**Medical University of Bialystok**


Doctoral dissertation in Medical Sciences

# Development and implementation of original bioinformatic pipelines for NGS data analysis in cancer and assisted reproduction research

Opracowanie i wdrożenie oryginalnych bioinformatycznych protokołów analitycznych do analizy danych NGS w badaniach nad nowotworami i rozrodem wspomaganym

**Bence Gálik**

Supervisor: Dr hab. Attila Gyenesei

Head of the Department: Prof. dr hab. Jacek Niklinski

Department of Clinical Molecular Biology

Medical University of Bialystok

2022 Bialystok

This research was conducted within the project which has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 754432 and the Polish Ministry of Education and Science, from the financial resources for science in 2018-2023 granted for the implementation of an international co-financed project.

## Table of content

## Abstract

An unprecedented wealth of biological data has been generated since the human and other genome projects were started. The huge request for raw data analysis and interpretation is being controlled by the evolving science of bioinformatics. Bioinformatics is defined as the application of computational tools and analysis approaches to capture and help to interpret results. It is an interdisciplinary field which harnesses computer science, mathematics, statistics, physics. It is essential for management of data in comprehensive medical research.

This Ph.D. thesis deals with the topic of bioinformatic pipelines implemented for end-to-end analysis of high-throughput sequencing data (DNA/RNA) in different research areas including cancer and assisted reproduction trough case studies.

First, example studies from primary central nervous system lymphoma (**PCNSL**) and chronic lymphocytic leukemia (**CLL**) research fields will be presented with the aim of identifying disease related somatic variants (SNPs, short INDELs) from custom targeted gene panels to create mutation profiles. Next, the reduced-representation bisulfite sequencing method and its related bioinformatic analysis will be demonstrated for DNA CpG methylation profiling to better understand one of the most aggressive cancers, known as glioblastoma multiforme (**GBM**). Additional to DNA related changes, microRNAs have emerged as promising biomarkers that can contribute more effective to early detection of lung cancer. The analysis of global miRNAome in early-stage non-small cell lung cancer (**NSCLC**) patients will be showed for the development of accurate predictive biomarkers of relapse, following surgery.

Infertility impacts the reproductive period of millions in the World and it influences their families and communities. *In vitro* fertilization helps in getting pregnant, embryo development and implantation, but also there is a high need for genetic examination before embryo transfer. Therefore, as the last application field, a comprehensive bioinformatic workflow for non-invasive pre-implantation genetic testing for aneuploidy (**NIPGT-A**) in assisted reproduction treatments will be introduced.

The described bioinformatic workflows in my thesis are essential in better understanding the molecular background of the previously mentioned disorders.



**Key words:** cancer, IVF, CLL, PCNSL, GBM, NCSCL, NIPGT-A, NGS, bioinformatics, workflow building, mutation profiling, methylome, CNVs, miRNAs

## Polish abstract

Niespotykane bogactwo danych biologicznych zostało wygenerowane od czasu rozpoczęcia Projektu poznania ludzkiego genomu (HGP, z ang. Human Genome Project) oraz innych projektów genomicznych. Odpowiedzią na ogromne zapotrzebowanie na analizę i interpretację surowych danych genetycznych jest dziedzina nauki nazwana bioinformatyką. Bioinformatyka jest definiowana jako zastosowanie narzędzi obliczeniowych i podejść analitycznych w celu generowania i pomocy w interpretacji wyników. Jest to interdyscyplinarna dziedzina nauki, która wykorzystuje matematykę, statystykę i fizykę. Jest niezbędna do zarządzania danymi w kompleksowych badaniach medycznych.

Niniejsza praca doktorska dotyczy tematu bioinformatycznych protokołów analitycznych wdrożonych do kompleksowej analizy danych z sekwencjonowania o dużej przepustowości (DNA/RNA) w różnych obszarach badawczych, w tym nowotworach i wspomaganym rozrodzie, na konkretnych przykładach.

W pierwszej kolejności przedstawione zostaną przykładowe badania z dziedziny pierwotnego chłoniaka ośrodkowego układu nerwowego (**PCNSL**) oraz przewlekłej białaczki limfocytowej (**PBL** ang. **CLL**), których celem jest identyfikacja somatycznych polimorfizmów pojedynczego nukleotydu (SNP) i małych insercji i delecji (INDELi) z paneli genów ukierunkowanych na konkretną chorobę w celu stworzenia profili mutacyjnych. Następnie zaprezentowana zostanie metoda analizy zmian w profilu i poziomie metylacji DNA (RRBS) i związana z nią analiza bioinformatyczna do profilowania metylacji DNA wysp CpG w celu lepszego zrozumienia jednego z najbardziej agresywnych nowotworów, znanego jako glejak wielopostaciowy (**GBM**). Oprócz zmian związanych z DNA, mikroRNA zostało uznane za obiecujące biomarkery, które mogą przyczynić się do bardziej efektywnego wykrywania raka płuc. Kolejny przedstawiony w pracy przykład to analiza mikroRNA u chorych na niedrobnokomórkowego raka płuca (**NSCLC**) we wczesnym stadium zaawansowania, która pozwoliła na opracowanie dokładnych biomarkerów predykcyjnych nawrotu choroby po operacji.

Niepłodność wpływa na psychofizyczne funkcjonowanie milionów ludzi na świecie, na ich rodziny i społeczności. Zapłodnienie in vitro pomaga w zajściu w ciążę, rozwoju zarodka i implantacji, ale istnieje duże zapotrzebowanie na badania genetyczne przed transferem zarodka. Dlatego też jako ostatni obszar zastosowań, zostanie przedstawiony kompleksowy

bioinformatyczny protokół analityczny dla nieinwazyjnych przed-implantacyjnych badań genetycznych w kierunku aneuploidii (**NIPGT-A**) w zabiegach wspomaganego rozrodu.

Opisane w mojej pracy dyplomowej bioinformatyczne protokoły analityczne są niezbędne w lepszym zrozumieniu molekularnego podłoża wcześniej wymienionych chorób i zaburzeń.

**Słowa kluczowe:** nowotwory, Zapłodnienie pozaustrojowe, zapłodnienie in vitro (IVF), przewlekła białaczka limfocytowa (PBL, ang. CLL chronic lymphocytic leukemia), pierwotny chłoniak ośrodkowego układu nerwowego (PCNSL), glejak wielopostaciowy (GBM; łac. glioblastoma multiforme), niedrobnokomórkowy rak płuca (NSCLC), aneuploidia (NIPGT-A), sekwencjonowanie nowej generacji (NGS), bioinformatyka, bioinformatyczne protokoły analityczne, tworzenie profili mutacji, metylom, zmienność liczby kopii DNA (CNV), mikroRNA

# Introduction

*Bioinformatics*

The significant decrease cost of high-throughput next-generation sequencing (NGS) during the past decade has made it available for a wider range of researchers. Using new NGS platforms, for instance Illumina NovaSeq 6000, such a large amount of data can be generated that cannot be processed with traditional methods. Bioinformatics workflows and pipelines are needed to extract information from sequencing data and obtain knowledge. The field of bioinformatics is growing every year almost exponentially as the number of applications and research data volume increases [Stephens et al., 2015]. Workflows consist of many separate steps starting with experimental design, through the initial check of raw data, and various quality control steps, all the way to final visualizations and results ready for interpretation. Workflows must be carefully designed, implemented and executed. A bioinformatics pipeline is a series of software algorithms that process raw data (within a workflow) and generate evaluable results from this data that researchers could interpret. The available basic workflows and best practices usually tackle only around one specific case and directly cannot be used elsewhere, or unique solution are needed. For this reason, workflows or algorithms specifically tailored to individual experiments need to be developed.

Bioinformatics can be divided in two major parts based on the aim of the research [Hagen, 2000]. So called “sequencing bioinformatics” focuses on data related to the DNA or RNA sequence or its modifications. Mainly it deals with genomics and transcriptomics data to understand biological hypothesis. This interdisciplinary field combines computer science, biology, mathematics, statistics and medicine. To correctly answer questions, one must know the biological background as well as ways on how to get to the final conclusions. Bioinformatics has become a central part of many areas of the current biology. Until recently, bioinformaticians played a support role for wet-lab biologists but in past few years the research started to change direction and more teams are becoming bioinformatics-centered [Perkel, 2016]. Also, a bioinformatician can be focused on pure informatics, such as software and database development, or algorithm design. Nowadays, a lot of software tools and databases are available and new ones are introduced every year.

A combination of bioinformatics tools and visualizations results into workflow or pipeline. Because of the high number of available tools, databases, and a vast number of biological applications it is not straightforward to create an appropriate workflow that correctly answers the hypothesis. In addition to basic and established workflows, many experiments require careful adjustments or completely novel construction of pipelines which requires deep knowledge and much experience in all the details of bioinformatics and biology as well. What is more important than the generation of raw data and results is the interpretation. Several information sources, biological knowledge and often a good hunch and teamwork need to be combined if we want to get to the definitive answer [Kanehisa and Bork, 2003].

The increasing complexity of results of omics analyses go together with problems about the reproducibility of experiments. When analysing large data sets, the main source of computational irreproducibility comes from a lack of good practice related to software and database usage [Brito et al., 2020; Masca et al., 2015; National Academies of Sciences, Engineering, and Medicine, 2016; Piccolo and Frampton, 2016]. Complexity grows even faster when all aspects of a given analysis are included. Small variations across computational platforms contribute to computational irreproducibility by producing numerical instability as well [Garijo et al., 2013]. This is particularly important to high-performance computational (HPC) environments that are routinely used for omics data analyses [Loman and Watson, 2013]. Handling many software packages at the same time, some of which may be incompatible, is a big challenge. The conflicting demands of frequent software updates and maintaining the reproducibility of original results add another unwanted source of problems. Together with these issues, high-throughput usage of complex pipelines can also be burdened by the hundreds of intermediate files often produced by individual tools. Hardware fluctuations in these types of pipelines, combined with poor error handling, could result in considerable readout instability. In silico workflow management systems like Nextflow [di Tommaso et al., 2017] are designed to overcome these problems and are capable of large-scale biological analyses. These systems enable faster implementation, prototyping and deployment of pipelines that combine complementary software packages.

Since, this thesis will focus on various NGS-based applications, their background should be briefly introduced.

*NGS applications*

One of the main advantages of NGS, in compared to other methods like Sanger sequencing or microarray, is that it is a sequence-it-all approach. Any kind and type of nucleotide sequence could be processed and read. NGS is not that sensitive to nucleotide quantity, quality and purity like current third-generation methods. Of course, generating good quality data from formalin-fixed paraffin-embedded (FFPE) tissue blocks, ultra-low input materials (e.g., single cells or ctDNA) or from samples with lower RNA integrity number could be challenging. Apart from the pros and cons these features make NGS a versatile tool for countless applications. Currently, there are more than 200 NGS applications [http://enseqlopedia.com/enseqlopedia/], and new applications are being published every year. Here, only the most common ones will be mentioned.

DNA-Seq is historically one of the first applications of NGS. If the desired outcome is the complete DNA sequence of a genome, we talk about whole genome sequencing (WGS). We can either focus on an unknown sequence and try to reconstruct it using *de novo* assembly or we can re-sequence something already known and look for genetic variability, such as single nucleotide variants (SNV), insertions or deletions (INDEL) [Nielsen et al., 2011]. We can also aim at larger structural variations such as copy-number variants (CNV) [Hayes et al., 2013] or structural variants (SV) [Tattini et al., 2011]. Some genomes are relatively big, and most of the bioinformatic solution rely on coverage, it is not necessary to sequence the entire genome. It is enough to focus on smaller, more specific parts of the genome. If we interested in the nucleotide order of a shorter and more specific DNA fragment, we talk about amplicon sequencing. If several genomic loci are enriched the application called as panel (a.k.a., targeted) sequencing. Targeted sequencing focuses on a specific subset of genes or genomic regions. This is often used in diagnostics where only a panel of disease related genes is selected [Bybee et al., 2011; Konnick et al., 2017]. Nowadays, some library preparation KITs of gene panels contain unique molecular identifiers (UMI) which help to determine the variant allele frequencies (VAF) more precisely [Crysup et al., 2022; Smith et al., 2017; Zhou and Swanstrom, 2020]. This means additional steps in the corresponding bioinformatic workflow as well. In the whole-exome sequencing (WES) application only coding parts of genome are enriched by hybridization probes [Rabbani et al., 2014]. This allows to focus on mutations presented only in coding parts of genes which often carry the disease-causing mutation. In

other cases, we are interested in gDNA modifications such as methylated cytosines. Bisulfite-seq (BS-Seq) or WGBS is a well-established protocol to provide single-base resolution of these changes [Kunde-Ramamoorthy et al., 2014; Yang and Mackenzie 2020].

RNA-Seq is another widely used application of NGS used mainly for a study of a gene/miRNA expression or build reference transcriptome. It can be targeted on several types of molecules like mRNA, short RNAs (miRNA, piRNA), long non-coding RNA, etc. [Wang et al., 2009]. In differential gene expression projects at least two or more groups of samples (min 3-5 biological replicates/group) are sequenced. The expression profiles of the groups are compared using statistical approaches. The results should reveal the differences based on the gene expression and potentially identify genes and pathways that are responsible for a given phenotype. RNA-Seq. To carry out basic RNA-Seq application transcriptome information and gene annotation, a list of genes, their composition (e.g., introns, exons and untranslated regions) and position in a reference genome should already know. The non-model organism's transcriptome could be assembled *de novo* which works on slightly different principles than *de novo* genome assembly as it must be able to assemble more isoforms of a single gene [Martin and Wang, 2011]. For these types of projects, third-generation sequencing methods are more suitable in a combination of NGS as well [Leonardi and Leger, 2021; Whang et al., 2021; Zhang et al., 2019].

As the widespread use of these tools makes possible to apply them in different clinical research studies. In the following sections example cases will be introduced.

*Somatic mutation profiling*

***CASE STUDY I.: CLL***

Chronic lymphocytic leukemia (CLL) is a type of cancer in blood and bone marrow. The term "chronic" comes from that this leukemia typically progresses more slowly than other ones. The name "lymphocytic" comes from the fact, that white blood cells, called lymphocytes are affected by the disease. CLL is characterized by substantial clinical and genetic heterogeneity. The latest WES/WGS studies undisclosed recurrently mutated driver genes, including *ATM*, *NOTCH1*, *SF3B1*, *BIRC3*, *NKFBIE*, *MYD88* and *TP53*, and identified clonal evolution as the major mechanism driving disease progression [Baliakas et al., 2015; Landau et al., 2013, 2015; Puente et al., 2011; Quesada et al., 2011; Schuh et al., 2012; Wang et al., 2011]. Patients with *TP53* aberrations have been characterized typically by refractoriness to standard therapies and particularly poor outcome with rapid selection of the resistant clones [Malcikova et al., 2015; Rossi et al., 2009]. The B-cell proliferation- and the irreversible Bruton's tyrosine kinase (BTK) inhibitor, called ibrutinib has been changing the treatment standards of CLL with remarkable outcomes in first line and in relapse [Ahn et al., 2018; Burger et al., 2015; Byrd et al., 2014; Farooqui et. al., 2015]. Regardless of the durable responses were observed in most patients (approx. 20%) develop resistance, with mutations in *BTK* and *PLCG2* representing the predominant mechanisms conferring secondary ibrutinib resistance [Furman et al., 2014; Woyach et al., 2014]. The loss of function in *BTK* Cys481 mutations are leading to impaired ibrutinib binding and/or the gain of function *PLCG2* (e.g*.,* Arg665Trp/Ser707Tyr/Leu845Phe) mutations are resulting in continuous B-cell receptor (BCR) signalling [Ahn et al., 2017; Maddocks et al., 2015; Woyach et al., 2017].

According to numerous studies, the above-mentioned mutations are commonly present in multiple independent subclones are proposing parallel clonal evolution and their emergence predates clinical progression and relapse [Ahn et al., 2017; 2015; Woyach et al., 2017]. Clonal shifts were identified by WES during the early periods of ibrutinib treatment in one third of the patients were associated with disease progression [Landau et al., 2017]. The comprehensive characterisation of the mechanisms underlying ibrutinib resistance and the related changes in the subclonal architecture have dominant clinical impact [Jain et al., 2015].

To dissect the clonal evolution affecting all relevant mutations in CLL, a temporal mutation profiling was performed by targeted analysis of 30 genes in paired pre- and post-treatment cohort of patients with ibrutinib therapy.

***CASE STUDY II.: PCNSL***

A primary central nervous system lymphoma (PCNSL) is a type of cancer originating from immune cells known as lymphocytes that develops in central nervous system (e.g., brain and/or spinal cord; CNS). It has rare malignancy with an exceptionally aggressive clinical course and a poor outcome. Histologically, it is manifested as diffuse large B-cell lymphoma (DLBCL), which is bound to the CNS structures [Hochberg et al., 1988; O'Neill et al., 1989; Kluin et al., 2017]. DLBCLs could be sub-classified into molecular subgroups including germinal center B-cell (GC) type or activated B-cell (ABC) type, with a small number of "unclassified" (UC) cases [Alizadeh et al., 2000]. These sub-classification methods have important prognostic and potential therapeutic implications [Alizadeh et al., 2000; Wright et al., 2003]. The GC/ABC classification of DLBCLs is based on gene expression patterns (GEP) of fresh or fresh-frozen tissues and Affymetrix became a "gold-standard" method. It was followed by the development of numerous formalin-fixed paraffin-embedded (FFPE) tissue-based immunohistochemistry (IHC) predictors, including the Hans algorithm [Choi et al., 2009; Hans et al., 2004; Meyer et al., 2011]. These methods showed poor efficacy in precise assignment of patients into subgroups. The Lymphoma Subtyping Test (LST) assay developed by NanoString Technologies, is an FFPE compatible, gene expression-based test for molecular subtyping of B-cell lymphomas. The assay is based on the expression of 15 target- and 5 housekeeping genes and begun a more accurate technique compared with standard IHC algorithms.

The molecular subtype of PCNSL has been studied by different methods resulted in conflicting conclusions. For instance, according to various IHC studies, an ABC-like immunophenotype is typical [Camilleri-Broet et al., 2006; Liu et al., 2017; Raoux et al., 2010], but immunoglobulin heavy chain gene (*IGHV*) mutational signatures also provide evidence for germinal center exposure [Larocca et al., 1998; Montesinos-Rongen et al., 1999; Thompsett et al., 1999]. In contrast, gene expression profiling studies indicate that PCNSLs are distributed among the spectrum of systemic DLBCL with roughly equal proportion of ABC and GC cases [Montesinos-Rongen et al., 2008; Rubenstein et al., 2006].

Recent studies profiling the genomic background of PCNSL have identified multiple mutated genes, which harbouring putative driver aberrations and others serving as aberrant somatic hypermutation (ASHM) targets [Braggio et al., 2015; Bruno et al., 2014; Chapuy et al., 2016; Fukumura et al., 2016; Nakamura et al., 2016; Vater et al., 2015; Zhou et al., 2018]. The most frequently mutated genes, overlap with the mutational targets identified in systemic DLBCL, are listed in Table 1.

| Pathway | Genes |
|---|---|
| B-cell receptor signalling | *MYD88*, *CD79B, CARD11* |
| Cell cycle/apoptosis regulation | *TP53*, *CCND3*, *BTG2*, *PIM1*, *CDKN2A*, *ATM* |
| Chromatin regulation | *KMT2D* |
| Transcriptional regulation | *C-MYC*, *PRDM1*, *TBL1XR1* |

**Table 1.** List of example genes.

Compared to nodal DLBCLs, the mutation landscape of PCNSLs of ABC and GC origin do not show considerable differences [Fukumura et al., 2016; Kraan et al., 2013; Yamada et al., 2015; Zhou et al., 2018] and treatment of the disease remains a significant clinical challenge. To overcome these difficulties NanoString LST-assay was used to precisely determine molecular subgroups of a large cohort of PCNSL and complementary targeted mutation profiling was applied to identify key genetic alterations on a subset of the patients.

*DNA methylation profiling*

***CASE STUDY III.: GBM***

Glioblastoma multiforme (GBM), also referred to as a grade IV astrocytoma, is one of the most aggressive and fast-growing brain tumors. It invades the nearby brain tissue, but generally does not spread to distant organs. Its exhibiting great variability at histopathological and molecular levels. The disease development is related to the accumulation of various types of mutations (e.g., somatic genomic rearrangements, SNVs, CNVs), accompanied by changes in epigenomic and gene expression profiles as well. Several studies showed genomic and transcriptomic characteristics of GBM [Brennan et al., 2013; Cancer Genome Atlas Research Network, 2008; Kim et al., 2015a, b; Patel et al., 2014; Sottoriva et al., 2013; Verhaak et al., 2010; Wang et al., 2016, 2017].

Nowadays, the background of GBM development is well described and the disease is divided into subgroups based on transcriptional and epigenomic profiles [Brennan et al., 2013; Cancer Genome Atlas Research Network, 2008; Noushmehr et al., 2010; Verhaak et al., 2010]. However, most studies involved cross-sectional cohorts, since the collection of sequential samples is difficult because of the aggressive progression of GBM. Investigate the methylome is an alternative to mRNA expression profiling in FFPE GBM specimens. The first comprehensive epigenomic analysis was reported by Noushmehr et al., 2010, followed by several ones [de Souza et al., 2018; Hu et al. 2016; Klughammer et al., 2018; Nagarajan et al., 2014;]. The early epigenomic studies determined levels of CpG methylation applying various methods but the evaluation of the results encountered difficulties [Hegi et al., 2005; Noushmehr et al., 2010]. These types of surveys became more feasible thanks to the recent availability of the reduced representation bisulfite sequencing (RRBS) method. Klughammer et al., (2018) reported single-CpG and single allele methylation profiles in the context of multidimensional clinical and molecular data applying RRBS. The involvement of the Wnt pathways in both cross-sectional and sequential FFPE GBM was previously reported [Tompa et al., 2018]. To further explore mechanisms of GBM, distribution of differentially methylated DNA CpG regions and pathways in 22 pairs of sequential FFPE GBM specimens were analysed.

*miRNome profiling*

***CASE STUDY IV.: NSCLC***

Lung cancer is one of the most common causes of cancer-related deaths worldwide. Non-small cell lung cancer (NSCLC) accounts for 85% of all lung cancers and represents a heterogenous group of malignancies comprised mainly of adenocarcinomas (ACs) and squamous cell carcinomas (SCCs) [Herbst et al., 2018; PDQ Adult Treatment Editorial Board, 2022; Siegel et al., 2019]. Recently, a numerous novel targeted therapy has been established as treatment options [Donington et al., 2011; Ettinger et al., 2017, 2022; Hirsch et al., 2017]. However, despite significant therapeutic progress, novel targeted anti-cancer drugs used in different NSCLC subtypes presented with differential levels of efficacy [Ferrara et al., 2021; Pasquali et al., 2018 Rekulapelli et al., 2022; Zhu et al., 2017]. Targeted therapies directed against specific cellular alterations were found most successful in patients with non-squamous tumors. Diagnosis of lung cancer based on histopathological analysis remains the gold standard, this method has several important limitations [IJC]. Recent advances in personalized targeted lung cancer therapies require not only accurate histological classification of NSCLC but need to be extended by a precise characterization of its molecular background [Gou et al., 2018; Schipper et al., 2022; Pilotto et al., 2015; Zhang et al., 2022].

miRNAs (microRNAs) constitute a group of endogenous short non-protein coding RNAs that regulate gene expression by degrading mRNA or by incomplete binding to a complementary sequence of a target mRNA. Recent studies demonstrated that aberrations in the profile of miRNAs expression can play a crucial role in carcinogenesis and progression of many human tumor types, including lung cancer [Chaturvedi and Som 2022; Rajakumar et al., 2022; Yan et al., 2022]. On the other hand, microRNAs have emerged as promising biomarkers that can contribute to more effective early detection of asymptomatic lung cancer and better prognostication of both disease course and efficacy of molecularly targeted therapies [Hua et al., 2022; Liang et al., 2022]. The fact miRNAs exhibit high tissue specificity, that is most likely associated with their significant role in the regulation of cell differentiation [Ghafouri-Fard et al., 2020; Goradel et al., 2019; Wang et al., 2020], become a base for numerous studies searching the use of miRNAs as putative molecular markers. Reports indicated that the analysis of global miRNome in early-stage NSCLC patients could become more useful tool

allowing for development of more accurate histotypic-associated markers to distinguish lung SCC from AC subtypes.

To precisely characterize the histopathological and molecular features of NSCLC, for more accurate identification of those patients that could benefit from novel molecular-targeted therapies, advanced omics technologies were applied for global miRNA expression profiling and biomarker research in a large and well-characterized group of patients with completely resected fresh-frozen early-stage lung tumours and blood samples.

*Assisted reproduction*

***CASE STUDY V.: NIPGT-A***

Embry selection clinical guidelines used today in *in vitro* fertilization (IVF) treatments are relied on non-invasive embryo morphology assessment. The grading criteria for standardized applicable oocyte and embryo assessment was latest updated in 2011 [Alpha Scientists in Reproductive Medicine and ESHRE Special Interest Group of Embryology, 2011]. In accordance with the development of morphological assessment [Gardner et al., 2015; Paternot et al., 2013] the number of non-invasive methods, based on the detection of molecular markers present in the spent culture media (SCM) of the embryo, are increasing.

There are several methods to assess embryo quality, ploidy and viability or monitor the catabolic activity in a less harmful procedure [Butler et al., 2013; Devreker et al., 2000; Gardner et al., 2001; Katz-Jaffe el al., 2006; Mains et al., 2011; Montskó et al., 20144-10]. Continuous development of these analytical techniques like ESI-MS fingerprinting, Nano-UHPLC MS/MS, MALDI-TOF, immunoassays, microarray and NGS approaches offer exceptional non-invasive way to profile the embryo from the SCM [Cortezzi et al., 2013; Hernandez-Vargas et al., 2020]. The use of minimal- or non-invasive methods have a major impact on the genetic composition assessment of the developing embryo. Pre-implantation genetic testing for aneuploidy (PGT-A) is integrated into many IVF programmes to achieve improvements in success outcome [Wells et al., 2010] emerging perspectives the need of non-invasive pre-implantation genetic testing for aneuploidy (NIPGT-A) [Huang et al., 2019, Shitara et al., 2021]. The growing scientific evidence emphasises the clinical applicability of SCM in NIPGT-A and the concordance of NIPGT-A with inner cell mass (ICM) or blastoderm biopsies [Handyside et al., 2016; Huang et al., 2019; Kuznyetsov et al., 2018; Rubio et al., 2020; Shitara et al., 2021;

Vera-Rodriguez et al., 2018]. NIPGT-A may have the potential to superior to TE biopsy for aneuploidy screening [Huang et al., 2019], but the major pitfall is that there are many well-defined sources of DNA contamination (e.g., polar bodies, cumulus cells, external fragmented DNA). These contamination mechanisms have been observed by independent studies [Huang et al., 2019; Rubio et al., 2020; Shitara et al., 2021; Vera-Rodriguez et al., 2018] and described as the key limitations of the method.

To address the above-named crucial limitations, the aim of the current case study was to develop a workflow based on NGS and the corresponding bioinformatic pipeline. During the workflow implementation, particular emphasis was placed to minimising the noise effect of the DNA contamination. The proposed methodology was tested on SMC droplets of morphologically good quality embryos to avoid false positive results from disproportionate embryonic cell divisions.

# Material and Methods

*Pipeline building*

For each pipeline, that was implemented within this thesis, the following strategy (Figure 1.) was applied to build standard and reproducible workflows that can be run in various environments (e.g., local, cloud or HPC). In the view of the known biological problem best tools were selected based on a literature search. Based on the chosen tool's manual potential useful parameters were selected in advance (before compiling it). Since, the installation of a software is not always straightforward and time-consuming Docker/Singularity containers [Merkel, 2014; Kurtzer et al., 2017] or conda [https://docs.anaconda.com/] predefined environments were preferred during the pipeline assembly. First, tools were tested separately using example dataset, if it was available to the tool, otherwise a small dataset was prepared manually. This test phase was important to define a general parameter set, input files and file types and debug the proper commands which will be run during the analysis.

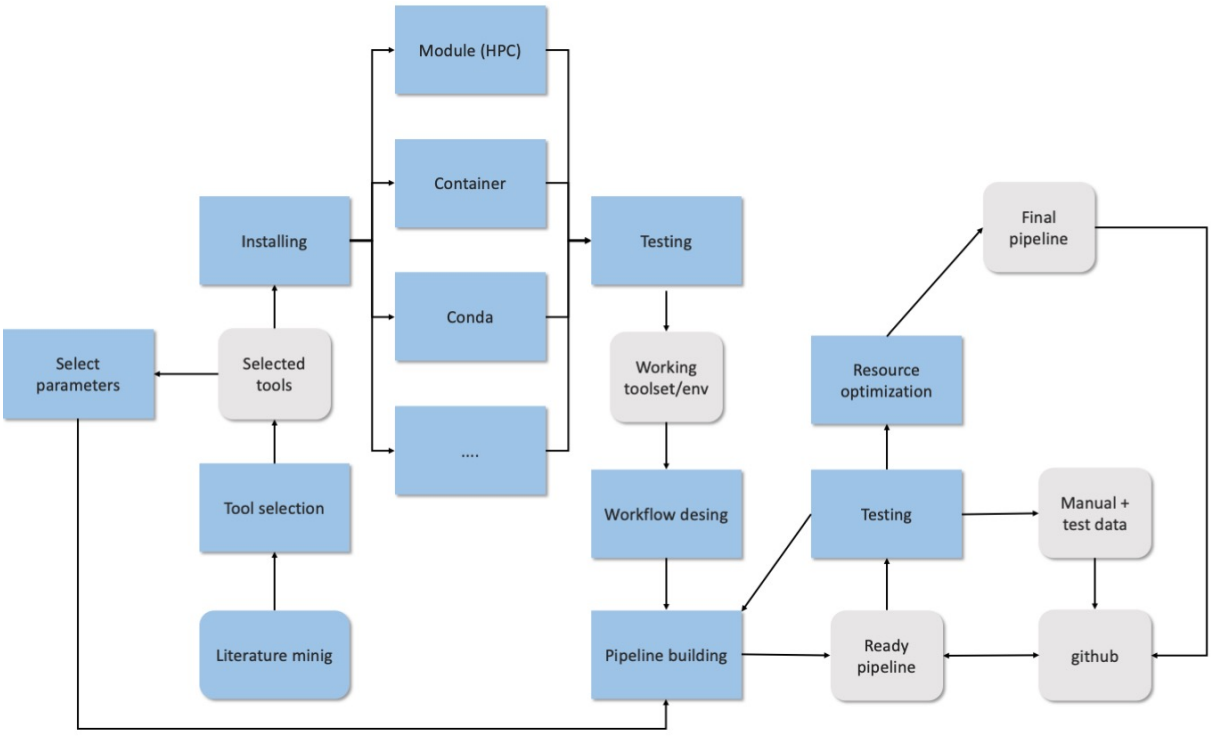



**Figure 1.** Sematic workflow of the pipeline building strategy.

In the next phase the actual workflow was designed using Nextflow workflow management system [Di Tommaso et al., 2017]. Nextflow uses Docker or Singularity technology for the multi-scale handling of containerized computation. It is designed to

address numerical instability, efficient parallel execution, error tolerance, execution provenance and traceability. It is a domain-specific language that enables rapid pipeline development through the adaptation of existing pipelines written in any scripting language. Nextflow enables users to run any current or previous version of a pipeline for any published and properly deposited analyses. The GitHub integration allows consistent tracking of software changes and versions, the containerization ensures numerical stability, and the cloud support provides rapid computation and effective scaling. Nextflow uses a functional reactive programming model in which each operation (typically a workflow task) is isolated in its own execution context. Outputs from one operation in during the run are streamed to other operations by dedicated channels in a process like UNIX pipes. Parallelization is an implicit consequence of the way inputs and outputs of each process are channelled into other processes. This approach spares users the need to implement an explicit parallelization strategy. After the first version of the pipeline, it was tested on an HPC platform using example and real data as well to fix bugs, finalize the parameters, inputs/outputs and utilized computational recourses for each process in the workflow. These instructions were stored in a separate config file that Nextflow can take as an input. An example of a nextflow script and the corresponding config file are presented on Figure 2.

A small readme file or manual was prepared about how to run the pipeline mentioning the crucial input parameters. Finally, if it was necessary the implemented pipeline and all the relevant files were uploaded to a git [Chacon and Straub, 2014] repository. For each case study the unique elements of the developed pipeline are described in the “Bioinformatics workflow” section.

**(a)**

```
/*
 * Define output folders
 */

fastqc_path = "FastQC_raw"
fastq_path = "Filtered_data"
filtered_path = "FastQC_filt"
metrics_path = "Metrics"
bamqc_path = "BamQC"
vars_path = "Mutect2_variants"
snpeff_path = "SnpEff"
anno_path = "Annovar"
multiqc_path = "MultiQC"
bam_path = "BAM_files"
stats_path = "Reports"
md5_path = "md5SUMs"

log.info """\

============================================

T P 5 3  S N P / I N D E L  -  G A T K  M U T E C T 2  P I P E L I N E  v 2.0

============================================

 input         : ${params.input}
 genome        : ${params.fasta}
 bwa index     : ${params.bwa}
 target        : ${params.target}
 know variants : ${params.vcf}
 dbsnp         : ${params.dbsnp}

============================================

"""

/*
 *  Parse the input parameters
 */

fasta_file = file(params.fasta)
bwa_index = file(params.bwa)
vcf_file = file(params.vcf)
idx_file = file(params.idx)
dbsnp_file = file(params.dbsnp)
target_file = file(params.target)
anno_file = file(params.annodb)
intervals = file(params.list)
bedpe_file = file(params.bedpe)
primers_file = file(params.primers)
amplicons_file = file(params.amplicons)
Channel
.fromFilePairs(params.input, flat: true)
.into{ data; fastqc; md5 }

/*
 * Process 1a: Run FastQC
 */

process '1a_run_fastqc'{
    publishDir "$fastqc_path", mode:'copy'

    input:
        set val(id), file(fastqc1), file(fastqc2) from fastqc

    output:
        file("*.zip") into fastqc_mqc
        file "*.html"
        file "*.zip"

    script:
        """
        module load fastqc/0.11.9

        export _JAVA_OPTIONS=-Xmx2048m

        fastqc -t ${task.cpus} ${fastqc1} ${fastqc2}
        """
}

/*
 * Process 1b: Run ptrimmer
 */

process '1b_ptrimmer'{
    publishDir "$fastq_path", pattern: '*.txt', mode:'copy'

    input:
        file(Amp) from amplicons_file
        set val(id), file(fastq1), file(fastq2) from data

    output:
        set val(id), file("${id}_cut_1.fq"), file("${id}_cut_2.fq") into (fastp_ch)
        file "*.txt"

    script:
        """
        module load pTrimmer/1.3.4

        pTrimmer-1.3.4 -t pair -a ${Amp} -f ${fastq2} -r ${fastq2} -d ${id}_cut_1.fq -e ${id}_cut_2.fq -s
{id}.ptrimmer.amps.txt

        cp .command.log ${id}.summary.txt
        """
}
```

**(b)**

```
/* Define the target files
 * and other params
 */

env.NXF_SINGULARITY_CACHEDIR = "/lustre01/singularity/"

params.fasta = "/lustre01/references/Genomes/Homo_sapiens/Ensembl/GRCh37/Sequence/WholeGenomeFasta/"
params.vcf = "/lustre01/references/Known_variants/dbSNP/dbsnp_151.hg19.vcf"
params.dbsnp = "/lustre01/references/Known_variants/dbSNP/dbsnp_151.hg19.vcf"
params.bwa = "/lustre01/references/Genomes/Homo_sapiens/Ensembl/GRCh37/Sequence/BWAIndex/"
params.target = "/lustre01/nextflow_scripts/TP53/TP53.bed"
params.run = null
params.input = null

process.executor = 'slurm'

process {
        withName:'1a_run_fastqc|1d_make_fastqc_report' {
        memory = 4.GB
        cpus = 2
    }
        withName:'1b_ptrimmer' {
        cpus = 2
    }
```

**Figure 2.** Partial example of a Nextflow script (a) including the first 2 steps, log section, general input parameters and (b) the corresponding config file.

*Somatic mutation profiling*

***CASE STUDY I.: CLL***

Patient samples

Table 2. summarize clinical characteristics of the samples that were added in this study. The cohort (20 consecutive patients, 12 males and 8 females) represented a pre-treated patient group with a median age of 63 and 2 (range: 1-5) lines of prior therapies. These patients were treated with ibrutinib via a case-by-case individual application process available in Hungary since July 2014, with full support by the National Health Insurance Fund from 2017.

| Patient ID | Age (years) | Gender | Prior therapy | Pre-ibrutinib FISH cytogenetics | IGHV (M, U) | CD38 expression | Binet stage | Number of prior therapies | TTFT (months) | Time from diagnosis to ibrutinib (months) | Duration of ibrutinib till NGS analysis (months) | Additional follow-up after the NGS analysis (months) |
|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1 | 60 | F | FC, ofatumumab, RFC, R-CHOP, RB | 12+/IgHdel | U | positive | B | 5 | 48 | 126 | 34 | 6 |
| 2 | 65 | M | RFC, R-CVP, RB, | 17p-/13q- | U | positive | C | 3 | 60 | 115 | 24 | 2 |
| 3 | 64 | M | FC, alemtuzumab, allo-SCT | 11q-/13q- | U | positive | B | 2 | 13 | 117 | 18 | 28 |
| 4 | 58 | F | RFC, RB | 11q-/6q-/13q- | M | positive | B | 2 | 3 | 39 | 12 | 28 |
| 5 | 63 | M | FC, R-CVP, RFC, RB | normal | U | positive | C | 4 | 6 | 103 | 32 | 12 |
| 6 | 85 | M | FC, RB, R-CVP, HDMP+R+RM | 13q- | M | positive | B | 4 | 51 | 119 | 29 | 0 |
| 7 | 48 | M | RFC | 11q-/13q- | U | positive | C | 1 | 36 | 58 | 27 | 13 |
| 8 | 50 | F | RFC, RB | 13q-/17p- | U | positive | C | 2 | 72 | 149 | 8 | 28 |
| 9 | 63 | F | RFC, R-CHOP, R-CNOP | 12+ | U | positive | C | 3 | 10 | 67 | 3 | 0 |
| 10 | 51 | M | RFC, R-CVP, RB | 13q- | U | positive | C | 3 | 1 | 41 | 21 | 0 |
| 11 | 65 | F | CVP, R-CVP, RB | 12+ | U | positive | C | 3 | 6 | 96 | 24 | 13 |
| 12 | 63 | M | RFC, R-CVP, RB | 13q- | U | positive | B | 3 | 50 | 99 | 26 | 12 |
| 13 | 69 | F | Chl, Cyc, RB | 17p-/12+ | U | positive | B | 2 | 15 | 86 | 8 | 27 |
| 14 | 66 | F | RFC, R-Chl | normal | U | positive | C | 2 | 20 | 53 | 28 | 13 |
| 15 | 63 | M | FCM, RB | 12+ | U | positive | B | 2 | 36 | 137 | 24 | 15 |
| 16 | 51 | M | FC, RB | 13q- | U | negative | C | 2 | 0,5 | 82 | 30 | 13 |
| 17 | 73 | F | RFC, RB, R-Chl, R-CVP, R-CHOP | 17p-/12+ | M | positive | C | 5 | 19 | 75 | 18 | 0 |
| 18 | 65 | M | RFC, RB | 17p- | U | positive | C | 2 | 0 | 19 | 3 | 29 |
| 19 | 64 | M | RFC,HDMP | 17p- | U | negative | B | 2 | 34 | 36 | 8 | 0 |
| 20 | 56 | M | RFC | 17p-/13q- | U | negative | C | 1 | 83 | 132 | 17 | 15 |

**Table 2.** Cohort characteristics. Allo-SCT: allogeneic stem cell transplantation; Chl: chlorambucil; Cyc: Cyclophosphamide F: female; FC: fludarabine, cyclophosphamide, FCM: fludarabine, cyclophosphamide, mitoxantrone; FISH: fluorescence in situ hybridization; HDMP: high dose methylprednisolone; M: Male; M: mutated; R: rituximab; RB: 24ituximab, bendamustin; R-Chl: rituximab, chlorambucile; R-CHOP: rituximab, cyclophosphamide, doxorubicin, vincristine and prednisolone; R-CNOP: rituximab, cyclophosphamide, mitoxantrone, vincristine, prednisone; R-CVP: rituximab, cyclophosphamide, vincristine, prednisolone; RFC: rituximab, fludarabine, cyclophosphamide; RM: ribomustin; U: unmutated [Gángó et al., 2019].

Pre-treatment peripheral blood mononuclear cells (PBMC) were available from all patients, with corresponding post-treatment samples as shown in Figure 3. The *IGHV* mutation conditions were determined according to the European Research Initiative on CLL recommendations [Rosenquist et al., 2017] (e.g., 13q, 11q and 17p deletions and trisomy 12) and analysed by interphase fluorescence *in situ* hybridization applying Vysis probe sets (Abbott Molecular, Lake Bluff, USA). The CLL cells rate in the samples was assessed by flow-cytometry using CD5/CD19/CD23/CD45 staining. Healthy volunteer set (n = 5) was used as negative controls. Written informed consent from all patients was obtained for the study which was conducted in accordance with the Declaration of Helsinki and approved by the Hungarian Medical Research Council.

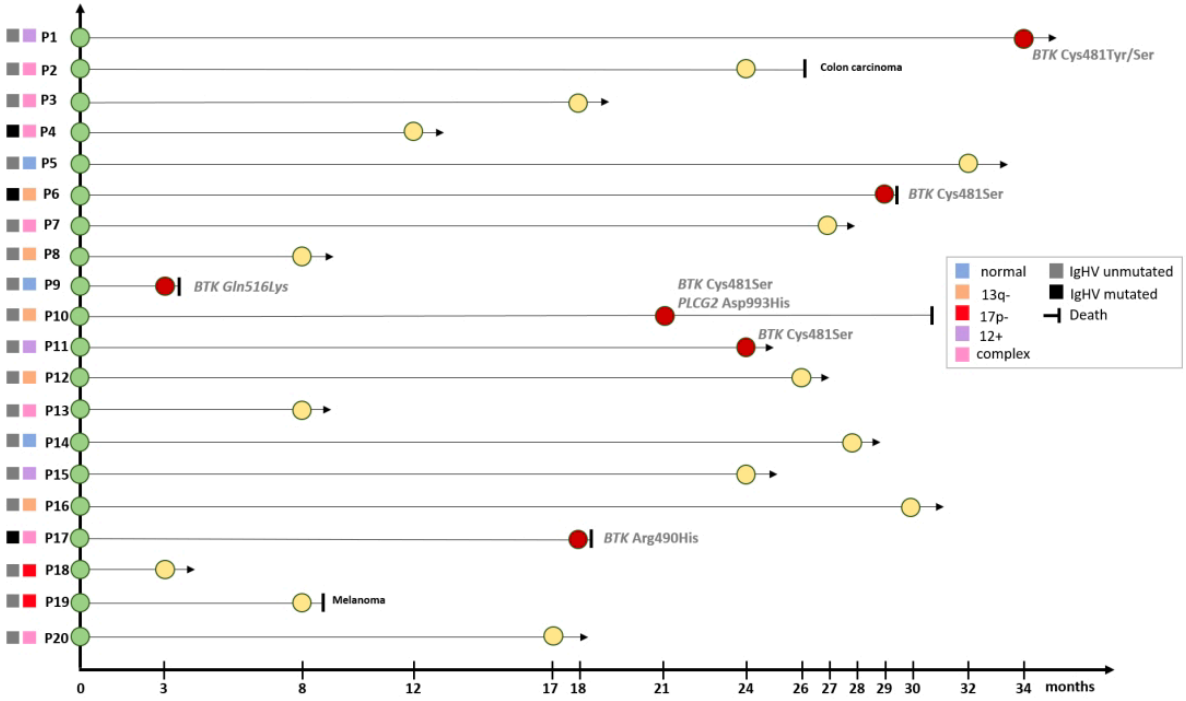



**Figure 3.** Timeline and basic cytogenetic features of the cohort (n = 20, P1-20) treated with ibrutinib. Red circles denote patients who progressed on ibrutinib with a *BTK* or *PLCG2* mutation as determined by NGS. Coloured squares indicating the *IGHV* status [Gángó et al., 2019].

