ALMA MATER STUDIORUM
UNIVERSITÀ DI BOLOGNA

DOTTORATO DI RICERCA IN

ONCOLOGIA, EMATOLOGIA E PATOLOGIA

Ciclo 37

Settore Concorsuale: 06/D3 - Malattie del sangue, Oncologia e reumatologia

Settore Scientifico Disciplinare: MED/15 Malattie del sangue

Multiomics data integration of acute myeloid leukemia cases to identify immunogenic profiles

Presentata da: Davide Angeli

Coordinatore Dottorato

Manuela Ferracin

Supervisore

Manuela Ferracin

Co-supervisore

Oriana Nanni

Esame finale anno 2025

1. Abstract

This study explores molecular mechanisms of 5-Azacytidine response in acute myeloid leukemia (AML) using an integrative multi-omics approach, combining data from whole exome sequencing (WES), RNA sequencing (RNA-seq), and neoantigen profiling. We aimed to characterize genetic mutations, transcriptional responses, and immunogenic profiles within AML patient clusters to identify potential biomarkers for predicting treatment efficacy. Samples were collected from patients before and 48 hours after 5-Azacytidine treatment, allowing for a comparative analysis of mutation and gene expression changes in response to therapy.

In WES data, we identified distinct mutational landscapes in two patient clusters. Cluster A showed genetic patterns favoring apoptosis and absence of KIT mutations, indicating a possible predisposition to 5-Azacytidine responsiveness. In contrast, Cluster B displayed alterations in key oncogenes such as KIT, ATM, and GNAS, which may contribute to cellular survival mechanisms and potential therapy resistance. Transcriptomic analysis revealed enriched metabolic and cell survival pathways in Cluster B, suggesting adaptive responses that could undermine the efficacy of 5-Azacytidine as a single agent. By contrast, Cluster A's transcriptional profile was associated with a more direct apoptotic response, aligning with its mutation profile. Neoantigen analysis further distinguished these clusters, with Cluster B showing a higher neoantigen load, suggesting an immunologically active environment that may enhance immune surveillance. This raises the potential for immunotherapy applications, particularly in Cluster B, where could complement 5-Azacytidine to overcome resistance mechanisms.

The integration of WES, RNA-seq, and neoantigen data underscores the heterogeneity within AML and provides insight into stratified therapy. Cluster A, with its apoptotic gene signatures and lower neoantigen burden, may respond more favorably to 5-Azacytidine alone, while Cluster B could benefit from combined targeted and immune therapies to address its adaptive resistance profile. This research highlights the importance of comprehensive molecular profiling in AML and lays the foundation for personalized therapeutic approaches based.

2. Introduction

2.1 Overview of Acute Myeloid Leukemia

Acute Myeloid Leukemia (AML) is a highly heterogeneous clonal hematologic malignancy characterized by the uncontrolled proliferation of immature myeloid progenitor cells, known as blasts. These cells accumulate in the bone marrow and peripheral blood, disrupting normal hematopoiesis and leading to life-threatening cytopenias, including anemia, neutropenia, and thrombocytopenia [1-4].

Consequently, patients commonly experience clinical symptoms such as fatigue, bleeding, and recurrent infections. These symptoms arise from the underproduction of normal blood cells; for instance, anemia leads to fatigue and weakness, while thrombocytopenia results in easy bruising and bleeding [5,6]. In some cases, organ infiltration may occur, causing abdominal fullness due to splenomegaly or hepatomegaly [2,3]. The rapid progression of AML, in the absence of immediate intervention, can lead to severe morbidity and mortality due to the failure of normal blood cell production and associated complications [5,7].

2.2 Epidemiology and Incidence

AML is the most common type of acute leukemia in adults, with an annual incidence of approximately 4.3 cases per 100,000 people in the United States, a figure that aligns with global incidence rates [8,9]. The median age at diagnosis for acute myeloid leukemia (AML) is 68 years, with a notable increase in incidence as individuals age. This trend underscores the aging populations in many developed countries [10,11]. Age is a critical factor in AML prognosis, as older patients tend to have poorer survival outcomes due to a higher prevalence of comorbidities

and the presence of adverse genetic mutations [1,3]. Although AML represents a relatively low percentage of overall cancer cases, accounting for roughly 1.1% of all cancers, the burden is increasing among older adults, largely attributable to increased life expectancy [8]. While AML is infrequent in children, it occurs more frequently in individuals with certain inherited genetic disorders, including Down syndrome, Fanconi anemia, and Bloom syndrome [1,5].

2.3 Classification Systems

The classification of AML has evolved significantly over the past decade, driven by advancements in understanding its molecular pathogenesis. Historically, the French-American-British (FAB) classification system was widely used, relying on morphologic and cytochemical criteria [9]. However, modern classification systems incorporate genetic and molecular markers, providing more accurate prognostic and therapeutic guidance. The three major classification frameworks currently employed are:

- World Health Organization (WHO) 2022: The latest WHO classification integrates genetic mutations and chromosomal abnormalities. It eliminates the 20% blast threshold in cases with AML-defining genetic abnormalities, reflecting a shift toward molecular diagnostics [12-14].
- European LeukemiaNet (ELN) 2022: ELN guidelines emphasize risk stratification based on specific genetic alterations, such as FLT3, NPM1, and TP53 mutations. These guidelines also incorporate measurable residual disease (MRD) as a monitoring tool [1,3,15].
- International Consensus Classification (ICC) 2022: This new classification framework further lowers the blast

count threshold to 10% in cases involving recurrent genetic abnormalities, underscoring the importance of molecular alterations in diagnosing AML [16].

These classification frameworks aim to tailor treatment based on individual patient risk, highlighting the critical role of genetics and molecular diagnostics in AML management [17].

2.4 Pathogenesis and Molecular Basis

The pathogenesis of AML is a multifactorial process involving genetic mutations, epigenetic alterations, and environmental exposures. Genomic studies have identified mutations in over 95% of AML cases, contributing to leukemogenesis [13]. Key mutations associated with AML include:

- FLT3: Internal tandem duplications (ITDs) and tyrosine kinase domain (TKD) mutations are among the most frequent genetic alterations in AML. FLT3-ITD mutations are especially associated with poor prognosis due to increased relapse rates [18].
- NPM1: Mutations in NPM1 are linked to a favorable prognosis, particularly when occurring without concurrent FLT3-ITD mutations [19].
- TP53: Mutations in the TP53 gene, which is crucial for tumor suppression, are more frequently associated with therapy-related AML and complex karyotypes, often leading to poor outcomes [20,21].
- IDH1/IDH2: Mutations in these genes, which play roles in the metabolic reprogramming of AML cells, are present in about 20% of cases and have become significant therapeutic targets due to the development of targeted inhibitors [22].
- CEBPA: Mutations in CEBPA, which encodes a transcription factor critical for hematopoiesis, are significant in AML

classification and prognosis, and are often associated with normal karyotype AML. These mutations are predictive of response to standard chemotherapy and are used to stratify risk in clinical practice [23].

