervous System: a summary. *Neuro Oncol*. 2021;23(8):1231-1251. doi:10.1093/neuonc/noab106
95. Becker AP, Sells BE, Haque SJ, Chakravarti A. Tumor Heterogeneity in Glioblastomas: From Light Microscopy to Molecular Pathology. *Cancers (Basel)*. 2021;13(4):761. doi:10.3390/cancers13040761
96. Kim H, Zheng S, Amini SS, et al. Whole-genome and multisector exome sequencing of primary and post-treatment glioblastoma reveals patterns of tumor evolution. *Genome Res*. 2015;25(3):316-327. doi:10.1101/gr.180612.114
97. LeBlanc VG, Trinh DL, Aslanpour S, et al. Single-cell landscapes of primary glioblastomas and matched explants and cell lines show variable retention of inter- and intratumor heterogeneity. *Cancer Cell*. 2022;40(4):379-392.e9. doi:https://doi.org/10.1016/j.ccell.2022.02.016
98. Szadkowska P, Roura AJ, Wojtas B, et al. Improvements in Quality Control and Library Preparation for Targeted Sequencing Allowed Detection of Potentially Pathogenic Alterations in Circulating Cell-Free DNA Derived from Plasma of Brain Tumor Patients. *Cancers (Basel)*. 2022;14(16). doi:10.3390/cancers14163902
99. Gozdz A, Wojtaś B, Szpak P, et al. Preservation of the Hypoxic Transcriptome in Glioblastoma Patient-Derived Cell Lines Maintained at Lowered Oxygen Tension. *Cancers (Basel)*. 2022;14(19):4852. doi:10.3390/cancers14194852

100. Dawson MA, Kouzarides T. Cancer Epigenetics: From Mechanism to Therapy. *Cell*. 2012;150(1):12-27. doi:https://doi.org/10.1016/j.cell.2012.06.013
101. Maleszewska M, Kaminska B. Is glioblastoma an epigenetic malignancy? *Cancers (Basel)*. 2013;5(3):1120-1139. doi:10.3390/cancers5031120
102. Tessarz P, Kouzarides T. Histone core modifications regulating nucleosome structure and dynamics. *Nat Rev Mol Cell Biol*. 2014;15(11):703-708. doi:10.1038/nrm3890
103. Gielniewski B, Poleszak K, Roura AJ, et al. The novel, recurrent mutation in the TOP2A gene results in the enhanced topoisomerase activity and transcription deregulation in glioblastoma. *bioRxiv*. Published online January 1, 2021:2020.06.17.158477. doi:10.1101/2020.06.17.158477
104. Liu X, Lang J, Li S, et al. Fragment Enrichment of Circulating Tumor DNA With Low-Frequency Mutations. *Front Genet*. 2020;11:147. doi:10.3389/fgene.2020.00147
105. Mouliere F, Chandrananda D, Piskorz AM, et al. Enhanced detection of circulating tumor DNA by fragment size analysis. *Sci Transl Med*. 2018;10(466):eaat4921. doi:10.1126/scitranslmed.aat4921
106. Bolger AM, Lohse M, Usadel B. Trimmomatic: a flexible trimmer for Illumina sequence data. Bioinformatics. 30 (15), 2114-2120. Published online 2014.
107. Sedlazeck FJ, Rescheneder P, von Haeseler A. NextGenMap: fast and accurate read mapping in highly polymorphic genomes. *Bioinformatics*. 2013;29(21):2790-2791. doi:10.1093/bioinformatics/btt468
108. Petersen KR, Streett DA, Gerritsen AT, Hunter SS, Settles ML. Super deduper, fast PCR duplicate detection in fastq files. In: *Proceedings of the 6th ACM Conference on Bioinformatics, Computational Biology and Health Informatics*. ; 2015:491-492.
109. Koboldt DC, Zhang Q, Larson DE, et al. VarScan 2: somatic mutation and copy number alteration discovery in cancer by exome sequencing. *Genome Res*. 2012;22(3):568-576. doi:10.1101/gr.129684.111
110. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic Acids Res*. 2010;38(16):e164-e164. doi:10.1093/nar/gkq603
111. Mayakonda A, Lin DC, Assenov Y, Plass C, Koeffler HP. Maftools: efficient and comprehensive analysis of somatic variants in cancer. *Genome Res*. 2018;28(11):1747-1756. doi:10.1101/gr.239244.118
112. do Valle ÍF, Giampieri E, Simonetti G, et al. Optimized pipeline of MuTect and GATK tools to improve the detection of somatic single nucleotide polymorphisms in whole-exome sequencing data. *BMC Bioinformatics*. 2016;17(12):341. doi:10.1186/s12859-016-1190-7
113. Jones DTW, Hutter B, Jäger N, et al. Recurrent somatic alterations of FGFR1 and NTRK2 in pilocytic astrocytoma. *Nat Genet*. 2013;45(8):927-932. doi:10.1038/ng.2682
114. Roura AJ, Gielniewski B, Pilanc P, et al. Identification of the immune gene expression signature associated with recurrence of high-grade gliomas. *J Mol Med*. 2021;99(2):241-255. doi:10.1007/s00109-020-02005-7
115. Arnedo-Pac C, Mularoni L, Muiños F, Gonzalez-Perez A, Lopez-Bigas N. OncodriveCLUSTL: a sequence-based clustering method to identify cancer drivers. *Bioinformatics*. 2019;35(22):4788-4790. doi:10.1093/bioinformatics/btz501