## Customized amplicon sequencing

Targeted NGS analysis of recurrently mutated genes listen in Table 3 with a published frequency of ≥2% [Baliakas et al., 2015; Puente et al., 2011; Quesada et al., 2011; Schuh et al., 2012; Wang et al., 2011] was performed by the TruSeq Custom Amplicon approach (Illumina, Inc., San Diego, USA) with maximum input of gDNA samples extracted from PBMCs. After quality control (QC) and equimolar pooling of samples, libraries were

sequenced on a HiSeq 4000 Instrument using 150bp paired-end chemistry (PE150). The variant allele frequencies (VAF) were normalized according to the proportion of CLL cells. During the sequencing, the median follow-up time was 22.5 months (range: 3-34 months).

| Gene | Analysed region | Transcript ID |
|---|---|---|
| *ATM* | whole CDS | ENST00000278616 |
| *BCOR* | Exons 4, 8-10, 12, 13 | ENST00000378444 |
| *BIRC3* | Exons 2, 6-9 | ENST00000615299 |
| *BRAF* | Exons 11-15 | ENST00000288602 |
| *BTK* | whole CDS | ENST00000308731 |
| *CHD2* | Exons 2-3, 8, 10, 13, 16, 17, 20-21, 25-27, 29-32, 35, 36 | ENST00000394196 |
| *DDX3X* | Exons 2, 3, 6, 7, 9-13 | ENST00000399959 |
| *EGR2* | Exons 1, 2 | ENST00000242480 |
| *EIF2A* | Exon 10 | ENST00000460851 |
| *EP300* | Exons 20, 28 | ENST00000263253 |
| *FBXW7* | Exons 5, 7, 9, 10, 11 | ENST00000281708 |
| *HIST1H1E* | Exon 1 | ENST00000304218 |
| *IGLL5* | whole CDS | ENST00000532223 |
| *KLHL6* | Exons 1, 2 | ENST00000341319 |
| *KMT2D* | Exon 39 | ENST00000301067 |
| *LRP1B* | Exons 7, 13, 32, 41 | ENST00000389484 |
| *MED12* | Exons 1, 2, 21 | ENST00000374080 |
| *MGA* | Exons 2, 3, 8, 11, 13, 15 | ENST00000219905 |
| *MYD88* | whole CDS | ENST00000396334 |
| *NFKBIE* | Exons 1, 2 | ENST00000275015 |
| *NOTCH1* | Exons 2, 4, 6, 7, 10, 13, 15, 17, 21, 22, 26, 31, 34 | ENST00000277541 |
| *PLCG2* | whole CDS | ENST00000564138 |
| *POT1* | Exons 5-10, 18 | ENST00000357628 |
| *RIPK1* | Exons 8, 10 | ENST00000380409 |
| *RPS15* | whole CDS | ENST00000592588 |
| *SAMHD1* | Exons 2-4, 6-13 | ENST00000262878 |
| *SF3B1* | Exons 14-18 | ENST00000335508 |
| *TP53* | whole CDS | ENST00000269305 |
| *XPO1* | Exons 2, 16, 20 | ENST00000401558 |
| *ZMYM3* | Exons 6, 15, 16, 20, 23, 24 | ENST00000314425 |

**Table 3.** Targeted genes and regions including Ensembl transcript IDs [Gángó et al., 2019].

Bioinformatics workflow

Applied bioinformatic pipeline is presented on Figure 4. As data preprocessing, filtered sequencing reads were mapped to the Ensembl Homo sapiens hg19/GRCh37 genome build using BWA v0.7.13 aligner [Li et al., 2010]. BAM files were sorted and indexed by SAMtools v1.7 [Danecek et al., 2021], GATK v4.0 BSQR tool [DePristo et al., 2011] was run on each sample to detect and correct systematic sequencing errors. SNV calling was performed with LoFreq v2.1 variant-caller [Wilm et al., 2012] that considers all dataset features (e.g., base-call qualities, mapping problems or base/INDEL misalignments) that are commonly ignored by other methods or only used just for filtering. Built in p-value calculation for each detected mutation granted a strict control of false positive findings. Raw variants were functionally, and database annotated using SnpEff v4.3i [Cingolani et al., 2012] and ANNOVAR v2017Jul17 tools [Wang et al., 2021], including up-to-date information from COSMIC, avSNP and CLINVAR databases. Variants in the TP53 coding region were additionally annotated using the ANNOVAR preformatted Seshat and IARC databases [Bouaoun et al., 2016; Tikkanen et al., 2018]. The raw sequencing data was uploaded to the European Nucleotide Archive (https://www.ebi.ac.uk/ena, Primary Accession: PRJEB32120, Secondary Accession: ERP114759).

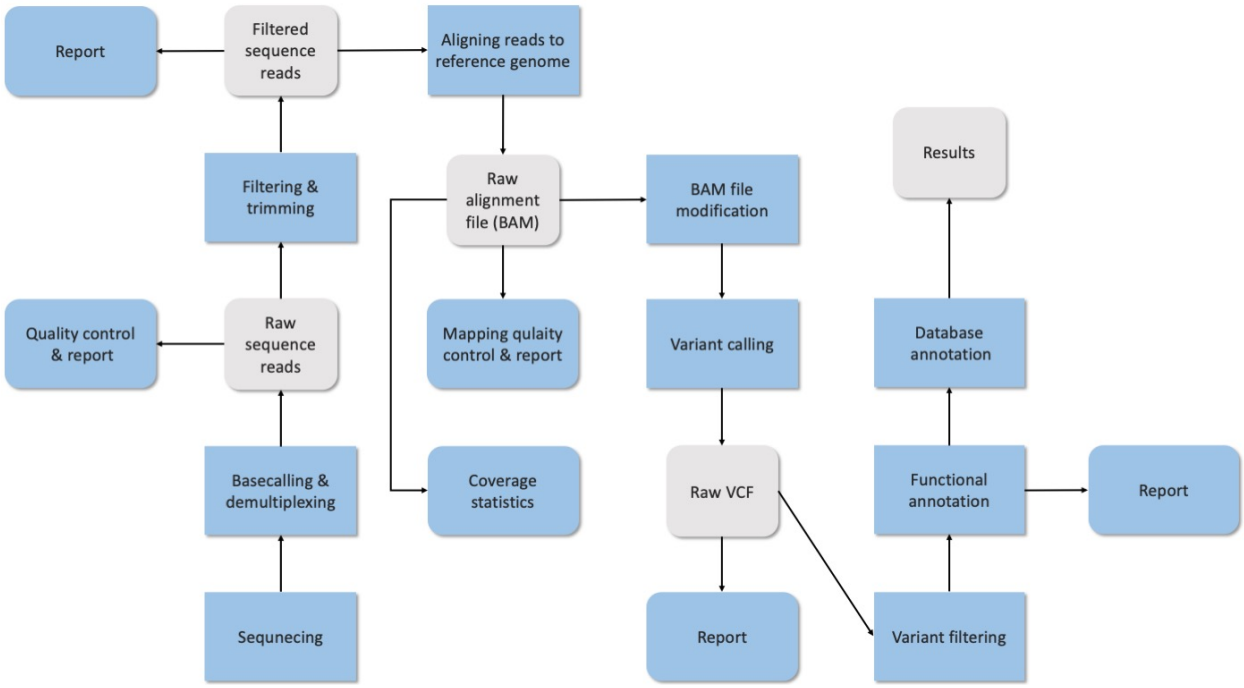



**Figure 4.** Used bioinformatic workflow (from raw data to annotated VCF file) including main steps and checkpoints.

Validation of somatic variants

Bidirectional Sanger sequencing was performed to validate all somatic variants with a VAF of >20% listed in Table 4.

| Patient number – month on ibrutinib | Gene | cDNA position | Amino acid | VAF (%) |
|---|---|---|---|---|
| P1 – M0 | *CHD2* | c.3349A>G | p.Arg1117Gly | 29.02 |
| P1 – M0 | *KMT2D* | c.12144_12145insGGGGCCG | p.Leu4049fs | 26.98 |
| P2 – M0 | *TP53* | c.376-2dupA | | 42.86 |
| P2 – M0 | *TP53* | c.742C>T | p.Arg248Trp | 21.19 |
| P2 – M24 | *TP53* | c.841G>A | p.Asp281Asn | 52.89 |
| P3 – M0 | *ATM* | c.8187A>T | p.Gln2729His | 66.26 |
| P3 – M0 | *DDX3X* | c.669_670dupTG | p.Ala224fs | 47.89 |
| P3 – M18 | *ATM* | c.6067G>A | p.Gly2023Arg | 22.16 |
| P3 – M18 | *ATM* | c.8187A>T | p.Gln2729His | 70.96 |
| P4 – M0 | *POT1* | c.284G>A | p.Gly95Asp | 43.01 |
| P4 – M12 | *POT1* | c.284G>A | p.Gly95Asp | 20.34 |
| P7 – M0 | *EGR2* | c.1066G>A | p.Glu356Lys | 41.86 |
| P7 – M0 | *POT1* | c.185T>C | p.Phe62Ser | 44.50 |
| P9 – M3 | *MED12* | c.97G>A | p.Glu33Lys | 41.68 |
| P10 – M0 | *ATM* | c.7177T>G | p.Phe2393Val | 34.24 |
| P10 – M0 | *NOTCH1* | c.7225C>T | p.Gln2409* | 32.92 |
| P10 – M21 | *ATM* | c.7177T>G | p.Phe2393Val | 27.62 |
| P10 – M21 | *BTK* | c.1442G>C | p.Cys481Ser | 21.59 |
| P10 – M21 | *NOTCH1* | c.7225C>T | p.Gln2409* | 25.62 |
| P13 – M0 | *EGR2* | c.1150C>A | p.His384Asn | 20.71 |
| P13 – M8 | *SF3B1* | c.2098A>G | p.Lys700Glu | 37.84 |
| P14 – M0 | *NOTCH1* | c.7541_7542delCT | p.Pro2514fs | 41.66 |
| P14 – M0 | *SAMHD1* | c.1015C>T | p.Arg339Cys | 84.88 |
| P14 – M28 | *NOTCH1* | c.7541_7542delCT | p.Pro2514fs | 33.55 |
| P14 – M28 | *SAMHD1* | c.1015C>T | p.Arg339Cys | 75.29 |
| P15 – M0 | *BRAF* | c.1801A>G | p.Lys601Glu | 21.23 |
| P18 – M0 | *TP53* | c.803A>T | p.Asn268Ile | 81.12 |
| P18 – M3 | *TP53* | c.803A>T | p.Asn268Ile | 87.83 |
| P20 – M0 | *NOTCH1* | c.7516G>T | p.Glu2506* | 45.48 |
| P20 – M0 | *TP53* | c.314G>T | p.Gly105Val | 24.73 |
| P20 – M17 | *NOTCH1* | c.7516G>T | p.Glu2506* | 30.75 |

**Table 4.** Mutation validated by Sanger sequencing [Gángó et al., 2019].

The *BTK* Cys481Ser and *PLCG2* Asp993His mutations were validated by droplet digital PCR (ddPCR). PCR Reactions were performed with 50 ng input DNA using locus-specific assays for the wild type and mutant targets (Table 5) following the manufacturer's protocol. Droplets were created by the QX200 Automated Droplet Generator and reading was completed with the QX200 ddPCR system (Bio-Rad, Hercules, CA, USA). Results were analysed using the Bio-Rad QuantaSoft software. The *BTK/PLCG2* mutation allelic burden was determined using the following equation FA = a/(a+b) (FA; fractional abundance, a; No. of mutant molecules, b; No. wild type molecules).

| **Target gene/mutation** | **Assay Name** | **Assay ID** |
|---|---|---|
| *PLCG2* Asp993His | PLCG2_G>C,D/H | dHsaMDS815970714 |
| *BTK* Cys481Ser | BTK_G>C,C/S | dsHsaMDS802598840 |
| | Sequence | |
| CGACCTCCTGAAGTACAATCAAAAGGGCCTGACCCGCGTCTACCCAAAGGGACAAAGAGTT[G/C]ACTCT TCAAACTACGACCCCTTCCGCCTCTGGCTGTGCGGTTCTCAGATGGTGGCACTCAA | | |
| ACATCTCTAGCAGCTGCTGAGTCTGGAAGCGGTGGCGCATCTCCCTCAGGTAGTTCAGGAG[G/C]CAGCC ATTGGCCATGTACTCAGTGATGATGAAGATGGGGCGCTGCTTGGTGCAGACGCCAT | | |

**Table 5.** Sequence of the digital droplet PCR assays [Gángó et al., 2019].

***CASE STUDY II.: PCNSL***

Patient samples

FFPE tissue samples from patients, included in the study, were obtained from the following centres:

(1) First Department of Pathology and Experimental Cancer Research, Semmelweis University, Budapest, Hungary

(2) Department of Pathology, University of Pécs, Pécs, Hungary

(3) Division of Neuropathology, The National Hospital for Neurology and Neurosurgery, University College London Hospitals, United Kingdom, through the UK Brain Archive Information Network (BRAIN UK)

Permissions to use the archived tissue have been obtained from the Local Ethical Committee (TUKEB-1552012) and from BRAIN UK (Ref.: 16/018), and the study was conducted in accordance with the Declaration of Helsinki.

Clinical data of the cohort (PCNSL n = 81, SCNSL n = 18) on the molecular subtype as determined by IHC during the routine diagnostic workup are summarized Table 6. Survival data were available for 65 PCNSL and 17 SCNSL cases, while treatment data available in 46 PCNSL and 12 SCNSL cases, respectively.

| Case No | Sex | Age at diagnosis (years) | OS (months) | Event | Molecular subtype IHC | Therapy |
|---|---|---|---|---|---|---|
| P1 | M | 75 | 4.6 | exit | ABC | NA |
| P2 | F | 64 | 3.9 | exit | ABC | MTX/Ara-C |
| P3 | F | 60 | 9.0 | alive | GC | MATRix |
| P4 | F | 73 | 2.7 | exit | ABC | Steroid |
| P5 | F | 77 | 7.5 | exit | ABC | MTX |
| P6 | M | 53 | 19.0 | exit | NA | MTX/Ara-C, RT, CEPP, VIM, Tem |
| P7 | F | 66 | 7.8 | exit | ABC | MTX |
| P8 | F | 59 | 2.1 | exit | ABC | MTX, WBRT |
| P9 | F | 72 | NA | NA | ABC | NA |
| P10 | F | 80 | NA | NA | ABC | NA |
| P11 | F | 51 | NA | NA | ABC | NA |
| P12 | M | 67 | 56.0 | alive | ABC | MTX/Ara-C/R, WBRT |
| P13 | F | 77 | 3.0 | exit | ABC | NA |
| P14 | M | 64 | 4.1 | exit | ABC | MTX/Ara-C/R, Steroid |
| P15 | M | 73 | 16.5 | exit | ABC | MTX/R, Steroid |
| P16 | M | 59 | 11.0 | exit | ABC | NA |
| P17 | M | 56 | 3.5 | alive | ABC | NA |
| P18 | F | 55 | 17.2 | exit | ABC | MTX/Ara-C, Steroid |
| P19 | F | 92 | 1.8 | exit | ABC | NA |
| P20 | M | 77 | 5.3 | exit | ABC | MTX |
| P21 | F | 54 | NA | NA | ABC | MTX/Ara-C |
| P22 | M | 67 | 39.5 | alive | ABC | MTX, CEPP |
| P23 | F | 78 | NA | NA | ABC | MTX |
| P24 | M | 72 | 5.7 | exit | ABC | NA |
| P25 | F | 47 | NA | NA | ABC | NA |
| P26 | F | 43 | 11.5 | exit | GC | NA |
| P27 | F | 78 | 23.0 | alive | ABC | MTX/Ara-C |
| P28 | F | 78 | 22.7 | alive | ABC | MTX, RT |
| P29 | F | 62 | 7.9 | alive | ABC | MTX/Ara-C/R |
| P30 | F | 70 | 15.5 | alive | ABC | MTX, RT |
| P31 | F | 59 | NA | NA | ABC | NA |
| P32 | F | 70 | 13.7 | alive | ABC | MTX |
| P33 | F | 43 | 12.0 | exit | ABC | MTX/Ara-C |
| P34 | M | 46 | 68.0 | exit | ABC | MTX/Ara-C/R, RT, Steroid |
| P35 | F | 68 | 6.9 | alive | ABC | MTX/Ara-C |
| P36 | M | 72 | NA | NA | ABC | NA |
| P37 | M | 71 | 0.5 | exit | ABC | MTX |
| P38 | M | 76 | 0.9 | exit | ABC | palliative |

| Case No | Sex | Age at diagnosis (years) | OS (months) | Event | Molecular subtype IHC | Therapy |
|---|---|---|---|---|---|---|
| P39 | M | 70 | 34.6 | exit | ABC | NA |
| P40 | F | 51 | 52.1 | alive | ABC | MTX/Ara-C |
| P41 | M | 59 | 88.2 | alive | ABC | NA |
| P42 | M | 70 | NA | NA | ABC | NA |
| P43 | F | 59 | 14.0 | exit | ABC | IDARAM, MTX/Ara-C/R, WBRT |
| P44 | F | 70 | 0.5 | exit | ABC | NA |
| P45 | F | 51 | 27.0 | alive | ABC | MTX/Ara-C |
| P46 | M | 66 | 2.2 | exit | ABC | MTX, CEPP |
| P47 | F | 81 | 0.01 | exit | ABC | NA |
| P48 | F | 65 | 2.1 | exit | ABC | MTX |
| P49 | M | 35 | NA | NA | ABC | NA |
| P50 | M | 68 | 8.9 | exit | NA | NA |
| P51 | M | 73 | 27.3 | alive | GC | MTX/Ara-C, R-IE |
| P52 | M | 61 | NA | NA | ABC | NA |
| P53 | F | 77 | NA | NA | ABC | NA |
| P54 | M | 58 | NA | NA | ABC | NA |
| P55 | M | 76 | 1.0 | exit | ABC | NA |
| P56 | F | 63 | 30.8 | exit | ABC | MTX |
| P57 | M | 63 | 54.9 | exit | ABC | MTX/Ara-C, RT, Steroid |
| P58 | F | 64 | 0.5 | exit | ABC | NA |
| P59 | M | 66 | 13.9 | exit | ABC | MTX |
| P60 | F | 82 | 3.5 | exit | ABC | NA |
| P61 | F | 64 | 2.1 | exit | ABC | NA |
| P62 | M | 71 | 6.3 | exit | ABC | NA |
| P63 | M | 56 | NA | NA | ABC | NA |
| P64 | F | 66 | NA | NA | ABC | NA |
| P65 | F | 85 | 0.9 | exit | ABC | NA |
| P66 | M | 75 | 0.5 | exit | ABC | MTX |
| P67 | F | 68 | 86.4 | alive | ABC | MTX |
| P68 | M | 70 | 0.5 | exit | ABC | MTX |
| P69 | F | 70 | 34.1 | exit | ABC | MTX, CEPP, RT |
| P70 | F | 57 | 48.2 | exit | ABC | MTX |
| P71 | F | 75 | 59.1 | alive | ABC | MTX |
| P72 | M | 69 | 39.4 | alive | ABC | MTX |
| P73 | M | 50 | 13.8 | exit | ABC | MTX, CEPP, RT |
| P74 | F | 68 | 43.1 | exit | ABC | MTX/Vumon/BCNU, RT, Ara-C |

| Case No | Sex | Age at diagnosis (years) | OS (months) | Event | Molecular subtype IHC | Therapy |
|---|---|---|---|---|---|---|
| P75 | F | 67 | NA | NA | ABC | NA |
| P76 | F | 20 | 116.4 | alive | GC | MTX, WBRT |
| P77 | M | 70 | 0.5 | exit | ABC | NA |
| P78 | F | 63 | 14.3 | exit | ABC | MTX, CEPP, RT |
| P79 | M | 68 | 0.1 | exit | ABC | NA |
| P80 | M | 75 | 13.8 | exit | ABC | NA |
| P81 | M | 65 | 2.3 | exit | ABC | MTX |
| S1 | F | 78 | NA | NA | NA | MTX |
| S2 | M | 72 | 0.6 | exit | GC | R-CHOP, R, WBRT |
| S3 | M | 63 | 127.3 | exit | ABC | MTX/Vumon/BCNU |
| S4 | M | 61 | 33.0 | alive | GC | R-IDARAM, R |
| S5 | F | 65 | 84.6 | exit | GC | NA |
| S6 | F | 25 | 5.0 | exit | GC | R-CODOX-M, R-IVAC, WBRT |
| S7 | F | 67 | 61.6 | alive | GC | R-CHOP |
| S8 | M | 56 | 0.5 | exit | GC | NA |
| S9 | F | 55 | 82.6 | alive | ABC | R-CHOP |
| S10 | M | 21 | 13.9 | alive | GC | MATRIX |
| S11 | F | 37 | 84.5 | alive | GC | NA |
| S12 | F | 69 | 7.2 | exit | GC | R-GCVP |
| S13 | F | 75 | 3.7 | exit | ABC | MTX, Steroid |
| S14 | F | 55 | 0.5 | exit | ABC | R-CHOP |
| S15 | F | 59 | 24.8 | exit | ABC | NA |
| S16 | M | 45 | 3.8 | exit | ABC | R-CHOP, MTX/Ara-C |
| S17 | F | 71 | 3.9 | exit | ABC | NA |
| S18 | M | 19 | 86.4 | alive | ABC | NA |

**Table 6.** Descriptive statistics of the PCNSL/SCNSL cohort. P[1-81] PCNSL cases; S[1-19]: SCNSL cases; ABC: activated B-cell; Ara-C: cytarabine; BCNU: 1,3-bis (2-chloroethyl)-1-nitroso-urea; CEPP: cyclophosphamide, etoposide, procarbazine and prednisone; F: female; GC: germinal center; IHC: immunohistochemistry; IVAC: ifosfamide, etoposide, and cytarabine; M: male; MATRIX: methotreate, cytarabine, thiotepa and rituximab; MTX: methotrexate; NA: not available; OS: overall survival; PCNSL: primary central nervous system lymphoma; R: rituximab; R-CHOP: rituximab, cyclophosphamide, doxorubicin, vincristine, and prednisolone; R-CODOX-M: cyclophosphamide, cytarabine, vincristine, doxorubicin, and methotrexate; R-GCVP: rituximsb, gemcitabine, cyclophosphamide, vincristine and prednisolone; R-IDARAM: rituximab, idarubicin, dexamethasone, cytarabine, and methotrexate; R-IE: rituximab, ifosfamide and etoposide; RT: radiotherapy; SCNSL: secondary central nervous system lymphoma; Tem: Temsirolimus; VIM: etoposide, ifosfamide and mitoxantrone; Vumon: teniposide; WBRT: whole brain radiation therapy [Bödör et al., 2020].

Molecular subtyping

RNA isolation from samples (PCNSL n = 77, SCNSL n = 17) was performed using the RecoverAllTM kit (Life Technologies/Ambion, Inc, Foster City USA) following the manufacturer's instructions. Molecular subtypes were determined using the Research Use Only version of the LST-assay on the nCounter® Analysis System (NanoString Technologies, Inc., Seattle, USA). Linear Predictor Score (LPS) was calculated using a weighted sum of the gene expression (15 signature- and 5 housekeeping genes) for all samples. The LPS is compared against thresholds that define value ranges for the assignment of ABC or GC subtype, or Unclassified within an equivocal zone.

Customized amplicon sequencing

Genomic DNA was extracted using the FFPE Tissue Kit (Qiagen, N.V., Venlo, Netherlands) following the manufacturer's protocol from 64 PCNSL and 12 SCNSL samples. Five non-malignant tissue specimens were used as negative controls. Mutation profiles of 14 genes (e.g., *CARD11, CCND3, CD79B, CSMD2, CSMD3, IRF4, KMT2D, C-MYC, MYD88, PAX5, PIM1*, *PRDM1, PTPRD* and *TP53*) were determined by targeted NGS using the TruSeq Custom Amplicon dual-strand approach (Illumina, Inc., San Diego, USA). This method is specifically designed for FFPE samples and utilize two mirrored sets of locus-specific primers generating matching complementary and strand-specific amplicon libraries. Separate preparation of the sample-specific libraries and sequencing with unique indexes allow for a subsequent bioinformatics correction of errors/bias caused by the by FFPE fixation. After QC and equimolar pooling, libraries were sequenced on a HiSeq 4000 Instrument using PE150 chemistry.

Bioinformatics workflow

Applied bioinformatic pipeline is presented on Figure 4. (see CLL case study). As data preprocessing, filtered sequencing reads were mapped to the Ensembl Homo sapiens hg19/GRCh37 genome build using BWA v0.7.13 aligner [Li et al., 2021]. BAM files were sorted and indexed by SAMtools v1.7 [Danecek et al., 2021], GATK v4.0 BSQR tool [DePristo et al., 2011] was run on each sample to detect and correct systematic sequencing errors. SNV calling was performed with LoFreq v2.1 variant-caller [Wilm et al., 2010]. Raw variants were functionally, and database annotated using SnpEff v4.3i and ANNOVAR v2017Jul17 tools, including up-to-date information from COSMIC, avSNP and CLINVAR databases [Cingolani et al., 2012; Wang et al., 2010]. After the bioinformatic analysis, somatic variants detected in sample-specific, matching individual libraries (A and B) were combined based on genomic position and allele type using an in-house R script (version 3.4.3 (2017-11-30)). Variants exclusively identified in both libraries A and B were considered as true aberrations. A subset of somatic variants with variant allele frequency of >20% was validated by bidirectional Sanger sequencing.

Statistical analysis

Kaplan-Meier survival curves and log-rank tests were performed to compare survival times between groups using GraphPad PRISM v. 5.0 software (GraphPad Software, San Diego, USA). Pearson Chi-square test or Fisher's exact test were used to analyse categorical data. P values 0.05 or below were considered statistically significant.

*DNA methylation profiling*

***CASE STUDY III.: GBM***

Patient samples

Surgically removed FFPE GBM specimens were obtained between 1999 and 2017. After routine histological work the leftover blocks were used for these molecular analyses according to the approval (Number: 7517 PTE 2018 and 2019) from the Regional Clinical Research Committee. The characteristics of patients and specimens are summarized in Table 7.

| PRIMARY | RRBS ID | RECURRENT | RRBS ID2 | Gender | Age at onset (years) | Age at death (years) | Treatment | T1-T2 (weeks) | Overall Survival (weeks) |
|---|---|---|---|---|---|---|---|---|---|
| 15043 | 1 | 9849 | R1 | man | 50 | 50 | No data | 31 | 41 |
| 9501 | 2 | 3624 | R2 | man | 52 | 53 | No data | 33 | 59 |
| 15916 | 4 | 9527 | R4 | woman | 63 | 64 | S+I 50 Gy | 30 | 43 |
| 9886 | 5 | 15289 | R5 | man | 41 | 43 | No data | 17 | 70 |
| 3094 | 6 | 15302 | R6 | man | 59 | 60 | S+I+TMZ | 35 | 65 |
| 5526 | 7 | 13808 | R7 | woman | 50 | 52 | S+I+TMZ | 77 | 88 |
| 13501 | 8 | 9614 | R8 | man | 39 | - | S+I+TMZ | 40 | - |
| 12732 | 9 | 17440 | R9 | man | 41 | 43 | S+I+TMZ; B+I | 117 | 149 |
| 17578 | 10 | 7779 | R10 | man | 63 | - | No data | 77 | - |
| 15466 | 11 | 16534 | R11 | man | 66 | - | S+I+TMZ | 56 | - |
| 10379 | 12 | 7536 | R12 | woman | 56 | 61 | STUPP + B/P | 199 | 287 |
| 14561 | 13 | 2315 | R13 | man | 45 | - | STUPP + B/P | 70 | - |
| 2525 | 14 | 1365 | R14 | man | 32 | 36 | S + TMZ, B, I | 177 | 203 |
| 14642 | 15 | 7990 | R15 | man | 43 | 46 | S+I+TMZ | 135 | 192 |
| 5693 | 16 | 612 | R16 | woman | 45 | 48 | S+I+TMZ | 143 | 169 |
| 7183 | 17 | 11956 | R17 | woman | 57 | 59 | S+I+TMZ | 51 | 95 |
| 6795 | 18 | 17545 | R18 | woman | 61 | 62 | S+I+TMZ | 31 | 54 |
| 16189 | 19 | 16742 | R19 | woman | 53 | 55 | S+I+TMZ | 55 | 69 |
| 8117 | 20 | 2908 | R20 | woman | 37 | 40 | S+I+TMZ | 92 | 106 |
| 3997 | 21 | 5120 | R21 | man | 62 | 63 | S+I+TMZ | 58 | 62 |
| 10776 | 23 | 2168 | R23 | man | 43 | 44 | S+I+TMZ | 29 | 46 |
| 13956 | 24 | 12107 | R24 | man | 60 | 62 | S+I+TMZ | 49 | 60 |

**Table 7.** Patient's characteristics. The table summarizes the gender, age at onset and age at death of patients, the treatment modalities and T1-T2 time. OS could not be calculated for four patients because the time of death was unavailable after extensive search of all electronic medical records. Therefore, instead of OS, the T1-T2 time values were used in the statistical analyses; TMZ temozolomide; S surgery; I irradiation; B bevacizumab; P placebo [Kraboth et al., 2020].

The diagnosis of primary GBM was established based on standard clinical and histopathological criteria [Louis et al., 2016]. After quality assessment, 22 pairs of isocitrate dehydrogenase (*IDH*)-1 R132H negative, initial (GBM1) and recurrent (GBM2) tumour blocks were identified. GBM1 specimens were taken before chemoradiation treatment, and GBM2 ones at recurrence after chemoradiation. Twenty-one patient received temozolomide-based chemo- and radiation therapy after the first surgery. In the first control group (CG1), six postmortem FFPE normal brain specimens were used from the tissue archive of the Pathology Institute, UP. This step was necessary because no surgically dissected normal brain or other neurological disease control FFPE specimens were available. In the second control group (CG2), DNA CpG methylation data of five brain specimens obtained during epilepsy surgery were included by downloading data from the EBI European genome–phenome archive (accession number: EGAS00001002538) [Klughammer et al., 2018]. DNA specimens of CG1 were processed by the same way as GBM1 and GBM2. DNA specimens of CG2 were also processed by RRBS but sequenced on Illumina HiSeq 3000 and 4000 machines [Klughammer et al., 2018)] Normal brain contamination could be excluded by the evaluation of a hematoxylin–eosin stained section from each tumour. The characteristics of the tumors are summarized in Table 8.

| RRBS ID | MI | MVP | Necrosis | Atypia | Cell | TIL | LG |
|---|---|---|---|---|---|---|---|
| 1 | 36 | high | none | high | astro/gemisto | focally med | no |
| 2 | 2 | low | palisade and geo | low | ependymoma-like | med | no |
| 4 | 10 | low | none | moderate | astro | no | no |
| 5 | 91 | high | geo | high | ependymoma-like | no | no |
| 6 | 120 | high | geo | low | small, spindle | many | no |
| 7 | 20 | low | geo | moderate | astro | few | yes |
| 8 | 13 | low | geo | high | melanoma-like | few | no |
| 9 | 2 | low | palisade and geo | low | ependymoma-like | med | no |
| 10 | 0 | no | extensive geo | moderate | spindle | many | no |
| 11 | 18 | high | geo | high | astro | many | yes |
| 12 | 30 | high | geo | high | small | many | no |
| 13 | 36 | high | palisade and geo | high | astro | few | no |
| 14 | 38 | high | palisade and geo | high | small/astro | few | yes |
| 15 | 78 | low | palisade and geo | moderate | small | many | no |
| 16 | 42 | low | palisade and geo | moderate to high | astro | few | no |
| 17 | 44 | high | palisade and geo | moderate | astro | few | no |
| 18 | 15 | low | geo | moderate | astro/gemisto | few | no |
| 19 | 24 | no | none | high | oligo | few | yes |
| 20 | 25 | high | none | high | astro | many | no |
| 21 | 12 | high | palisade | moderate | astro/spindle | few | no |
| 23 | 32 | high | geo | moderate | astro/gemisto | few | no |
| 24 | 32 | high | geo | high | astro/giant | few | no |
| R1 | 100 | med | palisade and geo | moderate | small | low | no |
| R2 | 2 | no | no | low | astro | many | no |
| R4 | 32 | high | palisade and geo | moderate | small | no | no |
| R5 | 94 | moderate | extensive geo | high | ependymoma-like | few | no |
| R6 | 4 | no | geo | focally high | spindle | few | no |
| R7 | 21 | no | no | moderate to high | astro | few | yes |

| RRBS ID | MI | MVP | Necrosis | Atypia | Cell | TIL | LG |
|---|---|---|---|---|---|---|---|
| R8 | 20 | no | palisade and geo | moderate | melanoma-like | few | no |
| R9 | 14 | no | geo | high | melanoma-like | few | no |
| R10 | 50 | low | palisade | moderate | astro | moderate | no |
| R11 | 14 | high | geo | high | giant/astro | many | yes |
| R12 | 62 | high | palisade | high | astro | many | no |
| R13 | 36 | high | palisade | high | astro | few | no |
| R14 | 40 | high | no | high | small/astro | many | no |
| R15 | 16 | low | palisade and geo | moderate | small | many | no |
| R16 | 12 | high | palisade and geo | high | astro | many | no |
| R17 | 22 | high | palisade | moderate | astro | few | no |
| R18 | 18 | no | palisade and geo | focally high | astro/spindle | few | no |
| R19 | 18 | no | palisade and geo | high | oligo | few | yes |
| R20 | 20 | yes | palisade | moderate | astro/spindle | no | no |
| R21 | 18 | high | palisade and geo | mild | astro | few | yes |
| R23 | 10 | no | no | moderate | small | many | yes |
| R24 | 16 | no | palisade | high | spindle | focally many | no |

**Table 8.** Summary of histopathological characteristics of GBM1 and GBM2. Histological parameters were assessed by manual eyeballing using low microscopic magnification (100x) and semiquantitative evaluation criteria published previously [Tompa et al., 2018]. In statistical analyses, semiquantitative determinants were replaced by numerical values: e.g., TIL: no = 0, sparse = 1, moderate = 2, dense = 3 MI mitotic index (number of mitoses per 10 high power fields), MVP microvascular proliferation, TIL tumor infiltrating lymphocytes [Kraboth et al., 2020]

DNA methylation profiling

Five cuts from each paraffin block were used for DNA extraction by the QIAamp DNA FFPE Tissue Kit (Qiagen GbmH, Hilden, Germany). DNA quality was measured using a Qubit™ 1X dsDNA HS Assay Kit on a Qubit 3 Fluorimeter (Invitrogen, Carlsbad, CA, USA). The distribution of the fragments was determined using an Agilent Genomic DNA ScreenTape Assay on an Agilent 4200 TapeStation System (Agilent Technologies, Santa Clara, CA, USA). The Premium RRBS kit 24x (Diagenode SA, Seraing, Belgium) was used to prepare the bisulfite libraries according to the manufacturer's instructions. To compensate for higher degrees of fragmentation input DNA was increased up to 350–400 ng. Next steps were DNA digestion by Msp1, fragment-end repair and adaptor ligation. Lbrary QC was determined using the Kapa

Sybr Fast qPCR kit (Kapabiosystems, Cape Town, South Africa) on a StepOnePlus Real-Time PCR System (Applied Biosystems, Foster City, CA, USA). Samples with similar Ct values were multiplexed in pools of eight. The pools were subjected to bisulfite conversion, followed by a second qPCR step to precisely set up the enrichment amplification cycles for the final PCR on a GeneAmp PCR Systems 9700 (Applied Biosystems, Foster City, CA, USA). After confirming the adequate fragment size distributions and the concentrations, the amplified libraries were sequenced using the NextSeq 500/550 High Output Kit v2.5 (single-end 75 cycles, SE75 chemistry) on a NextSeq 550 machine (Illumina, San Diego, CA, USA). Raw sequencing data were uploaded to the European Nucleotide Archive (https://www.ebi.ac.uk/ena, Primary Accession: PRJEB38380, Secondary Accession: ERP121800). The glioma CpG island methylator phenotype (G-CIMP) was excluded from the cohorts by adapting the eight gene method for bisulfite-converted sequence data [Noushmehr et al., 2010].

Bioinformatics workflow

Figure 5 summarize the main steps of the implemented workflow. First basecalling and demultiplexing steps were carried out and QC was run on raw FASTQ files using FastQC v0.11.5 [https://www.bioinformatics.babraham.ac.uk/projects/fastqc]. Sequences were filtered to remove low-quality bases and adapters by Trim-Galore [ https://www.bioinformatics.babraham.ac.uk/projects/trim_galore] using RRBS mode. Bisulfite-treated reads were aligned to the bisultife converted human (hg19/GRCh37) reference genome and methylation calls were performed using Bismark [Krueger and Andrews, 2011]. After obtaining the CpG calls, RnBeads [Müller et al., 2019] was run to identify differentially methylated sites, regions, and pathways in the cohorts. The Locus Overlap Analysis (LOLA) program, within RnBeads, was used for enrichment analysis of genomic region sets and regulatory elements [Sheffield and Bock, 2016].

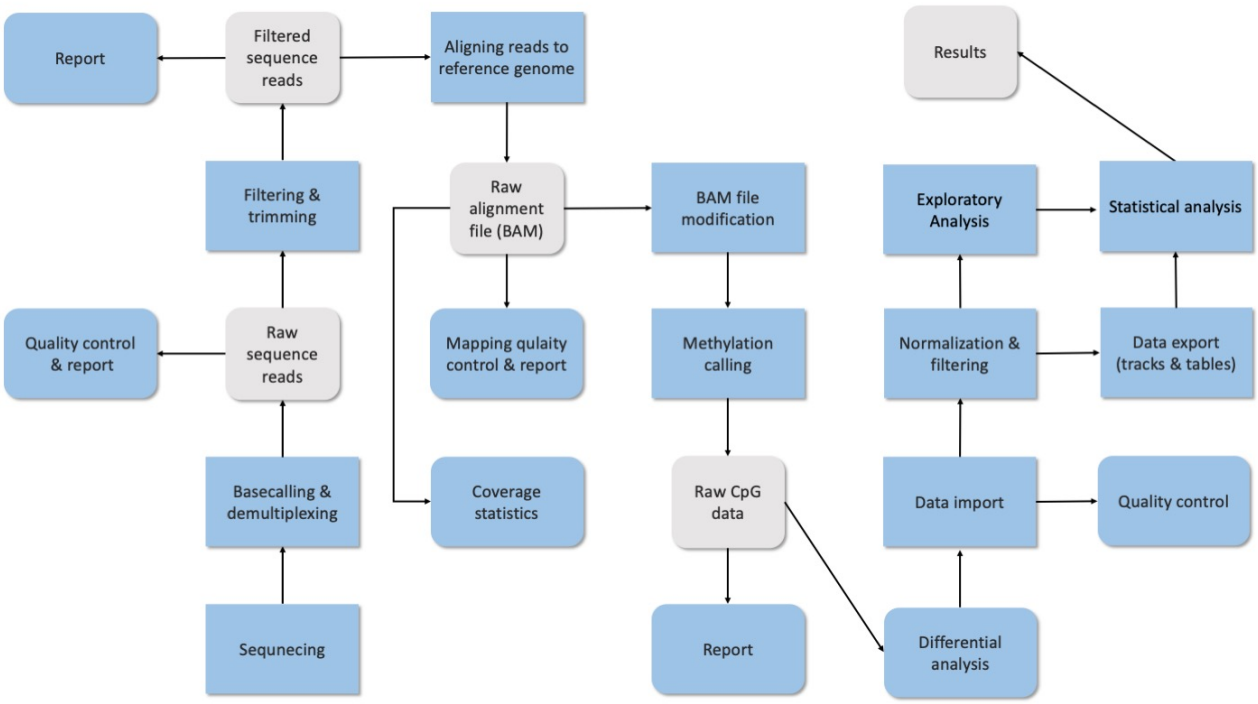



**Figure 5.** Used bioinformatic workflow including main steps and checkpoints.

## Statistical analysis

Patients' age, gender and time to recurrence (T1-T2) were correlated with histological characteristics using the Kruskal–Wallis and Mann–Whitney U tests, and Pearson's correlation.

*miRNome profiling*

***CASE STUDY IV.: NSCLC***

Patient samples

The study was approved by the Institutional Review Board of Medical University of Bialystok and Poznan and informed consent was obtained from each patient. The patients were recruited for the MOBIT project. A total of 177 cases of surgically resected NSCLC were used in this study. Inclusion criteria for this study were the following:

- diagnosis of lung AC or SCC based on histologic evidence
- completely resected tumor (free resection margins)
- stage I or stage II
- availability of representative fresh-frozen tumor specimens (at least 50% tumor cells)
- no neoadjuvant chemotherapy

In the first phase of the research miRNA profiling of all together 109 NSCLC tissue samples with matched controls (AC, n = 26; AC_c, n = 25; SCC, n = 30; SCC_c, n = 28) was done to describe the molecular background based on the DE miRNAs of the cohort and generate a "classification" set (Set 1) using all data and build prediction models. To confirm the results, miRNA expression levels as well as the molecular background were evaluated on an independent subset of 68 blood samples (AC, n = 32; SCC, n = 36) as a "validation" set (Set 2). On purpose, "Set 2" samples were collected from blood due to the limited number of tissue samples and development of future non-invasive methods. With respect to clinical characteristics (age, gender, disease stage and tumor histology), both groups were comparable (Table 9.).

Histologic diagnosis was rendered according to the most recent WHO classification of tumors of the lung [Travis, 2015] and the IASLC/ATS/ERS International Multidisciplinary Lung Adenocarcinoma Classification [Feng, 2012]. In case of any disagreement with the original diagnosis, the slides were evaluated immunohistochemically (IHC) for the expression of thyroid transcription factor-1 (TTF-1) (immunohistochemical profile) and tumor protein p63 (p63) (squamous immunophenotype). Additionally, all tumor slices were reviewed to evaluate the number of neoplastic cells for the RNA extraction.

| Characteristic | | Set 1, *n* = 109 | Set 2, *n* = 68 | All, *n* = 177 |
|---|---|---|---|---|
| Age (years) | Mean ± SD | 65.9 ± 6.65 | 66 ± 5.4 | 64 ± 7.1 |
| | Median | 65 | 64 | 65 |
| | Range | 51 - 81 | 49 - 80 | 49 - 81 |
| Gender | Female | 41 | 27 | 68 |
| | Male | 68 | 41 | 109 |
| Tumor stage | IA | 17 | 9 | 26 |
| | IB | 28 | 18 | 46 |
| | IIA | 16 | 8 | 24 |
| | IIB | 18 | 12 | 30 |
| | IIIA | 22 | 15 | 37 |
| | IIIB | 4 | 2 | 6 |
| Histology | AC | 51 | 32 | 83 |
| | SC | 58 | 36 | 94 |

**Table 9.** Patient characteristics for the classification set (*n* = 109) and the validation set (*n* = 68). *SD*, standard deviation.

Next-generation sequencing

Total RNA with small RNA fraction was isolated from fresh frozen tumor samples using mirVana™ miRNA Isolation Kit (Ambion, Poland) according to the manufacturer's instructions. RNA quantity and quality were assessed using a UV/VIS spectrophotometer NanoDrop 2000c (Thermo Scientific, Poland). The level of RNA integrity number required for analysis (RIN above 7) was determined for extracted total RNA using Agilent RNA 6000 Nano Kit on apparatus Bioanalyzer 2100 (Agilent Technologies, USA). Before constructing the RNA-seq libraries, the epicenter Ribo-ZeroTM Kit (Illumina, San Diego, CA, United States) was used to remove rRNA. Briefly, total RNA was purified by polyacryl-amide gel electrophoresis (PAGE) to enrich the sRNAs with lengths of 15–35 nt, then the sRNAs were ligated with adapters and amplified by RT-PCR. The amplification products were then separated by PAGE, and the transcriptome sequencing was performed on the HiSeq 2500 platform using SE50 chemistry.

Bioinformatics workflow

The data analysis process is shown on Figure 6. First, TrimGalore [https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/] was used for filter reads based on quality, discard sequences smaller than 13 bp from the original data and remove adapter contamination. The quality metrices of the sequences were checked before and after cleaning the data by using FastQC v0.11.5 [http://www.bioinformatics.babraham.ac.uk/projects/fastqc]. Then, the clean data was

mapped to the human reference genome (hg19/GRCh37) using STAR [Dobin et al., 2013] with miRNA specific parameter set. Finally, raw count matrix was generated using Rsubread package [Liao et al., 2019]. The quality of the mapping and sample relations are studied applying several different methods including visualization using in-house R v3.6.3 scripts. If low quality samples or data outliers are detected, they may be excluded from further analysis at this point. The data are also normalized to reduce systematic noise caused by non-biological sources and to improve the comparability of the samples.