These genetic alterations interact with epigenetic changes, such as mutations in DNMT3A, TET2, and ASXL1, which impact DNA methylation and histone modification, further driving leukemogenesis [24]. Understanding the complex interplay of these genetic and epigenetic factors is crucial for developing innovative therapeutic strategies.

2.5 Diagnosis of AML

The diagnosis of AML is primarily established by the presence of 20% or more myeloid blasts in the bone marrow or peripheral blood, as defined by the WHO classification [12]. Diagnosis is often augmented by morphological, cytogenetic, and molecular studies. Cytogenetic analysis, which examines chromosomal abnormalities, remains vital for defining prognosis and guiding treatment decisions. For example, translocations such as t(8;21) and inv(16) are associated with favorable outcomes, whereas monosomy 7 and complex karyotypes indicate a poor prognosis [7].

In recent years, next-generation sequencing (NGS) has become a cornerstone in AML diagnostics, enabling the identification of mutations in key genes, such as IDH1/2, DNMT3A, and RUNX1, which have both prognostic and therapeutic implications [25]. NGS has also facilitated the detection of MRD, allowing clinicians to monitor disease progression and relapse at a molecular level [26].

2.6 Treatment of AML: Challenges and Advances

The treatment of AML is complicated by its genetic and clinical heterogeneity, as well as the frequent occurrence of drug resistance and relapse. Current frontline therapy involves a combination of intensive chemotherapy, typically with cytarabine and an anthracycline, followed by consolidation therapy to achieve long-term remission. However, treatment outcomes in AML remain suboptimal, particularly in older adults and those with adverse cytogenetic or molecular features [27].

The development of drug resistance is a significant challenge in AML treatment with chemotherapy. Mutations in genes like FLT3, TP53, and DNMT3A can drive resistance to standard therapies, leading to relapse. Additionally, the presence of MRD, detected by molecular techniques, is a strong predictor of relapse. Approximately 50-60% of AML patients relapse after achieving initial remission, with poor prognosis upon relapse due to the clonal evolution of leukemic cells and the emergence of therapy-resistant subclones [28].

2.7 The Role of 5-Azacytidine in AML Therapy

5-Azacytidine, a DNA hypomethylating agent, has emerged as a vital therapeutic option for AML, particularly in older patients or those unfit for intensive chemotherapy. As an epigenetic modulator, 5-Azacytidine works by inhibiting DNA methyltransferases, thereby reducing the hypermethylation of tumor suppressor genes, leading to their reactivation and subsequent inhibition of leukemic cell proliferation. It also affects RNA metabolism and protein synthesis by incorporating it into RNA, further contributing to its anti-leukemic effects [29].

5-Azacytidine has shown meaningful clinical responses in AML, particularly in patients with high-risk cytogenetics, secondary AML, and TP53 mutations, where conventional chemotherapy has limited efficacy [30]. Clinical trials exploring the combination of 5-Azacytidine with novel agents like venetoclax (a BCL-2 inhibitor) have demonstrated enhanced therapeutic responses, leading to longer remissions in a subset of patients. This combination has been particularly beneficial in elderly patients and those with treatment-resistant disease [15].

Despite these advances, the development of resistance to 5-Azacytidine remains a major challenge, with mechanisms involving altered expression of DNA methyltransferase enzymes, changes in cellular uptake, and epigenetic reprogramming. Ongoing research focuses on understanding these resistance pathways and identifying biomarkers to predict response to 5-Azacytidine therapy [31,32].

3. Integrative Genomic and Transcriptomic Analysis

3.1 The Need for Multi-Omics Approaches in AML

AML's complex and variable nature necessitates multi-omics approaches that encompass Whole Exome Sequencing (WES) and RNA Sequencing (RNA-seq). These methodologies provide a holistic view of the tumor’s genetic landscape and the associated gene expression changes. Genomic studies have revealed that mutations are present in over 95% of AML cases, with key mutations, such as those in FLT3 and NPM1, playing critical roles in disease progression and therapeutic outcomes [33]. The integration of WES allows for the identification of somatic mutations that are essential for understanding the molecular

pathogenesis of AML [13]. RNA-seq, on the other hand, elucidates the expression profiles linked to different mutational backgrounds and therapy responses, revealing how specific genetic alterations drive leukemogenesis [7,34].

3.2 Whole Exome Sequencing in AML

Whole Exome Sequencing (WES) is a widely utilized technique for identifying coding mutations in the genome, particularly single nucleotide variants (SNVs) and small insertions or deletions (INDELs) that are frequently observed in AML. Key mutations in genes such as FLT3, NPM1, IDH1/IDH2, DNMT3A, and TP53 have been implicated in the development and progression of AML [34,35]. WES enables the identification of these driver mutations and the discovery of novel variants that may influence disease progression or therapeutic resistance. Studies have shown that mutations in FLT3 are particularly common and are associated with poor prognosis, emphasizing the critical role of WES in understanding the mutational landscape of AML [36,37].

3.3 RNA Sequencing and Its Role in AML

RNA Sequencing (RNA-seq) complements WES by providing both quantitative and qualitative data on gene expression levels, alternative splicing events, and fusion transcripts. RNA-seq is especially valuable in identifying aberrant gene expression patterns and transcriptional changes resulting from underlying genetic mutations or epigenetic dysregulation [38,39]. For instance, mutations in NPM1, although detectable at the DNA level, can lead to significant transcriptomic alterations that promote leukemogenesis [40]. Furthermore, RNA-seq can identify gene fusions, such as RUNX1-RUNX1T1 and CBFB-MYH11, which serve as important diagnostic and prognostic markers in AML [41].

One of the key advantages of RNA-seq is its ability to capture the dynamic nature of the transcriptome, revealing gene expression changes in response to treatment or disease progression [42]. By integrating RNA-seq data with WES findings, researchers can correlate genetic mutations with downstream effects on gene expression, offering deeper insights into the functional consequences of somatic alterations. This integration is essential for understanding the complex biology of AML and developing targeted therapeutic strategies [43].

3.4 Tumor Microenvironment and Neoantigens

The interplay between leukemic cells, the tumor microenvironment, and the immune system is critical in the progression of AML and its response to treatment. Integrative genomic and transcriptomic analysis is vital for elucidating this complex interaction, particularly through the identification of immune cell infiltration patterns and immune checkpoint expression [44]. A significant focus in recent AML research has been the detection of neoantigens—novel peptides generated by tumor-specific mutations. These neoantigens can be recognized by the immune system and serve as potential targets for immunotherapy [45,46].