116. Edwards AS, Scott JD. A-kinase anchoring proteins: protein kinase A and beyond. *Curr Opin Cell Biol*. 2000;12(2):217-221. doi:https://doi.org/10.1016/S0955-0674(99)00085-X
117. Kim YW, Koul D, Kim SH, et al. Identification of prognostic gene signatures of glioblastoma: a study based on TCGA data analysis. *Neuro Oncol*. 2013;15(7):829-839. doi:10.1093/neuonc/not024
118. Koutsioumpa M, Hatziapostolou M, Polytarchou C, et al. Lysine methyltransferase 2D regulates pancreatic carcinogenesis through metabolic reprogramming. *Gut*. 2019;68(7):1271. doi:10.1136/gutjnl-2017-315690
119. Knowles PP, Murray-Rust J, Kjær S, et al. Structure and Chemical Inhibition of the RET Tyrosine Kinase Domain*. *Journal of Biological Chemistry*. 2006;281(44):33577-33587. doi:https://doi.org/10.1074/jbc.M605604200
120. Trautmann M, Rehkämper J, Gevensleben H, et al. Novel pathogenic alterations in pediatric and adult desmoid-type fibromatosis – A systematic analysis of 204 cases. *Sci Rep*. 2020;10(1):3368. doi:10.1038/s41598-020-60237-6
121. Kovac M, Woolley C, Ribi S, et al. Germline <i>RET</i> variants underlie a subset of paediatric osteosarcoma. *J Med Genet*. 2021;58(1):20-24. doi:10.1136/jmedgenet-2019-106734
122. Rich JN, Bigner DD. Development of novel targeted therapies in the treatment of malignant glioma. *Nat Rev Drug Discov*. 2004;3(5):430-446. doi:10.1038/nrd1380
123. Yan H, Parsons DW, Jin G, et al. IDH1 and IDH2 Mutations in Gliomas. *New England Journal of Medicine*. 2009;360(8):765-773. doi:10.1056/NEJMoa0808710
124. McLendon R, Friedman A, Bigner D, et al. Comprehensive genomic characterization defines human glioblastoma genes and core pathways. *Nature*. 2008;455(7216):1061-1068. doi:10.1038/nature07385
125. Riddick G, Fine HA. Integration and analysis of genome-scale data from gliomas. *Nat Rev Neurol*. 2011;7(8):439-450. doi:10.1038/nrneurol.2011.100
126. Yang JM, Schiapparelli P, Nguyen HN, et al. Characterization of PTEN mutations in brain cancer reveals that pten mono-ubiquitination promotes protein stability and nuclear localization. *Oncogene*. 2017;36(26):3673-3685. doi:10.1038/onc.2016.493
127. Ohgaki H, Kleihues P. Epidemiology and etiology of gliomas. *Acta Neuropathol*. 2005;109(1):93-108. doi:10.1007/s00401-005-0991-y
128. Petitjean A, Achatz MIW, Borresen-Dale AL, Hainaut P, Olivier M. TP53 mutations in human cancers: functional selection and impact on cancer prognosis and outcomes. *Oncogene*. 2007;26(15):2157-2165. doi:10.1038/sj.onc.1210302
129. Hill R, Wu H. PTEN, stem cells, and cancer stem cells. *J Biol Chem*. 2009;284(18):11755-11759. doi:10.1074/jbc.R800071200
130. Saadeh FS, Mahfouz R, Assi HI. EGFR as a clinical marker in glioblastomas and other gliomas. *Int J Biol Markers*. 2017;33(1):22-32. doi:10.5301/ijbm.5000301
131. McLendon R, Friedman A, Bigner D, et al. Comprehensive genomic characterization defines human glioblastoma genes and core pathways. *Nature*. 2008;455(7216):1061-1068. doi:10.1038/nature07385
132. Qin T, Mullan B, Ravindran R, et al. ATRX loss in glioma results in dysregulation of cell-cycle phase transition and ATM inhibitor radio-sensitization. *Cell Rep*. 2022;38(2):110216. doi:https://doi.org/10.1016/j.celrep.2021.110216
133. Murphree AL, Benedict WF. Retinoblastoma: Clues to Human Oncogenesis. *Science (1979)*. 1984;223(4640):1028-1033. doi:10.1126/science.6320372