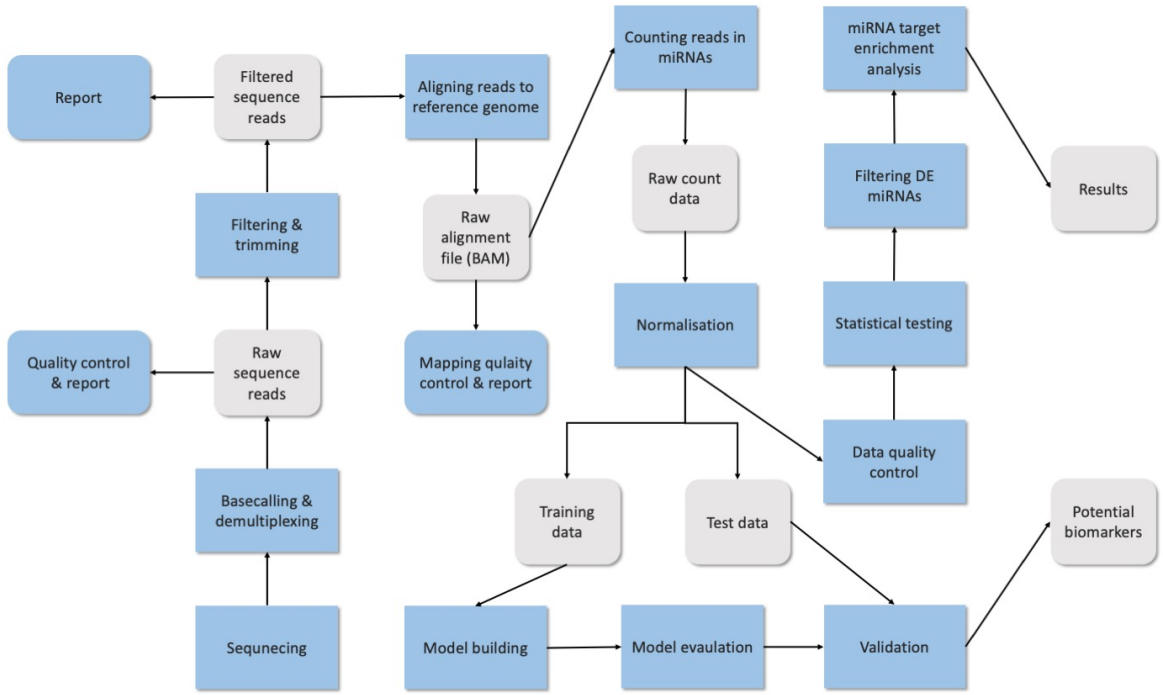



**Figure 6.** The analysis workflow for miRNA-seq data.

Statistical analysis

After the pre-processing, statistical testing is performed between the compared sample groups. The results from the testing are used to filter the so-called DE or differentially expressed miRNAs. The filtering is based on the statistical significance and the size of the difference in the mean expression levels between the sample groups. In NGS data analysis all pre-processing steps were executed within Rstudio 2020 using R v3.6.3 [http://www.rstudio.com/]. The probes were normalized applying quantile normalization method. As a result, a dataset with more than 200 miRNAs were obtained. R-package *limma* (Smyth, 2004, and Ritchie et al., 2015) was used to assess statistical significance of differences in miRNA expression between two histological subtypes of NSCLC (AC and SCC). The analysis of differential expression between AC and SCC patients was adjusted for gender and tumor stage. For each miRNA, a linear model with the histological subtypes, gender and stage as covariates was fitted. After fitting the models, the differences were tested with a t-test. The Benjamini-Hochberg procedure (Benjamini and Hochberg, 1995) was applied to correct for multiple testing with FDR = 0.05. The data analysis steps were conducted for the classification and validation set as well.

Finally, potential target genes of DE miRNAs were identified to carry out enrichment analysis to gain more functional information about molecular background of the subgroups. Functional analysis as a term covers all analyses of the expression results taking the functional annotations of genes into account. In this type of analysis, the focus is on pathways and other functional categorizations instead of single genes. Here the enrichment of functional terms and pathways within the differentially expressed miRNA target gene list has performed using the mirPath v.3 tool [Vlachos et al., 2015]. All analyses have been conducted against the KEGG [Kanehisa and Goto, 2000] databases which are the most commonly used databases for this purpose. Gene Ontology provides a hierarchical organization of genes into biological processes, molecular functions and cellular components whereas KEGG lists pathways for biological interactions. More information on these databases can be found on the KEGG websites.

Biomarker prediction

The CAncer bioMarker Prediction Pipeline (CAMPP) [Terkelsen et al., 2019] was run to identifiy possible biomarker using the classification dataset. Results were stratified for cancer staging. The pipeline can perform the following types of analysis:

- Differential expression/abundance analysis (limma [Ritchie et al., 2015])
- LASSO/Elastic-Netregression (glmnet)
- Weighed Gene Co-expression Network Analysis (WGCNA [Langfelder and Horvath, 2015])
- Correlation analysis (Pearson/Spearman)
- Survival analysis (Cox proportional hazard regression, survcomp [Schröder et al., 2011])
- Protein-protein/gene-miRNA interaction network analysis (multimiR [Ru et al., 2014] and the STRING [Jensen et al., 2008]).

In addition to the above-mentioned different types of analysis the pipeline performs missing value imputation, normalization, and transformation, along with data distributional checks.

Next, the identified potential biomarkers were used for classifying the training dataset. Decision tree model was built using rpart R package [https://CRAN.R-project.org/package=rpart]. Decision Tree is a supervised machine learning algorithm which can be used to perform both classification and regression on complex datasets. They are also known as Classification and Regression Trees (CART). Hence, it works for both continuous and categorical variables. Normalized miRNA data was used as input for model building. Finally, the fitted model was used to predict outcomes (AC or SCC) in the validation blood dataset and precision value was calculated using the confusion matrix approach.

*Assisted reproduction*

***CASE STUDY V.: NIPGT-A***

Proposed workflow

After the registered pregnancy outcome, SCM samples and corresponding blank culture media droplets were sequenced for CNV analysis. The developed comprehensive workflow shows the entire clinical procedure of IVF, the embryo culture and the wet-lab handling and dry-lab bioinformatics steps of sample processing as well (Figure 7.). The following sections briefly describe the main steps of the proposed workflow applied to the 40 selected samples.

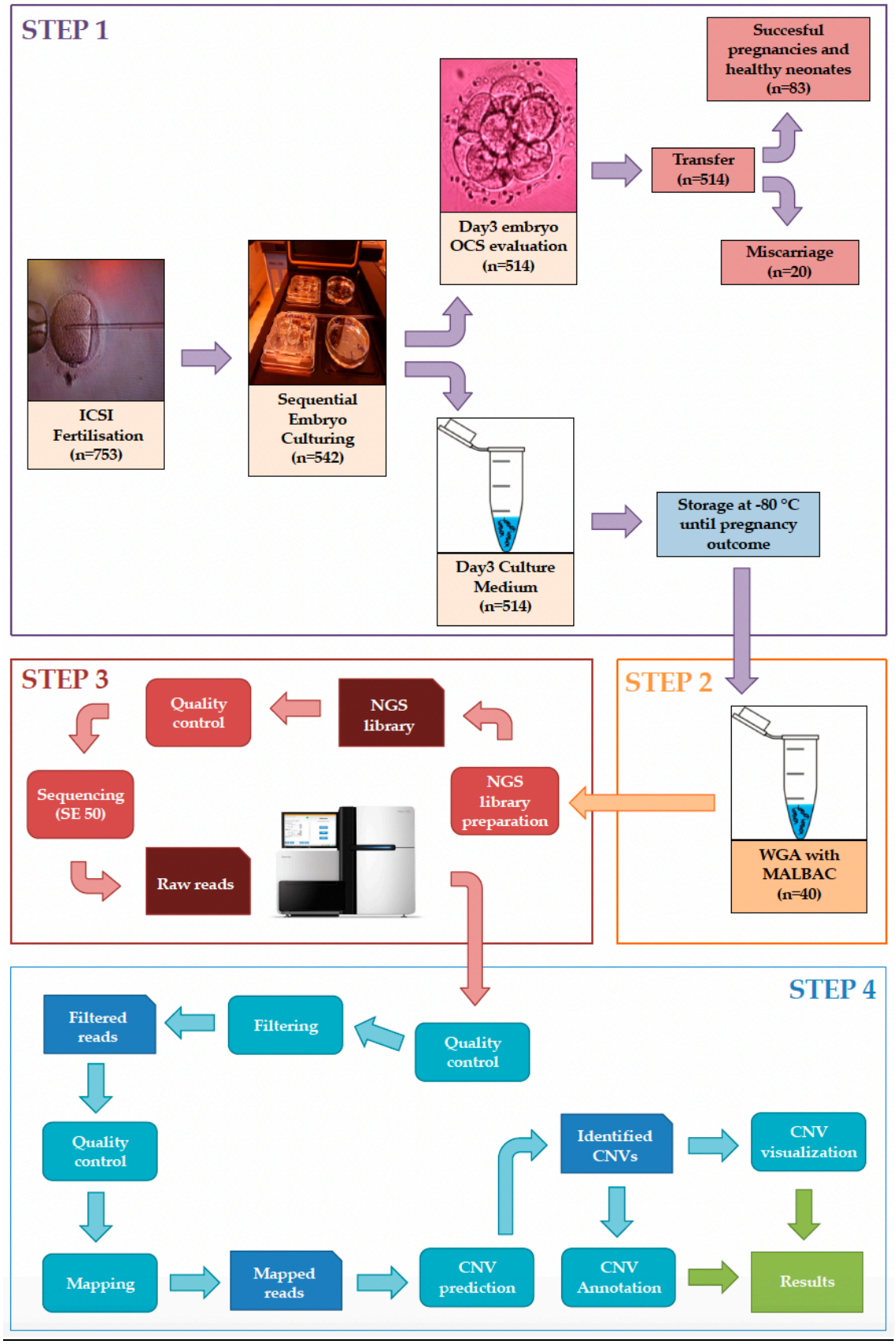



**Figure 7.** Representation of the entire workflow with all four main steps including Step 1: IVF procedure and sample collection, Step 2: Whole genome amplification. Step 3: Next-generation sequencing and Step 4: Bioinformatics analysis [Gombos et al., 2021].

Patient samples

To validate our workflow SCM dropltes (n = 7) of Day 3 embryos fertilised using intracytoplasmic sperm injection (ICSI) method and presenting good quality scores on morphology assessment were collected prospectively in the Assisted Reproduction Unit, Department of Obstetrics and Gynaecology, University of Pecs, Hungary. The work described here was approved by the Committee of Human Reproduction, National Science Council of Hungary: 5273-3-2012/HER, later superset by Public Health Officer Hungarian Government Office in Baranya County: BAR/006/58-2/2014). The research related to human use has been complied with all the relevant national regulations, institutional policies and in accordance with the tenets of the Helsinki Declaration.

The oocytes selected for ICSI were denuded carefully with hyaluronidase and assessed for maturity. Only metaphase II oocytes (n = 753), which had polar body, were chosen for fertilisation. ICSI was performed after oocyte recovery (3–6 h) in a bicarbonate-buffered medium (G-IVF, Vitrolife, Gothenburg, Sweden). Fertilisation was checked next day (24 h later) and embryos were transferred to G-1 v5 medium (Vitrolife) supplemented with human serum albumin (HSA; Vitrolife) in 5 mg/mL concentration. Embryos (n = 542) were cultured following a sequential culture protocol (~ 40 μL culture medium) and moved to fresh medium droplets on Day 3 (n = 514) and 20 μL of the SCM was collected. As negative control, we collected the same amount of blank culture medium and were collected from the same LOT of medium and HSA. All collected samples were frozen immediately in liquid nitrogen and stored at –80 °C. Further sample selection was based on the optimised criteria system (OCS) evaluation [7] and only embryos were chosen that fulfilled good composite score (e.g., high blastomere number ≥7, symmetric position, fragmented cell rate <10%). Selected embryo morphology parameters and parental gynaecological characteristics are summarized in Table 10.

| | Healthy Neonate (Group 1) | Miscarriage (Group 0) |
|---|---|---|
| Number of embryonic culture media samples sequenced | 20 | 20 |
| **ICCS Scoring parameters of D3 embryos** | **Group 1** | **Group 0** |
| average blastomere number | 8.2 | 8.6 |
| fragmentation by volume | <10% | <10% |
| blastomere symmetry | full | full |
| **Clinical characteristics** | **Group 1** | **Group 0** |
| female average age | 35.18 | 34.74 |
| cause of infertility -tubal factor | 27.27 | 22.5 |
| cause of infertility male factor | 45.45 | 42.5 |
| cause of infertility -other | 27.27 | 25 |
| basal FSH (Follicle Stimulating Hormone) cc (IU/μL) | 7.63 | 7.2 |
| previous miscarriage | 0 | 0 |
| oocyte collected | 9.3 | 8.6 |
| available embryos for culture | 2.5 | 2.5 |

**Table 10.** Embryo morphology parameters and parental gynaecological characteristics [Gombos et al., 2021].

After registration of pregnancy outcome in 184 cases, all spent embryo culture media samples were used for the downstream laboratory analysis. Twenty embryos were selected for the miscarriage Group 0. From embryos that developed to healthy neonates (n = 83), a matching number of 20 were randomly selected for group comparison and denoted as Group 1. Culture media samples were handled carefully to prevent media cross-contamination. Five μL from embryo's SCM was transferred into RNase–DNase-free PCR tubes mixed with 5 μL cell lysis buffer (Yikon Genomics, Beijing, China).

Next-generation sequencing

The multiple annealing and looping-based amplification (MALBAC) whole-genome amplification (WGA) method was applied to amplify DNA from the collected samples, following the manufacturer's protocol (Catalogue no. YK001B; Yikon Genomics, Beijing, China). Concentration of the WGA products were assessed using the Qubit 2.0 fluorometric quantitation system (Life Technologies, Carlsbad, CA, USA). Due to low sample quality, only 28 out of 40 samples were selected for the next processes. NGS libraries were prepared from

50 ng input material using the Nextera DNA Library Preparation Kit (Illumina, San Diego, CA, USA) with Nextera DNA Combinatorial Dual Indices. After QC step, individual libraries were diluted, equimolarly pooled, and sequenced on Illumina HiSeq 4000 using 50bp single-end (SE50) chemistry. The raw sequencing data was uploaded to the European Nucleotide Archive (https://www.ebi.ac.uk/ena, Primary Accession: PRJEB38821, Secondary Accession: ERP122272, 31 December 2020).

In real clinical practice a smaller sequencing instrument developed for clinical applications, such as MiSeq or iSeq, would be more practical and cost efficient to fulfil the requirements.

Bioinformatics workflow

During data pre-processing, overall quality metrics of raw sequencing reads were checked using FastQC v0.11.5 [http://www.bioinformatics.babraham.ac.uk/projects/fastqc]. Based on these results the dataset was cleaned by removing remaining adapters and low-quality (≤ Q30) parts using Cutadapt v1.18 [Marcel, 2011] and TrimGalore v0.4.1 [https://www.bioinformatics.babraham.ac.uk/projects/trim_galore] (Figure 7. STEP 4). Next, filtered sequences were mapped to the Homo sapiens hg19/GRCh37 reference genome using bwa mem algorithm of the BWA v0.7.13 aligner [Li et al, 2020]. BAM files were sorted and indexed by SAMtools v1.7 modules [Danecek et al., 2021]. For further QC mapping quality and alignment statiscic results were summarised for each sample using QualiMap bamqc v2.2.1 [Okonechnikov et al., 2016]. MutliQC v1. [Ewels et al., 2016] was run to combine mapping reports into one. Based on the mapping quality results 22 out of 28 samples were selected for further analysis.

The read-count-based CNV prediction tool cn.MOPS v1.30.0 [Klambauer et al., 2012] was optimized to carry out NIPGT-A analysis. Telomere and centromere regions were excluded from the analysis. Cause of the low sequencing coverage read numbers were counted in 1 Mb bin size along the whole genome. A copy number gain from two to three copies results in a 50% increase in read counts, whereas a copy number loss from two copies to one result in a 50% decrease in read counts. Results were exported in various formats (e.g., tabular and VCF). In the downstream analysis the identified alterations were visualized using R (version 3.4.3 (2017-11-30)) and functionally annotated by UNIQUE database [https://www.rarechromo.org, 31 December 2020], Genetic Alliance database

[https://www.geneticalliance.org.uk, 31 December 2020] and CDO database [https://chromodisorder.org, 31 December 2020].

Statistical analysis

To validate the statistical significance of the identified CNVs, ORs were calculated with 95% confidence intervals using the epi.2by2 function from the epiR R programming package [https://CRAN.R-project.org/package=epiR]. Two counting methods of CNV events were applied. First, CNVs were counted separately as simple events. Second, all events in one chromosome were merged into one large event. Applying the latter method, we could reduce the false positive CNVs that result from the low sequencing coverage. Results were visualised using the ggplot2 R package [https://CRAN.R-project.org/package=ggplot2].

# Results

*Somatic mutation profiling*

***CASE STUDY I.: CLL***

Mutation profile of the cohort

The bioinformatic analysis revealed a total of 211 relevant somatic variants in the 20 paired samples with an average coverage of 7500x across the 30 genes (Figure 8.). Most of the variants represented subclonal (157/211) with VAF of <10% and with remarkable heterogeneity across the cases. Average of 5 mutations (range: 0-19) detected in individual patients, affecting an average of 4 genes (range: 0-18). The most frequently mutated genes were:

- *NOTCH1* (70%)
- *ATM* (70%)
- *TP53* (65%)
- *BCOR* (55%)

All somatic variants with a VAF of >20% were successfully validated by Sanger sequencing.

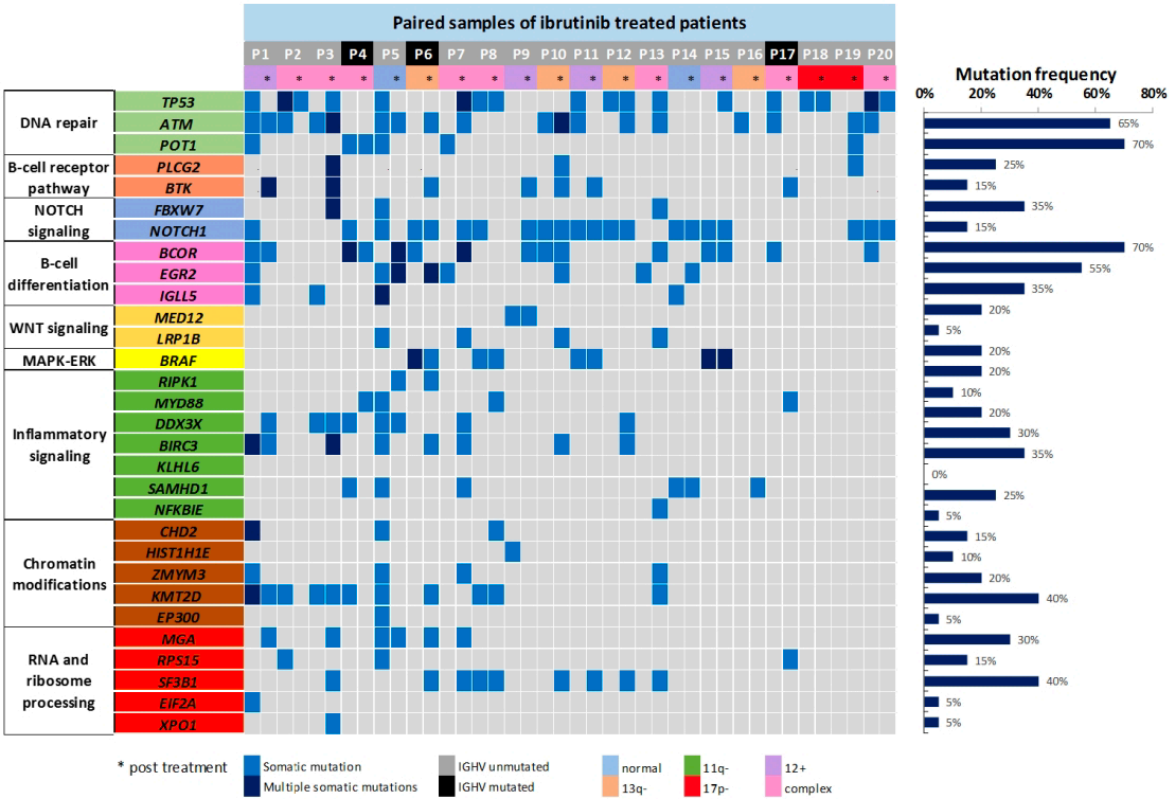



**Figure 8.** Heat map displaying the somatic variants detected in the 30 target genes analysed in the sequential samples of 20 patients treated with ibrutinib. Illustrated are the distribution of the somatic variants, mutation status of the IGHV gene, cytogenetic profile as determined by fluorescence *in situ* hybridization, as well as the mutation frequency of the individual genes for all cases [Gángó et al., 2019].

## Temporal dissection of the mutational profile

The post-treatment samples carried a slightly higher number of variants compared to the pre-treatment ones (118 vs 93), with an average of 5.9 mutations (range: 1-16) in the posttreatment specimens and on average 4.7 mutations in the pre-treatment samples (range: 0-19) (Figure 9.).

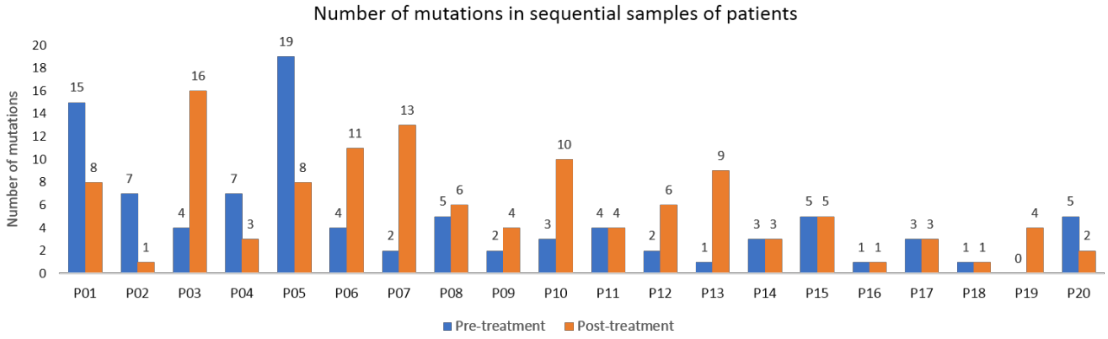



**Figure 9.** the number of mutations detected in the pre- and post-treatment samples of the 20 patients treated with ibrutinib [Gángó et al., 2019].

As mentioned before *NOTCH1*, *ATM*, *TP53* and *BCOR* represented the top four mutated target genes at baseline as well as post-treatment (Figure 10.). I contrast with that *IGLL5*, *EIF2A* and *EP300* mutations were eliminated from the post-treatment samples and the enrichment of *SF3B1* (5% vs 40%), *MGA* (5% vs 30%), *BIRC3* (10% vs 30%) mutations were observed in the post-treatment samples compared to the pre-treatment specimens (Figure 10.). *BTK*, *PLCG2*, *RIPK1*, *NFKBIE* and *XPO1* mutations were exclusively detected in the posttreatment samples in 35, 15, 10, 5 and 5% of the patients, respectively (Figure 10.).

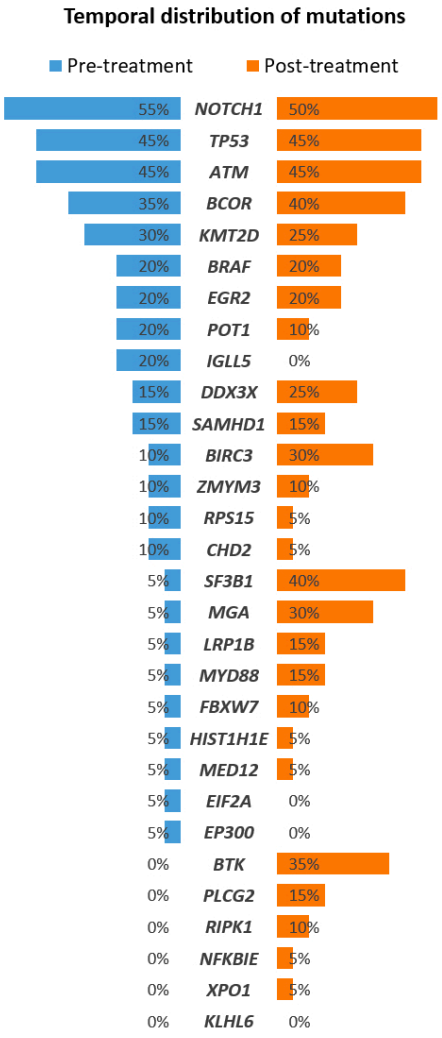



**Figure 10.** Comparison of the mutation frequency in the 30 genes analysed between the pre- and post-treatment specimens [Gángó et al., 2019].

Interestingly, multiple mutations in the same gene (convergent mutation evolution; CME) [Kiss et al., 2019] was identified in 40% of the genes, with 2-4 mutations per gene. Overall, CME was observed in 50% of patients and it was documented in both pre- and post-treatment samples in four patients and in either pre- or post-treatment samples of the three patients.

Subclonal dynamics

Mutations in associated with ibrutinib resistance were detected in 40% and 5% of the cases in *BTK* and *PLCG2* genes respectively, with mutations exclusively detected in the post-treatment samples. The *BTK* and *PLCG2* variants co-occurred in few patients carrying mutations in one of these genes with (1-4 variants/patients). In addition to the canonical *BTK* Cys481 and *PLCG2* Asp993 hotspots, four novel BTK mutations were identified (e.g., Arg28, Gly164, Arg490 and Gln516) (Figure 11a), with three previously unreported *PLCG2* mutations (e.g., Phe82, Arg694 and Ser1192) (Figure 11b), affecting four different patients.

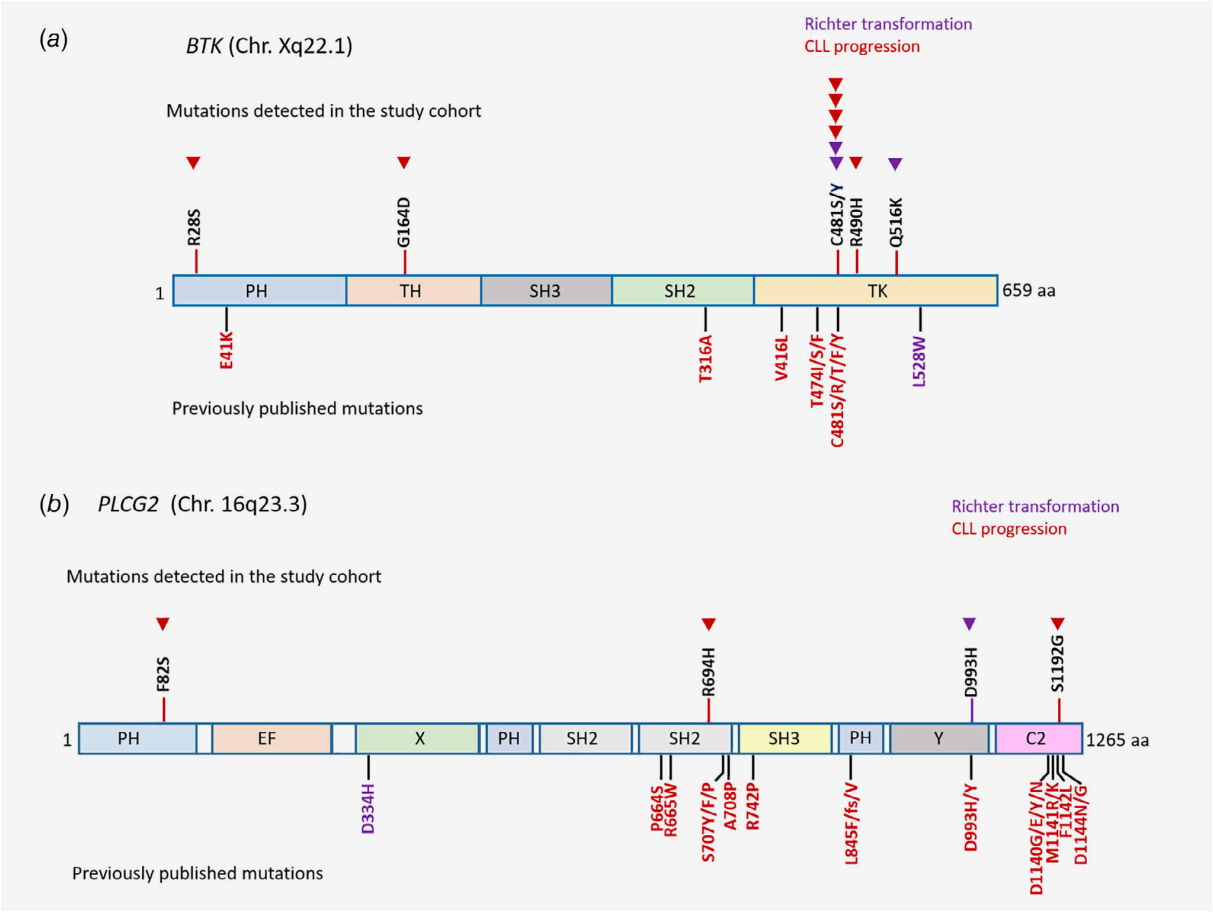



**Figure 11.** (a) Schematic domain structure of *BTK* with variants observed in our study cohort and/or identified by previous studies. We observed mutations within the PH, TH and TK domains of the protein. (PH: Pleckstrin homology; TH: Tec homology; SH3/2: Src homology 3/2; TK: Tyrosine kinase). Variants highlighted with red were detected in patients with CLL, while blue variants were observed in patients with Richter's transformation. *BTK* R28S, G164D, R490H and Q516K represent previously unreported variants, based on the COSMIC database. (b) Schematic domain structure *PLCG2* with variants observed in our study cohort and/or identified by previous studies. (PH: Pleckstrin homology; EF: EF-hand motifs; X: X domain; SH2/3: Src homology 2/3; Y: Y domain; C2: calcium-binding motif) Variants highlighted with red were detected in patients with CLL, while blue variants were observed in patients with Richter's transformation. *PLCG2* F82S and S1192G are previously unreported variants not annotated in COSMIC database. The PLCG2 R694H variant was previously reported in two colon cancer cases (COSM2693625); however, it represents a novel finding in CLL [Gángó et al., 2019].

Notably, an alternating dynamic of *BTK* and *TP53* mutations was observed in almost all patients. The emergence of *BTK* mutations upon ibrutinib treatment was accompanied by the concurrent decrease of *TP53* mutational abundance. Among the six patients harbouring BTK Cys481 mutations (Patients #1, #5, #6, #10, #11 and #20), all four patients carrying TP53 mutations (Patients #1, #5, #11, #20) demonstrated clonal elimination or reduction of the TP53 alteration in the post-treatment sample (Figure 12.).

Elimination of a TP53 mutation was also observed in Patient #17, acquiring a noncanonical BTK mutation. On the other hand, subclones carrying TP53 mutations persisted or expanded in 8/20. Also, other interesting patient specific events were obese during the study [Gángó et al., 2019] data was not show

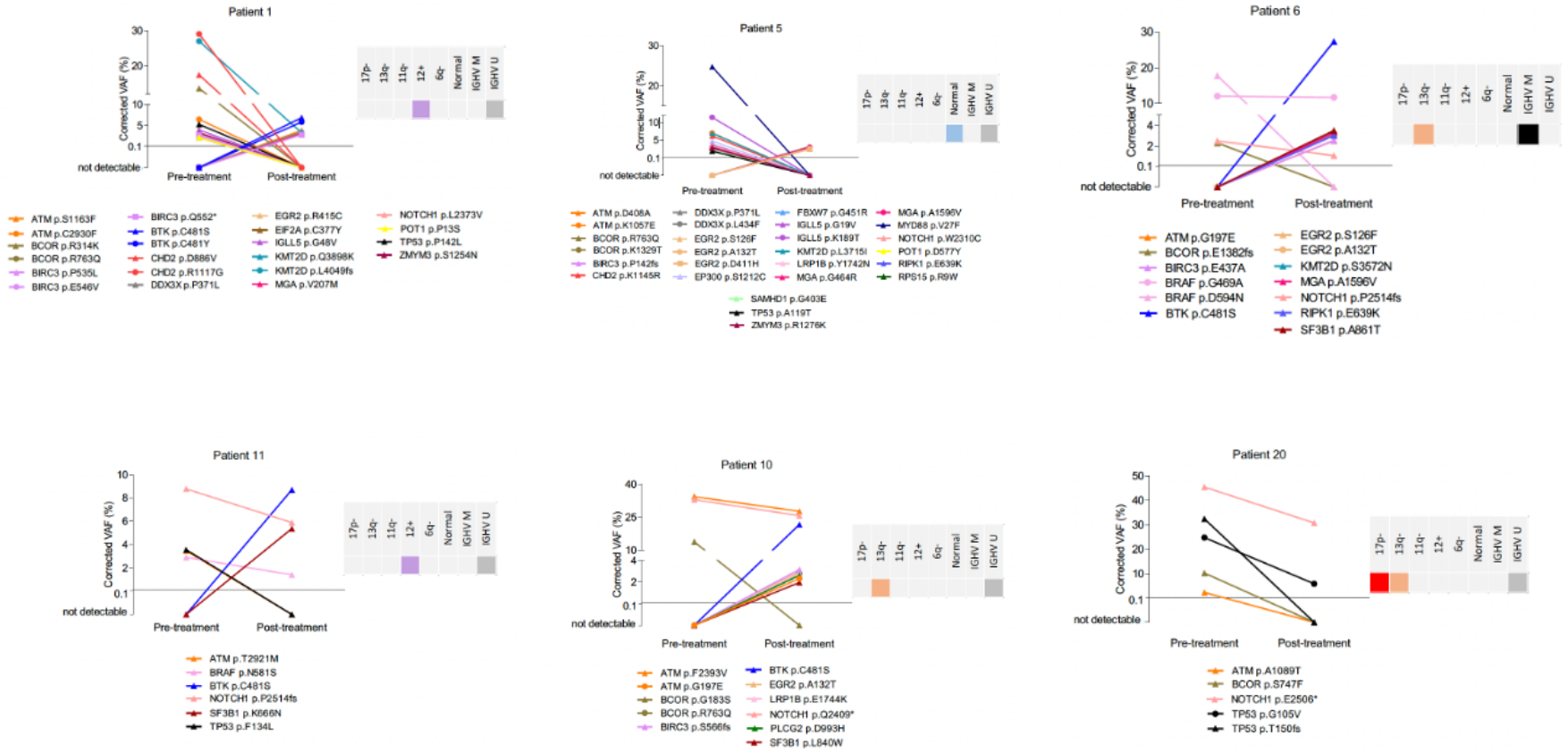



**Figure 12.** Dot plot illustration of the changes in mutational composition and heterogeneity between the pre- and post-treatment samples [Gángó et al., 2019].

***CASE STUDY II.: PCNSL***

Molecular subtypes

Using the Hans algorithm (Figure 13.) in the PCNSL cohort 95% of the cases showed ABC (non-GCB) and 5% of the cases showed GCB phenotype. In contrast, the LST-assay identified only 80.5% of the cases as ABC and 13% as GCB and 6.5% as UC subtypes, respectively. As for the SCNSL group, 47% classified as ABC and 53% as GCB phenotype. The ratio was identical using the LST-assay.

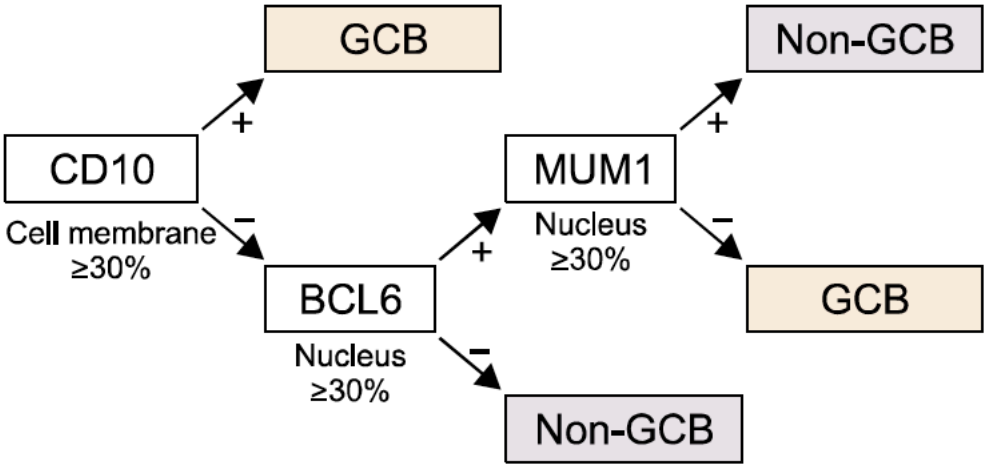



**Figure 13.** The Hans algorithm [Hwang et at., 2013].

The sub-classification obtained with the NanoString LST-assay showed discordant results in 16% of all cases (PCNSL, n = 13; SCNSL, n = 2) as compared to the IHC results. Twelve cases classified as ABC by the Hans algorithm showed a different readout using the LST-assay. Seven cases were assigned to the GCB group and 5 UC and only one IHC-GCB case was classified as ABC using the LST-assay (Figure 14). In the SCNSL group, only a single GC and a single ABC case did not match when comparing the classification results to the Hans algorithm. Overall, using the LST-assay, a significantly lower portion of the cases (80.5% versus 95%, p = 0.0219) were classified as ABC phenotype in PCNSL.

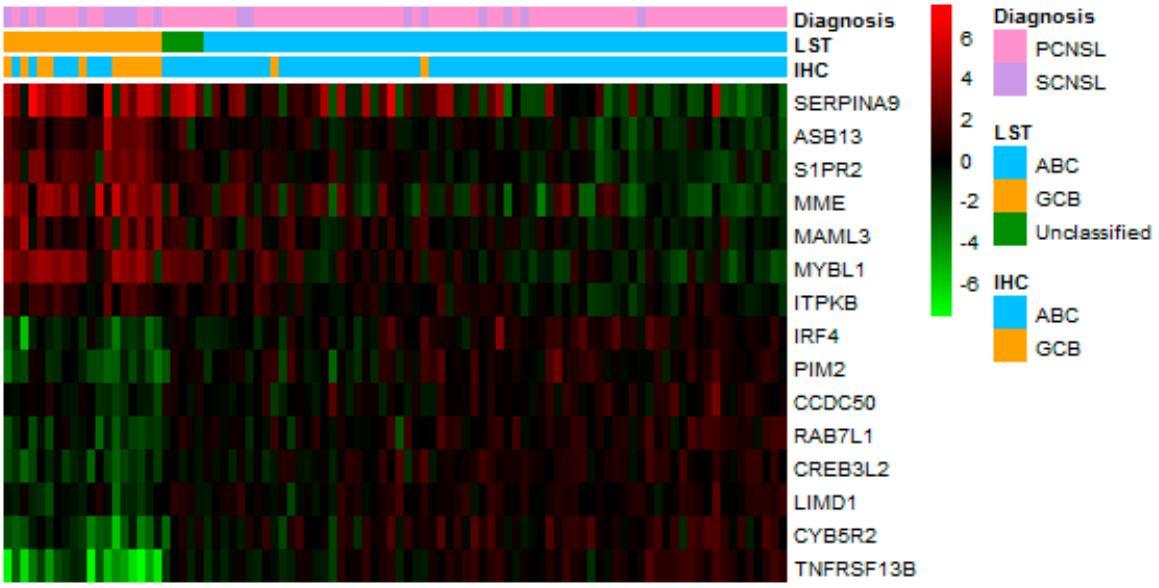



**Figure 14.** The NanoString LST readouts are illustrated in form of a gene expression heat map with the 15 target genes contributing to the model. Abbreviations: ABC: activated B-cell; GC: germinal centre; IHC: immunohistochemistry; PCNSL: primary central nervous system lymphoma; SCNSL: secondary central nervous system lymphoma; UC: unclassified [Bödör et al., 2020].

Mutation profiles of the cohort

A total of 239 relevant mutations were identified across the brain lymphomas (n = 76) with VAF min. 1.8 and max. 96.2% (mean: 41.4%). The majority (81%) of the mutations presented with a VAF ≥20%. A total of 210 somatic mutations were detected in the 64 PCNSL cases across the 14 target genes, with an avg. of 3.3 mutations/case (range: 0-10). Individual cases contained mutations in avg. 2.6 genes (range: 0-5). The distribution of the mutations was as follows:

- missense mutations: 75.2%
- mutations in 5'/3' prime UTR regions: 11.4%
- mutations at splice sites: 7.6%
- in frame deletions: 3.3%
- frameshift mutations: 1.9%
- nonsense mutations: 0.5%

The most frequently mutated genes in the PCNSL cohort were *MYD88* (66%), *PIM1* (41%), *KMT2D* (31%) and *PRDM1* (30%) (Figure 15.). No mutation was detected in *PTPRD*.

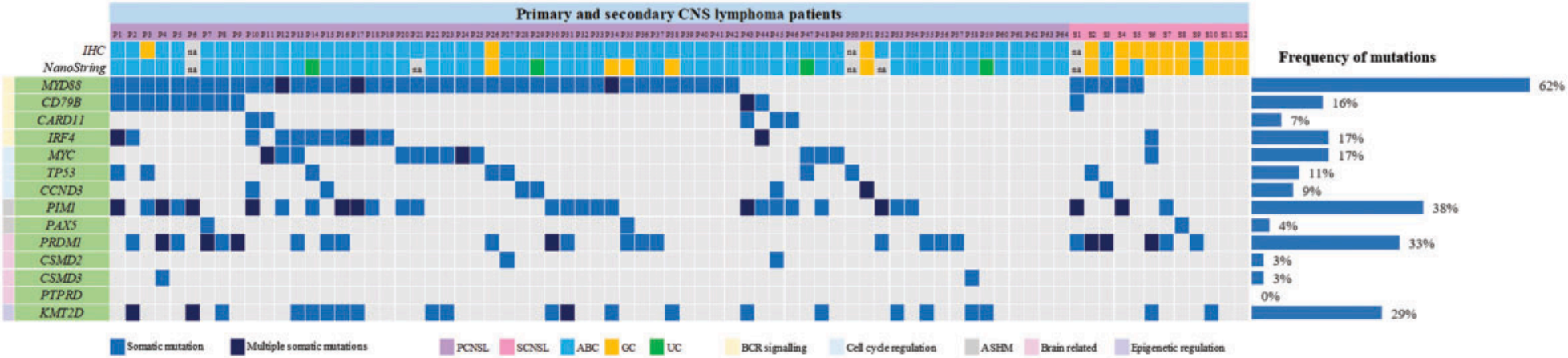



**Figure 15.** Illustrated are the mutation patterns of the 14 genes identified in 76 primary and secondary central nervous system lymphomas by next-generation sequencing, and molecular subtypes of 71 and 73 cases as defined by the NanoString LST-assay and the Hans algorithm using immunohistochemistry, respectively. Abbreviations: ABC: activated B-cell; ASHM: aberrant somatic hypermutation; BCR: B-cell receptor; GC: germinal center; IHC: immunohistochemistry; na: not available; PCNSL: primary central nervous system lymphoma; SCNSL: secondary central nervous system lymphoma; UC: unclassified [Bödör et al., 2020].

In the 12 SCNSL patients, a total of 29 somatic mutations were identified, with avg. 2.4 mutations (range: 0-5). The distribution of the mutations was as follows:

- missense mutations: 72.4%
- mutations in 5'/3' prime UTR regions: 20.7%
- frameshift mutations: 3.5%
- mutations at splice sites: 3.5%

Individual cases had mutations in avg. 1.8 genes (range: 0-4). The most frequently mutated target genes in the cohort were *PRDM1* (50%), *MYD88* (42%) and *PIM1* (25%). No mutation was identified in *CARD11*, *CSMD2*, *CSMD3* and *PTPRD* genes.

Correlation of mutation profiles and molecular subtypes

Considering all brain lymphomas, an enrichment was observed in *MYD88*, *PIM1*, *IRF4* and *MYC* in cases with ABC subtype, with mutations presented exclusively in *CD79B*, *CARD11*, *CSMD2* and *CSMD3* in ABC cases (19%, 9%, 4% and 4% vs 0% for the four genes, respectively). On the other hand, mutations of TP53 and PAX5 appeared to be more frequent in GC cases. As the results of the comparison of GC and ABC cases *PRDM1*, *KMT2D* and *CCND3* showed similar mutational frequencies (Figures 16.).