Neoantigens arise from somatic mutations, such as SNVs and small indels, that alter the protein-coding sequence, resulting in the presentation of non-self peptides on the surface of leukemic cells. By combining WES to identify tumor-specific mutations with RNA-seq to determine which mutated genes are actively transcribed, researchers can predict the presence of neoantigens that may be recognized by T cells [47,48].

Neoantigen detection is crucial for the development of personalized immunotherapies, including T-cell receptor (TCR)

therapies and cancer vaccines. Recent studies have demonstrated that AML patients with higher neoantigen loads may exhibit improved responses to immunotherapeutic interventions. Moreover, the identification of neoantigens has the potential to enhance immune surveillance by improving the efficacy of immune checkpoint inhibitors, which can reinvigorate T cells targeting these neoantigens [49,50].

3.5 Advantages of Multi-Omics Integration in AML Research

Integrating WES and RNA-seq data provides a deeper understanding of the molecular and transcriptional landscape of AML and offers significant advantages for neoantigen discovery and immunotherapy development.

By combining WES and RNA-seq, researchers can predict which somatic mutations lead to the generation of neoantigens. Mutations identified through WES, such as those in NPM1, FLT3, and IDH1/2, can be evaluated through RNA-seq to assess their transcriptional activity and potential to produce neoantigenic peptides [39]. Neoantigen load is also correlated with better immune recognition and may help stratify patients for immunotherapy [51].

Neoantigen analysis has revealed mechanisms of immune evasion in AML, such as loss of neoantigen expression or mutations that impair antigen presentation. RNA-seq data can be used to assess the expression of genes involved in antigen processing and presentation, such as HLA molecules, helping to identify cases where leukemic cells downregulate neoantigen presentation to escape immune surveillance [52].

Multi-omics data, including neoantigen detection, enables the design of personalized immunotherapies tailored to the specific mutation profile of each patient. Neoantigens

identified through WES-RNA-seq integration provide targets for TCR-based therapies and cancer vaccines [53]. These therapies can potentially improve patient outcomes by specifically targeting leukemic cells that present these tumor-specific antigens.

Neoantigen load, as detected through integrative WES and RNA-seq analysis, has been proposed as a biomarker for predicting response to immunotherapies such as immune checkpoint inhibitors (e.g., PD-1/PD-L1 inhibitors) [54,55]. Patients with a higher neoantigen burden may be more likely to respond favorably to such therapies, as their immune systems are more likely to recognize and attack tumor-specific antigens.

Incorporating neoantigen detection into the integrative analysis of WES and RNA-seq data enhances our ability to understand the immunogenicity of AML and provides a pathway for the development of novel immunotherapies. This strategy is essential for translating genetic and transcriptomic data into clinical applications that improve patient outcomes [56].

4. Material and Methods

4.1 Patient Recruitment

Patient recruitment was conducted as a prospective observational study involving adult patients diagnosed with AML. Recruitment occurred at multiple institutions, including Ospedale Santa Maria delle Croci - Ravenna, Ospedale degli Infermi - Rimini, ASST Spedali Civili - Brescia, Ospedale Policlinico San Martino IRCCS - Genova, Azienda Sanitaria Universitaria Friuli Centrale (ASUFC) - SOC Clinica Ematologica - Udine, Istituto Scientifico Romagnolo per lo Studio dei Tumori (IRST) "Dino Amadori" IRCCS - Meldola (FC),

Azienda ULSS 2 Marca Trevigiana - Distretto di Treviso - Treviso, IRCCS IOV - Castelfranco Veneto (TV), and AUO Maggiore della Carità - Novara.

Eligible participants included individuals aged 18 years or older with a confirmed diagnosis of AML according to WHO criteria. Patients scheduled to initiate treatment with 5-Azacytidine were specifically targeted to ensure a focus on treatment-naive individuals. Patients who had previously received any form of treatment for AML or had a history of other malignancies were excluded to minimize confounding factors that could impact the study's outcomes and interpretations. This recruitment protocol aligns with recent studies that underscore the need for standardized recruitment strategies in multi-site studies to enhance comparability and data integrity [57].

Clinical data, including patient demographics, cytogenetic abnormalities, and treatment responses, were collected from medical records. The study protocol was approved by the Institutional Ethics Committee (Comitato Etico della Romagna, protocol 983/2018). Informed consent was obtained from all participants prior to sample collection, ensuring they understood the study's purpose, procedures, potential risks, and benefits in accordance with the Declaration of Helsinki. Research has shown that clear communication and streamlined protocols can significantly enhance patient recruitment success by minimizing participant burden [58,59].

4.2 Sample Collection and Preparation

Bone marrow aspirates were collected from patients at both diagnosis and after treatment to assess changes in the bone marrow microenvironment and genetic profile over time. Approximately 10 mL of bone marrow aspirate was obtained using

standard sterile techniques to ensure patient safety and sample integrity. The procedure involved trained medical professionals using a sterile needle and syringe in a controlled clinical setting, with local anesthesia administered as appropriate to minimize discomfort. The collected samples were immediately placed in sterile tubes for subsequent analysis.

Blood samples were also collected from patients in EDTA-coated tubes, each consisting of approximately 10 mL of peripheral blood. This collection aimed to isolate peripheral blood mononuclear cells (PBMCs) for comprehensive immune profiling and gene expression analysis. Following collection, samples were processed within 2 hours to maintain cell viability, with PBMC isolation achieved through density gradient centrifugation using Ficoll-Paque solution.

For buccal swab collection, two sterile swabs were used to obtain DNA samples. Prior to the procedure, subjects were instructed to avoid eating, drinking, or using tobacco products for at least 30 minutes to ensure optimal sample quality. After collection, the swab was placed back into its original tube and labeled with patient identifiers, including full name and date of birth. This process was repeated using a second swab on the opposite side of the mouth to ensure adequate sample yield. Both labeled swabs were then transported at room temperature to the laboratory for processing on the same day as collection. A flow diagram illustrating the sample processing steps is depicted in Figure 1.