134. Kyritsis AP, Bondy ML, Rao JS, Sioka C. Inherited predisposition to glioma. *Neuro Oncol*. 2010;12(1):104-113. doi:10.1093/neuonc/nop011
135. Tan MH, Mester JL, Ngeow J, Rybicki LA, Orloff MS, Eng C. Lifetime cancer risks in individuals with germline PTEN mutations. *Clin Cancer Res*. 2012;18(2):400-407. doi:10.1158/1078-0432.CCR-11-2283
136. Tröger J, Moutty MC, Skroblin P, Klussmann E. A-kinase anchoring proteins as potential drug targets. *Br J Pharmacol*. 2012;166(2):420-433. doi:10.1111/j.1476-5381.2011.01796.x
137. Reggi E, Diviani D. The role of A-kinase anchoring proteins in cancer development. *Cell Signal*. 2017;40:143-155. doi:https://doi.org/10.1016/j.cellsig.2017.09.011
138. Will M, Qin ACR, Toy W, et al. Rapid Induction of Apoptosis by PI3K Inhibitors Is Dependent upon Their Transient Inhibition of RAS–ERK Signaling. *Cancer Discov*. 2014;4(3):334-347. doi:10.1158/2159-8290.CD-13-0611
139. Consortium APG. AACR Project GENIE: Powering Precision Medicine through an International Consortium. *Cancer Discov*. 2017;7(8):818-831. doi:10.1158/2159-8290.CD-17-0151
140. Rice T, Lachance DH, Molinaro AM, et al. Understanding inherited genetic risk of adult glioma - a review. *Neurooncol Pract*. 2016;3(1):10-16. doi:10.1093/nop/npv026
141. Chetan B, Mark S, J LR, et al. Detection of Circulating Tumor DNA in Early- and Late-Stage Human Malignancies. *Sci Transl Med*. 2014;6(224):224ra24-224ra24. doi:10.1126/scitranslmed.3007094
142. Underhill HR, Kitzman JO, Hellwig S, et al. Fragment Length of Circulating Tumor DNA. Kwiatkowski DJ, ed. *PLoS Genet*. 2016;12(7):e1006162. doi:10.1371/journal.pgen.1006162
143. Yuxuan W, Simeon S, Ming Z, et al. Detection of tumor-derived DNA in cerebrospinal fluid of patients with primary tumors of the brain and spinal cord. *Proceedings of the National Academy of Sciences*. 2015;112(31):9704-9709. doi:10.1073/pnas.1511694112
144. Chen WW, Balaj L, Liau LM, et al. BEAMing and Droplet Digital PCR Analysis of Mutant IDH1 mRNA in Glioma Patient Serum and Cerebrospinal Fluid Extracellular Vesicles. *Mol Ther Nucleic Acids*. 2013;2:e109. doi:https://doi.org/10.1038/mtna.2013.28
145. Bettegowda C, Sausen M, Leary RJ, et al. Detection of Circulating Tumor DNA in Early- and Late-Stage Human Malignancies. *Sci Transl Med*. 2014;6(224):224ra24-224ra24. doi:10.1126/scitranslmed.3007094
146. Zill OA, Banks KC, Fairclough SR, et al. The Landscape of Actionable Genomic Alterations in Cell-Free Circulating Tumor DNA from 21,807 Advanced Cancer Patients. *Clinical Cancer Research*. 2018;24(15):3528-3538. doi:10.1158/1078-0432.CCR-17-3837
147. Bagley SJ, Nabavizadeh SA, Mays JJ, et al. Clinical Utility of Plasma Cell-Free DNA in Adult Patients with Newly Diagnosed Glioblastoma: A Pilot Prospective Study. *Clinical Cancer Research*. 2020;26(2):397-407. doi:10.1158/1078-0432.CCR-19-2533
148. Muralidharan K, Yekula A, Small JL, et al. TERT Promoter Mutation Analysis for Blood-Based Diagnosis and Monitoring of Gliomas. *Clinical Cancer Research*. 2021;27(1):169-178. doi:10.1158/1078-0432.CCR-20-3083