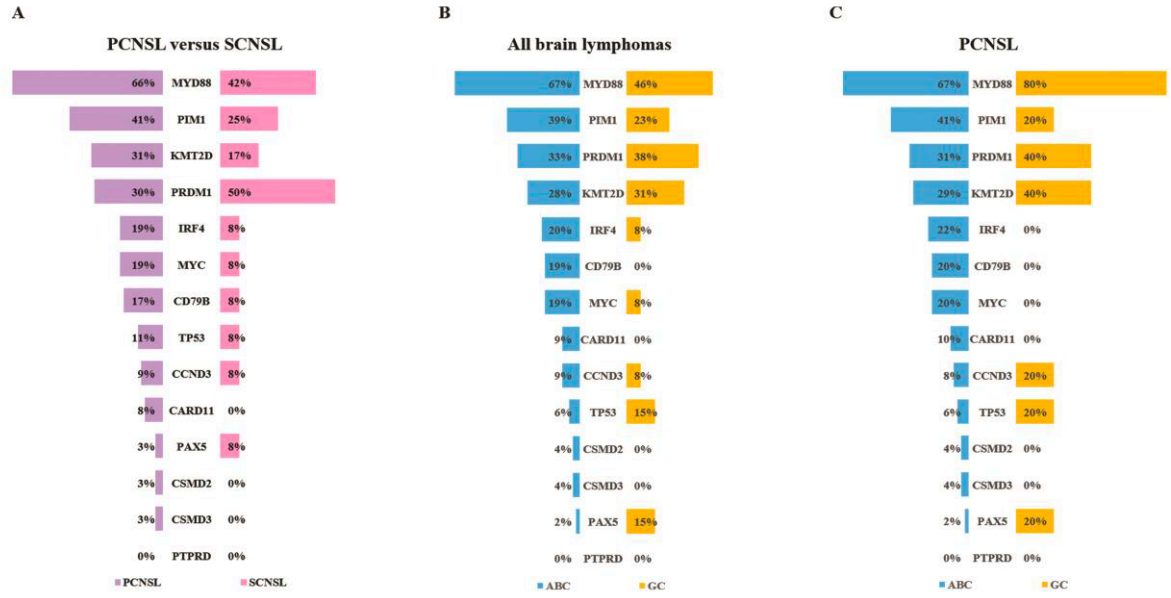



**Figure 16.** Comparison of mutation profiles between **(A)** primary and secondary central nervous system lymphomas, **(B)** all brain lymphomas of activated B-cell type (ABC) versus germinal center B-cell type (GC) and **(C)** primary brain lymphomas of ABC versus GC type. Abbreviations: ABC: activated B-cell; GC: germinal centre; NANO: NanoString; PCNSL: primary central nervous system lymphoma; SCNSL: secondary central nervous system lymphoma [Bödör et al., 2020].

In PCNSL, enrichment of *PIM1* mutations was observed (41% vs 20%) in ABC subtypes, with *IRF4*, *CD79B*, *MYC*, *CARD11*, *CSMD2* and *CSMD3* mutations being present exclusively (22%, 20%, 20%, 10%, 4% and 4% vs 0% for the six genes, respectively). In PCNSL group samples classified as GCB, mutations of *TP53* (20% vs 6%), *PAX5* (20% vs 2%) and *CCND3* (20% vs 8%) appeared to be more frequent compared to the ABC subtypes. Other genes like *YD88*, *PRDM1* and *KMT2D* showed similar mutational frequencies across ABC and GCB subtypes (Figures 16.). Regardless of the enrichment of these mutations, none of the above-mentioned differences were statistically significance when between the GC and ABC groups were compared (data not shown).

*DNA methylation profiling*

***CASE STUDY III.: GBM***

Methylation dataset

DNA CpG methylation patterns were compered in normal brain tissues and IDH-wild-type GBM specimens at initial diagnosis (GBM1) and at first recurrence (GBM2). Initially, two control groups (CG) were considered. The first group named CG1 included the methylomes of postmortem normal brain tissues (n = 6) from those who died from non-neurological causes. CG2 included the dataset of five FFPE brain tissues obtained during epilepsy surgery [Klughammer et al., 2018]. The main study group was represented as the 22 pairs of sequential surgically obtained FFPE GBM specimens in GBM1 and GBM2. According to the TapeStation analyses DNA fragmentation was slightly higher in GBM1 than in GBM2 (21.65% vs 25.10% of DNA ≥2000 bp). Cause of sample quality and quantity issues few samples were left out from the fragment analysis. In contrast with that, fragment rates were significantly different in freshly drawn total blood (87.15%) and in buffy coat (70.18%) (Table 11.).

| GBM1 | Average size 150-2000 bp | % of Total | Average size 2000-60000 bp | % of Total |
|---|---|---|---|---|
| 1_RRBS | 558 | 85% | 7014 | 10% |
| 2_RRBS | 700 | 66% | 7810 | 23% |
| 3_RRBS | 647 | 80% | 6571 | 16% |
| 4_RRBS | 745 | 76% | 6267 | 20% |
| 5_RRBS | 771 | 74% | 6954 | 21% |
| 6_RRBS | 603 | 80% | 7446 | 15% |
| 7_RRBS | 620 | 84% | 6073 | 12% |
| 8_RRBS | 660 | 76% | 7081 | 17% |
| 9_RRBS | 583 | 83% | 7598 | 13% |
| 10_RRBS | 810 | 75% | 6105 | 22% |
| 11_RRBS | 659 | 79% | 7003 | 17% |
| 12_RRBS | 767 | 73% | 6665 | 24% |
| 13_RRBS | 813 | 65% | 7275 | 28% |
| 14_RRBS | 770 | 77% | 6188 | 20% |
| 15_RRBS | 781 | 79% | 5699 | 18% |
| 16_RRBS | 682 | 70% | 8274 | 19% |
| 17_RRBS | 815 | 64% | 7404 | 26% |
| 18_RRBS | 504 | 84% | 7668 | 11% |
| 19_RRBS | 887 | 68% | 6184 | 28% |

| | | | | |
|---|---|---|---|---|
| 20_RRBS | 986 | 64% | 5428 | 30% |
| **MEAN** | **718.1** | **75.05%** | **683.4** | **19.54%** |
| **GBM2** | **Average size 150-2000 bp** | **% of Total** | **Average size 2000-60000 bp** | **% of Total** |
| 1R_RRBS | 796 | 66% | 7548 | 27% |
| 2R_RRBS | 731 | 65% | 8767 | 23% |
| 3R_RRBS | 658 | 85% | 5806 | 12% |
| 4R_RRBS | 890 | 59% | 7378 | 33% |
| 5R_RRBS | 735 | 79% | 6473 | 16% |
| 6R_RRBS | 494 | 85% | 8538 | 23% |
| 7R_RRBS | 771 | 64% | 7848 | 24% |
| 8R_RRBS | 726 | 79% | 6530 | 17% |
| 9R_RRBS | 699 | 81% | 5960 | 16% |
| 10R_RRBS | 948 | 60% | 6634 | 33% |
| 11R_RRBS | 929 | 59% | 6834 | 32% |
| 12R_RRBS | 758 | 66% | 7479 | 24% |
| 13R_RRBS | 848 | 75% | 5815 | 23% |
| 14R_RRBS | 783 | 77% | 5823 | 21% |
| 15R_RRBS | 876 | 60% | 7472 | 32% |
| 16R_RRBS | 1040 | 40% | 8204 | 43% |
| 17R_RRBS | 699 | 80% | 6894 | 17% |
| 18R_RRBS | 887 | 57% | 7231 | 34% |
| **MEAN** | **792.7** | **68.68%** | **7068.6** | **25.01%** |
| **CONTROL** | **Average size 150-2000 bp** | **% of Total** | **Average size 2000-60000 bp** | **% of Total** |
| Buffy Coat | 948 | 5.81% | 25535 | 70.18% |
| Total blood | 900 | 7.80% | 15207 | 87.15% |

**Table 11.** DNA fragmentation statistics [Kraboth et al., 2020].

The bisulfite conversion rates for all GBM1/2 samples are represented in Table 12. The mean underconversion rate was 1.32%, and the mean overconversion rate was 1.70% based on the spike-in controls.

| GBM1 | Unmeth conv. eff. (%) | Meth conv. eff. (%) | GBM2 | Unmeth conv. eff. (%) | Meth conv. eff. (%) |
|---|---|---|---|---|---|
| 1_RRBS | 99.77 | 98.82 | 1R_RRBS | 99.87 | 98.76 |
| 2_RRBS | 98.43 | 99.97 | 2R_RRBS | 95.22 | 98.17 |
| 4_RRBS | 99.36 | 99.94 | 4R_RRBS | 99.77 | 95.51 |
| 5_RRBS | 97.62 | 99.98 | 5R_RRBS | 96.95 | 99.66 |
| 6_RRBS | 97.80 | 99.09 | 6R_RRBS | 99.87 | 99.00 |
| 7_RRBS | 99.49 | 94.61 | 7R_RRBS | 99.55 | 99.19 |
| 8_RRBS | 96.14 | 99.84 | 8R_RRBS | 97.57 | 99.46 |
| 9_RRBS | 97.97 | 99.94 | 9R_RRBS | 96.93 | 95.83 |
| 10_RRBS | 99.46 | 99.58 | 10R_RRBS | 99.80 | 99.59 |
| 11_RRBS | 97.37 | 96.35 | 11R_RRBS | 98.94 | 98.62 |
| 12_RRBS | 96.75 | 98.77 | 12R_RRBS | 99.63 | 95.39 |
| 13_RRBS | 98.63 | 98.98 | 13R_RRBS | 95.97 | 94.62 |
| 14_RRBS | 99.86 | 96.69 | 14R_RRBS | 96.99 | 99.95 |
| 15_RRBS | 99.78 | 99.99 | 15R_RRBS | 99.29 | 99.93 |
| 16_RRBS | 99.65 | 98.63 | 16R_RRBS | 99.77 | 98.89 |
| 17_RRBS | 99.77 | 99.42 | 17R_RRBS | 99.61 | 94.33 |
| 18_RRBS | 99.28 | 97.32 | 18R_RRBS | 98.00 | 95.61 |
| 19_RRBS | 99.27 | 99.70 | 19R_RRBS | 99.32 | 99.38 |
| 20_RRBS | 99.02 | 99.03 | 20R_RRBS | 99.97 | 99.91 |
| 21_RRBS | 99.71 | 95.32 | 21R_RRBS | 96.73 | 99.65 |
| 23_RRBS | 99.81 | 99.89 | 23R_RRBS | 97.97 | 97.65 |
| 24_RRBS | 99.72 | 97.64 | 24R_RRBS | 99.24 | 96.51 |
| **Mean** | **98.85** | **98,61** | | **98,50** | **97.98** |

**Table 12.** Conversion rates or GBM1 and GBM2 samples [Kraboth et al., 2020].

In the nondeduplicated raw dataset the average mapping rate of the reads was 69% and the mean number of informative CpGs per sample was 20 741 979 (median: 16 574 809). Interestingly, these numbers are over ten times higher than expected ones due to duplications during library amplification. During RRBS data analysis deduplication is not recommended, because it could result in biases in the CpG representation. To overcome this issue 19 936 CpG sites with overlapping SNPs were removed and CpGs with extremely high coverage were filtered out for the correction. As was expected fewer informative CpGs could be identified in samples with lower quality. Due to the differences in surgical and postmortem FFPE specimens a higher mean CpG methylation rate (47.91%) was noted in CG2 compared to CG1 (32.31%).

CG1 data was abandoned cause of the extremely high level of DNA fragmentation (5.91% of DNA ≥2000 bp). Therefore, CG2 dataset was used as reference in all subsequent analyses. Overall, a shift toward hypomethylation was observed when comparing the controls and the sequential tumor samples. The mean CpG methylation levels were 47.91%, 41.34% and 31.6% in the CG2, GBM1 and GMB2, respectively. The methylation differences showed only a trend in the GBM1 vs CG2 comparison (Kruskal–Wallis test $p = 0.35$) but was significance in the GBM2 vs GBM1 ($p = 0.046$) and GBM2 vs CG2 ($p = 0.032$) comparisons.

Differential DNA methylation profiles in CG2, GBM1 and GBM2

The filtered and corrected data had a mean CpG site number of 60 169.48 and mean coverage of 366x. Apart from CpG sites, four regions were covered by the analyses like tiling, genes, promoters and CpG islands. Table 13. shows the detailed statistics of these regions. Group comparisons (CG2–GBM1, CG2–GBM2, and GBM1–GBM2) were focused only on differential methylation rates in gene promoters because the site and region levels revealed no FDR corrected p-values of ≤ 0.05. Detailed description of the results could be found in Tompa et al., 2018. Briefly, as the result of the GO analyses, hypermethylation was observed within promoter regions related to pathways of neuronal differentiation, morphogenesis, transcription and metabolic processes in GBM1 compared to CG2. The most significantly hypermethylated elements were linked to gastrulation regulation and cellular responses to the fibroblast growth factor. Other genes showed higher degrees of promoter methylation, but with lower degrees of significance.

| sampleName | sites num | sites covgMean | tiling num | tiling covgMean | tiling num SitesMean | genes num | genes covgMean | genes num SitesMean | promoters num | promoters covgMean | promoters num SitesMean | cpgislands num | cpgislands covgMean | cpgislands numSites Mean |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| 1R_RRBS | 50510 | 220,98 | 15098 | 732,92 | 0,1848 | 8586 | 1076,16 | 1,3675 | 4946 | 1094,57 | 0,9284 | 5057 | 1230,92 | 1,4566 |
| 1_RRBS | 30742 | 359,56 | 18856 | 579,07 | 0,1121 | 7748 | 1003,83 | 0,7016 | 1884 | 984,78 | 0,2188 | 1285 | 1438,76 | 0,3011 |
| 2R_RRBS | 30959 | 228,59 | 9199 | 761,67 | 0,1133 | 6013 | 988,53 | 0,8511 | 3402 | 1139,87 | 0,6125 | 3505 | 1234,35 | 0,9474 |
| 2_RRBS | 39535 | 259,68 | 10503 | 967,76 | 0,1445 | 6918 | 1262,41 | 1,1032 | 4406 | 1355,49 | 0,8333 | 4438 | 1513,95 | 1,2725 |
| 4R_RRBS | 38218 | 253,26 | 14835 | 645,49 | 0,1396 | 7966 | 983,17 | 0,9983 | 3858 | 1182,92 | 0,6145 | 3497 | 1357,90 | 0,8841 |
| 4_RRBS | 18117 | 277,09 | 5679 | 875,05 | 0,0662 | 3946 | 1081,85 | 0,4940 | 2085 | 1281,55 | 0,3426 | 2165 | 1375,27 | 0,5273 |
| 5R_RRBS | 57847 | 247,76 | 14352 | 988,46 | 0,2113 | 8761 | 1380,53 | 1,5945 | 5599 | 1375,77 | 1,1399 | 5890 | 1520,74 | 1,8105 |
| 5_RRBS | 43283 | 281,83 | 12055 | 1002,41 | 0,1583 | 7464 | 1377,59 | 1,1822 | 4451 | 1423,89 | 0,8299 | 4586 | 1499,06 | 1,2655 |
| 6R_RRBS | 53368 | 268,08 | 18837 | 752,11 | 0,1951 | 9513 | 1176,55 | 1,3935 | 5036 | 1154,92 | 0,8743 | 4735 | 1228,13 | 1,2972 |
| 6_RRBS | 111177 | 299,56 | 28171 | 1170,28 | 0,4064 | 13620 | 2014,69 | 3,0441 | 8977 | 1682,34 | 2,0405 | 8798 | 1890,77 | 3,1868 |
| 7R_RRBS | 3412 | 367,21 | 1624 | 760,60 | 0,0125 | 1117 | 892,78 | 0,0870 | 374 | 1294,15 | 0,0471 | 324 | 1813,19 | 0,0699 |
| 7_RRBS | 8420 | 384,32 | 3769 | 850,54 | 0,0308 | 2377 | 1013,99 | 0,2065 | 966 | 1308,47 | 0,1325 | 849 | 1459,82 | 0,1841 |
| 8R_RRBS | 38117 | 290,44 | 11215 | 977,61 | 0,1394 | 6967 | 1319,58 | 1,0315 | 3979 | 1405,86 | 0,7340 | 3992 | 1522,18 | 1,0977 |
| 8_RRBS | 27317 | 514,20 | 14134 | 982,16 | 0,0996 | 6906 | 1471,60 | 0,6621 | 2303 | 1471,44 | 0,3162 | 1869 | 1745,94 | 0,4397 |
| 9R_RRBS | 18029 | 407,66 | 7825 | 930,76 | 0,0659 | 4681 | 1212,12 | 0,4578 | 1964 | 1473,05 | 0,2758 | 1703 | 1673,87 | 0,3843 |
| 9_RRBS | 20207 | 434,13 | 8252 | 1049,90 | 0,0738 | 4791 | 1433,40 | 0,5254 | 1787 | 1750,77 | 0,2805 | 1775 | 1898,86 | 0,4355 |
| 10R_RRBS | 96021 | 287,65 | 19493 | 1404,02 | 0,3513 | 11896 | 2045,74 | 2,7435 | 8783 | 1909,30 | 2,1011 | 8598 | 2111,37 | 3,1481 |
| 10_RRBS | 71108 | 751,73 | 37724 | 1399,53 | 0,2594 | 12552 | 2875,73 | 1,6254 | 4640 | 2212,31 | 0,6551 | 3375 | 2719,11 | 0,8748 |
| 11R_RRBS | 25306 | 421,95 | 11512 | 917,43 | 0,0925 | 6276 | 1297,12 | 0,6414 | 2581 | 1376,30 | 0,3621 | 2255 | 1535,20 | 0,5067 |
| 11_RRBS | 27778 | 344,49 | 8921 | 1060,37 | 0,1014 | 5884 | 1348,56 | 0,7571 | 3141 | 1386,78 | 0,4920 | 3020 | 1522,92 | 0,7281 |
| 12R_RRBS | 15326 | 527,81 | 7520 | 1063,16 | 0,0560 | 4268 | 1431,14 | 0,3711 | 1476 | 1791,07 | 0,2011 | 1220 | 2124,16 | 0,2728 |
| 12_RRBS | 11805 | 578,29 | 6551 | 1031,23 | 0,0431 | 3700 | 1324,49 | 0,2787 | 1018 | 1627,21 | 0,1217 | 769 | 2054,10 | 0,1618 |
| 13R_RRBS | 82818 | 291,48 | 16698 | 1431,38 | 0,3029 | 10691 | 1970,68 | 2,3427 | 7856 | 1817,50 | 1,7680 | 7740 | 1958,58 | 2,6331 |
| 13_RRBS | 88785 | 248,03 | 21723 | 1004,54 | 0,3248 | 11858 | 1584,53 | 2,4451 | 7942 | 1421,61 | 1,6764 | 7908 | 1582,47 | 2,6066 |
| 14R_RRBS | 61306 | 384,07 | 14628 | 1594,19 | 0,2242 | 9160 | 2209,25 | 1,7117 | 6235 | 2076,15 | 1,2449 | 5918 | 2293,48 | 1,8319 |
| 14_RRBS | 32731 | 218,22 | 9720 | 727,17 | 0,1197 | 6295 | 957,71 | 0,8938 | 3701 | 1130,39 | 0,6563 | 3725 | 1175,22 | 0,9840 |
| 15R_RRBS | 44050 | 238,63 | 12073 | 861,43 | 0,1609 | 7452 | 1211,51 | 1,2069 | 4383 | 1201,65 | 0,8196 | 4348 | 1419,70 | 1,2579 |
| 15_RRBS | 11854 | 466,83 | 5587 | 979,12 | 0,0432 | 3321 | 1240,55 | 0,2912 | 1084 | 1682,84 | 0,1610 | 979 | 1802,21 | 0,2326 |
| 16R_RRBS | 61720 | 347,23 | 19366 | 1094,22 | 0,2254 | 10385 | 1685,57 | 1,6677 | 6131 | 1633,89 | 1,1163 | 5811 | 1764,14 | 1,6523 |
| 16_RRBS | 68191 | 340,55 | 28575 | 803,07 | 0,2490 | 11847 | 1478,68 | 1,7409 | 5956 | 1247,96 | 0,9765 | 5343 | 1437,77 | 1,4460 |
| 17R_RRBS | 24611 | 313,69 | 6458 | 1185,59 | 0,0901 | 4649 | 1429,63 | 0,6839 | 2823 | 1591,37 | 0,5110 | 2994 | 1685,21 | 0,7928 |
| 17_RRBS | 5047 | 953,14 | 3324 | 1433,37 | 0,0185 | 1853 | 1689,98 | 0,1097 | 293 | 1870,72 | 0,0301 | 185 | 2756,20 | 0,0395 |
| 18R_RRBS | 10764 | 272,53 | 3334 | 872,40 | 0,0394 | 2426 | 997,95 | 0,2848 | 1308 | 1273,57 | 0,2207 | 1354 | 1303,36 | 0,3252 |
| 18_RRBS | 24827 | 395,89 | 6150 | 1585,30 | 0,0909 | 4567 | 1829,62 | 0,6964 | 2913 | 1939,31 | 0,5147 | 3082 | 2065,22 | 0,8010 |
| 19R_RRBS | 217343 | 400,46 | 44399 | 1938,82 | 0,7943 | 17842 | 4008,00 | 5,8443 | 13416 | 3045,87 | 3,8547 | 12842 | 3383,32 | 5,9205 |
| 19_RRBS | 265825 | 468,59 | 56501 | 2182,79 | 0,9720 | 19713 | 5013,39 | 7,0185 | 14813 | 3353,04 | 4,3158 | 13639 | 3744,31 | 6,5372 |
| 20R_RRBS | 49194 | 312,05 | 10607 | 1434,51 | 0,1801 | 7204 | 1816,68 | 1,3574 | 5015 | 1767,30 | 1,0239 | 5110 | 1909,11 | 1,5676 |
| 20_RRBS | 12528 | 194,16 | 4669 | 515,85 | 0,0458 | 3143 | 652,97 | 0,3328 | 1562 | 917,81 | 0,2391 | 1409 | 1007,24 | 0,3198 |
| 21R_RRBS | 128 | 79,10 | 65 | 154,58 | 0,0005 | 44 | 156,66 | 0,0028 | 17 | 266,65 | 0,0019 | 13 | 355,54 | 0,0027 |
| 21_RRBS | 26017 | 378,72 | 7271 | 1342,36 | 0,0952 | 5222 | 1613,14 | 0,7294 | 3180 | 1689,07 | 0,5221 | 3204 | 1794,78 | 0,7834 |
| 22R_RRBS | 27010 | 408,52 | 8462 | 1292,35 | 0,0988 | 5457 | 1601,13 | 0,7088 | 3241 | 1662,41 | 0,5123 | 2995 | 1782,44 | 0,7251 |
| 22_RRBS | 16261 | 430,19 | 7634 | 902,52 | 0,0593 | 4401 | 1171,78 | 0,4026 | 1626 | 1448,89 | 0,2263 | 1388 | 1581,75 | 0,3133 |
| 24R_RRBS | 279444 | 440,42 | 56248 | 2164,87 | 1,0214 | 19911 | 5044,10 | 7,5322 | 15216 | 3621,80 | 4,9451 | 13839 | 3903,21 | 7,2098 |
| 24_RRBS | 400401 | 518,39 | 75191 | 2732,09 | 1,4635 | 22207 | 7578,81 | 10,7224 | 17712 | 4699,15 | 6,5458 | 16137 | 5344,45 | 9,9220 |
| **Mean** | **sites num** | **sites covgMean** | **tiling num** | **tiling covgMean** | **tiling num SitesMean** | **genes num** | **genes covgMean** | **genes num SitesMean** | **promoters num** | **promoters covgMean** | **promoters num SitesMean** | **cpgislands num** | **cpgislands covgMean** | **cpgislands numSites Mean** |
| **MEAN all** | 60169,48 | 366,07 | 16018,36 | 1116,71 | 0,2199 | 7763,59 | 1748,95 | 1,6101 | 4638,16 | 1660,04 | 1,0463 | 4401,55 | 1875,46 | 1,5718 |
| **Mean primary** | 61907,09 | 413,53 | 17316,36 | 1144,39 | 0,2262 | 7742,41 | 1864,51 | 1,6347 | 4383,64 | 1722,08 | 1,0058 | 4087,64 | 1973,19 | 1,5165 |
| **Mean recurrent** | 58431,86 | 318,62 | 14720,36 | 1089,03 | 0,2136 | 7784,77 | 1633,39 | 1,5855 | 4892,68 | 1598,00 | 1,0868 | 4715,45 | 1777,73 | 1,6270 |

**Table 13.** Sample coverage data in various regions (e.g., tiling, genes, promoter, CPG island) [Kraboth et al., 2020].

Additional, 17 different promoters in genes associated with nucleic acid-templated transcription had hypermethylation in GBM1 compared to CG2 (mean p = 0.0079). In 18 promoters associated with the regulation of different nucleobase-containing compound metabolic processes were hypermethylated (mean p = 0.0088). Moreover, there were 19 hypermethylated hits associated with pathways of neuron morphogenesis and differentiation in GBM1 vs CG2 comparison. Pathways with promoter hypomethylation in GBM1 compared to CG2 included genes that are related to synapse organization and assembly, neuronal ensheathment and endothelial cell proliferation.

The GBM2 vs CG2 comparison showed pathways with gene promoter hypermethylation associated with transcription regulation, cell adhesion and morphogenesis and embryonic development. Pathways which showed the most significant hypermethylation in promoters were associated with appendage morphogenesis and limb. Pathways with hypomethylated gene promoters in GBM2, compared to CG2, included a few associated with purine and pyrimidine nucleobase transports, Golgi transports and allantonin catabolic processes.

Comparing GBM1 to GBM2, the GO analysis identified several pathways of biological relevance. Pathways with hypermethylation in the recurrent compared to the primary tumors included genes related to regulation of the Wnt pathway, catecholamine secretion and transport, and cellular response, signaling and communication. Pathways with promoter hypomethylation in the GBM2 compared to the GBM1 included genes related to both the innate and adaptive immune responses, cellular processes and cell differentiation. The most significant p values were noted in pathways linked to the regulation of lymphocyte-mediated immunity, natural killer (NK) cell-mediated cytotoxicity and regulation of cell killing.

Figure 17. summarizes the potential mechanisms that play significant role in the GBM development and recurrence based on the results of the methylation analysis.

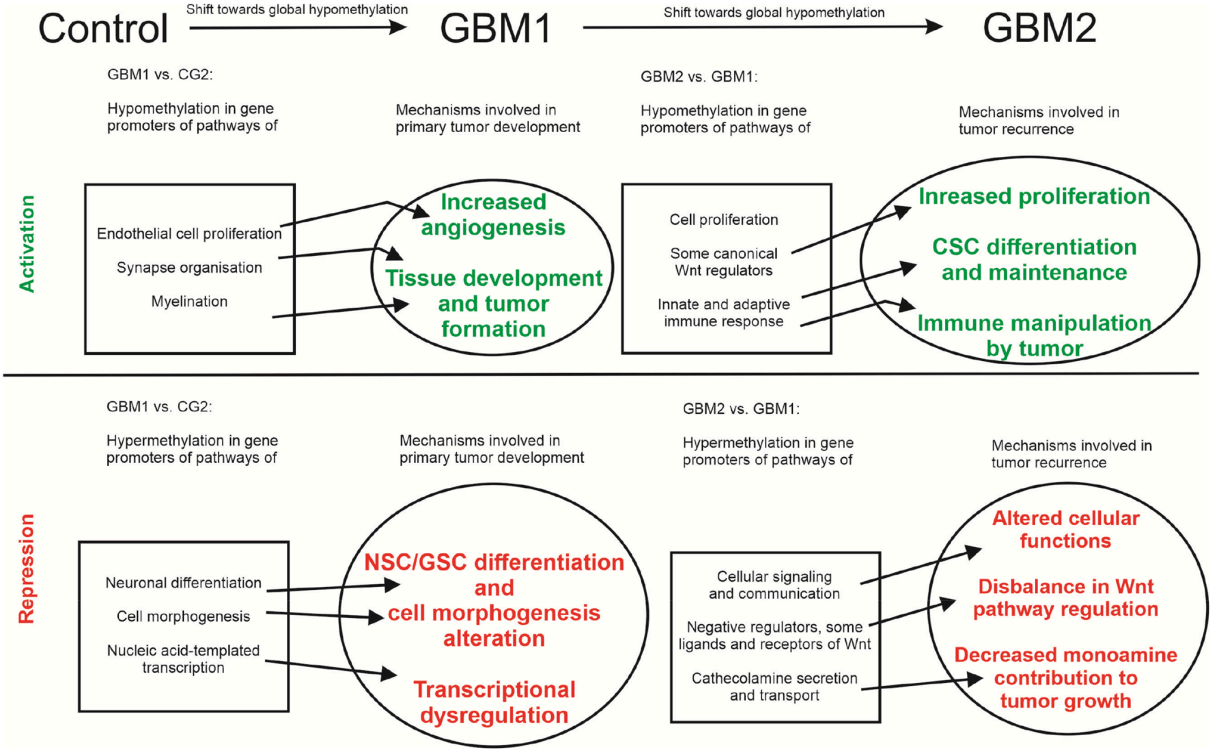



**Figure 17.** Mechanisms of GBM development and recurrence revealed by DNA CpG methylation. This figure provides a schematic depiction of molecular pathways and potential mechanisms contributing to GBM development and recurrence as revealed by RRBS of sequential GBM specimens [Kraboth et al., 2020].

## Enrichment analysis

The LOLA program was run to enrich for genomic region sets and regulatory elements relevant to the interpretation of functional epigenomics data [Sheffield and Bock, 2016]. Results of the top-ranking 1000 hypomethylated tiling regions were used. In both the CG2 vs GBM1/GBM2 comparisons, strong enrichment was identified in hypomethylated regions in the tumors for binding sites of transcription factors and histone proteins relevant to proper embryonic stem cell differentiation and lineage fidelity maintenance. In the GBM1 vs GBM2 comparison, GBM2 group showed enrichment in binding sites for transcription factors and histone proteins among the hypomethylated regions.

Cross-platform validation

Because the lack of available GBM dataset array-based DNA CpG methylation data of matched GBM samples (n = 12 pairs) from The Cancer Genome Atlas (TCGA) [https://www.cancer.gov/about-nci/organization/ccg/research/structural-genomics/tcga] was downloaded and used in a cross-platform analysis thanks to the RnBeads tool [Müller et al., 2019]. The results of the analysis revealed that promoters in genes of pathways involved in embryonic development, immune regulation and Wnt signaling were less methylated in the TCGA GBM samples than in the CG2 controls. The results of the second analysis showed less methylated promoters in genes of pathways involved in stem cell proliferation and cell dedifferentiation, intracellular regulatory and metabolic processes, negative regulation of apoptosis, cell adhesion and T cell polarity as well as migration in the TCGA recurrent compared to the primary samples. In contrast, promoters of genes in pathways involved in endothelial cell proliferation, negative regulation of the execution phase of apoptosis, T cell proliferation, cell–cell signaling, neuronal differentiation, and regulation of G protein-mediated signaling (including neurotransmitter, catecholamine and some Wnt receptor signaling, though with lower ranking in the list) were less methylated in the TCGA primary than in the recurrent samples.

Considering the technical limitations and interpretive difficulties when comparing data from various platforms and results from small cohorts, the outcome of the TCGA sample analyses is supporting the previously mentioned conclusions.

Correlation on clinical data

No association was detected between T1–T2 and gender or the age of patients, or T1–T2 and morphological subtype, mitotic rate, microvascular proliferation or necrosis of the tumors. However, a trend for association was found between T1–T2 and the amount of tumor infiltrating lymphocytes (TIL) in the GBM1 samples (Kruskal–Wallis test $p = 0.08$), but not in the GBM2 samples ($p = 0.737$). Neither Mann–Whitney nor Pearson's correlation analysis showed a link between TIL and mitotic rate.

*miRNome profiling*

***CASE STUDY IV.: NSCLC***

Data quality control

Quality control is a crucial step in the analysis to assess the overall quality of the samples, to see how well the replicates correlate with each other and to identify possible outliers. Here, quality control results of the "classification" dataset (Set 1; AC, n = 26; AC_c, n = 25; SCC, n = 30; SCC_c, n = 28, including matched control samples) will be introduced. Overall, more than 1600 miRNA were identified out of 2571 in the entire dataset. Elements with extreme low or 0 abundance were filtered out.

Between sample correlation values describe the similarity between the samples in a general level, when all measurement features of all samples are taken into consideration. In this analysis the so-called Spearman's metrics is used which describes the between sample similarity on a scale of 0-1. Value 0 means perfect uncorrelation between the samples whereas value 1 means perfect correlation between them (Figure 18., Table 14).

| GroupName | minCor | meanCor | medianCor | maxCor | corSD |
|---|---|---|---|---|---|
| **SCC** | 0.692 | 0.795 | 0.802 | 0.843 | 0.029 |
| **SCC_c** | 0.761 | 0.822 | 0.825 | 0.865 | 0.02 |
| **AC** | 0.703 | 0.791 | 0.795 | 0.837 | 0.024 |
| **AC_c** | 0.805 | 0.836 | 0.837 | 0.863 | 0.011 |

**Table 14.** Correlation values. AC_c control samples of AC group, SCC_c control samples of SCC group.

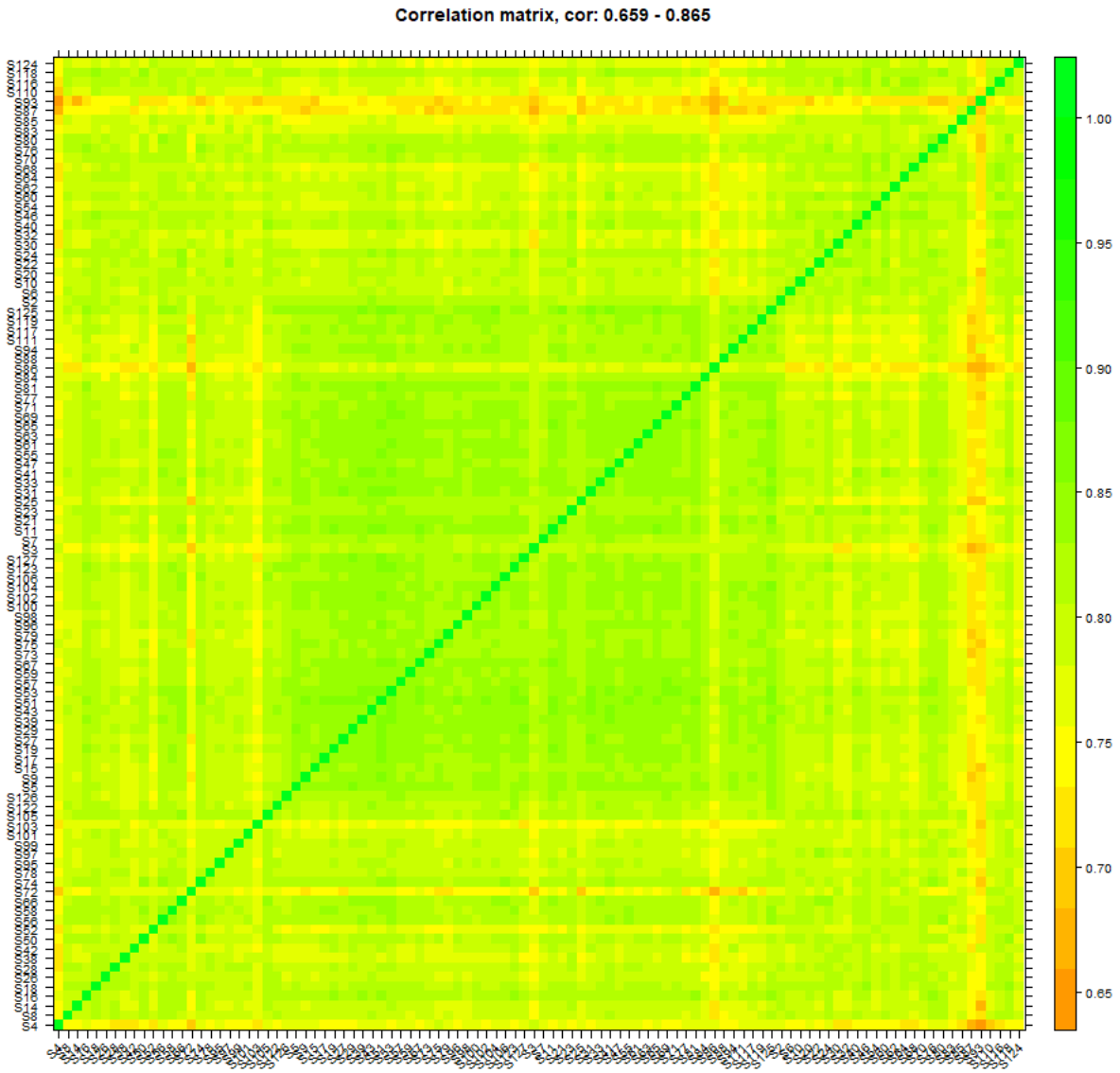



**Figure 18.** Correlation plot of the test "classification" dataset. Colours from orange to green indicating the correlation values (0.65-1)

Based on the analysis result the mean correlation for the AC and SCC samples (including control ones) are a little bit below the expectation (mean = 0.81). Considering the nature of the human dataset the overall correlation rates were good and only there are just a few possible outliers. These samples (e.g., S87 and S93) are noticeable on Figure 18. with a lot of darker orange squares compare to rest of the plot.

Next, in hierarchical clustering the samples are grouped according to their general similarity when all the measurements of all the samples are taken into consideration. Here, the samples were clustered with Euclidean metrics. The result of the cluster analysis can be visualized as a dendrogram, which is an out-branching graph where the most similar samples (in other words best correlating) can be found in the branches that are nearest to one another.

Dendrograms produced by cluster analysis for reads mapped to miRNA features are shown on the Figure 19. below.

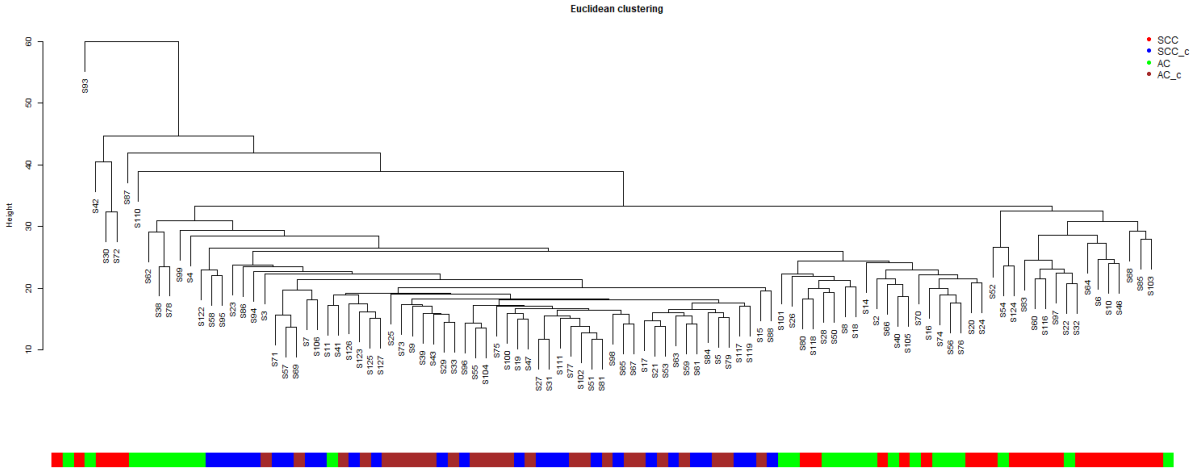



**Figure 19.** Hierarchical clustering results for AC and SCC samples.

The control samples are grouped in one big subcluster but with higher variation within the cluster, means the separation is not that clear for the AC_c and SCC_c samples. Rest of the tumor samples are in separate clusters on the two side of the dendrogram. Also, the previously mentioned possible outlier samples are alone in different branches, specially S93 (top left of the dendrogram).

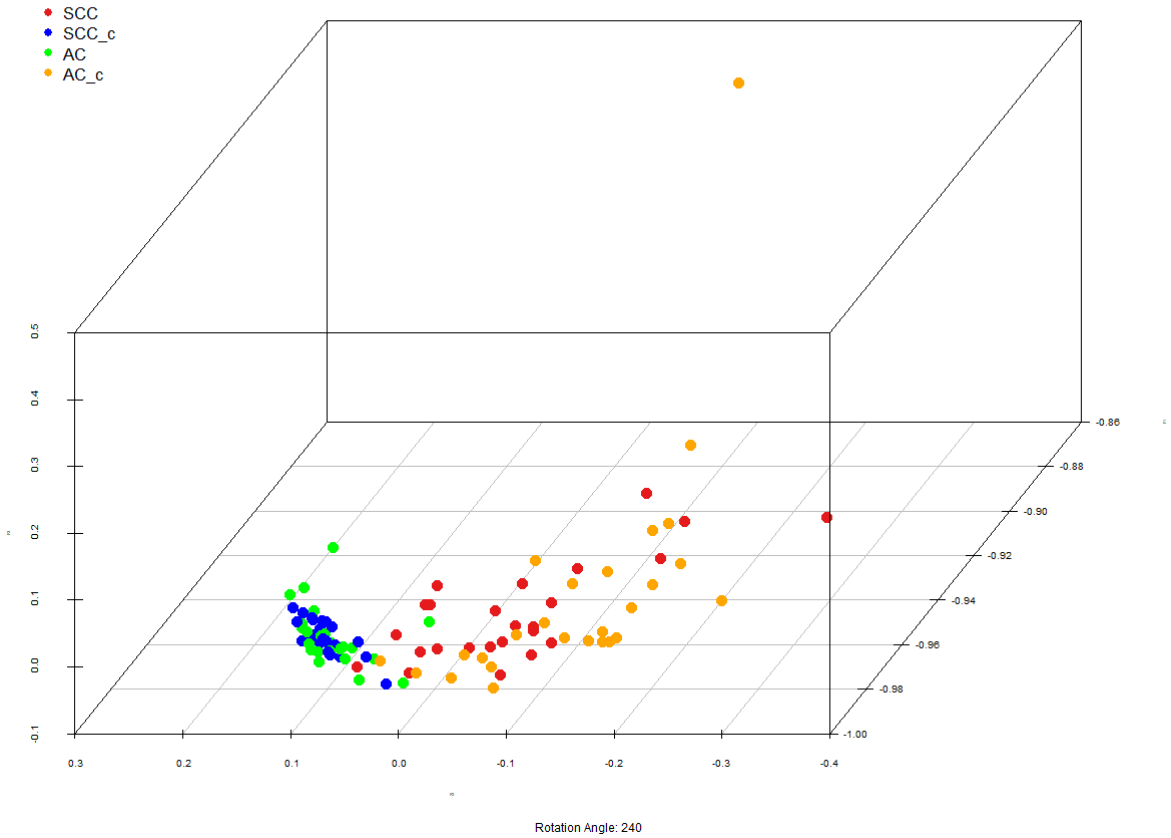



**Figure 20.** PCA plot for results of AC and SCC samples.

The sample relations can also be studied by the means of a principal component analysis (PCA) which is an ordination technique complementary to clustering. Ordination orders objects so that similar objects are placed near each other, and dissimilar objects are placed further from each other. In PCA analysis the sample relationships can be visualized in three-dimensional space (Figure 20.). The rotation angle is important because the plot itself and interpretation could be different. Based on patters on the PCA plot, from the actual rotation angle (240°), the control and tumor samples from the AC group are forming two separate clusters (green and orange dots). This is true for the SCC group as well (blue and red dots). Again, the possible outlier samplers are clearly visible. However, there is an overlap in the opposite direction between the groups than expected orientation. On the plot AC_c + SCC and AC + SCC_c dots are overlapping but we expected that the control sample will overlap. This attribute could be explained by the rotation angle and the higher variability within the real human dataset.

Differential expression analysis

Three different group comparisons (AC vs AC_c; SCC vs SCC_c; AC vs SCC) were carried out to get more information about the molecular background of NSCLC. When filtering up- and down-regulated (i.e., differentially expressed = DE) miRNAs between certain conditions (groups) fold changes and corrected p-values calculated during statistical testing were used as filtering criteria. All the measured miRNAs are filtered to list those that show the strongest evidence for being differentially expressed between the compared groups.

Fold change (FC) describes the size of the difference in gene expression between the compared groups. It is the results from linear modelling process performed with Limma package. Fold changes are often expressed as log2-transformed, where value 0 means 'no change' and 1 means doubled value and -1 means halved value. The values are always in relation to the group used as a base level group (reference or control).