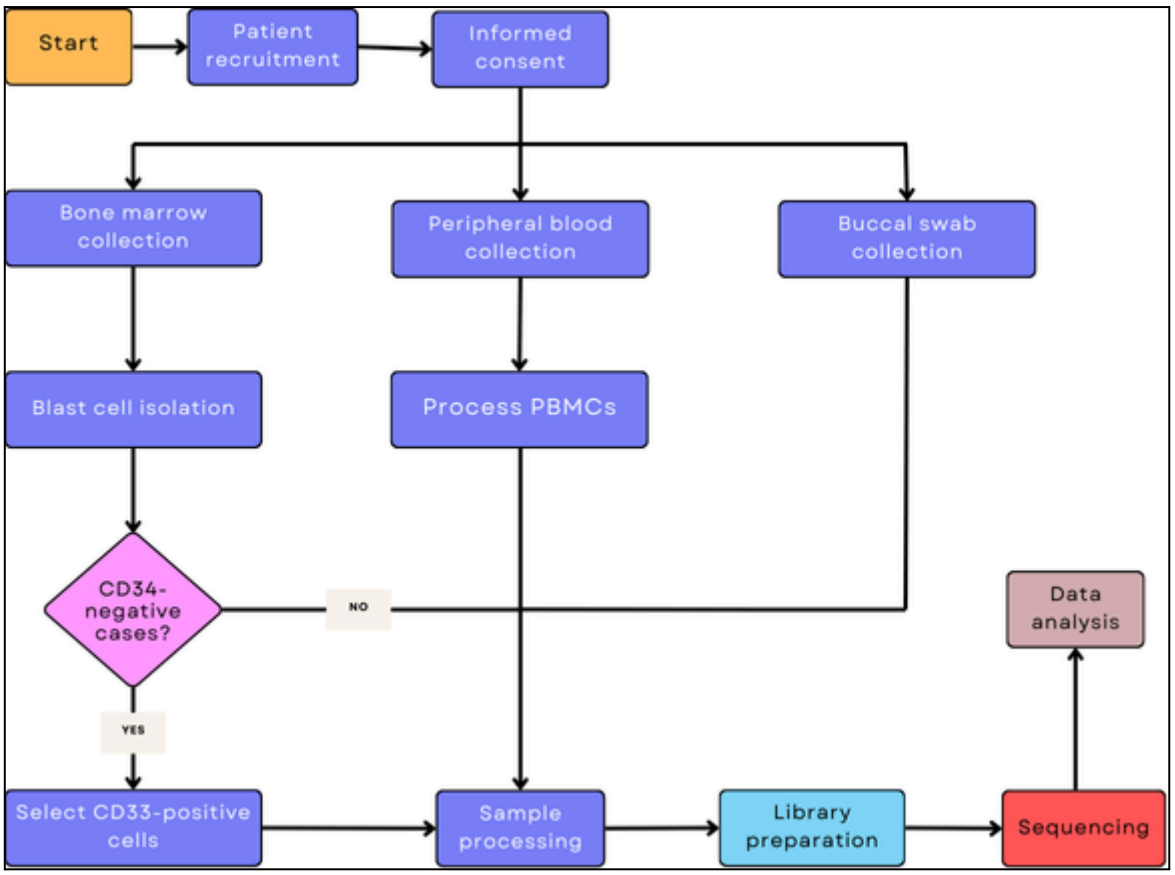

Figure 1: Sample Processing Steps. This flow diagram illustrates the methodology of the study, detailing the steps from patient recruitment to data analysis. The process begins with patient recruitment and proceeds to informed consent. Following consent, sample collection occurs, which includes bone marrow, peripheral blood, and buccal swab. Blast cells are isolated from bone marrow and cells are isolated by immunomagnetic separation in aun automated system (Automacs Pro, Miltenyi), base on expression of CD34 or CD44 with depletion of CD14- positive cells for CD34-negative cases. Then, sample are processed, undergo library preparation and are sequenced using either Whole Exome Sequencing or RNA Sequencing. The final step is data analysis, marking the completion of the study's methodology.

Total RNA was extracted using the Direct-Zol™ Mini/Micro DNA/RNA Kit, which allows for efficient purification of high-quality RNA directly from various sample types without phase separation or precipitation steps [60]. This method effectively isolates total RNA, including small RNAs, ensuring

unbiased recovery of all RNA species. Following extraction, single-stranded DNA (ssDNA) was obtained through reverse transcription. The ssDNA was then converted into double-stranded DNA (dsDNA) according to the protocol outlined in the Twist Total Nucleic Acids Library Preparation Enzymatic Fragmentation Kit 2.0, designed for high-throughput sequencing applications [61].

Tumor DNA was extracted using the Direct-Zol™ Mini/Micro DNA/RNA Kit, which allows for efficient purification of genomic DNA [59]. This method produced single-stranded DNA (ssDNA), which was subsequently converted into double-stranded DNA (dsDNA) using the Twist Total Nucleic Acids Library Preparation Enzimatic Fragmentation Kit 2.0 [62].

For normal samples, saliva was extracted using the Maxwell RSC Blood DNA Kit [61], while CD3-sorted cells from bone marrow were processed with the Qiagen QIAamp Micro DNA Kit, which utilizes column-based extraction techniques [63]. After extraction, all normal DNA samples were also quantified with Qubit to confirm their concentrations.

4.3 Library Preparation

The total input of 500 ng of RNA in 11 microliters was used during library preparation. RNA-seq was performed on matched samples from AML patients, specifically on baseline (t0) bone marrow mononuclear cells (BMMC) and PBMCs, and only PBMC samples collected at 48 hours post-treatment (t48) with 5-Azacytidine. The rationale for collecting samples 48 hours after treatment is based on feasibility, as it eliminates the need for patients to return for additional sample collection. Moreover, measurable changes in gene expression resulting from DNA demethylation induced by 5-Azacytidine are typically observed within 24 to 48 hours after treatment [29].

Using the Illumina Stranded Total RNA Prep with Ribo-Zero Plus Kit, rRNA was depleted from total RNA samples before reverse transcription into complementary DNA (cDNA). This method enriches for mRNA by removing ribosomal RNA, which is critical for improving the detection of gene expression levels [64].

Following RNA extraction, library preparation was carried out using a Hamilton Microlab STARlet-ivd automated liquid handling system which streamlined the workflow and minimized manual handling [65]. This automation reduced the risk of pipetting errors and improved reproducibility across samples. Key steps in the library preparation included:

- End Repair: Fragmented RNA underwent an end-repair process to create blunt ends.
- A-tailing: Adenine residues were added to the 3' ends to facilitate adapter ligation.
- Adapter Ligation: Indexed adapters were ligated to the fragments to enable multiplexing and sequencing compatibility.
- PCR Amplification: Adapter-ligated fragments were amplified to yield sufficient material for sequencing.

The prepared libraries were assessed using an Agilent Bioanalyzer to ensure appropriate size distribution and concentration.

For DNA library preparation, an input of 200 ng of DNA in 30 microliters was utilized. Genomic DNA was enzymatically fragmented into approximately 200-300 base pairs. Enzymatic fragmentation is a DNA shearing method used in sequencing and library preparation. It employs specific enzymes, endonucleases, to cleave DNA at controlled points, producing fragments of desired lengths. This method is gentler than physical shearing, reducing the risk of DNA damage, making it

particularly useful for sensitive or degraded samples. Indexed adapters were ligated to the fragments using the Illumina DNA Prep with Exome 2.0 Plus Enrichment Kit, specifically designed for high-throughput sequencing applications [66]. Key steps in the library preparation included:

- End Repair: Fragmented DNA underwent an end-repair process to create blunt ends.
- A-tailing: Adenine residues were added to the 3' ends to facilitate adapter ligation.
- Adapter Ligation: Indexed adapters were ligated to the fragments for multiplexing and sequencing compatibility.
- PCR Amplification: Adapter-ligated fragments were amplified to yield sufficient material for sequencing.
- Purification: Libraries were purified using magnetic bead-based cleanup to remove unligated adapters, enzymes, and other contaminants.
- Quality Control: The quality and size distribution of libraries were assessed using an Agilent Bioanalyzer.