149. Pan C, Diplas BH, Chen X, et al. Molecular profiling of tumors of the brainstem by sequencing of CSF-derived circulating tumor DNA. *Acta Neuropathol*. 2019;137(2):297-306. doi:10.1007/s00401-018-1936-6
150. Mair R, Mouliere F. Cell-free DNA technologies for the analysis of brain cancer. *Br J Cancer*. 2022;126(3):371-378. doi:10.1038/s41416-021-01594-5
151. Jung A, Kirchner T. Liquid Biopsy in Tumor Genetic Diagnosis. *Dtsch Arztebl Int*. 2018;115(10):169-174. doi:10.3238/arztebl.2018.0169
152. Esagian SM, Grigoriadou GI, Nikas IP, et al. Comparison of liquid-based to tissue-based biopsy analysis by targeted next generation sequencing in advanced non-small cell lung cancer: a comprehensive systematic review. *J Cancer Res Clin Oncol*. 2020;146(8):2051-2066. doi:10.1007/s00432-020-03267-x
153. Bando H, Nakamura Y, Taniguchi H, et al. Impact of a metastatic site on circulating tumor DNA (ctDNA) analysis in patients (pts) with metastatic colorectal cancer (mCRC). *Journal of Clinical Oncology*. 2021;39(15_suppl):3554. doi:10.1200/JCO.2021.39.15_suppl.3554
154. Darrigues L, Pierga JY, Bernard-Tessier A, et al. Circulating tumor DNA as a dynamic biomarker of response to palbociclib and fulvestrant in metastatic breast cancer patients. *Breast Cancer Research*. 2021;23(1):31. doi:10.1186/s13058-021-01411-0
155. Marsavela G, McEvoy AC, Pereira MR, et al. Detection of clinical progression through plasma ctDNA in metastatic melanoma patients: a comparison to radiological progression. *Br J Cancer*. Published online 2021. doi:10.1038/s41416-021-01507-6
156. Tu Q, Wu X, Le Rhun E, et al. CellSearch® technology applied to the detection and quantification of tumor cells in CSF of patients with lung cancer leptomeningeal metastasis. *Lung Cancer*. 2015;90(2):352-357. doi:https://doi.org/10.1016/j.lungcan.2015.09.008
157. Wang L, Balasubramanian P, Chen AP, Kummar S, Evrard YA, Kinders RJ. Promise and limits of the CellSearch platform for evaluating pharmacodynamics in circulating tumor cells. *Semin Oncol*. 2016;43(4):464-475. doi:10.1053/j.seminoncol.2016.06.004
158. Kwapisz D. The first liquid biopsy test approved. Is it a new era of mutation testing for non-small cell lung cancer? *Ann Transl Med*. 2017;5(3):46. doi:10.21037/atm.2017.01.32
159. Cui W, Milner-Watts C, O'Sullivan H, et al. Up-front cell-free DNA next generation sequencing improves target identification in UK first line advanced non-small cell lung cancer (NSCLC) patients. *Eur J Cancer*. 2022;171:44-54. doi:https://doi.org/10.1016/j.ejca.2022.05.012
160. Aggarwal C, Thompson JC, Black TA, et al. Clinical Implications of Plasma-Based Genotyping With the Delivery of Personalized Therapy in Metastatic Non–Small Cell Lung Cancer. *JAMA Oncol*. 2019;5(2):173-180. doi:10.1001/jamaoncol.2018.4305
161. Unger FT, Witte I, David KA. Prediction of individual response to anticancer therapy: historical and future perspectives. *Cellular and Molecular Life Sciences*. 2015;72(4):729-757. doi:10.1007/s00018-014-1772-3
162. Thompson SL, Compton DA. Chromosomes and cancer cells. *Chromosome Res*. 2011;19(3):433-444. doi:10.1007/s10577-010-9179-y
163. Kuijk E, Jager M, van der Roest B, et al. The mutational impact of culturing human pluripotent and adult stem cells. *Nat Commun*. 2020;11(1):2493. doi:10.1038/s41467-020-16323-4