The choice of the thresholds for p-value and fold change used for filtering the differentially expressed (DE) genes is not a trivial task. There is no one correct way or method to determine the thresholds, but the choice is based on various aspects of each study and comparisons. Different thresholds can also be used for filtering the data for different purposes. For example, often strict thresholds are chosen when the data is filtered to be included in a publication. Then the result list will contain very few false positive findings but

on the other hand many true positives are left outside the result set. Because of this, it is typically useful to use less stringent thresholds for filtering data for internal research purposes or functional analysis when a larger proportion of possible false positive findings can be tolerated. Table 15 shows the used filtering criteria and the number of identified up- and down-regulated DE miRNAs in details for each comparison.

| Comparison | FC | logFC | PType | P | Total | Up | Down |
|---|---|---|---|---|---|---|---|
| AC_vs_AC_c | 1.5 | 0.58 | adj.P.Val | 0.05 | 162 | 83 | 79 |
| SCC_vs_SCC_c | 1.5 | 0.58 | adj.P.Val | 0.05 | 232 | 138 | 94 |
| AC_vs_SCC | 1.3 | 0.38 | adj.P.Val | 0.05 | 31 | 4 | 27 |

**Table 15.** Filtering parameters and the number of DE elements.

A total number of 162, 232 and 31 DE miRNAs were identified from the AC vs AC_c, SCC vs SCC_c and AC vs SCC (tumor only) comparisons, respectively. The top 10 most significant elements based on average ranking value (based on both p value and fold change) are listed in Table 16.

| AC vs AC_c | SCC vs SCC_c | AC vs SCC |
|---|---|---|
| hsa-miR-490-3p | hsa-miR-451a | hsa-miR-944 |
| hsa-miR-144-3p | hsa-miR-144-3p | hsa-miR-205-5p |
| hsa-miR-451a | hsa-miR-4652-5p | hsa-miR-383-5p |
| hsa-miR-144-5p | hsa-miR-7974 | hsa-miR-3927-3p |
| hsa-miR-9-5p | hsa-miR-486-5p | hsa-miR-448 |
| hsa-miR-9-3p | hsa-miR-30a-3p | hsa-miR-3617-5p |
| hsa-miR-451b | hsa-miR-144-5p | hsa-miR-1911-5p |
| hsa-miR-30a-3p | hsa-miR-135a-5p | hsa-miR-1224-5p |
| hsa-miR-486-5p | hsa-miR-3180-3p | hsa-miR-205-3p |
| hsa-miR-196a-5p | hsa-miR-3180 | hsa-miR-6510-3p |

**Table 16.** Top 10 significant miRNAs based on average ranking.

The following Figure 21. show the results of the comparisons as Volcano plots. In a volcano plot the log10 of the p-values is on the y axis and the logFC calculated for the comparison group vs. base level group is on the x axis. In this plot it can be seen how the reliability values of the measurement features behave in relation to the fold change.

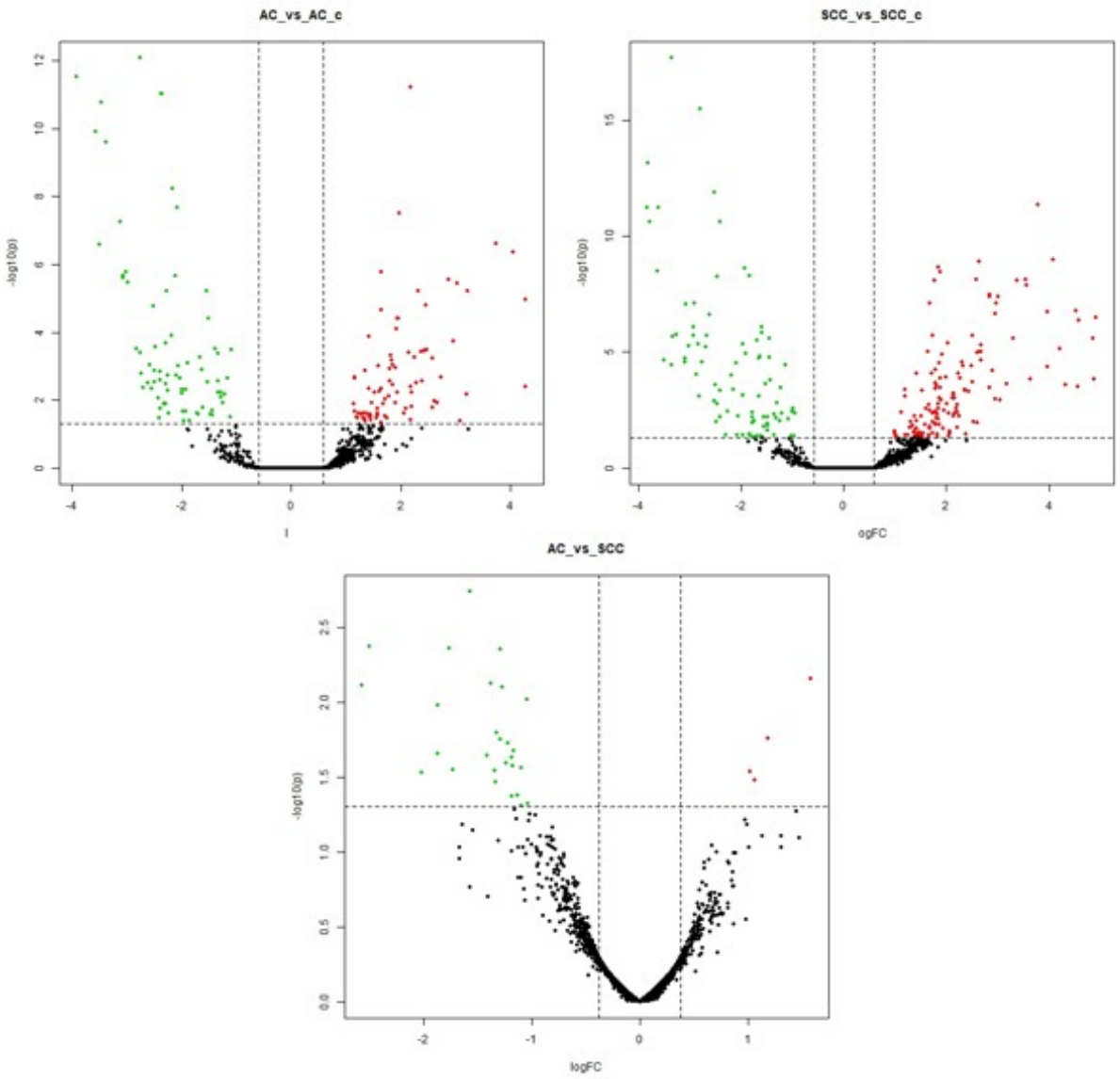



**Figure 21.** Volcano plots. The thresholds used in the filtering are marked in the plot with dashed lines, up-regulated genes are coloured red and down-regulated green.

There was a significant overlap between certain group comparisons. The “AC vs AC_c“ and “SCC vs SCC_c” dataset shares all together 121 miRNAs (Figure 22). The “The AC vs AC_c“ and “SCC vs SCC_c“ have 111 and 41 unique elements respectively for each group. In contrast with that, the “AC vs SCC” results share only one miRNA with “AC vs AC_c” and 16 miRNAs with “SCC vs SCC_c” out of 31 (data not shown).

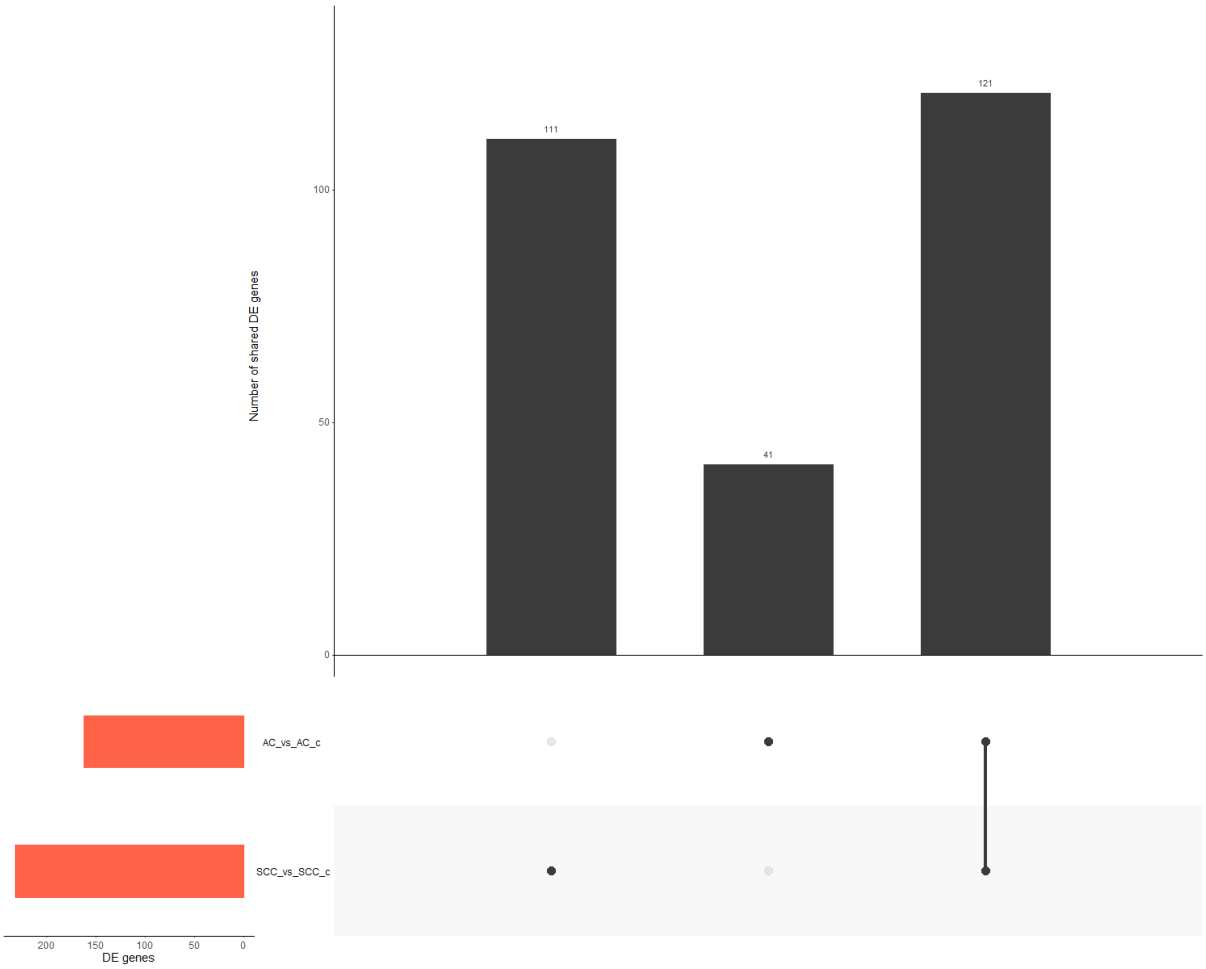



**Figure 22.** Shared DE miRNAs between the "AC vs AC_c" and "SCC vs SCC_c" dataset. Orange bars indicating the total number of DE elements, dark grey bars representing the number of unique and shared elements.

Cluster analysis of the filtered genes can also be used as a means for choosing the filtering thresholds. Such thresholds should be chosen, that the samples are grouping according to the known sample groups in the cluster analysis of the filtered genes. The result could be visualized on specific heatmaps. Figure 23. shows the heatmap clustering of the differentially expressed features for the comparisons. Pearson's metrics has been used in hierarchical clustering of the samples and filtered features. The clustering is based on the general expression measurement similarity. In the plot red colour means high expression and blue low expression. Each row represents one DE feature, and each column represents one sample. For "AC vs AC_c" and "SCC vs SCC_c" comparisons the sample separation is perfect, means the select thresholds were good and maybe more strict parameters could be applied. For the last analysis (Ac vs SCC) sample clustering and group separation is not that clear, compared to previous results, but still acceptable. These findings indicating that, the two groups are a little bit similar and the DE miRNAs not explaining enough variability.

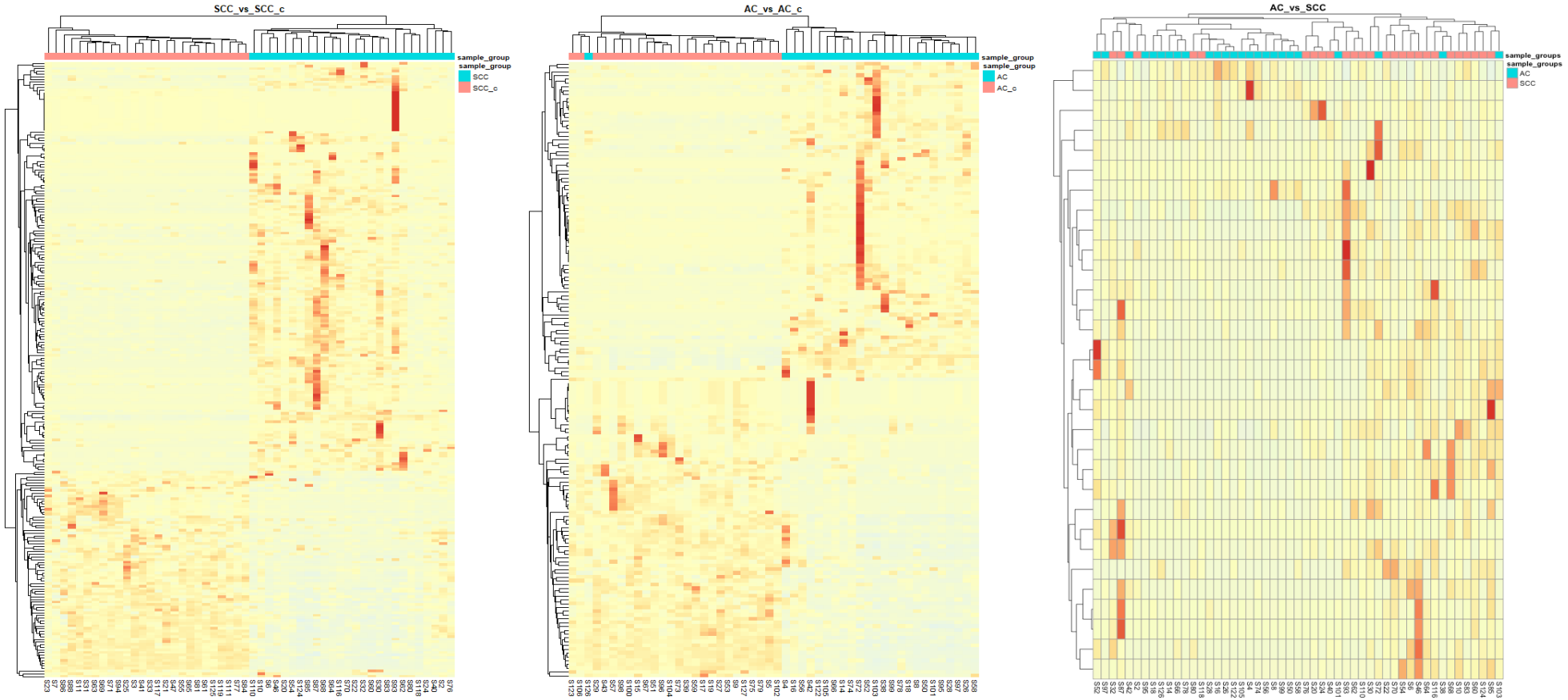



**Figure 23.** Heatmaps for all comparisons.

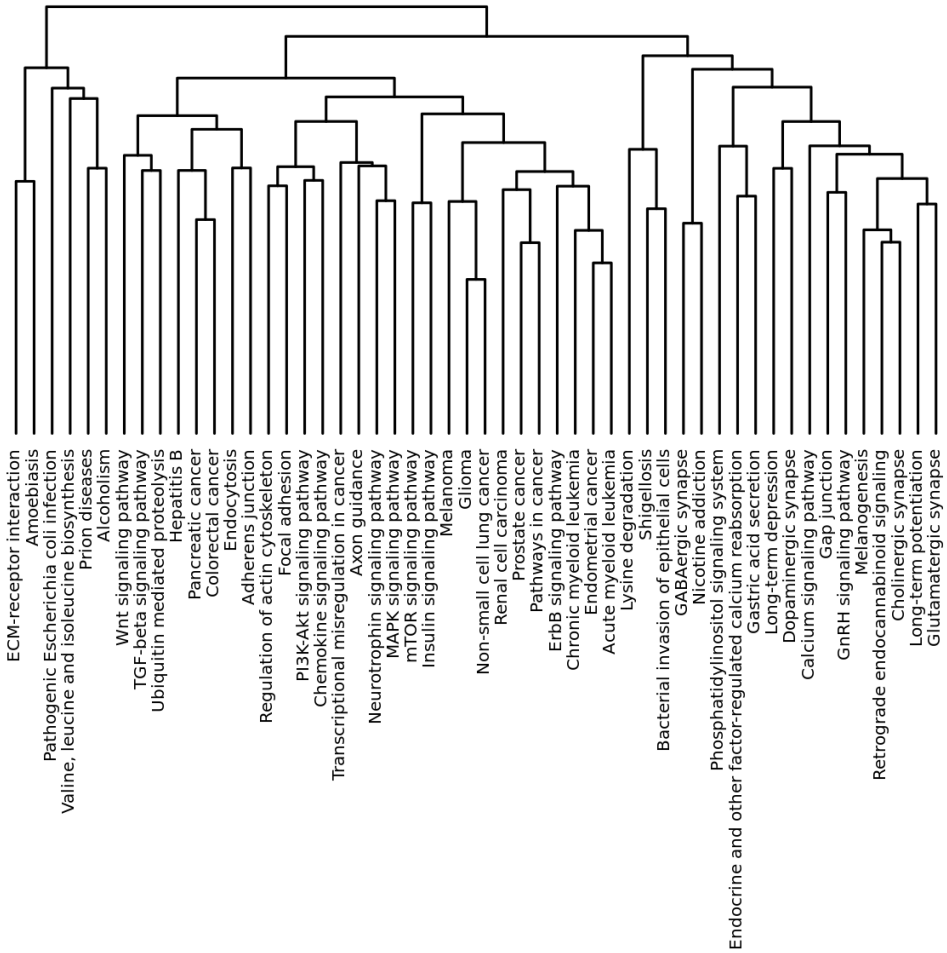
Enrichment analysis

The online available mirPath v.3 from DIANA tools was used to carry out enrichment analysis for KEGG pathways based on the miRNAs target gene results. Some of the miRNAs were excluded from the analysis because they were not included in the database. For “AC vs AC_c“ and “SCC vs SCC_c” comparisons analysis was done separately for the up- and down-regulated miRNAs. For “AC vs SCC” all elements were used in one search. Overall, 52, 180 and 81 unique pathways were significantly affected (corrected p-value $\leq 0.01$) in “AC vs AC_c“, “SCC vs SCC_c” and “AC vs SCC” comparisons, respectively. Most of the pathways are related to cancer or potentially connected with the disease development. Table 17. shows the top 10 pathways from each comparison. The DIANA algorithm can cluster together microRNAs targeting similar lists of pathways, as well as pathways, which are targeted by similar lists of microRNAs (Targeted Pathways Clusters) or take also into account the significance levels of the interactions (Significance Clusters) during the clustering process. This is especially useful during the interpretation phase. Figure 24. Represent the result of the clustering for all comparisons.

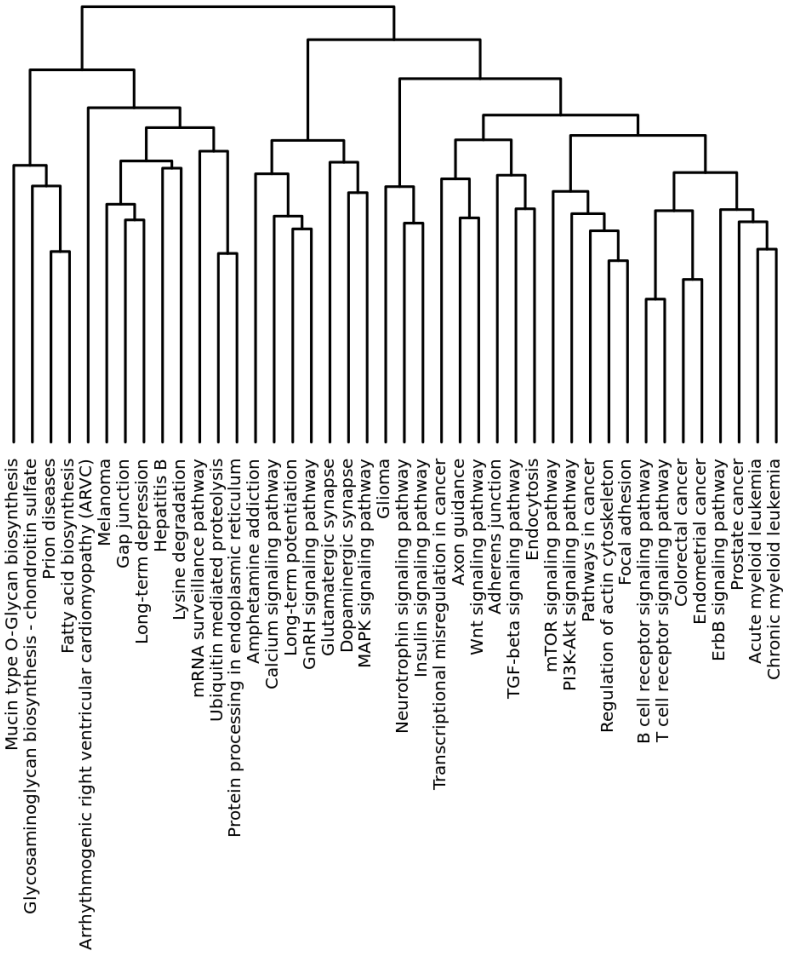
| AC vs AC_c UP | AC vs AC_c DOWN | SCC vs SCC UP | SCC vs SCC DOWN | AC vs SCC |
|---|---|---|---|---|
| ECM-receptor interaction | Prion diseases | Phenylalanine metabolism | Prion diseases | ErbB signaling pathway |
| Transcriptional misregulation in cancer | Ubiquitin mediated proteolysis | Sulfur relay system | Fatty acid biosynthesis | Endocytosis |
| PI3K-Akt signaling pathway | Dopaminergic synapse | Retinol metabolism | MAPK signaling pathway | Prostate cancer |
| TGF-beta signaling pathway | Axon guidance | Synaptic vesicle cycle | Ubiquitin mediated proteolysis | Wnt signaling pathway |
| Neurotrophin signaling pathway | Wnt signaling pathway | Morphine addiction | Dopaminergic synapse | Chronic myeloid leukemia |
| Regulation of actin cytoskeleton | Prostate cancer | Pancreatic secretion | Axon guidance | MAPK signaling pathway |
| Prostate cancer | Long-term potentiation | Lysine degradation | Long-term potentiation | Axon guidance |
| Gap junction | Neurotrophin signaling pathway | Calcium signaling pathway | Prostate cancer | Endometrial cancer |
| mTOR signaling pathway | TGF-beta signaling pathway | Regulation of actin cytoskeleton | PI3K-Akt signaling pathway | PI3K-Akt signaling pathway |
| ErbB signaling pathway | MAPK signaling pathway | Hematopoietic cell lineage | Wnt signaling pathway | Long-term potentiation |

**Table 17.** Top 10 significant KEGG pathways

(a)


ECM-receptor interaction
Amoebiasis
Pathogenic Escherichia coli infection
Valine, leucine and isoleucine biosynthesis
Prion diseases
Alcoholism
Wnt signaling pathway
TGF-beta signaling pathway
Ubiquitin mediated proteolysis
Hepatitis B
Pancreatic cancer
Colorectal cancer
Endocytosis
Adherens junction
Regulation of actin cytoskeleton
Focal adhesion
PI3K-Akt signaling pathway
Chemokine signaling pathway
Transcriptional misregulation in cancer
Axon guidance
Neurotrophin signaling pathway
MAPK signaling pathway
mTOR signaling pathway
Insulin signaling pathway
Melanoma
Glioma
Non-small cell lung cancer
Renal cell carcinoma
Prostate cancer
Pathways in cancer
ErbB signaling pathway
Chronic myeloid leukemia
Endometrial cancer
Acute myeloid leukemia
Lysine degradation
Shigellosis
Bacterial invasion of epithelial cells
GABAergic synapse
Nicotine addiction
Phosphatidylinositol signaling system
Endocrine and other factor-regulated calcium reabsorption
Gastric acid secretion
Long-term depression
Dopaminergic synapse
Calcium signaling pathway
Gap junction
GnRH signaling pathway
Melanogenesis
Retrograde endocannabinoid signaling
Cholinergic synapse
Long-term potentiation
Glutamatergic synapse

(b)


Mucin type O-Glycan biosynthesis
Glycosaminoglycan biosynthesis - chondroitin sulfate
Prion diseases
Fatty acid biosynthesis
Arrhythmogenic right ventricular cardiomyopathy (ARVC)
Melanoma
Gap junction
Long-term depression
Hepatitis B
Lysine degradation
mRNA surveillance pathway
Ubiquitin mediated proteolysis
Protein processing in endoplasmic reticulum
Amphetamine addiction
Calcium signaling pathway
Long-term potentiation
GnRH signaling pathway
Glutamatergic synapse
Dopaminergic synapse
MAPK signaling pathway
Glioma
Neurotrophin signaling pathway
Insulin signaling pathway
Transcriptional misregulation in cancer
Axon guidance
Wnt signaling pathway
Adherens junction
TGF-beta signaling pathway
Endocytosis
mTOR signaling pathway
PI3K-Akt signaling pathway
Pathways in cancer
Regulation of actin cytoskeleton
Focal adhesion
B cell receptor signaling pathway
T cell receptor signaling pathway
Colorectal cancer
Endometrial cancer
ErbB signaling pathway
Prostate cancer
Acute myeloid leukemia
Chronic myeloid leukemia

(d)


GABAergic synapse
Prion diseases
Fatty acid biosynthesis
Glycosaminoglycan biosynthesis - chondroitin sulfate
Insulin signaling pathway
Neurotrophin signaling pathway
Glioma
Long-term potentiation
GnRH signaling pathway
MAPK signaling pathway
Dopaminergic synapse
Phosphatidylinositol signaling system
Calcium signaling pathway
Amphetamine addiction
PI3K-Akt signaling pathway
mTOR signaling pathway
Focal adhesion
Pathways in cancer
p53 signaling pathway
Prostate cancer
Melanoma
Chronic myeloid leukemia
Non-small cell lung cancer
Endometrial cancer
ErbB signaling pathway
Colorectal cancer
Pancreatic cancer
Type II diabetes mellitus
Hepatitis B
B cell receptor signaling pathway
T cell receptor signaling pathway
mRNA surveillance pathway
Ubiquitin mediated proteolysis
Protein processing in endoplasmic reticulum
Glutamatergic synapse
Gap junction
Long-term depression
TGF-beta signaling pathway
Wnt signaling pathway
Regulation of actin cytoskeleton
Axon guidance
Transcriptional misregulation in cancer
Arrhythmogenic right ventricular cardiomyopathy (ARVC)
Hypertrophic cardiomyopathy (HCM)
Lysine degradation
Endocytosis

(c)


ECM-receptor interaction
Alcoholism
Metabolism of xenobiotics by cytochrome P450
Glycosaminoglycan biosynthesis - chondroitin sulfate
Systemic lupus erythematosus
Transcriptional misregulation in cancer
MAPK signaling pathway
Ubiquitin mediated proteolysis
Axon guidance
Lysine degradation
Wnt signaling pathway
Hedgehog signaling pathway
PI3K-Akt signaling pathway
Focal adhesion
Pathways in cancer
Prostate cancer
Endocytosis
RNA transport
mRNA surveillance pathway
GABAergic synapse
Long-term depression
Gap junction
Long-term potentiation
Neurotrophin signaling pathway
ErbB signaling pathway

(e)

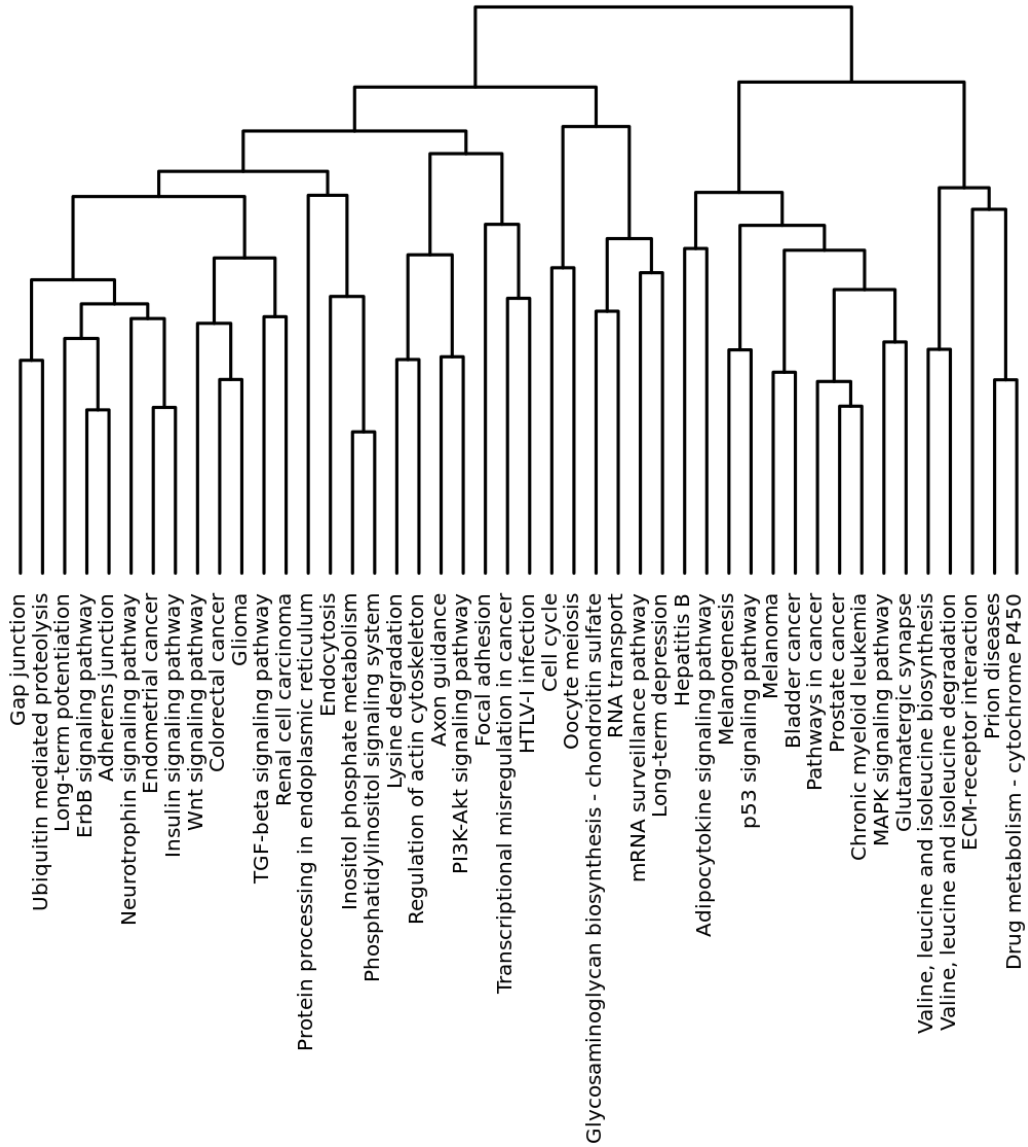



**Figure 24.** Targeted Pathways Clusters for the following comparisons: (a) AC vs AC_c UP; (b) AC vs AC_c DOWN; (c) SCC vs SCC_c UP; (d) SCC vs SCC_c DOWN; (e) AC vs SCC

One miRNA could target several genes and several miRNAs could target the same gene. Due to this fact and the high number of shared DE miRNAs between the comparisons there is a huge overlap between the results of the enrichment analyses (Figure 25.). There are 88 unique KEGG pathways that could be find in “SCC vs SCC_c” upregulated dataset. Also, the overlap is compelling between the “SCC vs SCC_c” and “AC vs SCC” groups (n = 37) and there are additional 24 pathways that is shared across all datasets. In the rest of the comparisons only 2 – 6 pathways are shred. Pathways that were affected both in the up- or down-regulated miRNA set within the same histological group should be removed from the results because it is difficult to predict which effect will prevail.

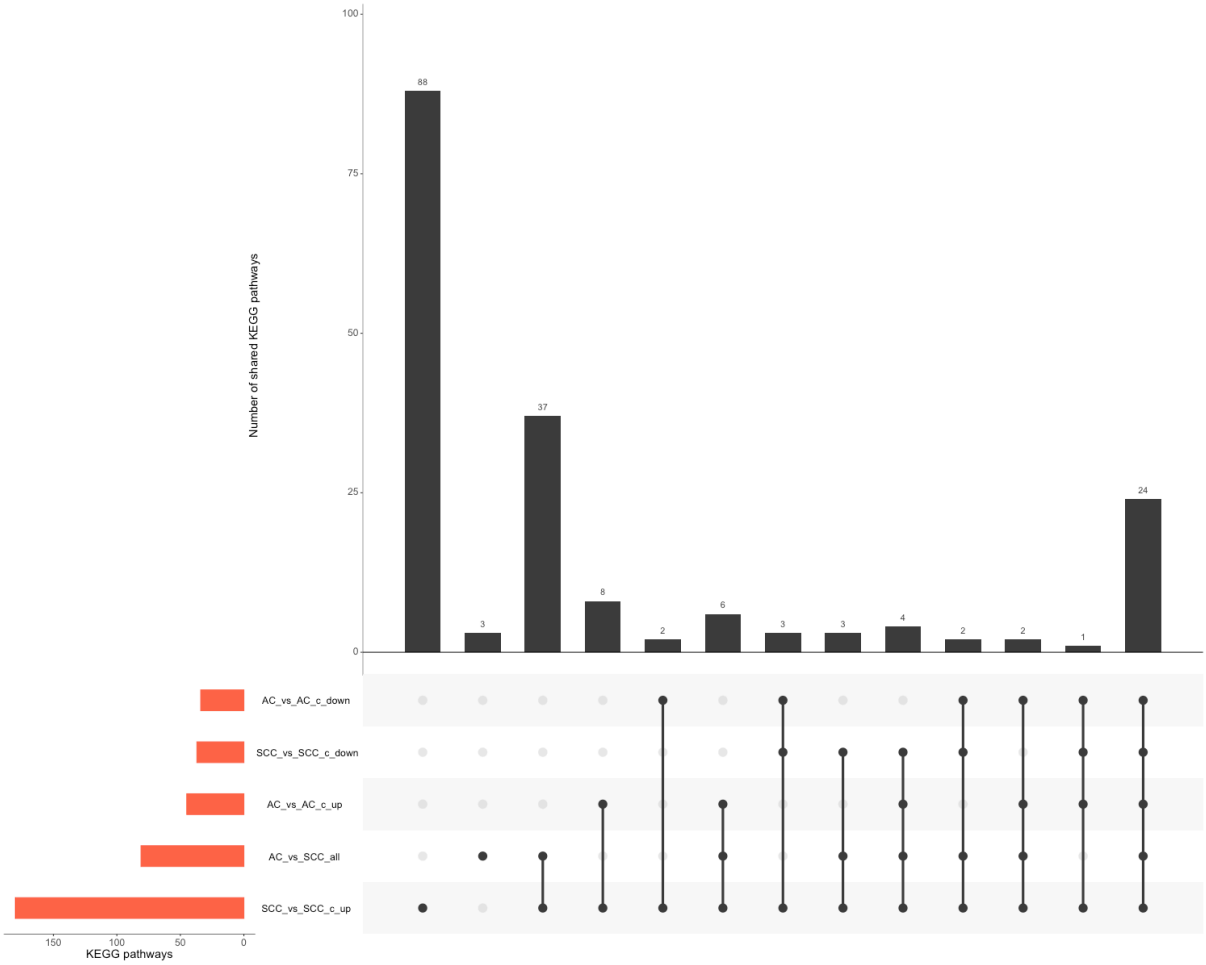



**Figure 25.** Number of shared KEGG pathways. Black dots marking the actual group and if there is a common element dots are connected

Development of the predictive model

Using the CAMPP toolset first 10 miRNAs were selected randomly to check the data distribution (Figure 26.) According to all data checks the input was normally distributed and good for further processing.

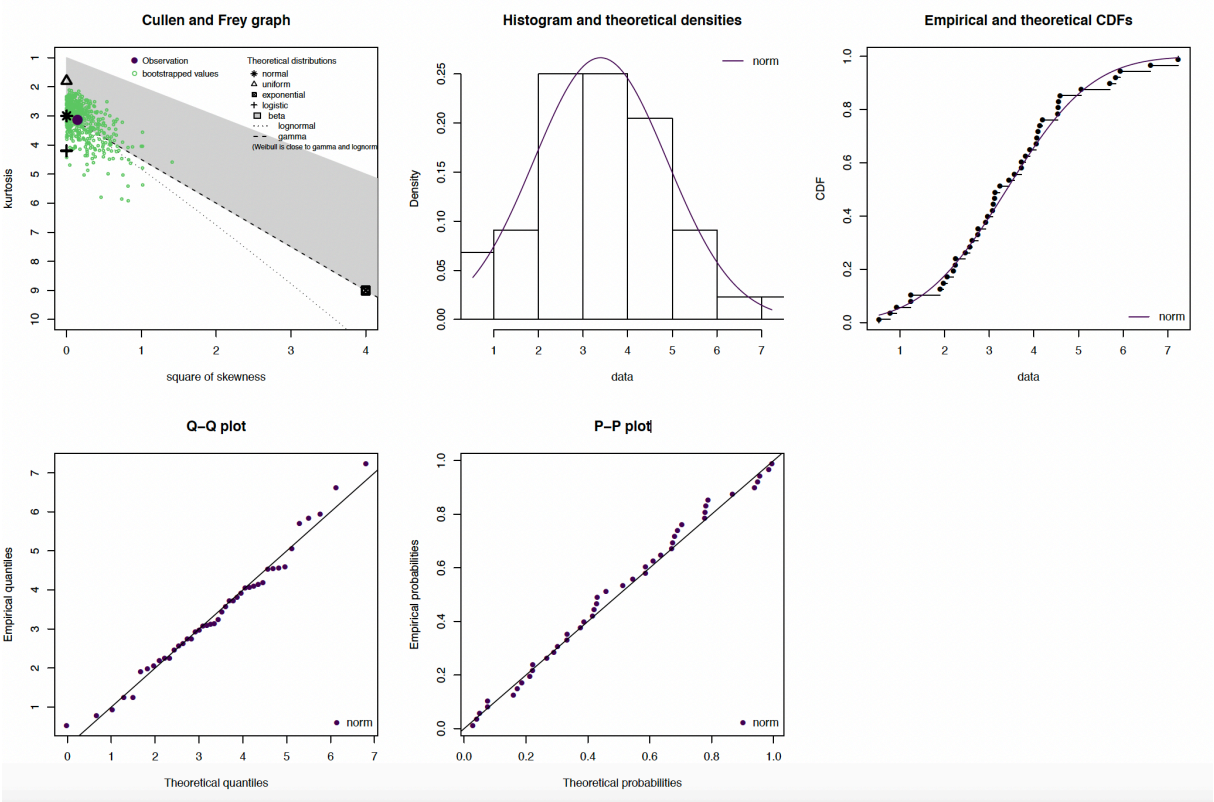



**Figure 26.** Data distributional check for randomly selected hsa_miR_138_5p miRNA from the dataset.

Overall, 17 miRNAs were identified as possible biomarkers (e.g., hsa_miR_1287_5p, hsa_miR_147b, hsa_miR_149_5p, hsa_miR_205_3p, hsa_miR_205_5p, hsa_miR_30b_5p, hsa_miR_326, hsa_miR_375, hsa_miR_450a_1_3p, hsa_miR_450a_5p, hsa_miR_4728_3p", hsa_miR_542_3p, hsa_miR_556_5p, hsa_miR_6510_3p, hsa_miR_653_3p, hsa_miR_7705, and hsa_miR_944). Multidimensional Scaling Plot (MDS) shows the separation of AC and SCC tumor samples based on miRNAs abundances (Figure 27). The components M1 and M2 in the plot below are those which best retained the distance relationship between samples in two dimensions. Based on these finding the groups separation is quite good, however the border is not that clear.

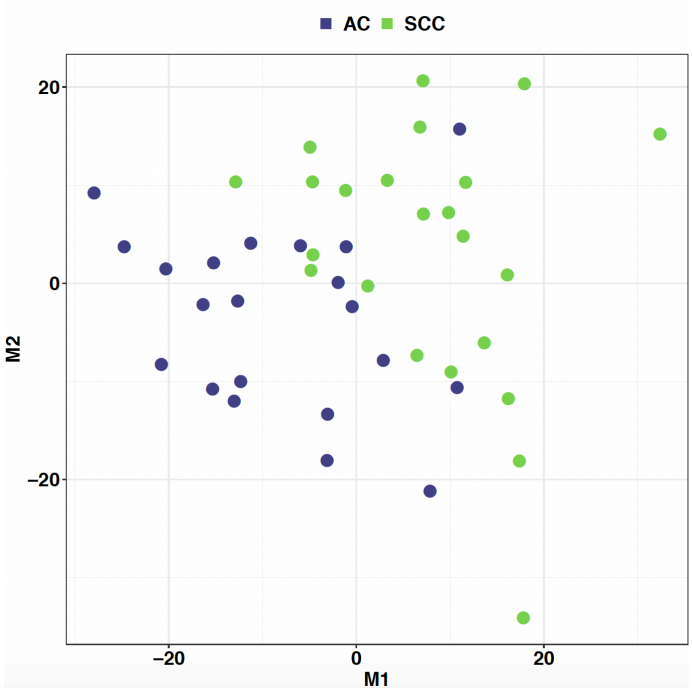



**Figure 27.** Multidimensional Scaling Plot.

In order to confirm the reliability of the identified histotypic-associated miRNAs, decision tree based classification model was built using 17 miRNAs. The fitted model was used for prediction in the blood dataset. The prediction accuracy was 96% in the training (tissue) and 74% in the test (blood) datasets, respectively. As expected, the model performs well on the training dataset. At the same time the model underachieves on blood data. The reason behind this that the abundance levels of the miRNAs in tissue and blood dataset are more different. For example, some of the miRNAs that are highly expressed in tissue they may not appear in the blood, or it is hard to sequence them. Therefore, the model needs further optimization, and the datasets need cleaning and scaling, focusing on the protentional huge differences in the abundances.

*Assisted reproduction*

***CASE STUDY V.: NIPGT-A***

NGS and data analysis

In prior to better understand the analysis outcome raw data and mapping QC results have been carefully investigated. After the successful WGA of 28 samples out of 40, the sequencing resulted in an average 12M SE50 reads per sample. Based on the QC analysis result (Figure 28a.) the dataset had high-quality sequences (as usually Illumina datasets). The sequence duplication level was generally low for a few samples but for rest of the cohort it was a little bit higher, but still acceptable, as depicted in yellow on Figure 28b. These marked samples were part of the control and the culture media droplets of healthy neonate groups. A similar trend could be observed in the GC content analysis (Figure 28c), where results are compared to a modelled normal distribution of 50% GC content (green line). The low sequencing coverage and the WGA method could cause these unusual patterns in the distribution. Finally, the adapter contamination showed the expected level considering the sample qualities and applied methods (Figure 28d.).

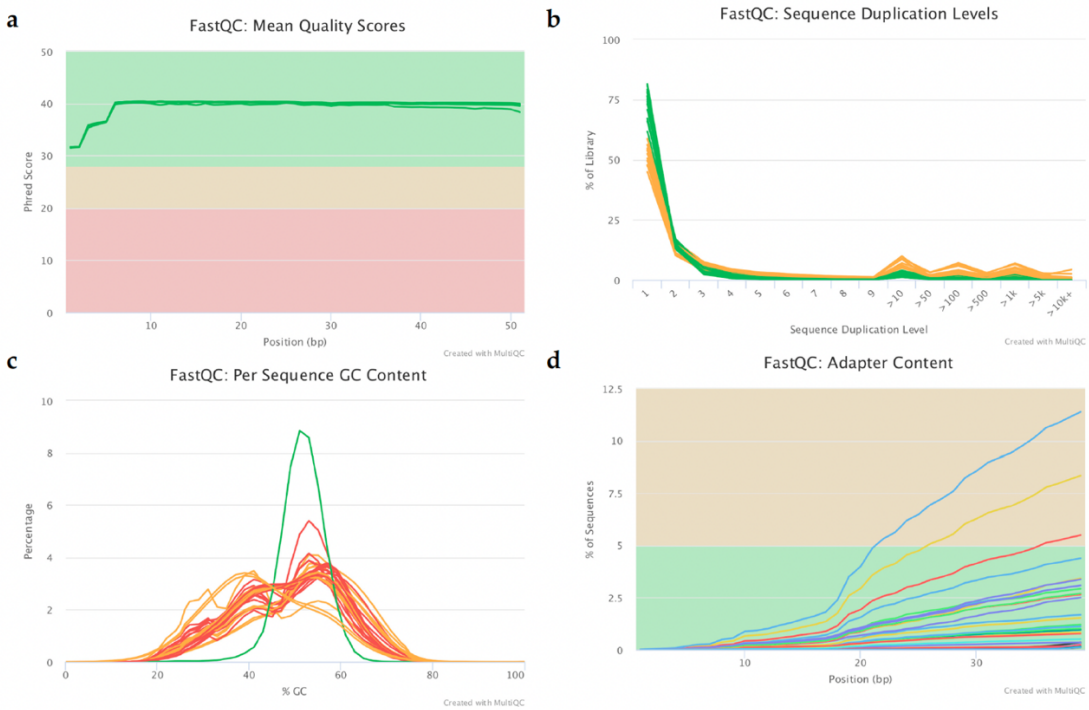



**Figure 28.** Representing plots from the raw data quality checking process with the following subfigures: (a) Sequence quality histogram, (b) Sequence duplication level, (c) Per sequence GC content, (d) Adapter content [Gombos et al., 2021].