The whole library preparation process was automated using Hamilton Microlab Starlet-IVD to ensure precision and reproducibility [65].

After purification, libraries were quantified using the Qubit dsDNA HS Assay to confirm that they met the required concentration for sequencing. Libraries were pooled to achieve a final concentration of 1.4 nM for loading onto the sequencer.

4.4 Sequencing

For RNA samples, the sequencing was conducted on the NovaSeq 6000 platform with 100 x 2 paired-end reads (Illumina, 2020), using a NovaSeq 6000 S2 Reagent Kit v1.5. This sequencing

approach ensures high-depth coverage, which is crucial for accurately capturing transcript levels and identifying lowly expressed genes. For WES, the tumor samples were processed on a NovaSeq 6000 S2 Reagent Kit v1.5 flow cell, while the normal samples were sequenced on a NovaSeq 6000 S1 Reagent Kit v1.5 flow cell. The choice of flow cells was based on prior sequencing experience to achieve a targeted average coverage of approximately 150X for tumor samples and 100X for normal samples, and more than 20 million reads for RNA samples.

We conducted three sequencing runs on the Illumina NovaSeq 6000, utilizing v1.5 flow cells: two for WES and one for RNA-seq. The first WES run focused on normal samples, comprising 27 paired-end samples and utilizing an S1 flow cell. The second WES run targeted tumor samples, also consisting of 27 paired-end samples, and was performed using an S2 flow cell. The RNA-seq run included a total of 66 paired-end samples and was executed on an S2 flow cell as well. Table 1 shows all the sequenced samples and their relations.

Table 1. List of Sequenced Samples. The following table summarizes the presence of various sample types collected from patients involved in this study. Each row represents a patient, while each column corresponds to a specific sample type. An "X" in a cell indicates that the sample is present for that particular patient.

patient	Tumor WES	Normal WES	BMMC RNA t0	PBMC RNA t0	PBMC RNA t48
01	X	X	X		X
02	X	X	X	X	X
03	X	X	X		

04	X	X	X	X	X
05	X	X	X		X
06	X	X	X	X	X
07	X	X	X	X	X
08	X	X	X		
09	X	X	X		X
10	X	X	X	X	X
11	X	X	X	X	X
12	X	X	X	X	X
13	X	X	X	X	
14			X		X
15			X		
16	X	X	X	X	X
17			X		
18	X	X	X		
19			X		
20	X	X		X	X
21	X	X	X	X	X
22			X	X	X
23			X	X	X
24			X	X	

25			X	X	X
26	X	X	X		
27	X	X	X		X
28	X	X		X	X
29			X	X	
30	X	X	X		
31	X	X	X		X
32	X	X	X		X
33	X	X		X	X
34	X	X	X		X
35	X	X	X	X	X
36	X	X	X		
TOT	**27**	**27**	**33**	**19**	**24**

All the runs exhibited high quality, as confirmed by analysis using the Sequence Analysis Viewer software from Illumina. The WES normal run achieved a Q30 score of 88% with a cluster passing filter rate of 83%. In the WES tumor run, we observed a Q30 score of 94% and a cluster passing filter rate of 77%. For the RNA-seq run, both the Q30 score and cluster passing filter rate were commendably high at 94% and 81%, respectively. These quality metrics indicate that the sequencing data is robust and suitable for subsequent analyses.

A total of 33 BMMC t0 samples, 19 PBMC t0 samples, and 24 PBMC t48 samples were subjected to RNA sequencing to investigate gene expression changes in response to 5-Azacytidine treatment. After quality checks, using a threshold on sequenced reads of 20 million, one of the PBMC t0 samples was discarded due to a low number of reads.

WES was conducted on 27 matched tumor-normal samples from patients with AML. After quality assessment, two germline samples were excluded due to insufficient quality, leaving 25 samples for further analysis. The quality filtering criteria were established based on coverage thresholds for both tumor and normal samples. Specifically, germline samples required a minimum coverage of 50X in at least 70% of the target regions, while tumor samples needed a minimum coverage of 100X in at least 70% of their target regions (Figure 2).

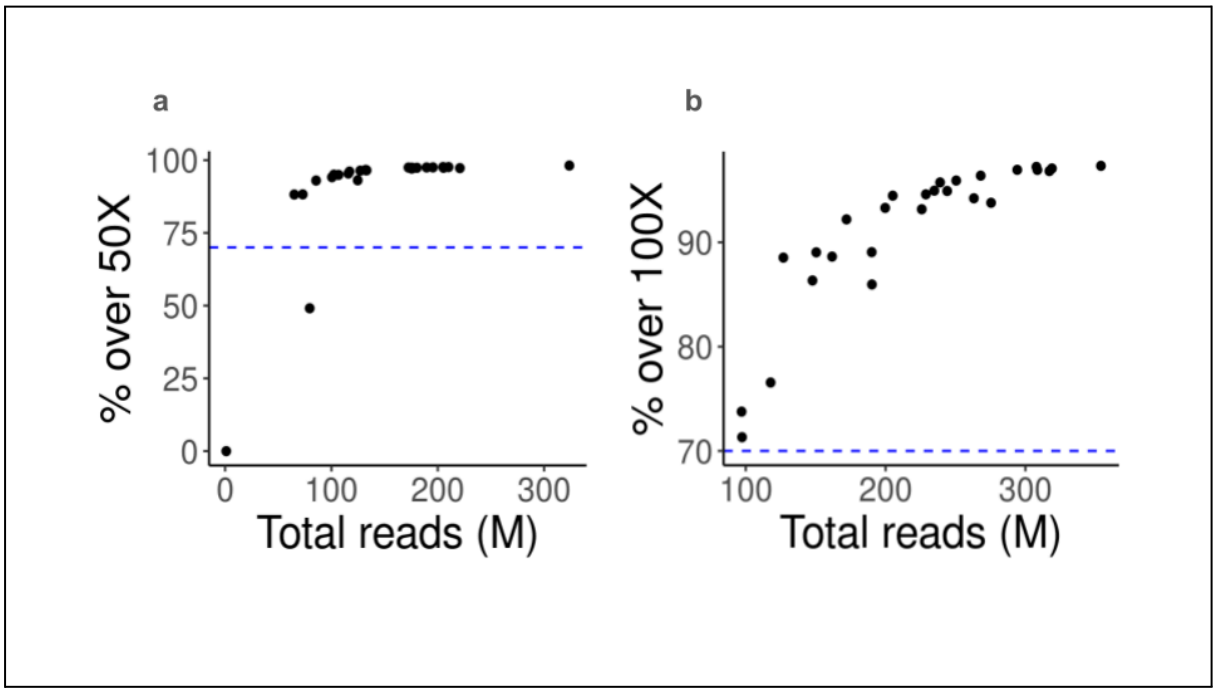

Figure 2. This figure illustrates the coverage metrics for each sample: the y-axis represents the percentage of bases with coverage exceeding 50X for normal samples and 100X for tumor samples, while the x-axis shows the total reads in millions. Both plots include a threshold line in blue at 70%, indicating the minimum acceptable coverage percentage necessary for reliable variant analysis. Panel a) displays the coverage metrics of normal samples, and panel b) presents the coverage metrics of tumor samples.