164. Goodspeed A, Heiser LM, Gray JW, Costello JC. Tumor-Derived Cell Lines as Molecular Models of Cancer Pharmacogenomics. *Molecular Cancer Research*. 2016;14(1):3-13. doi:10.1158/1541-7786.MCR-15-0189
165. Koutsioumpa M, Hatziapostolou M, Polytarchou C, et al. Lysine methyltransferase 2D regulates pancreatic carcinogenesis through metabolic reprogramming. *Gut*. 2019;68(7):1271. doi:10.1136/gutjnl-2017-315690
166. Borkin D, He S, Miao H, et al. Pharmacologic Inhibition of the Menin-MLL Interaction Blocks Progression of MLL Leukemia In Vivo. *Cancer Cell*. 2015;27(4):589-602. doi:https://doi.org/10.1016/j.ccell.2015.02.016
167. Feoli A, Viviano M, Cipriano A, Milite C, Castellano S, Sbardella G. Lysine methyltransferase inhibitors: where we are now. *RSC Chem Biol*. 2022;3(4):359-406. doi:10.1039/D1CB00196E
168. Meng X, Deng Y, Lv Z, et al. LncRNA SNHG5 Promotes Proliferation of Glioma by Regulating miR-205-5p/ZEB2 Axis. *Onco Targets Ther*. 2019;12:11487-11496. doi:10.2147/OTT.S228439
169. Williams K, Sobol RW. Mutation research/fundamental and molecular mechanisms of mutagenesis: special issue: DNA repair and genetic instability. *Mutat Res*. 2013;743-744:1-3. doi:10.1016/j.mrfmmm.2013.04.009
170. Wang LC, Gautier J. The Fanconi anemia pathway and ICL repair: implications for cancer therapy. *Crit Rev Biochem Mol Biol*. 2010;45(5):424-439. doi:10.3109/10409238.2010.502166
171. Woo HY, Na K, Yoo J, et al. Glioblastomas harboring gene fusions detected by next-generation sequencing. *Brain Tumor Pathol*. 2020;37(4):136-144. doi:10.1007/s10014-020-00377-9
172. Yu Z, Li H, Wang M, Luo W, Xue Y. GDNF regulates lipid metabolism and glioma growth through RET/ERK/HIF-1/SREBP-1. *Int J Oncol*. 2022;61(3):109. doi:10.3892/ijo.2022.5399
173. Tilak M, Holborn J, New LA, Lalonde J, Jones N. Receptor Tyrosine Kinase Signaling and Targeting in Glioblastoma Multiforme. *Int J Mol Sci*. 2021;22(4):1831. doi:10.3390/ijms22041831
174. Arighi E, Borrello MG, Sariola H. RET tyrosine kinase signaling in development and cancer. *Cytokine Growth Factor Rev*. 2005;16(4):441-467. doi:https://doi.org/10.1016/j.cytogfr.2005.05.010
175. Bordeaux MC, Forcet C, Granger L, et al. The RET proto-oncogene induces apoptosis: a novel mechanism for Hirschsprung disease. *EMBO J*. 2000;19(15):4056-4063. doi:10.1093/emboj/19.15.4056
176. Pottier C, Fresnais M, Gilon M, Jérusalem G, Longuespée R, Sounni NE. Tyrosine Kinase Inhibitors in Cancer: Breakthrough and Challenges of Targeted Therapy. *Cancers (Basel)*. 2020;12(3):731. doi:10.3390/cancers12030731
177. Jensen SA, Day ES, Ko CH, et al. Spherical Nucleic Acid Nanoparticle Conjugates as an RNAi-Based Therapy for Glioblastoma. *Sci Transl Med*. 2013;5(209):209ra152-209ra152. doi:10.1126/scitranslmed.3006839