After filtered reads were aligned to the reference genome mapping quality metrics were checked (Figure 29.). Samples, that showed good results were selected (n = 22) for further analysis (Table 18.). On average, 6.55% of the genome had at least 1x coverage and 0.5% had at least 5x coverage across all the samples (Figure 29a,b).

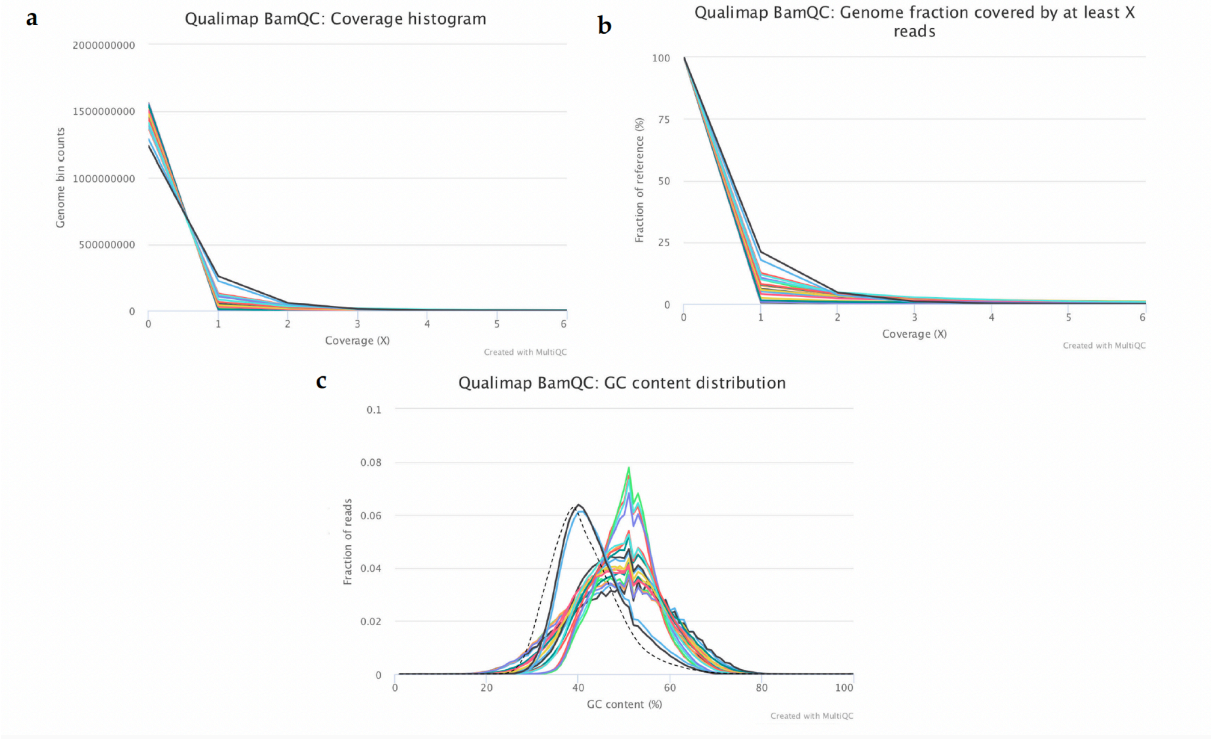



**Figure 29.** Representing plots from the analysis of mapping quality metrics of the selected samples. The subfigures represent the following results: (a) Coverage histogram showing the genomics bin counts with the corresponding coverage, (Cumulative coverage genome fractions showing the fraction (%) of the genome which has at least "X" coverage, (c) GC content distribution of the mapped reads where the dashed lines correspond to the theoretical distribution [Gombos et al., 2021].

Majority of the reference had coverage between 0–1x for all samples. According to the GC distribution of the mapped reads samples could be split into two groups (Figure 29c). The first group consisted of only the two samples that were taken from cord blood with mean GC: 40%. This value is close to the pre-calculated GC distribution for the reference genome (Figure 29c) marked with dashed line. These samples were used as controls with known CNVs for data analysis optimization and had extremely good quality thanks to the sample nature compared to the droplets. The remaining samples clustered into the second group and had an average of 49% mean GC content, which is slightly higher than the expected mean value. Repeatedly the low sequencing coverage, the original DNA quality and the WGA is the cause of the bizarre shape. Since DNA was fragmented in the culture medium and the fragments origins were not uniformly distributed compared to the DNA that have been isolated from pure tissue or a

small number of cells. The lower mapping percentages in the control culture media samples (35–44%) and the ratio of genomic regions that have at least 1x coverage compared to the other samples are supporting the fact there is a known DNA contamination source from HSA that were added to the culture media. This background contamination could cause difficulties in the downstream analysis.

| Sample Name | group | Avg. GC | ≥ 1X | ≥ 5X | Median coverage | % Aligned |
|---|---|---|---|---|---|---|
| **G1_plus_HSA1** | c | 49% | 0.9% | 0.3% | 0.0X | 40.9% |
| **G1_plus_HSA2** | c | 49% | 0.5% | 0.2% | 0.0X | 39.7% |
| **G1_plus_HSA4** | c | 47% | 0.8% | 0.2% | 0.0X | 35.1% |
| **G1_plus_HSA5** | c | 48% | 0.7% | 0.3% | 0.0X | 36.2% |
| **G1_plus_HSA6** | c | 48% | 0.8% | 0.3% | 0.0X | 44.2% |
| **7567_1A** | 0 | 50% | 10.6% | 0.4% | 0.0X | 91.9% |
| **7567_1B** | 0 | 48% | 4.0% | 1.1% | 0.0X | 76.0% |
| **7010_1A** | 0 | 49% | 1.4% | 0.4% | 0.0X | 41.7% |
| **7010_1B** | 0 | 50% | 12.6% | 0.4% | 0.0X | 96.2% |
| **7301_1A** | 0 | 50% | 11.9% | 0.4% | 0.0X | 95.2% |
| **7301_1B** | 0 | 50% | 5.0% | 1.0% | 0.0X | 84.7% |
| **7316_1A** | 0 | 49% | 6.1% | 1.0% | 0.0X | 86.4% |
| **7316_1B** | 0 | 50% | 9.9% | 0.3% | 0.0X | 95.7% |
| **7370_1B** | 0 | 50% | 7.5% | 0.9% | 0.0X | 87.4% |
| **6341_4B** | 1 | 49% | 1.5% | 0.4% | 0.0X | 40.4% |
| **6341_4C** | 1 | 50% | 2.4% | 0.5% | 0.0X | 47.9% |
| **7793_1A** | 1 | 49% | 5.7% | 1.3% | 0.0X | 83.1% |
| **7793_1B** | 1 | 50% | 7.8% | 1.1% | 0.0X | 87.8% |
| **7938_1A** | 1 | 50% | 8.1% | 1.1% | 0.0X | 88.8% |
| **7938_1C** | 1 | 49% | 9.9% | 1.1% | 0.0X | 92.2% |
| **A7Down** | 2 | 44% | 17.8% | 0.0% | 0.0X | 98.0% |
| **A8Down** | 2 | 44% | 21.1% | 0.0% | 0.0X | 98.0% |

**Table 18.** Mapping quality metrics of the selected control media samples (c.), culture media droplets from embryos of miscarriage (0), culture media droplets of healthy neonates (1), cord blood sample with known CNVs (2) [Gombos et al., 2021].

CNV analysis and Statistical testing

The number of read counts served as the basis of CNV analysis. Reads were counted and visualized almost all along on the genome (telomer and centromere regions were excluded from the analysis) in 1 Mb bin size because of the low sequencing coverage. The Cn.MOPS algorithm [Klambauer et al., 2012] was used to identify chromosomal alterations in the samples. Odds ratios (OR) were calculated in two different ways, between missed, healthy and control media groups, respectively. In the first case (v1) overall CNV occurrence was counted as one main simple event on a chromosome to reduce the bias caused by the false positives. In the second case (v2) every single CNV was counted separately on a chromosome. Both OR calculation methods confirmed statistically significant differences between the culture media droplets of aborted embryos (marked as "Missed") and the control media (marked as "Media") (Figure 30.). In contrast, results did not show significant difference between the SCM droplets of healthy neonates ("Healthy") and the control media ("Media") group. The gDNA features, like fragmentation and quality, of the healthy and culture media groups were very similar and clinically relevant CNVs could not be identified could explain the results. Also, this is supported by the difference in the embryonic gDNA content and quality found in the SCM droplets of the cleavage-stage embryos that developed to healthy neonates compared to the group of embryos that were aborted. The explanation of these observations comes from the fact that the culture media of the "Healthy" embryos mostly contains only fragmented gDNA which particularly comes from HSA. Most likely there is some gDNA from the embryo as well, but it is hard to separate the HSA and embryonic sequences even with bioinformatic approaches. Other reason is that the WGA efficiency is lower when the gDNA is very fragmented and short fragments are discarded during the sample preparation. In contrast with that the SCM of "Missed" embryos may contain more gDNA mainly form the dead cells of the embryo. These DNA is less fragmented, and the WGA works better on it. Therefore, evaluable results could be predicted only from the missed aborted group.

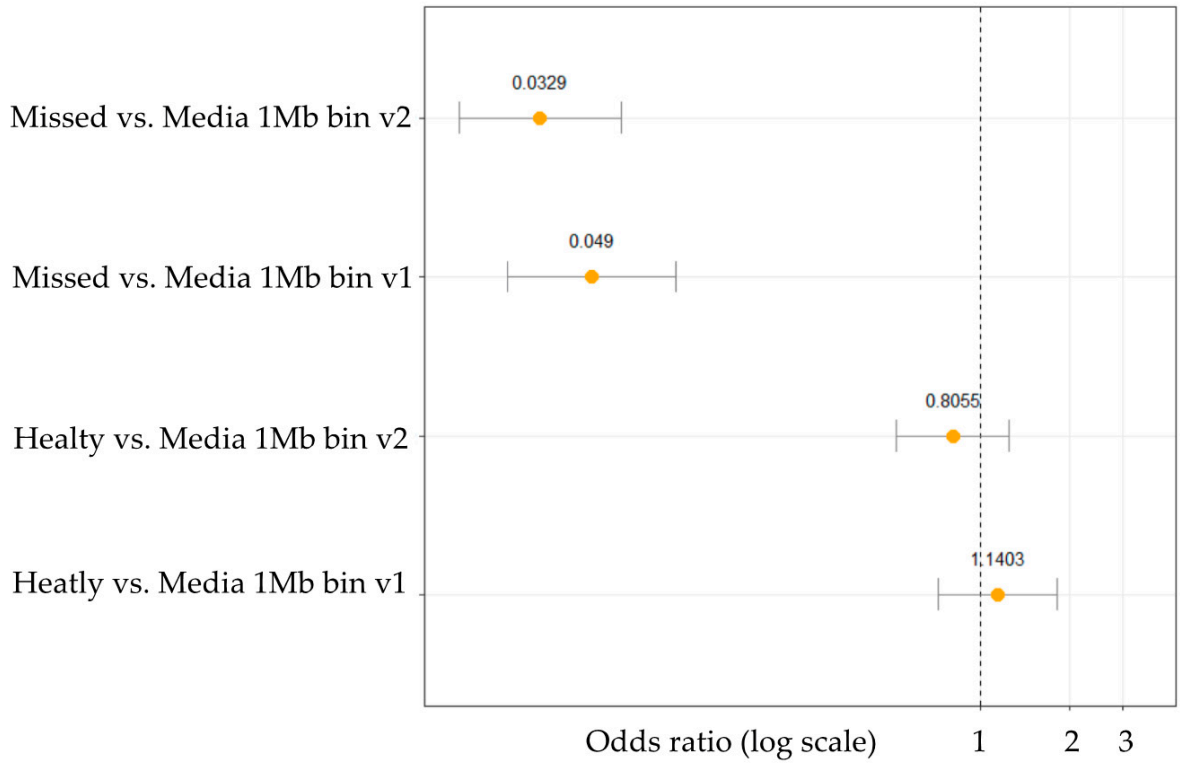



**Figure 30.** Odds ratio analysis for CNV in culture media droplets of aborted embryos (Missed) compared to control media and culture media droplets of healthy neonates (Healthy) compared to control media. Odds ratios are in log transformed scale for better visualization [Gombos et al., 2021].

Further annotation of the predicted CNVs using UNIQUE database [https://www.rarechromo.org, 31 December 2020], Genetic Alliance database [https://www.geneticalliance.org.uk, 31 December 2020] and CDO database [https://chromodisorder.org, 31 December 2020] revealed 17 relevant chromosomal alterations. All of these occurred only in the aborted embryo group and were related to registered chromosomal alterations and major developmental impairments. Table 19. lists all the identified CNVs and Figure 31. displays the variations on a karyogram (marked with blue lines). Clinically significant CNVs were could not be predicted in two of the SCMs from the aborted embryos. The remaining 9 SCM samples were positive for multiple chromosomal abnormalities.

In particular, analysis of DNA profiles of Day 3 spent media demonstrated that higher gDNA copy number is associated with impaired intrauterine development and indicated miscarriage outcomes, while low gDNA of embryonic origin in the culture medium was found to be characteristic of healthy pregnancy and live birth.


**Figure 31.** Karyogram representing clinically relevant autosomal alterations identified based on the NGS analysis of the gDNA content from the culture media of the 9 aborted embryos. Dark red bands showing the centromeres, green bands above the chromosomes are indicating gains and dark blue bands are showing losses [Gombos et al., 2021].

| Chromosomal location | Type of alteration | Function |
|---|---|---|
| 2q35 | deletion | XRCC5 gene inactivation- defect in DNA repair function |
| 2q37 | 2,3-2,4 mb deletion | IGFBP2 inactivation |
| 3p25.3-p25.1 | deletion | miR-885 inactivation, impaired differentiation |
| 4p16.3-p16.1 | duplication | CNV identified with chromosomal microarray in individuals with developmental disabilities or congenital anomalies (ISCA) |
| 8q24.3 | duplication | MYC proto-oncogene gene desert in GWAS (Genome Wide Association Studies) studies |
| 9p12-p11.2 | deletion | ANKRD20A3 gene inactivation syndromic hydrocephalus due to diffuse hyperplasia of choroid plexus, glioma |
| 10q22.1 | duplication | COLl13A1 frameshift with pathogenic interpretation (ClinVar) |
| 11q23.1-23.3 | duplication | Beckwith-Wiedemann syndrome |
| 14q31.1-q31.3 | deletion | autosomal dominant disorder (HPPD)involving hypertelorism and deafness |
| 14q32.2-q32.33 | deletion | FOXG1 inactivation, impaired development and structural brain abnormalities |
| 15q13.3 | deletion | MTMR10, FAN1 frameshift associated with karyomegalic interstitial nephritis |
| 16q23.3-24.3 | duplication | APRT, FOXC2 indel, adenine phosphoribosyl transferase deficiency, disichtiasis lympoedema syndrome |
| 17q22-p23.2 | deletion | Ateleiotic dwarfism, isolated growth hormone deficiency |
| 20p12.2-p12.1 | deletion | JAG1 related Alagille syndrome |
| 20q13.31-q13.33 | duplication | PKC1, phosphoenolpyruvate carboxikinase deficiency |
| 21q22.3 | duplication | RIPK4, PCNT popliteal pterygum syndrome, lethal type |
| 21p13-p11.2 | deletion | short arm loss monosomy |
| 22q13.2-13.31 | duplication | SCO2 cardioencephalomyopathy due to cytochrome c oxidase deficiency, fatal |

**Table 19.** Chromosomal alterations found in missed aborted embryos and listed in human genetic databases (UNIQUE, Genetic Alliance and CDO) [Gombos et al., 2021].

# Discussion

*Bioinformatics*

The pipeline development was easy and relatively fast using the Nextflow workflow management environment, but it requires basic programming skills. The introduced NGS data processing pipelines contains similar steps (e.g., quality control, trimming, filtering, adapter removing, aligning and mapping statistics) some of the steps or parts could be used as scaffolds in other workflows, thereby speeding up the future pipeline implementation. One of the useful features of Nextflow is that failed runs, after fixing the error, could be continue from certain step. This is particularly important during pipeline building and in the test phases, also sample specific errors could occur while real data analysis. Therefore, we do not have to start from the beginning, and we could save some time. Running such pipeline need less hands-on time and people with less bioinformatic experience could run it too. Since, the designed workflow is sticked to certain tools and input files with dedicated version results will be reproducible if reanalysis is needed. Another key specification of Nextflow is its integration with software repositories (e.g., GitHub or BitBucket) and its native support for cloud systems. Again, some of the practical implications of this integration are relevant to computational reproducibility. The impact of GitHub has been recently highlighted as a driving force behind data sharing [Perkel, 2016].

*Somatic mutation profiling*

***CASE STUDY I.: CLL***

Ibrutinib has changed the CLL therapy by inducing strong responses in patients with previously relapsed/refractory disease as well as in high-risk patients harbouring TP53 aberrations [Ahn et al., 2018]. Nonetheless, the consequences of subclonal changes occurring under the selective pressure during the treatment have not been completely investigated so far. Applying targeted deep NGS analysis of real-world patient cohort treated with ibrutinib helped to reveal the dynamics of clonal selection emerging on the ground of a profound subclonal heterogeneity. Also, commonly conferring convergent evolution was observed in most patients. The clonal landscape revealed that all individual cases were characterized by

unique combinations of mutations as well as different patterns of clonal variegation upon the selective pressure caused by ibrutinib. In contrast with that no mutations were found in the main driver genes that are involved in pathogenesis of CLL. This fact supporting the selective fitness in patients to ibrutinib.

This case study and some of the most recent data in the literature suggest that comprehensive deep sequencing of cancer driver genes may have clinical benefit in the future. In addition, genomic profiling of various sites (e.g., peripheral blood, bone marrow and lymph nodes) that provide distinct microenvironments for parallel clonal evolution, should be considered to achieve a more precise characterization of CLL in individual patients [Kiss et al., 2018].

***CASE STUDY II.: PCNSL***

The treatment of PCNSL patients is still challenging [Grommes et al., 2019], with considerably worse outcomes compared to nodal DLBCLs, indicating the need for novel biomarkers and therapies. Precise classification of individual cases into the proper ABC or GCB molecular subgroup was described by Alizadeh *et al.*, 2000, complemented with the mutation profiling of target genes may help to develop a more effective patient stratification method and new strategies for personalized therapies [Coutinho et al., 2013, Karmali et al., 2017]. Recently, the NanoString LST-assay emerged as one of the most reliable and highly reproducible approach to classify the subtypes from FFPE material. It was successfully demonstrated on a large cohort of DLBCLs [Scott et al., 2014] but not used in routine application.

In this PCNSL case study the NanoString LST-assay was successfully applied for the for molecular subtyping of samples from a large cohort of patients with brain lymphoma. Applying this molecular subtyping method revealed a higher proportion of cases with a GCB subtype compared to the traditional IHC analysis (13% vs 5%).

The genomic profile of PCNSL has only been analysed recently, using various NGS technologies. These studies included only smaller patient cohorts the results revealed a similar mutational burden and profile to nodal DLBCLs, with predominant mutations of the BCR/NFKB pathway [Braggio et al., 2015; Bruno et al., 2014; Chapuy et al., 2016; Fukumura et al., 2016; Nakamura et al., 2016; Vater et al. 2015; Zhou et al., 2018]. Here, a targeted genomic profiling of 14 genes was carried out in a relatively big cohort (PCNSL = 64, SCNSL = 12),

focusing on genes with potential prognostic impact. In both brain lymphoma cohorts, the most frequently mutated genes were *MYD88*, *PIM1*, *KMT2D* and *PRDM1,* followed by *IRF4, MYC* and *CD79B*. Mutation frequencies of the individual genes observed in this study are in line with already published results but with a wider range of frequencies [Braggio et al., 2015; Bruno et al., 2014; Chapuy et al., 2016; Courts et al., 2008; Fukumura et al., 2016; Gonzalez-Aguilar et al., 2012; Nakamura et al., 2016; Vater et al. 2015; Zheng et al., 2017; Zhou et al., 2018]. This can be explained by the heterogeneity in the type and depth of the used sequencing methods, type of the analysed material and the difference between the applied bioinformatics pipelines. In this study the dual-strand approach was utilized additionally to increase the sensitivity and accuracy of variant detection from FFPE samples. Comparing the mutation frequencies between the GCB and ABC subtypes defined by the LST-assay, the mutation patterns observed in PCNSL (as well as SCNSL) do not follow the ones documented earlier in nodal DLBCLs [Davis et al., 2010; Dubois et al., 2017; Kraan et al., 2013; Kuo et al., 2016]. Some of the *CD79B*, *CARD11*, *CSMD2* and *CSMD3* were exclusively detected in ABC. Statistical analysis did not reveal significance differences in mutation frequencies (e.g., *MYD88*, *PIM1* and *KMT2D*) between the GCB and ABC groups. This may support the hypothesis that PCNSL represents a distinct clinical entity irrespective of the cell of origin classification as proposed by Fukumura *et al.*, 2016.

Considering the collected knowledge about the genomic complexity of PCNSL in the GCB and ABC patient groups, precise assignment of molecular subtypes using routinely available FFPE tissues and complementary mutation analysis of the actionable mutation targets will most likely support and drive personalized therapeutic decisions during the management of the disease.

*DNA methylation profiling*

***CASE STUDY III.: GBM***

In this study, the main goal was to identify molecular drivers and pathways that are essential for GBM development and recurrence from a different perspective. Therefore, the genome-wide DNA CpG methylation patterns were analysed to infer the expression of genes defining the most critical pathways in the GBM cohort. Based on the quality assessment results DNA specimens from surgically removed FFPE samples were significantly more

fragmented than that of freshly obtained blood samples but worked well in RRBS. The CG2 included DNA CpG methylomes of five brain specimens obtained during epilepsy surgery [Klughammer et al., 2018]. The DNA controls from the small and heterogeneous populations of all normal and some degenerative cell types of adult brains may not be perfect for the methylome from transformed glial tumor cells of GBM. As no ideal control tissue is available for human GBM, control brain methylomes that have been successfully applied in a similar epigenomic analysis [Klughammer et al. 2018] was selected for the analysis.

A shift toward global DNA hypomethylation was observed when comparing CG2, GBM1 and GBM2, these findings are in line with already published results [Brothman et al., 2005; Ehrlich 2009; Feinberg et al., 1988; Hansen et al., 2011; Makos et al., 1992; Nagarajan et al., 2014]. Comparisons of differential methylation data at site and region levels revealed no significant results in any of the three pairwise comparisons but the GO analyses highlighted several pathways with biological relevance.

In the comparison of GBM1 vs. CG2 significant hypomethylation was found, possibly activation in the following pathways like synapse formation and myelination, positive regulation of endothelial cell proliferation, a factor contributing to angioneogenesis, which promoting GBM growth [Ameratunga et al., 2018; Etcheverry et al., 2010; Fisher et al., 2005; Roth et al., 2020]. In the same GBM1 vs CG2 comparison hypermethylation (repression) was identified in pathways related to related to neuronal differentiation, nucleic acid-templated transcription and different nucleobase containing metabolic processes, which affect multiple genes whose abnormal function may modify cell function and define subtype formation [Cuperlovic-Culf et al. 2012; Marziali et al. 2016]. These findings reflect a disturbed balance in elements of a normal neuronal differentiation underlying the distorted patterns which was observed by other investigators as well in cancer stem cells (CSCs) and GBM [Etcheverry et al., 2010; Silvestris et al., 2019].

Comparing differential promoter methylation in GBM2 vs CG2, the hypomethylated pathways were primarily related to intracellular function and transport, offering new targets for experimental intervention [Fallacara et al., 2019]. The hypermethylated pathways included transcriptional regulation, cell adhesion and embryonic development, which may also contribute to a distortion of normal neuronal differentiation and abnormal proliferation of pluripotent neuroepithelial cells, thereby defining progression of GBM [Bradshaw et al., 2016a, b; Etcheverry et al., 2010].

The comparison in GBM2 vs GBM1 identified several changes involving essential cellular functions that may contribute to GBM development. Higher gene expression and activity were inferred from the lower methylation of elements essential in cell response, signaling and communication in GBM1 than in GBM2. Elements of the canonical Wnt signaling pathway, particularly those regulating endothelial cell migration, cell adhesion or wound healing also appeared more active in GBM1 compared to GBM2.

The above-mentioned results are overlapping with results described in other publications [Anastas and Moon 2013; Etcheverry et al., 2010; Hu et al., 2016; Klughammer et al., 2018; Lamb et al., 2013; Mazieres et al., 2005; Tompa et al., 2018] and was supported by comparing the array-based DNA CpG methylation data of TCGA GBMs to the sequence-based methylomes of CG2 controls [Klughammer et al., 2018], and the sample pairs of the array-based methylation data to each other (REF) as well. The weaknesses of the analyses are the heterogeneous tumor biology, differences in cohorts' sizes, distributions of patients' age, gender and ethnic background, and the reduced representation of methylome itself. Apart from these differences these methylome analyses revealed important molecular pathways and mechanisms contributing to the occurrence and progression of GBM.

*miRNome profiling*

***CASE STUDY IV.: NSCLC***

Several recent studies showed that profiling miRNA expression could be exploited for both histological and prognostic characterization (classification) of NSCLC [Hua et al., 2022; Liang et al., 2022; Yan et al., 2022]. In the current study, NGS-based technology was used in order to discover miRNA expression differences among lung cancer histologies and identify possible biomarker as well.

Our study of global miRNA expression profiling in lung cancer allowed for identification of set of miRNAs of which expression profiles differed significantly between AC and SCC and control samples. Overall, it was demonstrated that 138 miRNAs were significantly overexpressed in SCC while 83 miRNAs were found overexpressed in AC samples. Majority of these differentially expressed miRNAs were also identified in other studies [Joshi et al., 2014; Wani et la., 2022]. Landi and colleagues [Landi et al., 2014] in the unadjusted analysis

demonstrated the expression profile of 127 miRNAs which were significantly and consistently different between lung AC and SCC patients.

Altogether, our and other studies demonstrated that specific miRNAs can be considered molecular drivers determining the histology of NSCLC. The patterns of differential expression of miRNAs in lung AC and SCC may indicate the occurrence of different microRNA-related signaling pathways underlying pathogenesis of these histologies. Our results clearly showed that several miRNAs can be strongly associated with lung cancer histology.

According to previous reports [Arechaga-Ocampo et al., 2017; Lv et al., 2020; Sun et al., 2018], miR-29c can function as tumor suppressors or epigenetic normalizators in lung AC tumors. Plaisier and colleagues showed [Plaisier et al., 2012] that the miR-29 family inhibited specific genes accounting for invasion and metastasis of lung AC. Fabbri and colleagues [Fabbri et al., 2007] observed that the members of miR-29 family induced DNA hypomethylation and led to re-expression of certain tumor-suppressor genes such as *PTEN* and *WWOX*. In other studies, miR-29s has been shown to upregulate p53 levels and activate apoptosis in a p53-dependent manner [Nguyen et al.,2022].

Similarly, several recent studies demonstrated that members of miR-34 family were strongly downregulated in NSCLC tumors as compared to normal tissues indicating a protective role of miR-34 in lung tumorigenesis. miR-34 family members were reported to act as tumor-suppressor miRNAs targeting many oncogenes related to proliferation, apoptosis and metastasis. The expression of miR-34a is directly trans-activated by tumor suppressor p53, and its activity is frequently reduced in p53 mutant tumors [Gallardo et al., 2009]. Notably, the frequency of p53 mutations is significantly higher in lung SCC than in lung AC. Our findings demonstrated over-expression of miR-34a-3p and miR-34a-5p in lung AC tumors, suggesting a potential synergism between miR-34a expression and activation of p53, the mechanism being relevant for SCC tumorigenesis.

Recently, the members of miR-181 family were demonstrated to play a role in inhibition of the growth, migration, and invasion of NSCLC cells. Huang and colleagues recently showed that miR-181 reduction was associated with increased Bcl-2 levels, indicating its proapoptotic function of this molecule in lung cancer pathogenesis.

Our group has found that the high co-expression of hsa-miR-30 family in lung AC may be of clinical significance in classification of NSCLC subtypes. According to Chen and colleagues [Chen et al., 2015], upregulation of mir-30d-5p inhibited tumor cell proliferation and motility

by direct targeting of cyclin E2 (*CCNE2*). Interestingly, several studies demonstrated that miR-30d-5p was downregulated in lung SCC, in concordance with data derived from Cancer Genome Atlas (TCGA) miRNASeq database. This may suggest that miR-30d-5p can play a critical role as a tumor suppressor thus modulating the development and progression of NSCLC.

In our study, we clearly showed that the specific patterns of miRNA expression may reflect biological distinctions of lung SCC. To the best of our knowledge, this is the first report to show that miR-31-3p and miR-31-5p are significantly up-regulated in SCC and AC as compared to controls. Thus, the members of miR-31 family can become useful diagnostic markers allowing for more detailed discrimination between lung AC and SCC. In a study [Okudela et al., 2014], it was demonstrated that the miR-31 expression was deregulated in lung cancer through either the amplification or loss of the host gene locus. It was also shown that the loss of miR-31 expression was observed mainly in lung AC tumors. Taken together above findings, we suggest that miR-31 can play an oncogenic role by promoting carcinogenesis, especially of lung SCC.

In conclusion, we identified here and validated novel histology-specific miRNA patterns that can be further exploited diagnostically to precise subclassification of lung AC and SCC. Our results indicated that miRNA expression profiles in early-stages NSCLC may help elucidate histological distinctions of NSCLC tumors through the identification of different microRNA-mediated signaling pathways involved in the pathogenesis of histologically distinct tumors.

*Assisted reproduction*

***CASE STUDY V.: NIPGT-A***

SCM is found to be a potentially useful liquid biopsy sample that represents embryonic genetic material but the well-established clinical applications of PGT are still part of the routine. The NIPGT-A methods are evolving, since spent embryonic media collection does not require excess intervention and manipulation of the human embryo or any subsequent modification of the clinical routine and it has the potential to reflect chromosomal composition of the developing embryo [Fang et al., 2019, Farra et al., 2018; Gianaroli et al., 2014; Handyside et al., 2016; Hammond et al., 2016, 2017; Ho et al., 2018; Huang et al., 2019;

Kuznyetsov et al., 2018; Rubio et al. 2020; Palini et al., 2013; Shitara et al., 2021; Stigliani et al., 2013; Vera-Rodriguez et al., 2018; Wu et al., 2015; Xu et al., 2016; Yeung et al., 2019].

The selection of embryos for transfer is a frequently appearing relevant clinical dilemma; therefore, main goal of this study was to complete a generally applicable non-invasive embryo selection strategy combined with same-cycle transfer and follow the already existing clinical routine concerning IVF methods, embryo culture and transfer conditions in cases of the genetic-disease-free population of women of average age 35. The routine sequential culturing and collected spent embryonic culture media after assisted hatching (AH) on Day 3 was applied, after embryos were morphologically evaluated and moved to fresh G2 media. In contrast with the facts that NIPGT and PGT, shows better results on Day 5 due to higher ICM mass of the embryos and a greater amount of leaked gDNA the current study was focused on the gDNA content on Day 3 of the cleavage-stage embryos' culture media. This was because the aim was to complete the introduced NGS workflow within 48 h, when embryo assessment results are summarised for embryo selection for SET to achieve fresh, same-cycle embryo transfer.

This is important for IVF protocols that do not include embryo cryopreservation and vitrification procedures. However, the promoted workflow may give better result on the SCM of Day 5 embryos combining it with "freeze-all" or "elective frozen embryo" strategies latest on the $6^{th}$ day to give descent time for the NGS and a bioinformatic analysis. Generally, it can be fitted into most of the currently used IVF strategies. Choosing the right sequencing platform is also an important aspect of the workflow. Small NGS platforms like MiSeq and iSeq could be more cost effective and more suitable for real clinical practice. Moreover, there is possibility for time-lapse morphology evaluation in the time between Day 3 and Day 5 embryo culture, and additional verification of the selection decision can be gained during the sequential culturing methods. The demonstrated study design also enables the collection of multiplex data about the developing embryo, since around 5 μL of culture media of the total 20 μL is enough for NGS analysis. The remainder could be used in other methods like proteomic and miRNA analysis, which can also be integrated into a complex embryo assessment strategy.

# Conclusion

Bioinformatics is a recent scientific discipline that has undergone strong and rapid progression and evolution [Ouzounis, 2012]. The use of bioinformatics analyses and support in biological studies for example in cancer research, metagenomics [Fomitcheva-Khartchenko et al., 2022; Hurwitz et al., 2014] and is now more and more accepted and viewed as normal. In the discipline of bioinformatics and computational biology, there are numerous ways in which curricula and specially autodidacticism can be designed to achieve the desired educational outcomes. In this thesis only just the tip of the iceberg has been presented.

In conclusion, subclonal heterogeneity, dynamic clonal selection and various patterns of clonal variegation in **CLL** were identified with novel resistance-associated BTK mutations in individual patients treated with ibrutinib.

The mutational landscapes of 14 target genes in **PCNSL** were determined applying deep NGS. The LST-assay was successfully used for molecular classification and a significantly lower proportion of cases displayed ABC phenotype compared to the traditional IHC-based characterization. The described workflow could lead to a more precise patient stratification with mutation profiling potentially applicable in the diagnostic algorithm of PCNSL.

Methylation analyses in sequential **GBM** specimens revealed hypomethylation in certain pathways such as neuronal tissue development and angiogenesis likely involved in early tumor development and growth, while suggested altered regulation in catecholamine secretion and transport, Wnt expression and immune response contributing to recurrence. These pathways merit further investigations and may represent novel therapeutic targets.

The study of global miRNA expression profiling in **NSCLC** allowed for identification of set of miRNAs of which expression profiles differed significantly between AC and SCC samples. The analyses demonstrated that specific miRNAs can be considered molecular drivers determining the histology of NSCLC. The miRNA expression profiles in early-stages NSCLC may help elucidate histological distinctions of NSCLC tumors through the identification of different microRNA-mediated signaling pathways involved in the pathogenesis of histologically distinct tumors.

In the last case study, an optimized **NIPTG-A** workflow was proposed which combines low gDNA input based NGS application and downstream bioinformatic analyses to identify CNVs from SCM droplets. The workflow can be applied in same-cycle transfer IVF cases. The

study has clear limitations due to the sample size and the lack of comparison of spent blastocyst culture media with corresponding TE and ICM NGS analysis to accurately describe embryonic chromosomal composition. This was due to ethical regulation of the IVF centre, which limits all invasiveness during embryo culture.

As our NGS analysis permitted deep CNV evaluation, chromosomal compositions of the embryos were also detected. We found clinically significant autosomal ploidy alterations only among the aborted embryos—this affected 75% of them. In some cases, the chromosomal ploidy aberration was found to be multiple, which can be irreconcilable with healthy embryonic development and embryonic viability.

## Acknowledgements

I am grateful to my promoter, Attila Gyenesei and my supervisor Jacek Nikliński for her contribution and help during the years of my PhD and the thesis preparation. I would like to thank Professor Jacek Nikliński, who provided me an opportunity to join the Department of Clinical Molecular Biology, at the Medical University of Bialystok. I am grateful to my colleagues and my collaboration partners for their work during the experiments. Last but not least, I would like to thank my family: my parents and my wife, Zita Gálik-Oláh, and my children Sári and Eszter for supporting me both in my life and in my experiment.

## Funding

This research was conducted within the project which has received funding from the European Union's Horizon 2020 research and innovation programme under the Marie Skłodowska-Curie grant agreement No 754432 and the Polish Ministry of Education and Science, from the financial resources for science in 2018-2023 granted for the implementation of an international co-financed project.

**CLL**

This work was funded by the KH17-126718, K_16 #119950, NVKP_16-1-2016-0004 and NVKP_16-1-2016-0005 grants of the Hungarian National Research, Development and Innovation Office (NKFIH), the Momentumgrant (LP-95021) and the János Bolyai Research Scholarship Program (BO/00320/18/5) of the Hungarian Academy of Sciences, the ÚNKP- 18-4-SE-62, ÚNKP-18-3-I-SE-48 and EFOP-3.6.3-VEKOP-16-2017-00009 grants of the Ministry of Human Capacities and the Higher Education Institutional Excellence Programme of the Ministry of Human Capacities in Hungary, within the framework of the Molecular Biology Thematic Programme of the Semmelweis University.

**PCNSL**

This study was funded by the Hungarian Science Foundation (OTKAPD115792 to LR), Hungarian National Research, Development and Innovation Office (NKFIH) (KH17-126718 to CsB, NVKP_16-1-2016-0004 to AM as well as K_16 #119950), the Momentum grant (LP-95021 to CsB). DA was supported by Janos Bolyai Research Scholarship (BO/00320/18/5) of the Hungarian Academy of Sciences. Furthermore, the study was supported by the UNKP-19-4-SE-77 and UNKP-19-2-I-SE-47 grants of the New National Excellence Program of the Ministry for Innovation and Technology, and the Complementary Research Excellence Program of Semmelweis University (EFOP-3.6.3-VEKOP-16-2017-00009 to BB, as well as by the Higher Education Institutional Excellence Programme of the Ministry of Human Capacities in Hungary within the framework of the Molecular Biology thematic programme of the Semmelweis University to CsB, the Hungarian Brain Research Program (2017-1.2.1-NKP-2017-00002 to LR) and by a research grant of the University of Pecs (ID: KA-2019-32).

**GBM**

The study was supported by Hungarian state funds administered through the graduate studies program of the University of Pecs, the Dr. Janos Szolcsanyi Research Fund by the University of Pecs, School of Medicine (Nr. KA-2019-42), and by a private donation. The research was performed in collaboration with the Genomics and Bioinformatics Core Facility at the Szentagothai Research Center of University of Pecs. Bioinformatics infrastructure was supported by ELIXIR Converge (Nr. 871075)

**NSCLC**

MOBIT Study: "Development of personalized diagnostic of malignant tumours based on tumour heterogeneity and integrated genomic, transcriptomic, metabolomics and imaging PET/MRI analysis. Getting ready for individualized therapy." The project funded by the National Centre for Research and Development in the framework of Programme "Prevention practices and treatment of civilization diseases" - STRATEGMED (contract no. STRATEGMED2/266484/2/NCBR/2015). Implementation period: 2016 – 2020. This project with around 5 million Euro budget is to our knowledge the first-in-Europe work wherein lung cancer patients living in the region have their resected tumors biobanked in highly standardized way and then subjected to large scale genomic, transcriptomic, metabolomic, proteomic analysis in order to provide a personalized diagnostics and prepare the patient for personalized therapy. The –omics data are integrated with each other, clinical data of patients and data from PET/MR examination. It project provides a substantial scientific background for execution of objectives described in this thesis.

**NIPTG-A**

This research was partly supported by NRDIO-EDIOP-2.3.2-15-2016-00021 ‘The use of chip-technology in increasing the effectiveness of human in vitro fertilization’ and NRDIOK/115394/2015 ‘Early biochemical indicators of embryo viability’ and NRDIO-KA-2016-04 grants. A.G. were supported by the grants GINOP-2.3.4-15-2020-00010, GINOP-2.3.1-20-2020-00001 and ERASMUS+-2019-0-HU01-KA203-061251. Bioinformatics infrastructure was supported by ELIXIR Hungary (http://elixir-hungary.org). The funding organizations played no role in the conceptualization, design, data collection and analysis or preparation and submission of the manuscript.