4.5 RNA-seq Analysis

The analysis workflow for RNA-seq data begins with the demultiplexing process using bcl2fastq Conversion software v2.20, a tool developed by Illumina [67]. This process converts binary base call files (BCL files) into FASTQ files while separating multiplexed samples based on their unique index sequences, as specified in a sample sheet.

Following demultiplexing, the raw sequencing reads are processed using Kallisto v0.46.2, a fast and efficient tool for quantifying gene expression levels [65]. Kallisto employs a pseudo-alignment method to map RNA-seq reads to a transcriptome, providing accurate abundance estimates for each transcript. The analysis workflow includes the following key steps:

1. Read Pseudomapping: In this step, raw sequencing reads are pseudomapped to the indexed transcriptome using Kallisto. Instead of performing traditional alignments that determine the exact positions of reads within transcripts, Kallisto identifies which transcripts are compatible with each read based on shared k-mers (short sequences of nucleotides).

- K-mer Decomposition: Kallisto breaks down each transcript into smaller sequences called k-mers and constructs a colored de Bruijn graph (T-DBG) where each node corresponds to a k-mer and retains information about the transcripts that contain it.
- Pseudomapping Process: During analysis, Kallisto decomposes each read into its constituent k-mers and uses these to navigate the T-DBG. It identifies compatible transcripts by finding intersections among the k-compatibility classes of the read's k-mers, allowing it to infer which transcripts could have generated the read without needing direct alignment.
- Efficiency and Speed: This method enables Kallisto to process large datasets rapidly. By bypassing the computationally intensive alignment step, Kallisto provides fast and reliable quantification of transcript abundance.
- Equivalence Classes: Kallisto groups reads into equivalence classes based on their compatibility with sets of transcripts, facilitating efficient estimation of transcript abundance.

2. Quantification: Kallisto quantifies expression at the transcript level, yielding estimates of transcripts per million (TPM) for each transcript across samples. This metric normalizes raw counts to account for differences in sequencing depth and gene length.
3. Data Summarization with tximport: After quantification, raw counts at the transcript level are collapsed to the gene level using tximport v1.12.1 [69]. This step aggregates the transcript-level data into gene-level counts,

facilitating downstream analyses such as differential expression testing.

4. Data Normalization: The normalization process in DESeq2 is crucial for accurate comparisons of gene expression between samples. DESeq2 uses a method known as the median of ratios to normalize raw count data for sequencing depth and RNA composition differences [70].
    - Size Factors Calculation: DESeq2 calculates size factors for each sample by taking the median ratio of observed counts to geometric means across all samples. This approach helps account for differences in library sizes, ensuring that samples with higher sequencing depths do not skew results.
    - Normalization Formula: The normalized counts are obtained by dividing the raw counts by their respective size factors. This adjustment allows for fair comparisons between samples by standardizing counts relative to their sequencing depth.
    - Importance of Normalization: Normalization is essential not only for differential expression analysis but also for exploratory data analysis and visualization. It ensures that observed differences in expression levels are biologically meaningful rather than artifacts of varying library sizes or sequencing depths.
5. Hierarchical Clustering: After these steps, hierarchical clustering is performed to identify groups based on changes in gene expression between matched PBMC samples at t48 and t0:
    - Z-Score Calculation: Z-scores provide standardized measures of expression level changes, allowing for comparisons across genes.

- Distance Metric: The maximum distance metric measures similarity between samples based on their gene expression profiles.
- Clustering Algorithm: The Ward D2 algorithm minimizes variance within clusters while maximizing variance between them [71]. This approach results in a dendrogram visualization that illustrates relationships among samples based on their expression profile changes.

6. Gene Set Enrichment Analysis (GSEA): GSEA is performed using the Python package GSEApy v1.1.3 [72] on BMMC t0 samples matched to PBMC t0 samples, grouped according to clusters identified in the PBMC t48/t0 analysis. GSEA works through several steps:
   - Ranking Genes: A ranked list of genes is generated based on a metric reflecting differential expression, the fold change, calculated using DESeq2, to compare the two sample groups.
   - Predefined Gene Sets: GSEA utilizes predefined gene sets from databases that group genes involved in specific biological pathways or functions. In our case, we used Reactome v.2022, as our source for gene sets. Reactome is a highly curated and comprehensive pathway database that covers many biological processes, including signaling pathways, metabolism, immune responses, and cell cycle regulation, making it particularly suited for studying cancer-related pathways relevant to AML. Its manually curated, regularly updated pathways ensure high accuracy and relevance, which is crucial for studying complex diseases like AML. Reactome is designed to provide a clear overview of biological

processes and interactions, helping to identify high-level pathway changes rather than isolated gene-level alterations. Compared to broader or more generalized databases, Reactome often contains more cancer-specific pathways directly linked to AML mechanisms and emphasizes pathway interconnections, offering insights into how biological processes interact and influence AML progression or response to treatments like 5-Azacytidine. Finally, focusing on a single database like Reactome can simplify result interpretation, reducing the redundancy and overlap that can arise from using multiple databases [73].

- Enrichment Score Calculation: For each gene set, GSEA calculates an enrichment score by performing a running sum statistic as it traverses the ranked list. If a gene belongs to the gene set, it adds to a running total; if not, it subtracts. The maximum value of this running sum indicates how enriched that gene set is within the ranked list.
- Statistical Significance: GSEA employs permutation testing to assess whether the observed enrichment score is significant compared to a null distribution generated by randomly permuting class labels.
- Results Interpretation: The output includes normalized enrichment scores and p-values, allowing researchers to identify pathways significantly associated with differentially expressed genes [74].

4.6 WES Analysis

Post-sequencing, raw data from both tumor and normal samples were processed using the Illumina DRAGEN DNA Pipeline (DRAGEN

Host Software Version 4.3.6), a highly optimized computational tool designed for NGS analysis. This pipeline was specifically configured to take matched tumor-normal samples as input, enabling the removal of germline variants and ensuring the identification of somatic mutations exclusive to tumor cells (Figure 3).