Publications

Identification of the immune gene expression signature associated with recurrence of high-grade gliomas.

Roura AJ, Gielniewski B, Pilanc P, **Szadkowska P**, Maleszewska M, Krol SK, Czepko R, Kaspera W, Wojtas B, Kaminska B.

J Mol Med, 2020 Nov 19. https://doi:10.1007/s00109-020-02005-7 [114]

Improvements in Quality Control and Library Preparation for Targeted Sequencing Allowed Detection of Potentially Pathogenic Alterations in Circulating Cell-Free DNA Derived from Plasma of Brain Tumor Patients.

Szadkowska P, Roura AJ, Wojtas B, Wojnicki K, Licholai S, Waller T, Gubala T, Zukowski K, Karpeta M, Wilkus K, Kaspera W, Nawrocki S, Kaminska B.

Cancers 2022, *14*, 3902. https://doi.org/10.3390/cancers14163902 [98]

Preservation of the Hypoxic Transcriptome in Glioblastoma Patient-Derived Cell Lines Maintained at Lowered Oxygen Tension.

Gozdz A, Wojtaś B, Szpak P, **Szadkowska P**, Czernicki T, Marchel A, Wójtowicz K, Kaspera W, Ladzinski P, Szopa W, Niedbala M, Nawrocki S, Kaminska B, Kalaszczynska I.

Cancers 2022; 14(19):4852. https://doi.org/10.3390/cancers14194852 [99]

Exploring Novel Therapeutic Opportunities for Glioblastoma Using Patient-Derived Cell Cultures.

Ciechomska I, Wojnicki K, Wojtas B, **Szadkowska P**, Poleszak K, Kaza B, Jaskula K, et al.

Cancers 2023; 15 (5): 1562. https://doi.org/10.3390/cancers15051562.

Regulatory Networks Driving Expression of Genes Critical for Glioblastoma Are Controlled by the Transcription Factor C-Jun and the Pre-Existing Epigenetic Modifications.

Roura A-J, **Szadkowska P**, Poleszak K, Dabrowski M J, Ellert-Miklaszewska A, Wojnicki K, Ciechomska I, Stepniak K, Kaminska B, and Wojtas B.

*Clinical Epigenetics 2023;*15(1):29. https://doi.org/10.1186/s13148-023-01446-4.

Targeted sequencing of cancer related genes reveals a recurrent TOP2A variant which affects DNA binding and coincides with global transcriptional changes in glioblastoma.

Gielniewski B, Poleszak K, Roura A, **Szadkowska P**, Jacek K, Krol SK, Guzik R, Wiechecka P, Maleszewska M, Kaza B, Marchel A, Czernicki, et al.

International Journal of Cancer; 20 June 2023; https://doi.org/10.1002/ijc.34631