# References


Ahn IE, Underbayev C, Albitar A, Herman SE, Tian X, Maric I, Arthur DC, Wake L, Pittaluga S, Yuan CM, Stetler-Stevenson M, Soto S, Valdez J, Nierman P, Lotter J, Xi L, Raffeld M, Farooqui M, Albitar M, Wiestner A. Clonal evolution leading to ibrutinib resistance in chronic lymphocytic leukemia. Blood. 2017 Mar 16;129(11):1469-1479. doi: 10.1182/blood-2016-06-719294. Epub 2017 Jan 3. PubMed [citation] PMID: 28049639, PMCID: PMC5356450

Ahn IE, Farooqui MZH, Tian X, Valdez J, Sun C, Soto S, Lotter J, Housel S, Stetler-Stevenson M, Yuan CM, Maric I, Calvo KR, Nierman P, Hughes TE, Saba NS, Marti GE, Pittaluga S, Herman SEM, Niemann CU, Pedersen LB, Geisler CH, Childs R, et al. Depth and durability of response to ibrutinib in CLL: 5-year follow-up of a phase 2 study. Blood. 2018 May 24;131(21):2357-2366. doi: 10.1182/blood-2017-12-820910. Epub 2018 Feb 26. PubMed [citation] PMID: 29483101, PMCID: PMC5969380

Alizadeh AA, Eisen MB, Davis RE, Ma C, Lossos IS, Rosenwald A, Boldrick JC, Sabet H, Tran T, Yu X, Powell JI, Yang L, Marti GE, Moore T, Hudson J Jr, Lu L, Lewis DB, Tibshirani R, Sherlock G, Chan WC, Greiner TC, Weisenburger DD, et al. Distinct types of diffuse large B-cell lymphoma identified by gene expression profiling. Nature. 2000 Feb 3;403(6769):503-11. PubMed [citation] PMID: 10676951

Alpha Scientists in Reproductive Medicine and ESHRE Special Interest Group of Embryology. The Istanbul consensus workshop on embryo assessment: proceedings of an expert meeting. Hum Reprod. 2011 Jun;26(6):1270-83. doi: 10.1093/humrep/der037. Epub 2011 Apr 18. PubMed [citation] PMID: 21502182

Ameratunga M, Pavlakis N, Wheeler H, Grant R, Simes J, Khasraw M. Anti-angiogenic therapy for high-grade glioma. Cochrane Database Syst Rev. 2018 Nov 22;11:CD008218. doi: 10.1002/14651858.CD008218.pub4. PubMed [citation] PMID: 30480778, PMCID: PMC6516839

Anastas JN, Moon RT. WNT signalling pathways as therapeutic targets in cancer. Nat Rev Cancer. 2013 Jan;13(1):11-26. doi: 10.1038/nrc3419. Review. PubMed [citation] PMID: 23258168

Arechaga-Ocampo E, Lopez-Camarillo C, Villegas-Sepulveda N, Gonzalez-De la Rosa CH, Perez-Añorve IX, Roldan-Perez R, Flores-Perez A, Peña-Curiel O, Angeles-Zaragoza O, Rangel Corona R, Gonzalez-Barrios JA, Bonilla-Moreno R, Del Moral-Hernandez O, Herrera LA, Garcia-Carranca A. Tumor suppressor miR-29c regulates radioresistance in lung cancer cells. Tumour Biol. 2017 Mar;39(3):1010428317695010. doi: 10.1177/1010428317695010. PubMed [citation] PMID: 28345453

Baliakas P, Hadzidimitriou A, Sutton LA, Rossi D, Minga E, Villamor N, Larrayoz M, Kminkova J, Agathangelidis A, Davis Z, Tausch E, Stalika E, Kantorova B, Mansouri L, Scarfò L, Cortese D, Navrkalova V, Rose-Zerilli MJ, Smedby KE, Juliusson G, Anagnostopoulos A, Makris AM, et al. Recurrent mutations refine prognosis in chronic lymphocytic leukemia. Leukemia. 2015 Feb;29(2):329-36. doi: 10.1038/leu.2014.196. Epub 2014 Jun 19. PubMed [citation] PMID: 24943832

Bödör C, Alpár D, Marosvári D, Galik B, Rajnai H, Bátai B, Nagy Á, Kajtár B, Burján A, Deák B, Schneider T, Alizadeh H, Matolcsy A, Brandner S, Storhoff J, Chen N, Liu M, Ghali N, Csala I, Bagó AG, Gyenesei A, Reiniger L. Molecular Subtypes and Genomic Profile of Primary Central Nervous System Lymphoma. J Neuropathol Exp Neurol. 2020 Feb 1;79(2):176-183. doi: 10.1093/jnen/nlz125. PubMed [citation] PMID: 31886867

Bouaoun L, Sonkin D, Ardin M, Hollstein M, Byrnes G, Zavadil J, Olivier M. TP53 Variations in Human Cancers: New Lessons from the IARC TP53 Database and Genomics Data. Hum Mutat. 2016 Sep;37(9):865-76. doi: 10.1002/humu.23035. Epub 2016 Jul 8. PubMed [citation] PMID: 27328919

Bradshaw A, Wickremesekera A, Brasch HD, Chibnall AM, Davis PF, Tan ST, Itinteang T. Cancer Stem Cells in Glioblastoma Multiforme. Front Surg. 2016 Aug 26;3:48. doi: 10.3389/fsurg.2016.00048. eCollection 2016. PubMed [citation] PMID: 27617262, PMCID: PMC5001191

Bradshaw A, Wickremsekera A, Tan ST, Peng L, Davis PF, Itinteang T. Cancer Stem Cell Hierarchy in Glioblastoma Multiforme. Front Surg. 2016 Apr 15;3:21. doi: 10.3389/fsurg.2016.00021. eCollection 2016. Review. PubMed [citation] PMID: 27148537, PMCID: PMC4831983

Braggio E, Van Wier S, Ojha J, McPhail E, Asmann YW, Egan J, da Silva JA, Schiff D, Lopes MB, Decker PA, Valdez R, Tibes R, Eckloff B, Witzig TE, Stewart AK, Fonseca R, O'Neill BP. Genome-Wide Analysis Uncovers Novel Recurrent Alterations in Primary Central Nervous System Lymphomas. Clin Cancer Res. 2015 Sep 1;21(17):3986-94. doi: 10.1158/1078-0432.CCR-14-2116. Epub 2015 May 19. PubMed [citation] PMID: 25991819, PMCID: PMC4558226

Brennan CW, Verhaak RG, McKenna A, Campos B, Noushmehr H, Salama SR, Zheng S, Chakravarty D, Sanborn JZ, Berman SH, Beroukhim R, Bernard B, Wu CJ, Genovese G, Shmulevich I, Barnholtz-Sloan J, Zou L, Vegesna R, Shukla SA, Ciriello G, Yung WK, Zhang W, et al. The somatic genomic landscape of glioblastoma. Cell. 2013 Oct 10;155(2):462-77. doi: 10.1016/j.cell.2013.09.034. Erratum in: Cell. 2014 Apr 24;157(3):753. PubMed [citation] PMID: 24120142, PMCID: PMC3910500

Brito JJ, Li J, Moore JH, Greene CS, Nogoy NA, Garmire LX, Mangul S. Recommendations to enhance rigor and reproducibility in biomedical research. Gigascience. 2020 Jun 1;9(6). pii: giaa056. doi: 10.1093/gigascience/giaa056. Erratum in: Gigascience. 2020 Sep 17;9(9). PubMed [citation] PMID: 32479592, PMCID: PMC7263079

Brothman AR, Swanson G, Maxwell TM, Cui J, Murphy KJ, Herrick J, Speights VO, Isaac J, Rohr LR. Global hypomethylation is common in prostate cancer cells: a quantitative predictor for clinical outcome? Cancer Genet Cytogenet. 2005 Jan 1;156(1):31-6. PubMed [citation] PMID: 15588853

Bruno A, Boisselier B, Labreche K, Marie Y, Polivka M, Jouvet A, Adam C, Figarella-Branger D, Miquel C, Eimer S, Houillier C, Soussain C, Mokhtari K, Daveau R, Hoang-Xuan K. Mutational analysis of primary central nervous system lymphoma. Oncotarget. 2014 Jul 15;5(13):5065-75. PubMed [citation] PMID: 24970810, PMCID: PMC4148122

Burger JA, Tedeschi A, Barr PM, Robak T, Owen C, Ghia P, Bairey O, Hillmen P, Bartlett NL, Li J, Simpson D, Grosicki S, Devereux S, McCarthy H, Coutre S, Quach H, Gaidano G, Maslyak Z, Stevens DA, Janssens A, Offner F, Mayer J, et al. Ibrutinib as Initial Therapy for Patients with Chronic Lymphocytic Leukemia. N Engl J Med. 2015 Dec 17;373(25):2425-37. doi: 10.1056/NEJMoa1509388. Epub 2015 Dec 6. PubMed [citation] PMID: 26639149, PMCID: PMC4722809

Butler SA, Luttoo J, Freire MO, Abban TK, Borrelli PT, Iles RK. Human chorionic gonadotropin (hCG) in the secretome of cultured embryos: hyperglycosylated hCG and hCG-free beta subunit are potential markers for infertility management and treatment. Reprod Sci. 2013 Sep;20(9):1038-45. doi: 10.1177/1933719112472739. Epub 2013 Feb 25. PubMed [citation] PMID: 23439616

Bybee SM, Bracken-Grissom H, Haynes BD, Hermansen RA, Byers RL, Clement MJ, Udall JA, Wilcox ER, Crandall KA. Targeted amplicon sequencing (TAS): a scalable next-gen approach to multilocus, multitaxa phylogenetics. Genome Biol Evol. 2011;3:1312-23. doi: 10.1093/gbe/evr106. Epub 2011 Oct 13. PubMed [citation] PMID: 22002916, PMCID: PMC3236605

Byrd JC, Brown JR, O'Brien S, Barrientos JC, Kay NE, Reddy NM, Coutre S, Tam CS, Mulligan SP, Jaeger U, Devereux S, Barr PM, Furman RR, Kipps TJ, Cymbalista F, Pocock C, Thornton P, Caligaris-Cappio F, Robak T, Delgado J, Schuster SJ, Montillo M, et al. Ibrutinib versus ofatumumab in previously treated chronic lymphoid leukemia. N Engl J Med. 2014 Jul 17;371(3):213-23. doi: 10.1056/NEJMoa1400376. Epub 2014 May 31. PubMed [citation] PMID: 24881631, PMCID: PMC4134521

Camilleri-Broët S, Crinière E, Broët P, Delwail V, Mokhtari K, Moreau A, Kujas M, Raphaël M, Iraqi W, Sautès-Fridman C, Colombat P, Hoang-Xuan K, Martin A. A uniform activated B-cell-like immunophenotype might explain the poor prognosis of primary central nervous system lymphomas: analysis of 83 cases. Blood. 2006 Jan 1;107(1):190-6. Epub 2005 Sep 8. PubMed [citation] PMID: 16150948

Chacon, S., & Straub, B. (2014). Pro git. Apress.

Chapuy B, Roemer MG, Stewart C, Tan Y, Abo RP, Zhang L, Dunford AJ, Meredith DM, Thorner AR, Jordanova ES, Liu G, Feuerhake F, Ducar MD, Illerhaus G, Gusenleitner D, Linden EA, Sun HH, Homer H, Aono M, Pinkus GS, Ligon AH, Ligon KL, et al. Targetable genetic features of primary testicular and primary central nervous system lymphomas. Blood. 2016 Feb 18;127(7):869-81. doi: 10.1182/blood-2015-10-673236. Epub 2015 Dec 23. PubMed [citation] PMID: 26702065, PMCID: PMC4760091

Chaturvedi A, Som A. The LCNetWork: An electronic representation of the mRNA-lncRNA-miRNA regulatory network underlying mechanisms of non-small cell lung cancer in humans, and its explorative analysis. Comput Biol Chem. 2022 Dec;101:107781. doi: 10.1016/j.compbiolchem.2022.107781. Epub 2022 Oct 22. PubMed [citation] PMID: 36327779

Chen D, Guo W, Qiu Z, Wang Q, Li Y, Liang L, Liu L, Huang S, Zhao Y, He X. MicroRNA-30d-5p inhibits tumour cell proliferation and motility by directly targeting CCNE2 in non-small cell lung cancer. Cancer Lett. 2015 Jul 1;362(2):208-17. doi: 10.1016/j.canlet.2015.03.041. Epub 2015 Apr 2. PubMed [citation] PMID: 25843294

Cingolani P, Platts A, Wang le L, Coon M, Nguyen T, Wang L, Land SJ, Lu X, Ruden DM. A program for annotating and predicting the effects of single nucleotide polymorphisms, SnpEff: SNPs in the genome of Drosophila melanogaster strain w1118; iso-2; iso-3. Fly (Austin). 2012 Apr-Jun;6(2):80-92. doi:10.4161/fly.19695. PubMed [citation] PMID: 22728672, PMCID: PMC3679285

Cortezzi SS, Cabral EC, Trevisan MG, Ferreira CR, Setti AS, Braga DP, Figueira Rde C, Iaconelli A Jr, Eberlin MN, Borges E Jr. Prediction of embryo implantation potential by mass spectrometry fingerprinting of the culture medium. Reproduction. 2013 Apr 29;145(5):453-62. doi: 10.1530/REP-12-0168. Print 2013 May. PubMed [citation] PMID: 23404850

Courts C, Montesinos-Rongen M, Brunn A, Bug S, Siemer D, Hans V, Blümcke I, Klapper W, Schaller C, Wiestler OD, Küppers R, Siebert R, Deckert M. Recurrent inactivation of the PRDM1 gene in primary central nervous system lymphoma. J Neuropathol Exp Neurol. 2008 Jul;67(7):720-7. doi: 10.1097/NEN.0b013e31817dd02d. PubMed [citation] PMID: 18596541

Crysup B, Mandape S, King JL, Muenzler M, Kapema KB, Woerner AE. Using unique molecular identifiers to improve allele calling in low-template mixtures. Forensic Sci Int Genet. 2022 Nov 24;63:102807. doi: 10.1016/j.fsigen.2022.102807. [Epub ahead of print] PubMed [citation] PMID: 36462297

Cuperlovic-Culf M, Ferguson D, Culf A, Morin P Jr, Touaibia M. 1H NMR metabolomics analysis of glioblastoma subtypes: correlation between metabolomics and gene expression characteristics. J Biol Chem. 2012 Jun 8;287(24):20164-75. doi: 10.1074/jbc.M111.337196. Epub 2012 Apr 23. PubMed [citation] PMID: 22528487, PMCID: PMC3370199

Danecek P, Bonfield JK, Liddle J, Marshall J, Ohan V, Pollard MO, Whitwham A, Keane T, McCarthy SA, Davies RM, Li H. Twelve years of SAMtools and BCFtools. Gigascience. 2021 Feb 16;10(2). pii: giab008. doi: 10.1093/gigascience/giab008. PubMed [citation] PMID: 33590861, PMCID: PMC7931819

Davis RE, Ngo VN, Lenz G, Tolar P, Young RM, Romesser PB, Kohlhammer H, Lamy L, Zhao H, Yang Y, Xu W, Shaffer AL, Wright G, Xiao W, Powell J, Jiang JK, Thomas CJ, Rosenwald A, Ott G, Muller-Hermelink HK, Gascoyne RD, Connors JM, et al. Chronic active B-cell-receptor signalling in diffuse large B-cell lymphoma. Nature. 2010 Jan 7;463(7277):88-92. doi: 10.1038/nature08638. PubMed [citation] PMID: 20054396, PMCID: PMC2845535

de Sena Brandine G, Smith AD. Falco: high-speed FastQC emulation for quality control of sequencing data. Version 2. F1000Res. 2019 Nov 7 [revised 2021 Jan 1];8:1874. doi: 10.12688/f1000research.21142.2. eCollection 2019. PubMed [citation] PMID: 33552473, PMCID: PMC7845152

de Souza CF, Sabedot TS, Malta TM, Stetson L, Morozova O, Sokolov A, Laird PW, Wiznerowicz M, Iavarone A, Snyder J, deCarvalho A, Sanborn Z, McDonald KL, Friedman WA, Tirapelli D, Poisson L, Mikkelsen T, Carlotti CG Jr, Kalkanis S, Zenklusen J, Salama SR, Barnholtz-Sloan JS, et al. A Distinct DNA Methylation Shift in a Subset of Glioma CpG Island Methylator Phenotypes during Tumor Recurrence. Cell Rep. 2018 Apr 10;23(2):637-651. doi: 10.1016/j.celrep.2018.03.107. PubMed [citation] PMID: 29642018, PMCID: PMC8859991

DePristo MA, Banks E, Poplin R, Garimella KV, Maguire JR, Hartl C, Philippakis AA, del Angel G, Rivas MA, Hanna M, McKenna A, Fennell TJ, Kernytsky AM, Sivachenko AY, Cibulskis K, Gabriel SB, Altshuler D, Daly MJ. A framework for variation discovery and genotyping using next-generation DNA sequencing data. Nat Genet. 2011 May;43(5):491-8. doi: 10.1038/ng.806. Epub 2011 Apr 10. PubMed [citation] PMID: 21478889, PMCID: PMC3083463

Devreker F, Hardy K, Van den Bergh M, Winston J, Biramane J, Englert Y. Noninvasive assessment of glucose and pyruvate uptake by human embryos after intracytoplasmic sperm injection and during the formation of pronuclei. Fertil Steril. 2000 May;73(5):947-54. PubMed [citation] PMID: 10785219

Di Tommaso P, Chatzou M, Floden EW, Barja PP, Palumbo E, Notredame C. Nextflow enables reproducible computational workflows. Nat Biotechnol. 2017 Apr 11;35(4):316-319. doi: 10.1038/nbt.3820. No abstract available. PubMed [citation] PMID: 28398311

Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, Gingeras TR. STAR: ultrafast universal RNA-seq aligner. Bioinformatics. 2013 Jan 1;29(1):15-21. doi: 10.1093/bioinformatics/bts635. Epub 2012 Oct 25. PubMed [citation] PMID: 23104886, PMCID: PMC3530905

Donington JS, Koo CW, Ballas MS. Novel therapies for non-small cell lung cancer. J Thorac Imaging. 2011 May;26(2):175-85. doi: 10.1097/RTI.0b013e3182161709. Review. PubMed [citation] PMID: 21508738

Dubois S, Viailly PJ, Bohers E, Bertrand P, Ruminy P, Marchand V, Maingonnat C, Mareschal S, Picquenot JM, Penther D, Jais JP, Tesson B, Peyrouze P, Figeac M, Desmots F, Fest T, Haioun C, Lamy T, Copie-Bergman C, Fabiani B, Delarue R, Peyrade F, et al. Biological and Clinical Relevance of Associated Genomic Alterations in MYD88 L265P and non-L265P-Mutated Diffuse Large B-Cell Lymphoma: Analysis of 361 Cases. Clin Cancer Res. 2017 May 1;23(9):2232-2244. doi: 10.1158/1078-0432.CCR-16-1922. Epub 2016 Dec 6. PubMed [citation] PMID: 27923841

Ehrlich M. DNA hypomethylation in cancer cells. Epigenomics. 2009 Dec;1(2):239-59. doi: 10.2217/epi.09.33. Review. PubMed [citation] PMID: 20495664, PMCID: PMC2873040

Etcheverry A, Aubry M, de Tayrac M, Vauleon E, Boniface R, Guenot F, Saikali S, Hamlat A, Riffaud L, Menei P, Quillien V, Mosser J. DNA methylation in glioblastoma: impact on gene expression and clinical outcome. BMC Genomics. 2010 Dec 14;11:701. doi: 10.1186/1471-2164-11-701. PubMed [citation] PMID: 21156036, PMCID: PMC3018478

Ettinger DS, Wood DE, Aisner DL, Akerley W, Bauman J, Chirieac LR, D'Amico TA, DeCamp MM, Dilling TJ, Dobelbower M, Doebele RC, Govindan R, Gubens MA, Hennon M, Horn L, Komaki R, Lackner RP, Lanuti M, Leal TA, Leisch LJ, Lilenbaum R, Lin J, et al. Non-Small Cell Lung Cancer, Version 5.2017, NCCN Clinical Practice Guidelines in Oncology. J Natl Compr Canc Netw. 2017 Apr;15(4):504-535. PubMed [citation] PMID: 28404761

Ettinger DS, Wood DE, Aisner DL, Akerley W, Bauman JR, Bharat A, Bruno DS, Chang JY, Chirieac LR, D'Amico TA, DeCamp M, Dilling TJ, Dowell J, Gettinger S, Grotz TE, Gubens MA, Hegde A, Lackner RP, Lanuti M, Lin J, Loo BW, Lovly CM, et al. Non-Small Cell Lung Cancer, Version 3.2022, NCCN Clinical Practice Guidelines in Oncology. J Natl Compr Canc Netw. 2022 May;20(5):497-530. doi: 10.6004/jnccn.2022.0025. PubMed [citation] PMID: 35545176

Ewels P, Magnusson M, Lundin S, Käller M. MultiQC: summarize analysis results for multiple tools and samples in a single report. Bioinformatics. 2016 Oct 1;32(19):3047-8. doi: 10.1093/bioinformatics/btw354. Epub 2016 Jun 16. PubMed [citation] PMID: 27312411, PMCID: PMC5039924

Fabbri M, Garzon R, Cimmino A, Liu Z, Zanesi N, Callegari E, Liu S, Alder H, Costinean S, Fernandez-Cymering C, Volinia S, Guler G, Morrison CD, Chan KK, Marcucci G, Calin GA, Huebner K, Croce CM. MicroRNA-29 family reverts aberrant methylation in lung cancer by targeting DNA methyltransferases 3A and 3B. Proc Natl Acad Sci U S A. 2007 Oct 2;104(40):15805-10. Epub 2007 Sep 21. PubMed [citation] PMID: 17890317, PMCID: PMC2000384

Fallacara AL, Zamperini C, Podolski-Renić A, Dinić J, Stanković T, Stepanović M, Mancini A, Rango E, Iovenitti G, Molinari A, Bugli F, Sanguinetti M, Torelli R, Martini M, Maccari L, Valoti M, Dreassi E, Botta M, Pešić M, Schenone S. A New Strategy for Glioblastoma Treatment: In Vitro and In Vivo Preclinical Characterization of Si306, a Pyrazolo[3,4-d]Pyrimidine Dual Src/P-Glycoprotein Inhibitor. Cancers (Basel). 2019 Jun 19;11(6). pii: E848. doi: 10.3390/cancers11060848. PubMed [citation] PMID: 31248184, PMCID: PMC6628362

Fang R, Yang W, Zhao X, Xiong F, Guo C, Xiao J, Chen L, Song X, Wang H, Chen J, Xiao X, Yao B, Cai LY. Chromosome screening using culture medium of embryos fertilised in vitro: a pilot clinical study. J Transl Med. 2019 Mar 8;17(1):73. doi: 10.1186/s12967-019-1827-1. PubMed [citation] PMID: 30849973, PMCID: PMC6408780

Farooqui MZ, Valdez J, Martyr S, Aue G, Saba N, Niemann CU, Herman SE, Tian X, Marti G, Soto S, Hughes TE, Jones J, Lipsky A, Pittaluga S, Stetler-Stevenson M, Yuan C, Lee YS, Pedersen LB, Geisler CH, Calvo KR, Arthur DC, Maric I, et al. Ibrutinib for previously untreated and relapsed or refractory chronic lymphocytic leukaemia with TP53 aberrations: a phase 2, single-arm trial. Lancet Oncol. 2015 Feb;16(2):169-76. doi: 10.1016/S1470-2045(14)71182-9. Epub 2014 Dec 31. PubMed [citation] PMID: 25555420, PMCID: PMC4342187

Farra C, Choucair F, Awwad J. Non-invasive pre-implantation genetic testing of human embryos: an emerging concept. Hum Reprod. 2018 Dec 1;33(12):2162-2167. doi: 10.1093/humrep/dey314. Review. Erratum in: Hum Reprod. 2019 Mar 1;34(3):590. PubMed [citation] PMID: 30357338

Feinberg AP, Gehrke CW, Kuo KC, Ehrlich M. Reduced genomic 5-methylcytosine content in human colonic neoplasia. Cancer Res. 1988 Mar 1;48(5):1159-61. PubMed [citation] PMID: 3342396

Feng RE. [IASLC/ATS/ERS international multidisciplinary new classification of lung adenocarcinoma and its clinical significance]. Zhonghua Jie He Hu Xi Za Zhi. 2012 Feb;35(2):95-6. Chinese. PubMed [citation] PMID: 22455963

Ferrara R, Imbimbo M, Malouf R, Paget-Bailly S, Calais F, Marchal C, Westeel V. Single or combined immune checkpoint inhibitors compared to first-line platinum-based chemotherapy with or without bevacizumab for people with advanced non-small cell lung cancer. Cochrane Database Syst Rev. 2021 Apr 30;4:CD013257. doi: 10.1002/14651858.CD013257.pub3. PubMed [citation] PMID: 33930176, PMCID: PMC8092423

Fischer I, Gagner JP, Law M, Newcomb EW, Zagzag D. Angiogenesis in gliomas: biology and molecular pathophysiology. Brain Pathol. 2005 Oct;15(4):297-310. Review. PubMed [citation] PMID: 16389942, PMCID: PMC8096031

Fomitcheva-Khartchenko A, Kashyap A, Geiger T, Kaigala GV. Space in cancer biology: its role and implications. Trends Cancer. 2022 Dec;8(12):1019-1032. doi: 10.1016/j.trecan.2022.07.008. Epub 2022 Aug 20. Review. PubMed [citation] PMID: 35995681

Fukumura K, Kawazu M, Kojima S, Ueno T, Sai E, Soda M, Ueda H, Yasuda T, Yamaguchi H, Lee J, Shishido-Hara Y, Sasaki A, Shirahata M, Mishima K, Ichimura K, Mukasa A, Narita Y, Saito N, Aburatani H, Nishikawa R, Nagane M, Mano H. Genomic characterization of primary central nervous system lymphoma. Acta Neuropathol. 2016 Jun;131(6):865-75. doi: 10.1007/s00401-016-1536-2. Epub 2016 Jan 12. PubMed [citation] PMID: 26757737

Furman RR, Cheng S, Lu P, Setty M, Perez AR, Guo A, Racchumi J, Xu G, Wu H, Ma J, Steggerda SM, Coleman M, Leslie C, Wang YL. Ibrutinib resistance in chronic lymphocytic leukemia. N Engl J Med. 2014 Jun 12;370(24):2352-4. doi: 10.1056/NEJMc1402716. Epub 2014 May 28. No abstract available. Erratum in: N Engl J Med. 2014 Jun 26;370(26):2547. Perez, Alijandro R [corrected to Perez, Alexendar R]. PubMed [citation] PMID: 24869597, PMCID: PMC4512173

Gallardo E, Navarro A, Viñolas N, Marrades RM, Diaz T, Gel B, Quera A, Bandres E, Garcia-Foncillas J, Ramirez J, Monzo M. miR-34a as a prognostic marker of relapse in surgically resected non-small-cell lung cancer. Carcinogenesis. 2009 Nov;30(11):1903-9. doi: 10.1093/carcin/bgp219. Epub 2009 Sep 7. PubMed [citation] PMID: 19736307

Gángó A, Alpár D, Galik B, Marosvári D, Kiss R, Fésüs V, Aczél D, Eyüpoglu E, Nagy N, Nagy Á, Krizsán S, Reiniger L, Farkas P, Kozma A, Ádám E, Tasnády S, Réti M, Matolcsy A, Gyenesei A, Mátrai Z, Bödör C. Dissection of

subclonal evolution by temporal mutation profiling in chronic lymphocytic leukemia patients treated with ibrutinib. Int J Cancer. 2020 Jan 1;146(1):85-93. doi: 10.1002/ijc.32502. Epub 2019 Jun 25. PubMed [citation] PMID: 31180577

Gardner DK, Meseguer M, Rubio C, Treff NR. Diagnosis of human preimplantation embryo viability. Hum Reprod Update. 2015 Nov-Dec;21(6):727-47. doi: 10.1093/humupd/dmu064. Epub 2015 Jan 6. Review. PubMed [citation] PMID: 25567750

Gardner DK, Lane M, Stevens J, Schoolcraft WB. Noninvasive assessment of human embryo nutrient consumption as a measure of developmental potential. Fertil Steril. 2001 Dec;76(6):1175-80. PubMed [citation] PMID: 11730746

Garijo D, Kinnings S, Xie L, Xie L, Zhang Y, Bourne PE, Gil Y. Quantifying reproducibility in computational biology: the case of the tuberculosis drugome. PLoS One. 2013 Nov 27;8(11):e80278. doi: 10.1371/journal.pone.0080278. eCollection 2013. PubMed [citation] PMID: 24312207, PMCID: PMC3842296

Ghafouri-Fard S, Niazi V, Taheri M. Role of miRNAs and lncRNAs in hematopoietic stem cell differentiation. Noncoding RNA Res. 2020 Dec 19;6(1):8-14. doi: 10.1016/j.ncrna.2020.12.002. eCollection 2021 Mar. Review. PubMed [citation] PMID: 33385102, PMCID: PMC7770514

Gianaroli L, Magli MC, Pomante A, Crivello AM, Cafueri G, Valerio M, Ferraretti AP. Blastocentesis: a source of DNA for preimplantation genetic testing. Results from a pilot study. Fertil Steril. 2014 Dec;102(6):1692-9.e6. doi: 10.1016/j.fertnstert.2014.08.021. Epub 2014 Sep 23. Erratum in: Fertil Steril. 2015 Aug;104(2):498. PubMed [citation] PMID: 25256935

Gombos K, Gálik B, Kalács KI, Gödöny K, Várnagy Á, Alpár D, Bódis J, Gyenesei A, Kovács GL. NGS-Based Application for Routine Non-Invasive Pre-Implantation Genetic Assessment in IVF. Int J Mol Sci. 2021 Feb 28;22(5). pii: 2443. doi: 10.3390/ijms22052443. PubMed [citation] PMID: 33671014, PMCID: PMC7957524

Gonzalez-Aguilar A, Idbaih A, Boisselier B, Habbita N, Rossetto M, Laurenge A, Bruno A, Jouvet A, Polivka M, Adam C, Figarella-Branger D, Miquel C, Vital A, Ghesquières H, Gressin R, Delwail V, Taillandier L, Chinot O, Soubeyran P, Gyan E, Choquet S, Houillier C, et al. Recurrent mutations of MYD88 and TBL1XR1 in primary central nervous system lymphomas. Clin Cancer Res. 2012 Oct 1;18(19):5203-11. doi: 10.1158/1078-0432.CCR-12-0845. Epub 2012 Jul 26. PubMed [citation] PMID: 22837180

Goradel NH, Mohammadi N, Haghi-Aminjan H, Farhood B, Negahdari B, Sahebkar A. Regulation of tumor angiogenesis by microRNAs: State of the art. J Cell Physiol. 2019 Feb;234(2):1099-1110. doi: 10.1002/jcp.27051. Epub 2018 Aug 2. Review. PubMed [citation] PMID: 30070704

Grommes C, Rubenstein JL, DeAngelis LM, Ferreri AJM, Batchelor TT. Comprehensive approach to diagnosis and treatment of newly diagnosed primary CNS lymphoma. Neuro Oncol. 2019 Feb 19;21(3):296-305. doi: 10.1093/neuonc/noy192. Review. PubMed [citation] PMID: 30418592, PMCID: PMC6380418

Guo Q, Wang J, Xiao J, Wang L, Hu X, Yu W, Song G, Lou J, Chen J. Heterogeneous mutation pattern in tumor tissue and circulating tumor DNA warrants parallel NGS panel testing. Mol Cancer. 2018 Aug 28;17(1):131. doi: 10.1186/s12943-018-0875-0. PubMed [citation] PMID: 30153823, PMCID: PMC6114875

Hagen JB. The origins of bioinformatics. Nat Rev Genet. 2000 Dec;1(3):231-6. doi: 10.1038/35042090. PubMed [citation] PMID: 11252753

Hammond ER, Shelling AN, Cree LM. Nuclear and mitochondrial DNA in blastocoele fluid and embryo culture medium: evidence and potential clinical use. Hum Reprod. 2016 Aug;31(8):1653-61. doi: 10.1093/humrep/dew132. Epub 2016 Jun 6. Review. PubMed [citation] PMID: 27270971

Hammond ER, McGillivray BC, Wicker SM, Peek JC, Shelling AN, Stone P, Chamley LW, Cree LM. Characterizing nuclear and mitochondrial DNA in spent embryo culture media: genetic contamination identified. Fertil Steril. 2017 Jan;107(1):220-228.e5. doi: 10.1016/j.fertnstert.2016.10.015. Epub 2016 Nov 16. PubMed [citation] PMID: 27865449

Handyside AH. Noninvasive preimplantation genetic testing: dream or reality? Fertil Steril. 2016 Nov;106(6):1324-1325. doi: 10.1016/j.fertnstert.2016.08.046. Epub 2016 Sep 16. No abstract available. PubMed [citation] PMID: 27645293

Hans CP, Weisenburger DD, Greiner TC, Gascoyne RD, Delabie J, Ott G, Müller-Hermelink HK, Campo E, Braziel RM, Jaffe ES, Pan Z, Farinha P, Smith LM, Falini B, Banham AH, Rosenwald A, Staudt LM, Connors JM, Armitage JO, Chan WC. Confirmation of the molecular classification of diffuse large B-cell lymphoma by immunohistochemistry using a tissue microarray. Blood. 2004 Jan 1;103(1):275-82. Epub 2003 Sep 22. PubMed [citation] PMID: 14504078

Hansen KD, Timp W, Bravo HC, Sabunciyan S, Langmead B, McDonald OG, Wen B, Wu H, Liu Y, Diep D, Briem E, Zhang K, Irizarry RA, Feinberg AP. Increased methylation variation in epigenetic domains across cancer types. Nat Genet. 2011 Jun 26;43(8):768-75. doi: 10.1038/ng.865. PubMed [citation] PMID: 21706001, PMCID: PMC3145050

Hayes JL, Tzika A, Thygesen H, Berri S, Wood HM, Hewitt S, Pendlebury M, Coates A, Willoughby L, Watson CM, Rabbitts P, Roberts P, Taylor GR. Diagnosis of copy number variation by Illumina next generation sequencing is comparable in performance to oligonucleotide array comparative genomic hybridisation. Genomics. 2013 Sep;102(3):174-81. doi: 10.1016/j.ygeno.2013.04.006. Epub 2013 Apr 15. PubMed [citation] PMID: 23598253

Hegi ME, Diserens AC, Gorlia T, Hamou MF, de Tribolet N, Weller M, Kros JM, Hainfellner JA, Mason W, Mariani L, Bromberg JE, Hau P, Mirimanoff RO, Cairncross JG, Janzer RC, Stupp R. MGMT gene silencing and benefit from temozolomide in glioblastoma. N Engl J Med. 2005 Mar 10;352(10):997-1003. PubMed [citation] PMID: 15758010

Herbst RS, Morgensztern D, Boshoff C. The biology and management of non-small cell lung cancer. Nature. 2018 Jan 24;553(7689):446-454. doi: 10.1038/nature25183. Review. PubMed [citation] PMID: 29364287

Hernández-Vargas P, Muñoz M, Domínguez F. Identifying biomarkers for predicting successful embryo implantation: applying single to multi-OMICs to improve reproductive outcomes. Hum Reprod Update. 2020 Feb 28;26(2):264-301. doi: 10.1093/humupd/dmz042. Review. PubMed [citation] PMID: 32096829

Hirsch FR, Scagliotti GV, Mulshine JL, Kwon R, Curran WJ Jr, Wu YL, Paz-Ares L. Lung cancer: current therapies and new targeted treatments. Lancet. 2017 Jan 21;389(10066):299-311. doi: 10.1016/S0140-6736(16)30958-8. Epub 2016 Aug 27. Review. PubMed [citation] PMID: 27574741

Ho JR, Arrach N, Rhodes-Long K, Ahmady A, Ingles S, Chung K, Bendikson KA, Paulson RJ, McGinnis LK. Pushing the limits of detection: investigation of cell-free DNA for aneuploidy screening in embryos. Fertil Steril. 2018 Aug;110(3):467-475.e2. doi: 10.1016/j.fertnstert.2018.03.036. Epub 2018 Jun 28. PubMed [citation] PMID: 29960707

Hochberg FH, Miller DC. Primary central nervous system lymphoma. J Neurosurg. 1988 Jun;68(6):835-53. Review. PubMed [citation] PMID: 3286832

Hu B, Wang Q, Wang YA, Hua S, Sauvé CG, Ong D, Lan ZD, Chang Q, Ho YW, Monasterio MM, Lu X, Zhong Y, Zhang J, Deng P, Tan Z, Wang G, Liao WT, Corley LJ, Yan H, Zhang J, You Y, Liu N, et al. Epigenetic Activation of WNT5A Drives Glioblastoma Stem Cell Differentiation and Invasive Growth. Cell. 2016 Nov 17;167(5):1281-1295.e18. doi: 10.1016/j.cell.2016.10.039. PubMed [citation] PMID: 27863244, PMCID: PMC5320931

Hua J, Liu J, Ma M, Xie L, Tian J. MicroRNA in the diagnosis of lung cancer: An overview of ten systematic reviews. Ann Clin Biochem. 2022 Nov 12:45632221128684. doi: 10.1177/00045632221128684. [Epub ahead of print] Review. PubMed [citation] PMID: 36085569

Huang L, Bogale B, Tang Y, Lu S, Xie XS, Racowsky C. Noninvasive preimplantation genetic testing for aneuploidy in spent medium may be more reliable than trophectoderm biopsy. Proc Natl Acad Sci U S A. 2019 Jul 9;116(28):14105-14112. doi: 10.1073/pnas.1907472116. Epub 2019 Jun 24. PubMed [citation] PMID: 31235575, PMCID: PMC6628824

Hurwitz BL, Westveld AH, Brum JR, Sullivan MB. Modeling ecological drivers in marine viral communities using comparative metagenomics and network analyses. Proc Natl Acad Sci U S A. 2014 Jul 22;111(29):10714-9. doi: 10.1073/pnas.1319778111. Epub 2014 Jul 7. PubMed [citation] PMID: 25002514, PMCID: PMC4115555

Hwang HS, Yoon DH, Suh C, Park CS, Huh J. Prognostic value of immunohistochemical algorithms in gastrointestinal diffuse large B-cell lymphoma. Blood Res. 2013 Dec;48(4):266-73. doi: 10.5045/br.2013.48.4.266. Epub 2013 Dec 24. PubMed [citation] PMID: 24466551, PMCID: PMC3894385

Jain P, Keating M, Wierda W, Estrov Z, Ferrajoli A, Jain N, George B, James D, Kantarjian H, Burger J, O'Brien S. Outcomes of patients with chronic lymphocytic leukemia after discontinuing ibrutinib. Blood. 2015 Mar 26;125(13):2062-7. doi: 10.1182/blood-2014-09-603670. Epub 2015 Jan 8. PubMed [citation] PMID: 25573991, PMCID: PMC4467871

Jensen LJ, Kuhn M, Stark M, Chaffron S, Creevey C, Muller J, Doerks T, Julien P, Roth A, Simonovic M, Bork P, von Mering C. STRING 8--a global view on proteins and their functional interactions in 630 organisms. Nucleic Acids Res. 2009 Jan;37(Database issue):D412-6. doi: 10.1093/nar/gkn760. Epub 2008 Oct 21. PubMed [citation] PMID: 18940858, PMCID: PMC2686466

Joshi P, Middleton J, Jeon YJ, Garofalo M. MicroRNAs in lung cancer. World J Methodol. 2014 Jun 26;4(2):59-72. doi: 10.5662/wjm.v4.i2.59. eCollection 2014 Jun 26. Review. PubMed [citation] PMID: 25332906, PMCID: PMC4202482

Kanehisa M, Bork P. Bioinformatics in the post-sequence era. Nat Genet. 2003 Mar;33 Suppl:305-10. Review. PubMed [citation] PMID: 12610540

Kanehisa M, Goto S. KEGG: kyoto encyclopedia of genes and genomes. Nucleic Acids Res. 2000 Jan 1;28(1):27-30. PubMed [citation] PMID: 10592173, PMCID: PMC102409

Katz-Jaffe MG, Gardner DK, Schoolcraft WB. Proteomic analysis of individual human embryos to identify novel biomarkers of development and viability. Fertil Steril. 2006 Jan;85(1):101-7. PubMed [citation] PMID: 16412738

Kechin A, Boyarskikh U, Kel A, Filipenko M. cutPrimers: A New Tool for Accurate Cutting of Primers from Reads of Targeted Next Generation Sequencing. J Comput Biol. 2017 Nov;24(11):1138-1143. doi: 10.1089/cmb.2017.0096. Epub 2017 Jul 17. PubMed [citation] PMID: 28715235

Kim H, Zheng S, Amini SS, Virk SM, Mikkelsen T, Brat DJ, Grimsby J, Sougnez C, Muller F, Hu J, Sloan AE, Cohen ML, Van Meir EG, Scarpace L, Laird PW, Weinstein JN, Lander ES, Gabriel S, Getz G, Meyerson M, Chin L, Barnholtz-Sloan JS, et al. Whole-genome and multisector exome sequencing of primary and post-treatment glioblastoma reveals patterns of tumor evolution. Genome Res. 2015 Mar;25(3):316-27. doi: 10.1101/gr.180612.114. Epub 2015 Feb 3. PubMed [citation] PMID: 25650244, PMCID: PMC4352879

Kim J, Lee IH, Cho HJ, Park CK, Jung YS, Kim Y, Nam SH, Kim BS, Johnson MD, Kong DS, Seol HJ, Lee JI, Joo KM, Yoon Y, Park WY, Lee J, Park PJ, Nam DH. Spatiotemporal Evolution of the Primary Glioblastoma Genome. Cancer Cell. 2015 Sep 14;28(3):318-28. doi: 10.1016/j.ccell.2015.07.013. PubMed [citation] PMID: 26373279

Kiss R, Alpár D, Gángó A, Nagy N, Eyupoglu E, Aczél D, Matolcsy A, Csomor J, Mátrai Z, Bödör C. Spatial clonal evolution leading to ibrutinib resistance and disease progression in chronic lymphocytic leukemia. Haematologica. 2019 Jan;104(1):e38-e41. doi: 10.3324/haematol.2018.202085. Epub 2018 Sep 27. No abstract available. PubMed [citation] PMID: 30262564, PMCID: PMC6312015

Klambauer G, Schwarzbauer K, Mayr A, Clevert DA, Mitterecker A, Bodenhofer U, Hochreiter S. cn.MOPS: mixture of Poissons for discovering copy number variations in next-generation sequencing data with a low false discovery rate. Nucleic Acids Res. 2012 May;40(9):e69. doi: 10.1093/nar/gks003. Epub 2012 Feb 1. PubMed [citation] PMID: 22302147, PMCID: PMC3351174

Klughammer J, Kiesel B, Roetzer T, Fortelny N, Nemc A, Nenning KH, Furtner J, Sheffield NC, Datlinger P, Peter N, Nowosielski M, Augustin M, Mischkulnig M, Ströbel T, Alpar D, Ergüner B, Senekowitsch M, Moser P, Freyschlag CF, Kerschbaumer J, Thomé C, Grams AE, et al. The DNA methylation landscape of glioblastoma disease progression shows extensive heterogeneity in time and space. Nat Med. 2018 Oct;24(10):1611-1624. doi: 10.1038/s41591-018-0156-x. Epub 2018 Aug 27. PubMed [citation] PMID: 30150718, PMCID: PMC6181207

Konnick E, Lockwood CM, Wu D. Targeted Next-Generation Sequencing of Acute Leukemia. Methods Mol Biol. 2017;1633:163-184. doi: 10.1007/978-1-4939-7142-8_11. PubMed [citation] PMID: 28735487

Kraan W, Horlings HM, van Keimpema M, Schilder-Tol EJ, Oud ME, Scheepstra C, Kluin PM, Kersten MJ, Spaargaren M, Pals ST. High prevalence of oncogenic MYD88 and CD79B mutations in diffuse large B-cell lymphomas presenting at immune-privileged sites. Blood Cancer J. 2013 Sep 6;3:e139. doi: 10.1038/bcj.2013.28. PubMed [citation] PMID: 24013661, PMCID: PMC3789201

Kraboth Z, Galik B, Tompa M, Kajtar B, Urban P, Gyenesei A, Miseta A, Kalman B. DNA CpG methylation in sequential glioblastoma specimens. J Cancer Res Clin Oncol. 2020 Nov;146(11):2885-2896. doi: 10.1007/s00432-020-03349-w. Epub 2020 Aug 10. PubMed [citation] PMID: 32779022, PMCID: PMC7519911

Krueger F, Andrews SR. Bismark: a flexible aligner and methylation caller for Bisulfite-Seq applications. Bioinformatics. 2011 Jun 1;27(11):1571-2. doi: 10.1093/bioinformatics/btr167. Epub 2011 Apr 14. PubMed [citation] PMID: 21493656, PMCID: PMC3102221

Kunde-Ramamoorthy G, Coarfa C, Laritsky E, Kessler NJ, Harris RA, Xu M, Chen R, Shen L, Milosavljevic A, Waterland RA. Comparison and quantitative verification of mapping algorithms for whole-genome bisulfite sequencing. Nucleic Acids Res. 2014 Apr;42(6):e43. doi: 10.1093/nar/gkt1325. Epub 2014 Jan 3. PubMed [citation] PMID: 24391148, PMCID: PMC3973287

Kuo HP, Ezell SA, Hsieh S, Schweighofer KJ, Cheung LW, Wu S, Apatira M, Sirisawad M, Eckert K, Liang Y, Hsu J, Chen CT, Beaupre D, Chang BY. The role of PIM1 in the ibrutinib-resistant ABC subtype of diffuse large B-cell lymphoma. Am J Cancer Res. 2016 Nov 1;6(11):2489-2501. eCollection 2016. PubMed [citation] PMID: 27904766, PMCID: PMC5126268

Kurtzer GM, Sochat V, Bauer MW. Singularity: Scientific containers for mobility of compute. PLoS One. 2017 May 11;12(5):e0177459. doi: 10.1371/journal.pone.0177459. eCollection 2017. PubMed [citation] PMID: 28494014, PMCID: PMC5426675

Kuznyetsov V, Madjunkova S, Antes R, Abramov R, Motamedi G, Ibarrientos Z, Librach C. Evaluation of a novel non-invasive preimplantation genetic screening approach. PLoS One. 2018 May 10;13(5):e0197262. doi: 10.1371/journal.pone.0197262. eCollection 2018. PubMed [citation] PMID: 29746572, PMCID: PMC5944986

Lamb R, Ablett MP, Spence K, Landberg G, Sims AH, Clarke RB. Wnt pathway activity in breast cancer sub-types and stem-like cells. PLoS One. 2013 Jul 4;8(7):e67811. doi: 10.1371/journal.pone.0067811. Print 2013. PubMed [citation] PMID: 23861811, PMCID: PMC3701602

Landau DA, Sun C, Rosebrock D, Herman SEM, Fein J, Sivina M, Underbayev C, Liu D, Hoellenriegel J, Ravichandran S, Farooqui MZH, Zhang W, Cibulskis C, Zviran A, Neuberg DS, Livitz D, Bozic I, Leshchiner I, Getz G, Burger JA, Wiestner A, Wu CJ. The evolutionary landscape of chronic lymphocytic leukemia treated with ibrutinib targeted therapy. Nat Commun. 2017 Dec 19;8(1):2185. doi: 10.1038/s41467-017-02329-y. PubMed [citation] PMID: 29259203, PMCID: PMC5736707

Landau DA, Tausch E, Taylor-Weiner AN, Stewart C, Reiter JG, Bahlo J, Kluth S, Bozic I, Lawrence M, Böttcher S, Carter SL, Cibulskis K, Mertens D, Sougnez CL, Rosenberg M, Hess JM, Edelmann J, Kless S, Kneba M, Ritgen M, Fink A, Fischer K, et al. Mutations driving CLL and their evolution in progression and relapse. Nature. 2015 Oct 22;526(7574):525-30. doi: 10.1038/nature15395. Epub 2015 Oct 14. PubMed [citation] PMID: 26466571, PMCID: PMC4815041

Landau DA, Carter SL, Stojanov P, McKenna A, Stevenson K, Lawrence MS, Sougnez C, Stewart C, Sivachenko A, Wang L, Wan Y, Zhang W, Shukla SA, Vartanov A, Fernandes SM, Saksena G, Cibulskis K, Tesar B, Gabriel S, Hacohen N, Meyerson M, Lander ES, et al. Evolution and impact of subclonal mutations in chronic lymphocytic leukemia. Cell. 2013 Feb 14;152(4):714-26. doi: 10.1016/j.cell.2013.01.019. PubMed [citation] PMID: 23415222, PMCID: PMC3575604

Landi MT, Zhao Y, Rotunno M, Koshiol J, Liu H, Bergen AW, Rubagotti M, Goldstein AM, Linnoila I, Marincola FM, Tucker MA, Bertazzi PA, Pesatori AC, Caporaso NE, McShane LM, Wang E. MicroRNA expression differentiates histology and predicts survival of lung cancer. Clin Cancer Res. 2010 Jan 15;16(2):430-41. doi: 10.1158/1078-0432.CCR-09-1736. Epub 2010 Jan 12. PubMed [citation] PMID: 20068076, PMCID: PMC3163170

Langfelder P, Horvath S. WGCNA: an R package for weighted correlation network analysis. BMC Bioinformatics. 2008 Dec 29;9:559. doi: 10.1186/1471-2105-9-559. PubMed [citation] PMID: 19114008, PMCID: PMC2631488