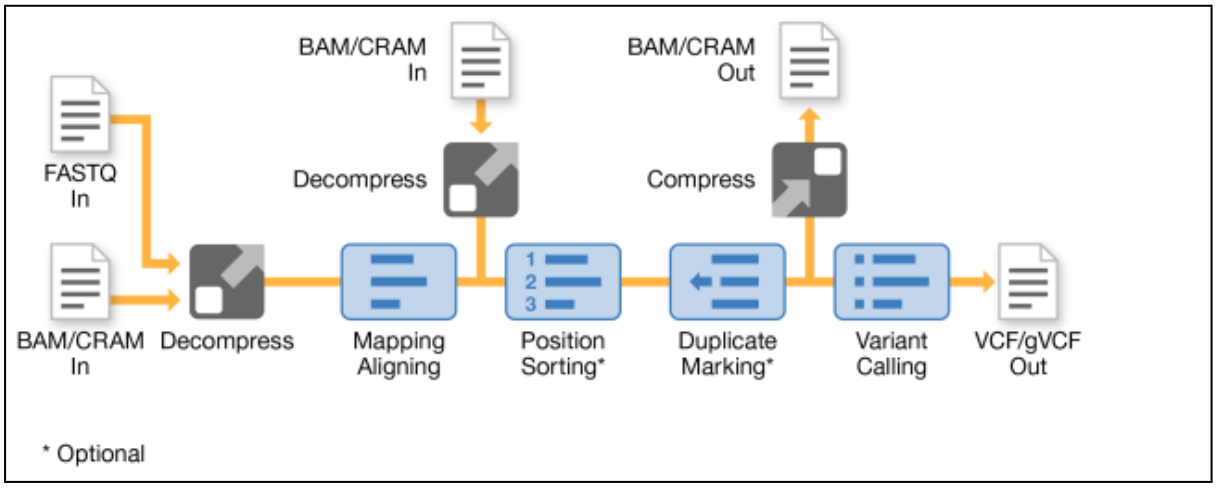

Figure 3. Illumina DRAGEN DNA Pipeline. This overview depicts key steps in the bioinformatic workflow from raw read processing to variant calling [75].

The key steps in the analysis are as follows:

- Demultiplexing: Raw sequencing data were demultiplexed using Illumina's bcl2fastq tool, assigning pooled reads to individual samples based on unique barcodes.
- DNA Mapping: Demultiplexed reads were aligned to the human reference genome (GRCh38) using the DRAGEN aligner, with its high-speed and precise read mapping.
- Read Trimming: Low-quality bases and adapter sequences were trimmed to improve overall data quality and eliminate potential biases.
- Quality Control: DRAGEN FastQC was employed to assess read quality, evaluating factors such as sequencing depth, GC content, and quality score distribution.

- Sorting and Duplicate Marking: Aligned reads were sorted by genomic coordinates, and PCR duplicates were marked to avoid inflating variant counts.
- Small Variant Calling: Variants, including SNVs and INDELs, were called across exonic regions. The analysis used matched normal samples to filter out germline variants, leaving only somatic mutations for further investigation.
- Copy Number Variant Calling: The pipeline also performed copy number variant (CNV) analysis, identifying large-scale genomic alterations such as gene amplifications or deletions in the tumor samples.
- HLA Typing: Accurate HLA typing is essential for predicting which peptides, including neoantigens, can be presented by HLA molecules on the surface of cells for immune recognition. The DRAGEN DNA Pipeline provides high-resolution genotyping of Class I HLA alleles (HLA-A, HLA-B, and HLA-C) from WES data. This pipeline is optimized for analyzing complex genomic regions like the highly polymorphic HLA region. The WES data from tumor and normal samples were first aligned to the reference genome, and the DRAGEN pipeline's HLA typing module was used to compare aligned reads against known HLA reference sequences. The pipeline accurately predicted specific alleles at a 4-digit resolution (e.g., HLA-A*02:01), allowing precise identification of allele subtypes within each HLA gene group. This level of precision enhances neoantigen prediction by allowing more accurate peptide-binding predictions.

Starting from the variants derived by DRAGEN DNA pipeline output, we performed two steps of annotation. At first, we used ANNOVAR (version 2019Oct24) for a Gene-based annotation

[76]. It utilizes reference gene databases, particularly Ensembl, to classify variants based on their positions relative to genes:

- Gene Names and HUGO Gene Symbols: ANNOVAR provides the names of the genes associated with each variant, including the corresponding HUGO Gene Symbols.
- Exonic Variants: Variants located within coding sequences are identified and categorized as synonymous, nonsynonymous (missense and nonsense), in-frame, or frameshift alterations.
- Intronic Variants: Variants located within introns are annotated with special attention given to those that may affect splicing or gene regulation.
- UTR Variants: Variants in the 5' and 3' untranslated regions (UTRs) are also classified.

In addition to ANNOVAR, VarSome (version 11.18.0) was employed to enhance the clinical interpretation of somatic mutations identified in the analysis. VarSome is a comprehensive genomic variant database that integrates multiple sources of information to provide insights into the potential pathogenicity and clinical relevance of genetic variants [77]. VarSome aggregates data from various databases, including gnomAD, ClinVar, and COSMIC, facilitating a thorough evaluation of each variant's significance.

Variants are classified according to the American College of Medical Genetics and Genomics (ACMG) guidelines, which recommend a structured five-tier classification system:

a. Pathogenic: Variants that are known to cause disease.
b. Likely Pathogenic: Variants with substantial evidence suggesting a disease-causing effect but not yet fully confirmed.

c. Variant of Uncertain Significance (VUS): Variants for which there is insufficient evidence to determine their impact on disease.
d. Likely Benign: Variants that are unlikely to be associated with disease based on available evidence.
e. Benign: Variants that are well-established as not causing disease.

The ACMG guidelines include 28 specific criteria that help classify variants based on various types of evidence, such as population frequency data, computational predictions, functional studies, and segregation analysis [78]. These criteria are categorized into two main groups:

- Pathogenic Criteria: Sixteen criteria evaluate evidence toward a pathogenic interpretation, including:
    a. Very Strong (PVS1): Evidence indicating that the variant is a loss-of-function mutation in a gene where loss-of-function is known to cause disease.
    b. Strong (PS1–4): Multiple types of strong evidence supporting pathogenicity.
    c. Moderate (PM1–6): Evidence that supports pathogenicity but is not as robust as strong criteria
    d. Supporting (PP1–5): Additional evidence that supports the classification but is less definitive.
- Benign Criteria: Twelve criteria assess evidence toward benign classification, including:
    a. Stand-alone (BA1): Evidence indicating that the variant is common in the general population.
    b. Strong (BS1–4): Robust evidence suggesting that the variant does not affect protein function or is not associated with disease.

c. Supporting (BP1–7): Additional evidence suggesting that the variant is benign.

This structured approach allows for consistent interpretation across studies and supports clinical decision-making by providing clear classifications based on established criteria.