Larocca LM, Capello D, Rinelli A, Nori S, Antinori A, Gloghini A, Cingolani A, Migliazza A, Saglio G, Cammilleri-Broet S, Raphael M, Carbone A, Gaidano G. The molecular and phenotypic profile of primary central nervous system lymphoma identifies distinct categories of the disease and is consistent with histogenetic derivation from germinal center-related B cells. Blood. 1998 Aug 1;92(3):1011-9. PubMed [citation] PMID: 9680371

Leonardi T, Leger A. Nanopore RNA Sequencing Analysis. Methods Mol Biol. 2021;2284:569-578. doi: 10.1007/978-1-0716-1307-8_31. PubMed [citation] PMID: 33835464

Li H, Durbin R. Fast and accurate long-read alignment with Burrows-Wheeler transform. Bioinformatics. 2010 Mar 1;26(5):589-95. doi: 10.1093/bioinformatics/btp698. Epub 2010 Jan 15. PubMed [citation] PMID: 20080505, PMCID: PMC2828108

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R; 1000 Genome Project Data Processing Subgroup. The Sequence Alignment/Map format and SAMtools. Bioinformatics. 2009 Aug 15;25(16):2078-9. doi: 10.1093/bioinformatics/btp352. Epub 2009 Jun 8. PubMed [citation] PMID: 19505943, PMCID: PMC2723002

Liang X, Wu Q, Wang Y, Li S. MicroRNAs as early diagnostic biomarkers for non-small cell lung cancer (Review). Oncol Rep. 2023 Jan;49(1). pii: 8. doi: 10.3892/or.2022.8445. Epub 2022 Nov 16. Review. PubMed [citation] PMID: 36382661

Liao Y, Smyth GK, Shi W. The R package Rsubread is easier, faster, cheaper and better for alignment and quantification of RNA sequencing reads. Nucleic Acids Res. 2019 May 7;47(8):e47. doi: 10.1093/nar/gkz114. PubMed [citation] PMID: 30783653, PMCID: PMC6486549

Liu J, Wang Y, Liu Y, Liu Z, Cui Q, Ji N, Sun S, Wang B, Wang Y, Sun X, Liu Y. Immunohistochemical profile and prognostic significance in primary central nervous system lymphoma: Analysis of 89 cases. Oncol Lett. 2017 Nov;14(5):5505-5512. doi: 10.3892/ol.2017.6893. Epub 2017 Sep 6. PubMed [citation] PMID: 29113178, PMCID: PMC5656017

Loman N, Watson M. So you want to be a computational biologist? Nat Biotechnol. 2013 Nov;31(11):996-8. doi: 10.1038/nbt.2740. No abstract available. PubMed [citation] PMID: 24213777

Louis DN, Perry A, Reifenberger G, von Deimling A, Figarella-Branger D, Cavenee WK, Ohgaki H, Wiestler OD, Kleihues P, Ellison DW. The 2016 World Health Organization Classification of Tumors of the Central Nervous System: a summary. Acta Neuropathol. 2016 Jun;131(6):803-20. doi: 10.1007/s00401-016-1545-1. Epub 2016 May 9. Review. PubMed [citation] PMID: 27157931

Lv T, Jiang L, Kong L, Yang J. MicroRNA-29c-3p acts as a tumor suppressor gene and inhibits tumor progression in hepatocellular carcinoma by targeting TRIM31. Oncol Rep. 2020 Mar;43(3):953-964. doi: 10.3892/or.2020.7469. Epub 2020 Jan 17. PubMed [citation] PMID: 32020206, PMCID: PMC7041178

Maddocks KJ, Ruppert AS, Lozanski G, Heerema NA, Zhao W, Abruzzo L, Lozanski A, Davis M, Gordon A, Smith LL, Mantel R, Jones JA, Flynn JM, Jaglowski SM, Andritsos LA, Awan F, Blum KA, Grever MR, Johnson AJ, Byrd JC, Woyach JA. Etiology of Ibrutinib Therapy Discontinuation and Outcomes in Patients With Chronic Lymphocytic Leukemia. JAMA Oncol. 2015 Apr;1(1):80-7. doi: 10.1001/jamaoncol.2014.218. PubMed [citation] PMID: 26182309, PMCID: PMC4520535

Mains LM, Christenson L, Yang B, Sparks AE, Mathur S, Van Voorhis BJ. Identification of apolipoprotein A1 in the human embryonic secretome. Fertil Steril. 2011 Aug;96(2):422-427.e2. doi: 10.1016/j.fertnstert.2011.05.049. Epub 2011 Jun 15. PubMed [citation] PMID: 21676393

Makos M, Nelkin BD, Lerman MI, Latif F, Zbar B, Baylin SB. Distinct hypermethylation patterns occur at altered chromosome loci in human lung and colon cancer. Proc Natl Acad Sci U S A. 1992 Mar 1;89(5):1929-33. PubMed [citation] PMID: 1347428, PMCID: PMC48567

Malcikova J, Stano-Kozubik K, Tichy B, Kantorova B, Pavlova S, Tom N, Radova L, Smardova J, Pardy F, Doubek M, Brychtova Y, Mraz M, Plevova K, Diviskova E, Oltova A, Mayer J, Pospisilova S, Trbusek M. Detailed analysis of therapy-driven clonal evolution of TP53 mutations in chronic lymphocytic leukemia. Leukemia.2015 Apr;29(4):877-85. doi: 10.1038/leu.2014.297. Epub 2014 Oct 28. PubMed [citation] PMID: 25287991, PMCID: PMC4396398

Marcel M. Cutadapt removes adapter sequences from high-throughput sequencing reads. EMBnet J. 2011;17:10–12. doi: 10.14806/ej.17.1.200. – DO

Martin JA, Wang Z. Next-generation transcriptome assembly. Nat Rev Genet. 2011 Sep 7;12(10):671-82. doi: 10.1038/nrg3068. Review. PubMed [citation] PMID: 21897427

Marziali G, Signore M, Buccarelli M, Grande S, Palma A, Biffoni M, Rosi A, D'Alessandris QG, Martini M, Larocca LM, De Maria R, Pallini R, Ricci-Vitiani L. Metabolic/Proteomic Signature Defines Two Glioblastoma Subtypes With DifferentClinical Outcome. Sci Rep. 2016 Feb 9;6:21557. doi: 10.1038/srep21557. PubMed [citation] PMID: 26857460, PMCID: PMC4746700

Masca NG, Hensor EM, Cornelius VR, Buffa FM, Marriott HM, Eales JM, Messenger MP, Anderson AE, Boot C, Bunce C, Goldin RD, Harris J, Hinchliffe RF, Junaid H, Kingston S, Martin-Ruiz C, Nelson CP, Peacock J, Seed PT, Shinkins B, Staples KJ, Toombs J, et al. RIPOSTE: a framework for improving the design and analysis of laboratory-based research. Elife. 2015 May 7;4. doi: 10.7554/eLife.05519. PubMed [citation] PMID: 25951517, PMCID: PMC4461852

Mazieres J, He B, You L, Xu Z, Jablons DM. Wnt signaling in lung cancer. Cancer Lett. 2005 May 10;222(1):1-10. Review. PubMed [citation] PMID: 15837535

Merkel, D. (2014). Docker: lightweight linux containers for consistent development and deployment. Linux Journal, 2014(239), 2.

Meyer PN, Fu K, Greiner TC, Smith LM, Delabie J, Gascoyne RD, Ott G, Rosenwald A, Braziel RM, Campo E, Vose JM, Lenz G, Staudt LM, Chan WC, Weisenburger DD. Immunohistochemical methods for predicting cell of origin and survival in patients with diffuse large B-cell lymphoma treated with rituximab. J Clin Oncol. 2011 Jan 10;29(2):200-7. doi: 10.1200/JCO.2010.30.0368. Epub 2010 Dec 6. PubMed [citation] PMID: 21135273, PMCID: PMC3058275

Montesinos-Rongen M, Brunn A, Bentink S, Basso K, Lim WK, Klapper W, Schaller C, Reifenberger G, Rubenstein J, Wiestler OD, Spang R, Dalla-Favera R, Siebert R, Deckert M. Gene expression profiling suggests primary central nervous system lymphomas to be derived from a late germinal center B cell. Leukemia. 2008 Feb;22(2):400-5. Epub 2007 Nov 8. PubMed [citation] PMID: 17989719, PMCID: PMC6053313

Montesinos-Rongen M, Küppers R, Schlüter D, Spieker T, Van Roost D, Schaller C, Reifenberger G, Wiestler OD, Deckert-Schlüter M. Primary central nervous system lymphomas are derived from germinal-center B cells and show a preferential usage of the V4-34 gene segment. Am J Pathol. 1999 Dec;155(6):2077-86. PubMed [citation] PMID: 10595937, PMCID: PMC1866926

Montskó G, Zrínyi Z, Janáky T, Szabó Z, Várnagy Á, Kovács GL, Bódis J. Noninvasive embryo viability assessment by quantitation of human haptoglobin alpha-1 fragment in the in vitro fertilization culture medium: an additional tool to increase success rate. Fertil Steril. 2015 Mar;103(3):687-93. doi: 10.1016/j.fertnstert.2014.11.031. Epub 2015 Jan 7. PubMed [citation] PMID: 25577461

Montskó G, Gödöny K, Herczeg R, Várnagy Á, Bódis J, Kovács GL. Alpha-1 chain of human haptoglobin as viability marker of in vitro fertilized human embryos: information beyond morphology. Syst Biol Reprod Med. 2019 Apr;65(2):174-180. doi: 10.1080/19396368.2018.1518499. Epub 2018 Sep 17. PubMed [citation] PMID: 30222008

Müller F, Scherer M, Assenov Y, Lutsik P, Walter J, Lengauer T, Bock C. RnBeads 2.0: comprehensive analysis of DNA methylation data. Genome Biol. 2019 Mar 14;20(1):55. doi: 10.1186/s13059-019-1664-9. PubMed [citation] PMID: 30871603, PMCID: PMC6419383

Nagarajan RP, Zhang B, Bell RJ, Johnson BE, Olshen AB, Sundaram V, Li D, Graham AE, Diaz A, Fouse SD, Smirnov I, Song J, Paris PL, Wang T, Costello JF. Recurrent epimutations activate gene body promoters in primary glioblastoma. Genome Res. 2014 May;24(5):761-74. doi: 10.1101/gr.164707.113. Epub 2014 Apr 7. PubMed [citation] PMID: 24709822, PMCID: PMC4009606

Nakamura T, Tateishi K, Niwa T, Matsushita Y, Tamura K, Kinoshita M, Tanaka K, Fukushima S, Takami H, Arita H, Kubo A, Shuto T, Ohno M, Miyakita Y, Kocialkowski S, Sasayama T, Hashimoto N, Maehara T, Shibui S, Ushijima T, Kawahara N, Narita Y, et al. Recurrent mutations of CD79B and MYD88 are the hallmark of primary central nervous system lymphomas. Neuropathol Appl Neurobiol. 2016 Apr;42(3):279-90. doi: 10.1111/nan.12259. Epub 2015 Jul 20. PubMed [citation] PMID: 26111727

National Academies of Sciences, Engineering, and Medicine; Policy and Global Affairs; Committee on Science, Engineering, Medicine, and Public Policy; Board on Research Data and Information; Division on Engineering and Physical Sciences; Committee on Applied and Theoretical Statistics; Board on Mathematical Sciences and Analytics; Division on Earth and Life Studies; Nuclear and Radiation Studies Board; Division of Behavioral and Social Sciences and Education; Committee on National Statistics; Board on Behavioral, Cognitive, and Sensory Sciences; Committee on Reproducibility and Replicability in Science. Reproducibility and Replicability in Science. Washington (DC): National Academies Press (US); 2019 May 7. PubMed [citation] PMID: 31596559

Nguyen TTP, Suman KH, Nguyen TB, Nguyen HT, Do DN. The Role of miR-29s in Human Cancers-An Update. Biomedicines. 2022 Aug 29;10(9). pii: 2121. doi: 10.3390/biomedicines10092121. Review. PubMed [citation] PMID: 36140219, PMCID: PMC9495592

Nielsen R, Paul JS, Albrechtsen A, Song YS. Genotype and SNP calling from next-generation sequencing data. Nat Rev Genet. 2011 Jun;12(6):443-51. doi: 10.1038/nrg2986. Review. PubMed [citation] PMID: 21587300, PMCID: PMC3593722

Noushmehr H, Weisenberger DJ, Diefes K, Phillips HS, Pujara K, Berman BP, Pan F, Pelloski CE, Sulman EP, Bhat KP, Verhaak RG, Hoadley KA, Hayes DN, Perou CM, Schmidt HK, Ding L, Wilson RK, Van Den Berg D, Shen H, Bengtsson H, Neuvial P, Cope LM, et al. Identification of a CpG island methylator phenotype that defines a distinct subgroup of glioma. Cancer Cell. 2010 May 18;17(5):510-22. doi: 10.1016/j.ccr.2010.03.017. Epub 2010 Apr 15. PubMed [citation] PMID: 20399149, PMCID: PMC2872684

O'Neill BP, Illig JJ. Primary central nervous system lymphoma. Mayo Clin Proc. 1989 Aug;64(8):1005-20. Review. PubMed [citation] PMID: 2677529

Okonechnikov K, Conesa A, García-Alcalde F. Qualimap 2: advanced multi-sample quality control for high-throughput sequencing data. Bioinformatics. 2016 Jan 15;32(2):292-4. doi: 10.1093/bioinformatics/btv566. Epub 2015 Oct 1. PubMed [citation] PMID: 26428292, PMCID: PMC4708105

Okudela K, Tateishi Y, Umeda S, Mitsui H, Suzuki T, Saito Y, Woo T, Tajiri M, Masuda M, Miyagi Y, Ohashi K. Allelic imbalance in the miR-31 host gene locus in lung cancer--its potential role in carcinogenesis. PLoS One. 2014 Jun 30;9(6):e100581. doi: 10.1371/journal.pone.0100581. eCollection 2014. PubMed [citation] PMID: 24978700, PMCID: PMC4076198

Ouzounis CA. Rise and demise of bioinformatics? Promise and progress. PLoS Comput Biol. 2012;8(4):e1002487. doi: 10.1371/journal.pcbi.1002487. Epub 2012 Apr 26. PubMed [citation] PMID: 22570600, PMCID: PMC3343106

Palini S, Galluzzi L, De Stefani S, Bianchi M, Wells D, Magnani M, Bulletti C. Genomic DNA in human blastocoele fluid. Reprod Biomed Online. 2013 Jun;26(6):603-10. doi: 10.1016/j.rbmo.2013.02.012. Epub 2013 Mar 13. PubMed [citation] PMID: 23557766

Pasquali S, Hadjinicolaou AV, Chiarion Sileni V, Rossi CR, Mocellin S. Systemic treatments for metastatic cutaneous melanoma. Cochrane Database Syst Rev. 2018 Feb 6;2:CD011123. doi: 10.1002/14651858.CD011123.pub2. Review. PubMed [citation] PMID: 29405038, PMCID: PMC6491081

Patel AP, Tirosh I, Trombetta JJ, Shalek AK, Gillespie SM, Wakimoto H, Cahill DP, Nahed BV, Curry WT, Martuza RL, Louis DN, Rozenblatt-Rosen O, Suvà ML, Regev A, Bernstein BE. Single-cell RNA-seq highlights intratumoral heterogeneity in primary glioblastoma. Science. 2014 Jun 20;344(6190):1396-401. doi: 10.1126/science.1254257. Epub 2014 Jun 12. PubMed [citation] PMID: 24925914, PMCID: PMC4123637

Paternot G, Debrock S, De Neubourg D, D'Hooghe TM, Spiessens C. Semi-automated morphometric analysis of human embryos can reveal correlations between total embryo volume and clinical pregnancy. Hum Reprod. 2013 Mar;28(3):627-33. doi: 10.1093/humrep/des427. Epub 2013 Jan 12. PubMed [citation] PMID: 23315063

PDQ Adult Treatment Editorial Board. Non-Small Cell Lung Cancer Treatment (PDQ®): Health Professional Version. 2022 Nov 3. PDQ Cancer Information Summaries [Internet]. Bethesda (MD): National Cancer Institute (US); 2002-. PubMed [citation] PMID: 26389304

Perkel J. Democratic databases: science on GitHub. Nature. 2016 Oct 6;538(7623):127-128. doi: 10.1038/538127a. No abstract available. Erratum in: Nature. 2016 Oct 31;539(7627):126. PubMed [citation] PMID: 27708327

Perkel JM. How bioinformatics tools are bringing genetic analysis to the masses. Nature. 2017 Feb 28;543(7643):137-138. doi: 10.1038/543137a. No abstract available. PubMed [citation] PMID: 28252089

Piccolo SR, Frampton MB. Tools and techniques for computational reproducibility. Gigascience. 2016 Jul 11;5(1):30. doi: 10.1186/s13742-016-0135-4. Review. PubMed [citation] PMID: 27401684, PMCID: PMC4940747

Pilotto S, Molina-Vila MA, Karachaliou N, Carbognin L, Viteri S, González-Cao M, Bria E, Tortora G, Rosell R. Integrating the molecular background of targeted therapy and immunotherapy in lung cancer: a way to explore the impact of mutational landscape on tumor immunogenicity. Transl Lung Cancer Res. 2015 Dec;4(6):721-7. doi: 10.3978/j.issn.2218-6751.2015.10.11. Review. PubMed [citation] PMID: 26798581, PMCID: PMC4700230

Plaisier CL, Pan M, Baliga NS. A miRNA-regulatory network explains how dysregulated miRNAs perturb oncogenic processes across diverse cancers. Genome Res. 2012 Nov;22(11):2302-14. doi: 10.1101/gr.133991.111. Epub 2012 Jun 28. PubMed [citation] PMID: 22745231, PMCID: PMC3483559

Polyatskin IL, Artemyeva AS, Krivolapov YA. [Revised WHO classification of tumors of hematopoietic and lymphoid tissues, 2017 (4th edition): lymphoid tumors]. Arkh Patol. 2019;81(3):59-65. doi: 10.17116/patol20198103159. Russian. PubMed [citation] PMID: 31317932

Puente XS, Pinyol M, Quesada V, Conde L, Ordóñez GR, Villamor N, Escaramis G, Jares P, Beà S, González-Díaz M, Bassaganyas L, Baumann T, Juan M, López-Guerra M, Colomer D, Tubío JM, López C, Navarro A, Tornador C, Aymerich M, Rozman M, Hernández JM, et al. Whole-genome sequencing identifies recurrent mutations in chronic lymphocytic leukaemia. Nature. 2011 Jun 5;475(7354):101-5. doi: 10.1038/nature10113. PubMed [citation] PMID: 21642962, PMCID: PMC3322590

Quesada V, Conde L, Villamor N, Ordóñez GR, Jares P, Bassaganyas L, Ramsay AJ, Beà S, Pinyol M, Martínez-Trillos A, López-Guerra M, Colomer D, Navarro A, Baumann T, Aymerich M, Rozman M, Delgado J, Giné E, Hernández JM, González-Díaz M, Puente DA, Velasco G, et al. Exome sequencing identifies recurrent mutations

of the splicing factor SF3B1 gene in chronic lymphocytic leukemia. Nat Genet. 2011 Dec 11;44(1):47-52. doi: 10.1038/ng.1032. PubMed [citation] PMID: 22158541

Rabbani B, Tekin M, Mahdieh N. The promise of whole-exome sequencing in medical genetics. J Hum Genet. 2014 Jan;59(1):5-15. doi: 10.1038/jhg.2013.114. Epub 2013 Nov 7. Review. PubMed [citation] PMID: 24196381

Rajakumar S, Jamespaulraj S, Shah Y, Kejamurthy P, Jaganathan MK, Mahalingam G, Ramya Devi KT. Long non-coding RNAs: an overview on miRNA sponging and its co-regulation in lung cancer. Mol Biol Rep. 2022 Nov 28. doi: 10.1007/s11033-022-07995-w. [Epub ahead of print] Review. PubMed [citation] PMID: 36441373

Raoux D, Duband S, Forest F, Trombert B, Chambonnière ML, Dumollard JM, Khaddage A, Gentil-Perret A, Péoc'h M. Primary central nervous system lymphoma: immunohistochemical profile and prognostic significance. Neuropathology. 2010 Jun;30(3):232-40. doi: 10.1111/j.1440-1789.2009.01074.x. Epub 2009 Nov 18. PubMed [citation] PMID: 19925562

Rekulapelli A, E Flausino L, Iyer G, Balkrishnan R. Effectiveness of immunological agents in non-small cell lung cancer. Cancer Rep (Hoboken). 2022 Oct 26:e1739. doi: 10.1002/cnr2.1739. [Epub ahead of print] Review. PubMed [citation] PMID: 36289059

Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res. 2015 Apr 20;43(7):e47. doi: 10.1093/nar/gkv007. Epub 2015 Jan 20. PubMed [citation] PMID: 25605792, PMCID: PMC4402510

Rosenquist R, Ghia P, Hadzidimitriou A, Sutton LA, Agathangelidis A, Baliakas P, Darzentas N, Giudicelli V, Lefranc MP, Langerak AW, Belessi C, Davi F, Stamatopoulos K. Immunoglobulin gene sequence analysis in chronic lymphocytic leukemia: updated ERIC recommendations. Leukemia. 2017 Jul;31(7):1477-1481. doi: 10.1038/leu.2017.125. Epub 2017 Apr 25. No abstract available. PubMed [citation] PMID: 28439111, PMCID: PMC5508071

Rossi D, Cerri M, Deambrogi C, Sozzi E, Cresta S, Rasi S, De Paoli L, Spina V, Gattei V, Capello D, Forconi F, Lauria F, Gaidano G. The prognostic value of TP53 mutations in chronic lymphocytic leukemia is independent of Del17p13: implications for overall survival and chemorefractoriness. Clin Cancer Res. 2009 Feb 1;15(3):995-1004. doi: 10.1158/1078-0432.CCR-08-1630. PubMed [citation] PMID: 19188171

Roth P, Hottinger AF, Hundsberger T, Läubli H, Schucht P, Reinert M, Mamot C, Roelcke U, Pesce G, Hofer S, Weller M. A contemporary perspective on the diagnosis and treatment of diffuse gliomas in adults. Swiss Med Wkly. 2020 Jun 18;150:w20256. doi: 10.4414/smw.2020.20256. eCollection 2020 Jun 1. PubMed [citation] PMID: 32557428

Ru Y, Kechris KJ, Tabakoff B, Hoffman P, Radcliffe RA, Bowler R, Mahaffey S, Rossi S, Calin GA, Bemis L, Theodorescu D. The multiMiR R package and database: integration of microRNA-target interactions along with their disease and drug associations. Nucleic Acids Res. 2014;42(17):e133. doi: 10.1093/nar/gku631. Epub 2014 Jul 24. PubMed [citation] PMID: 25063298, PMCID: PMC4176155

Rubenstein JL, Fridlyand J, Shen A, Aldape K, Ginzinger D, Batchelor T, Treseler P, Berger M, McDermott M, Prados M, Karch J, Okada C, Hyun W, Parikh S, Haqq C, Shuman M. Gene expression and angiotropism in primary CNS lymphoma. Blood. 2006 May 1;107(9):3716-23. Epub 2006 Jan 17. PubMed [citation] PMID: 16418334, PMCID: PMC1895776

Rubio C, Navarro-Sánchez L, García-Pascual CM, Ocali O, Cimadomo D, Venier W, Barroso G, Kopcow L, Bahçeci M, Kulmann MIR, López L, De la Fuente E, Navarro R, Valbuena D, Sakkas D, Rienzi L, Simón C. Multicenter prospective study of concordance between embryonic cell-free DNA and trophectoderm biopsies from 1301 human blastocysts. Am J Obstet Gynecol. 2020 Nov;223(5):751.e1-751.e13. doi: 10.1016/j.ajog.2020.04.035. Epub 2020 May 26. PubMed [citation] PMID: 32470458

Silvestris F, D'Oronzo S, Lovero D, Palmirotta R, Dammacco F. Bone metastases from solid tumors: in search of predictive biomarkers for clinical translation. Oncogenomics. Cambridge: Academic Press; 2019. pp. 141–163.

Schipper LJ, Samsom KG, Snaebjornsson P, Battaglia T, Bosch LJW, Lalezari F, Priestley P, Shale C, van den Broek AJ, Jacobs N, Roepman P, van der Hoeven JJM, Steeghs N, Vollebergh MA, Marchetti S, Cuppen E, Meijer GA, Voest EE, Monkhorst K. Complete genomic characterization in patients with cancer of unknown primary origin in routine diagnostics. ESMO Open. 2022 Dec 1;7(6):100611. doi: 10.1016/j.esmoop.2022.100611. [Epub ahead of print] PubMed [citation] PMID: 36463731

Schröder MS, Culhane AC, Quackenbush J, Haibe-Kains B. survcomp: an R/Bioconductor package for performance assessment and comparison of survival models. Bioinformatics. 2011 Nov 15;27(22):3206-8. doi: 10.1093/bioinformatics/btr511. Epub 2011 Sep 7. PubMed [citation] PMID: 21903630, PMCID: PMC3208391

Schuh A, Becq J, Humphray S, Alexa A, Burns A, Clifford R, Feller SM, Grocock R, Henderson S, Khrebtukova I, Kingsbury Z, Luo S, McBride D, Murray L, Menju T, Timbs A, Ross M, Taylor J, Bentley D. Monitoring chronic lymphocytic leukemia progression by whole genome sequencing reveals heterogeneous clonal evolution patterns. Blood. 2012 Nov 15;120(20):4191-6. doi: 10.1182/blood-2012-05-433540. Epub 2012 Aug 22. PubMed [citation] PMID: 22915640

Scott DW, Wright GW, Williams PM, Lih CJ, Walsh W, Jaffe ES, Rosenwald A, Campo E, Chan WC, Connors JM, Smeland EB, Mottok A, Braziel RM, Ott G, Delabie J, Tubbs RR, Cook JR, Weisenburger DD, Greiner TC, Glinsmann-Gibson BJ, Fu K, Staudt LM, et al. Determining cell-of-origin subtypes of diffuse large B-cell lymphoma using gene expression in formalin-fixed paraffin-embedded tissue. Blood. 2014 Feb 20;123(8):1214-7. doi: 10.1182/blood-2013-11-536433. Epub 2014 Jan 7. PubMed [citation] PMID: 24398326, PMCID: PMC3931191

Sheffield NC, Bock C. LOLA: enrichment analysis for genomic region sets and regulatory elements in R and Bioconductor. Bioinformatics. 2016 Feb 15;32(4):587-9. doi: 10.1093/bioinformatics/btv612. Epub 2015 Oct 27. PubMed [citation] PMID: 26508757, PMCID: PMC4743627

Shitara A, Takahashi K, Goto M, Takahashi H, Iwasawa T, Onodera Y, Makino K, Miura H, Shirasawa H, Sato W, Kumazawa Y, Terada Y. Cell-free DNA in spent culture medium effectively reflects the chromosomal status of embryos following culturing beyond implantation compared to trophectoderm biopsy. PLoS One. 2021 Feb 11;16(2):e0246438. doi: 10.1371/journal.pone.0246438. eCollection 2021. PubMed [citation] PMID: 33571233, PMCID: PMC7877764

Siegel RL, Miller KD, Jemal A. Cancer statistics, 2019. CA Cancer J Clin. 2019 Jan;69(1):7-34. doi: 10.3322/caac.21551. Epub 2019 Jan 8. PubMed [citation] PMID: 30620402

Smith T, Heger A, Sudbery I. UMI-tools: modeling sequencing errors in Unique Molecular Identifiers to improve quantification accuracy. Genome Res. 2017 Mar;27(3):491-499. doi: 10.1101/gr.209601.116. Epub 2017 Jan 18. PubMed [citation] PMID: 28100584, PMCID: PMC5340976

Stephens ZD, Lee SY, Faghri F, Campbell RH, Zhai C, Efron MJ, Iyer R, Schatz MC, Sinha S, Robinson GE. Big Data: Astronomical or Genomical? PLoS Biol. 2015 Jul 7;13(7):e1002195. doi: 10.1371/journal.pbio.1002195. eCollection 2015 Jul. PubMed [citation] PMID: 26151137, PMCID: PMC4494865

Stigliani S, Anserini P, Venturini PL, Scaruffi P. Mitochondrial DNA content in embryo culture medium is significantly associated with human embryo fragmentation. Hum Reprod. 2013 Oct;28(10):2652-60. doi: 10.1093/humrep/det314. Epub 2013 Jul 25. PubMed [citation] PMID: 23887072

Sun DM, Tang BF, Li ZX, Guo HB, Cheng JL, Song PP, Zhao X. MiR-29c reduces the cisplatin resistance of non-small cell lung cancer cells by negatively regulating the PI3K/Akt pathway. Sci Rep. 2018 May 22;8(1):8007. doi: 10.1038/s41598-018-26381-w. PubMed [citation] PMID: 29789623, PMCID: PMC5964122

Tattini L, D'Aurizio R, Magi A. Detection of Genomic Structural Variants from Next-Generation Sequencing Data. Front Bioeng Biotechnol. 2015 Jun 25;3:92. doi: 10.3389/fbioe.2015.00092. eCollection 2015. Review. PubMed [citation] PMID: 26161383, PMCID: PMC4479793

Terkelsen T, Krogh A, Papaleo E. CAncer bioMarker Prediction Pipeline (CAMPP)-A standardized framework for the analysis of quantitative biological data. PLoS Comput Biol. 2020 Mar 16;16(3):e1007665. doi: 10.1371/journal.pcbi.1007665. eCollection 2020 Mar. PubMed [citation] PMID: 32176694, PMCID: PMC7108742

Thompsett AR, Ellison DW, Stevenson FK, Zhu D. V(H) gene sequences from primary central nervous system lymphomas indicate derivation from highly mutated germinal center B cells with ongoing mutational activity. Blood. 1999 Sep 1;94(5):1738-46. PubMed [citation] PMID: 10477699

Tikkanen T, Leroy B, Fournier JL, Risques RA, Malcikova J, Soussi T. Seshat: A Web service for accurate annotation, validation, and analysis of TP53 variants generated by conventional and next-generation sequencing. Hum Mutat. 2018 Jul;39(7):925-933. doi: 10.1002/humu.23543. Epub 2018 May 17. PubMed [citation] PMID: 29696732

Tompa M, Kalovits F, Nagy A, Kalman B. Contribution of the Wnt Pathway to Defining Biology of Glioblastoma. Neuromolecular Med. 2018 Dec;20(4):437-451. doi: 10.1007/s12017-018-8514-x. Epub 2018 Sep 26. Review. PubMed [citation] PMID: 30259273

Travis WD. The 2015 WHO classification of lung tumors. Pathologe. 2014 Nov;35 Suppl 2:188. doi: 10.1007/s00292-014-1974-3. Review. No abstract available. PubMed [citation] PMID: 25394966

Vater I, Montesinos-Rongen M, Schlesner M, Haake A, Purschke F, Sprute R, Mettenmeyer N, Nazzal I, Nagel I, Gutwein J, Richter J, Buchhalter I, Russell RB, Wiestler OD, Eils R, Deckert M, Siebert R. The mutational pattern of primary lymphoma of the central nervous system determined by whole-exome sequencing. Leukemia. 2015 Mar;29(3):677-85. doi: 10.1038/leu.2014.264. Epub 2014 Sep 5. PubMed [citation] PMID: 25189415

Vera-Rodriguez M, Diez-Juan A, Jimenez-Almazan J, Martinez S, Navarro R, Peinado V, Mercader A, Meseguer M, Blesa D, Moreno I, Valbuena D, Rubio C, Simon C. Origin and composition of cell-free DNA in spent medium from human embryo culture during preimplantation development. Hum Reprod. 2018 Apr 1;33(4):745-756. doi: 10.1093/humrep/dey028. PubMed [citation] PMID: 29471395

Verhaak RG, Hoadley KA, Purdom E, Wang V, Qi Y, Wilkerson MD, Miller CR, Ding L, Golub T, Mesirov JP, Alexe G, Lawrence M, O'Kelly M, Tamayo P, Weir BA, Gabriel S, Winckler W, Gupta S, Jakkula L, Feiler HS, Hodgson JG, James CD, et al. Integrated genomic analysis identifies clinically relevant subtypes of glioblastoma characterized by abnormalities in PDGFRA, IDH1, EGFR, and NF1. Cancer Cell. 2010 Jan 19;17(1):98-110. doi: 10.1016/j.ccr.2009.12.020. PubMed [citation] PMID: 20129251, PMCID: PMC2818769

Vlachos IS, Zagganas K, Paraskevopoulou MD, Georgakilas G, Karagkouni D, Vergoulis T, Dalamagas T, Hatzigeorgiou AG. DIANA-miRPath v3.0: deciphering microRNA function with experimental support. Nucleic

Acids Res. 2015 Jul 1;43(W1):W460-6. doi: 10.1093/nar/gkv403. Epub 2015 May 14. PubMed [citation] PMID: 25977294, PMCID: PMC4489228

Wang J, Cazzato E, Ladewig E, Frattini V, Rosenbloom DI, Zairis S, Abate F, Liu Z, Elliott O, Shin YJ, Lee JK, Lee IH, Park WY, Eoli M, Blumberg AJ, Lasorella A, Nam DH, Finocchiaro G, Iavarone A, Rabadan R. Clonal evolution of glioblastoma under therapy. Nat Genet. 2016 Jul;48(7):768-76. doi: 10.1038/ng.3590. Epub 2016 Jun 6. PubMed [citation] PMID: 27270107, PMCID: PMC5627776

Wang JH, Gouda-Vossos A, Dzamko N, Halliday G, Huang Y. DNA extraction from fresh-frozen and formalin-fixed, paraffin-embedded human brain tissue. Neurosci Bull. 2013 Oct;29(5):649-54. doi: 10.1007/s12264-013-1379-y. Epub 2013 Aug 30. PubMed [citation] PMID: 23996594, PMCID: PMC5562648

Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic Acids Res. 2010 Sep;38(16):e164. doi: 10.1093/nar/gkq603. Epub 2010 Jul 3. PubMed [citation] PMID: 20601685, PMCID: PMC2938201

Wang L, Lawrence MS, Wan Y, Stojanov P, Sougnez C, Stevenson K, Werner L, Sivachenko A, DeLuca DS, Zhang L, Zhang W, Vartanov AR, Fernandes SM, Goldstein NR, Folco EG, Cibulskis K, Tesar B, Sievers QL, Shefler E, Gabriel S, Hacohen N, Reed R, et al. SF3B1 and other novel cancer genes in chronic lymphocytic leukemia. N Engl J Med. 2011 Dec 29;365(26):2497-506. doi: 10.1056/NEJMoa1109016. Epub 2011 Dec 12. PubMed [citation] PMID: 22150006, PMCID: PMC3685413

Wang Q, Hu B, Hu X, Kim H, Squatrito M, Scarpace L, deCarvalho AC, Lyu S, Li P, Li Y, Barthel F, Cho HJ, Lin YH, Satani N, Martinez-Ledesma E, Zheng S, Chang E, Sauvé CG, Olar A, Lan ZD, Finocchiaro G, Phillips JJ, et al. Tumor Evolution of Glioma-Intrinsic Gene Expression Subtypes Associates with Immunological Changes in the Microenvironment. Cancer Cell. 2017 Jul 10;32(1):42-56.e6. doi: 10.1016/j.ccell.2017.06.003. Erratum in: Cancer Cell. 2018 Jan 8;33(1):152. PubMed [citation] PMID: 28697342, PMCID: PMC5599156

Wang SH, Zhang C, Wang Y. microRNA regulation of pluripotent state transition. Essays Biochem. 2020 Dec 7;64(6):947-954. doi: 10.1042/EBC20200028. Review. PubMed [citation] PMID: 33034348

Wang Y, Zhao Y, Bollas A, Wang Y, Au KF. Nanopore sequencing technology, bioinformatics and applications. Nat Biotechnol. 2021 Nov;39(11):1348-1365. doi: 10.1038/s41587-021-01108-x. Epub 2021 Nov 8. Review. PubMed [citation] PMID: 34750572, PMCID: PMC8988251

Wang Z, Gerstein M, Snyder M. RNA-Seq: a revolutionary tool for transcriptomics. Nat Rev Genet. 2009 Jan;10(1):57-63. doi: 10.1038/nrg2484. Review. PubMed [citation] PMID: 19015660, PMCID: PMC2949280

Wani JA, Majid S, Imtiyaz Z, Rehman MU, Alsaffar RM, Shah NN, Alshehri S, Ghoneim MM, Imam SS. MiRNAs in Lung Cancer: Diagnostic, Prognostic, and Therapeutic Potential. Diagnostics (Basel). 2022 Jul 1;12(7). pii: 1610. doi: 10.3390/diagnostics12071610. Review. PubMed [citation] PMID: 35885514, PMCID: PMC9322918

Wells D. Embryo aneuploidy and the role of morphological and genetic screening. Reprod Biomed Online. 2010 Sep;21(3):274-7. doi: 10.1016/j.rbmo.2010.06.035. Epub 2010 Jun 30. Review. PubMed [citation] PMID: 20688567

Wilm A, Aw PP, Bertrand D, Yeo GH, Ong SH, Wong CH, Khor CC, Petric R, Hibberd ML, Nagarajan N. LoFreq: a sequence-quality aware, ultra-sensitive variant caller for uncovering cell-population heterogeneity from high-throughput sequencing datasets. Nucleic Acids Res. 2012 Dec;40(22):11189-201. doi: 10.1093/nar/gks918. Epub 2012 Oct 12. PubMed [citation] PMID: 23066108, PMCID: PMC3526318

Woyach JA, Ruppert AS, Guinn D, Lehman A, Blachly JS, Lozanski A, Heerema NA, Zhao W, Coleman J, Jones D, Abruzzo L, Gordon A, Mantel R, Smith LL, McWhorter S, Davis M, Doong TJ, Ny F, Lucas M, Chase W, Jones JA, Flynn JM, et al. BTK(C481S)-Mediated Resistance to Ibrutinib in Chronic Lymphocytic Leukemia. J Clin Oncol. 2017 May 1;35(13):1437-1443. doi: 10.1200/JCO.2016.70.2282. Epub 2017 Feb 13. PubMed [citation] PMID: 28418267, PMCID: PMC5455463

Woyach JA, Furman RR, Liu TM, Ozer HG, Zapatka M, Ruppert AS, Xue L, Li DH, Steggerda SM, Versele M, Dave SS, Zhang J, Yilmaz AS, Jaglowski SM, Blum KA, Lozanski A, Lozanski G, James DF, Barrientos JC, Lichter P, Stilgenbauer S, Buggy JJ, et al. Resistance mechanisms for the Bruton's tyrosine kinase inhibitor ibrutinib. N Engl J Med. 2014 Jun 12;370(24):2286-94. doi: 10.1056/NEJMoa1400029. Epub 2014 May 28. PubMed [citation] PMID: 24869598, PMCID: PMC4144824

Wright G, Tan B, Rosenwald A, Hurt EH, Wiestner A, Staudt LM. A gene expression-based method to diagnose clinically distinct subgroups of diffuse large B cell lymphoma. Proc Natl Acad Sci U S A. 2003 Aug 19;100(17):9991-6. Epub 2003 Aug 4. PubMed [citation] PMID: 12900505, PMCID: PMC187912

Wu H, Ding C, Shen X, Wang J, Li R, Cai B, Xu Y, Zhong Y, Zhou C. Medium-based noninvasive preimplantation genetic diagnosis for human α-thalassemias-SEA. Medicine (Baltimore). 2015 Mar;94(12):e669. doi: 10.1097/MD.0000000000000669. PubMed [citation] PMID: 25816038, PMCID: PMC4554004

Xu J, Fang R, Chen L, Chen D, Xiao JP, Yang W, Wang H, Song X, Ma T, Bo S, Shi C, Ren J, Huang L, Cai LY, Yao B, Xie XS, Lu S. Noninvasive chromosome screening of human embryos by genome sequencing of embryo culture medium for in vitro fertilization. Proc Natl Acad Sci U S A. 2016 Oct 18;113(42):11907-11912. Epub 2016 Sep 29. PubMed [citation] PMID: 27688762, PMCID: PMC5081593

Yamada S, Ishida Y, Matsuno A, Yamazaki K. Primary diffuse large B-cell lymphomas of central nervous system exhibit remarkably high prevalence of oncogenic MYD88 and CD79B mutations. Leuk Lymphoma. 2015 Jul;56(7):2141-5. doi: 10.3109/10428194.2014.979413. Epub 2015 Jan 14. PubMed [citation] PMID: 25347427

Yan H, Tang S, Tang S, Zhang J, Guo H, Qin C, Hu H, Zhong C, Yang L, Zhu Y, Zhou H. miRNAs in anti-cancer drug resistance of non-small cell lung cancer: Recent advances and future potential. Front Pharmacol. 2022 Oct 25;13:949566. doi: 10.3389/fphar.2022.949566. eCollection 2022. Review. PubMed [citation] PMID: 36386184, PMCID: PMC9640411

Yang X, Mackenzie SA. Approaches to Whole-Genome Methylome Analysis in Plants. Methods Mol Biol. 2020;2093:15-31. doi: 10.1007/978-1-0716-0179-2_2. PubMed [citation] PMID: 32088886

Yeung QSY, Zhang YX, Chung JPW, Lui WT, Kwok YKY, Gui B, Kong GWS, Cao Y, Li TC, Choy KW. A prospective study of non-invasive preimplantation genetic testing for aneuploidies (NiPGT-A) using next-generation sequencing (NGS) on spent culture media (SCM). J Assist Reprod Genet. 2019 Aug;36(8):1609-1621. doi: 10.1007/s10815-019-01517-7. Epub 2019 Jul 10. PubMed [citation] PMID: 31292818, PMCID: PMC6707994

Zhang G, Sun M, Wang J, Lei M, Li C, Zhao D, Huang J, Li W, Li S, Li J, Yang J, Luo Y, Hu S, Zhang B. PacBio full-length cDNA sequencing integrated with RNA-seq reads drastically improves the discovery of splicing transcripts in rice. Plant J. 2019 Jan;97(2):296-305. doi: 10.1111/tpj.14120. Epub 2018 Dec 3. PubMed [citation] PMID: 30288819

Zhang M, Feng Y, Qu C, Meng M, Li W, Ye M, Li S, Li S, Ma Y, Wu N, Jia S. Comparison of the somatic mutations between circulating tumor DNA and tissue DNA in Chinese patients with non-small cell lung cancer. Int J Biol Markers. 2022 Dec;37(4):386-394. doi: 10.1177/03936155221099036. Epub 2022 Jul 6. PubMed [citation] PMID: 35791673

Zheng M, Perry AM, Bierman P, Loberiza F Jr, Nasr MR, Szwajcer D, Del Bigio MR, Smith LM, Zhang W, Greiner TC. Frequency of MYD88 and CD79B mutations, and MGMT methylation in primary central nervous system diffuse large B-cell lymphoma. Neuropathology. 2017 Dec;37(6):509-516. doi: 10.1111/neup.12405. Epub 2017 Aug 30. PubMed [citation] PMID: 28856744

Zhou S, Swanstrom R. Fact and Fiction about 1%: Next Generation Sequencing and the Detection of Minor Drug Resistant Variants in HIV-1 Populations with and without Unique Molecular Identifiers. Viruses. 2020 Aug 4;12(8). pii: E850. doi: 10.3390/v12080850. PubMed [citation] PMID: 32759675, PMCID: PMC7472098

Zhou Y, Liu W, Xu Z, Zhu H, Xiao D, Su W, Zeng R, Feng Y, Duan Y, Zhou J, Zhong M. Analysis of Genomic Alteration in Primary Central Nervous System Lymphoma and the Expression of Some Related Genes. Neoplasia. 2018 Oct;20(10):1059-1069. doi: 10.1016/j.neo.2018.08.012. Epub 2018 Sep 15. PubMed [citation] PMID: 30227305, PMCID: PMC6141698

Zhu J, Li R, Tiselius E, Roudi R, Teghararian O, Suo C, Song H. Immunotherapy (excluding checkpoint inhibitors) for stage I to III non-small cell lung cancer treated with surgery or radiotherapy with curative intent. Cochrane Database Syst Rev. 2017 Dec 16;12:CD011300. doi: 10.1002/14651858.CD011300.pub2. Review. Update in: Cochrane Database Syst Rev. 2021 Dec 6;12:CD011300. PubMed [citation] PMID: 29247502, PMCID: PMC6486009

https://www.bioinformatics.babraham.ac.uk/projects/fastqc/, 2019.02.12

https://www.bioinformatics.babraham.ac.uk/projects/trim_galore/, 2019.02.12

https://CRAN.R-project.org/package=epiR, 2020.02.03

https://CRAN.R-project.org/package=ggplot2, 2020.02.03

https://CRAN.R-project.org/package=rpart, 2022.05.15

https://docs.anaconda.com/, 2019.01.03