VarSome also incorporates the Association for Molecular Pathology (AMP) Tier classification system to evaluate the clinical relevance of variants based on the strength of evidence supporting their association with disease [79]. This system provides a structured framework that categorizes variants into four distinct tiers, facilitating consistent interpretation and aiding clinical decision-making:

a. Tier 1 (Strong Clinical Significance): Variants have robust evidence supporting their pathogenicity. These variants are often linked to established cancer genes and have clear implications for diagnosis, prognosis, or treatment decisions.
b. Tier 2 (Potential Clinical Significance): Variants possess moderate evidence suggesting they may be clinically relevant. While they may not have definitive associations with disease, there is enough data to warrant further investigation or consideration in clinical contexts.
c. Tier 3 (Unknown Clinical Significance): Variants in this category lack sufficient evidence to classify them definitively as either pathogenic or benign. They may represent true unknowns or variants with conflicting evidence, making them challenging to interpret in a clinical setting.
d. Tier 4 (Benign or Likely Benign): Variants that are well-established as not causing disease. These variants typically have high frequency in the general population

and are not associated with any known clinical implications.

The AMP classification system is based on a combination of molecular, functional, statistical, and clinical data. Each variant is evaluated according to specific criteria that consider various types of evidence, such as: population allele frequencies, functional studies, clinical correlations, and family history data. This multi-faceted approach allows clinicians and researchers to assess the potential impact of variants on patient care effectively. The AMP Tier classification system enhances the interpretation of genetic test results by providing clear guidelines that support clinical decision-making. By categorizing variants based on their clinical significance, it aids in identifying actionable mutations that may inform treatment strategies and improve patient outcomes.

Through the integration of ANNOVAR and VarSome, along with adherence to ACMG and AMP guidelines, a comprehensive understanding of the functional and clinical significance of somatic mutations was achieved, allowing for a deeper interpretation of the genetic landscape in patients.

After these annotations, to ensure high-confidence calls and relevance of identified somatic mutations, rigorous filtering steps were applied:

- Quality Filtering: A minimum sequencing depth of 100X across targeted exonic regions was established for reliable variant calls. Variants with a variant allele frequency (VAF) of ≥5% were retained.
- Germline Variant Removal: Matched tumor-normal pairs were utilized to identify and remove germline variants shared between normal and tumor samples.

- Functional Impact Filtering: Synonymous or intronic variants were filtered out unless there was strong evidence suggesting an impact on splicing or gene regulation.
- ACMG Verdict Filtering: Variants classified as benign or likely benign according to ACMG guidelines were removed from further analysis.

To effectively communicate the results of the variant analysis, an oncoprint was generated using the ComplexHeatmap R package [80]. This visualization focuses on AML-related genes, providing a comprehensive overview of mutation profiles across AML patients.

4.7 Multi-Omics Integration

Multi-omics integration was performed to comprehensively evaluate the immune landscape and therapeutic potential in AML patients. This approach focused on the combined analysis of WES and RNA-seq data, and immuno-profiling to identify neoantigens and assess the immune response to treatment.

The identification of neoantigens—novel peptides presented by HLA molecules due to tumor-specific mutations—is crucial for developing effective cancer immunotherapies. Neoantigens arise from non-synonymous mutations in tumor cells, leading to the production of unique peptides that can be recognized by the immune system [82,83]. Their discovery is essential for designing personalized vaccines and immunotherapies that target these specific tumor-associated antigens [84,85].

To predict the binding affinities of the identified neoantigens to patient-specific HLA molecules, we employed NetMHCpan, a robust tool for peptide-MHC binding prediction [85,86]. NetMHCpan is an extension of the NetMHC algorithm,

specifically designed to predict binding affinities for a wide variety of MHC class I alleles, including both common and rare variants. The tool is built upon a sophisticated artificial neural network (ANN) that integrates a large dataset of experimentally determined peptide-MHC interactions, allowing for high accuracy in predicting binding affinities, even for peptides not included in the training data.

NetMHCpan is trained on over 850,000 quantitative binding affinity (BA) data points and mass-spectrometry eluted ligands (EL) from more than 170 MHC molecules. This extensive training enables the model to generalize well across different alleles and peptide sequences. The tool employs a unique feature extraction method that captures important biochemical properties of both the peptide and MHC class I alleles. This leads to a refined understanding of peptide-MHC interactions, which significantly enhances the prediction accuracy for neoantigen identification. By accurately identifying which tumor-derived peptides are likely to bind MHC molecules and be presented on the cell surface, researchers can better assess the potential immunogenicity of these peptides. NetMHCpan can accommodate novel or uncharacterized alleles, making it a versatile choice for diverse populations. This is particularly important in cancer immunotherapy, where individual patient HLA types can vary significantly [87].

Our neoantigen analysis workflow integrates various data sources to comprehensively predict potential neoantigen candidates. Here's an overview of the approach:

- Input Data Collection: We used non-synonymous variants identified and annotated from WES data as the starting point, selecting these mutations due to their potential to alter amino acid sequences in a way that could produce immunogenic peptides. Simultaneously, gene expression

levels were quantified from RNA-seq data using TPM to prioritize mutations in actively expressed genes, as these are more likely to generate immune-recognized neoantigens. HLA Class I typing was performed on each sample using the DRAGEN pipeline, allowing for the identification of patient-specific HLA alleles that play a critical role in determining which peptides could potentially bind to HLA molecules.

- Peptide Generation and Prediction: In this step, all possible 8- to 11-mer peptides containing each variant were generated from the WES data, creating a comprehensive set of candidate peptides.
- Binding Affinity Prediction and Filtering: The ANN model, refined through extensive datasets of experimentally validated peptide-MHC interactions, offers a significant advantage in precision. By predicting the binding affinity of each peptide-HLA pair, we gain insight into which peptides are most likely to be presented on the cell surface and thus recognized by T-cells, while peptides that did not meet established binding affinity thresholds were filtered out, ensuring that only those with significant binding potential were retained. This step is crucial as only peptides with strong or moderate binding affinities to HLA molecules are prioritized, as they have the highest likelihood of functioning as true neoantigens in AML.
- High-Confidence Candidate Selection: The final list was refined by retaining peptides that exhibited both high expression (TPM ≥ 1) and strong binding affinities, maximizing the likelihood of selecting immunogenic neoantigens.

5. Results

5.1 Transcriptomic Profiling

Hierarchical clustering was performed on the 15 matched PBMC samples to discern distinct groups based on gene expression profiles between t0 and t48. Z-scores were computed for each gene to standardize the expression levels across samples, facilitating comparative analysis. The maximum distance metric was employed to assess the similarity between samples, and the Ward D2 clustering algorithm was utilized to minimize variance within clusters while maximizing variance between them. The clustering analysis yielded two main distinct clusters: Cluster A, which included 7 samples, and Cluster B, which comprised 8 samples (Figure 4).

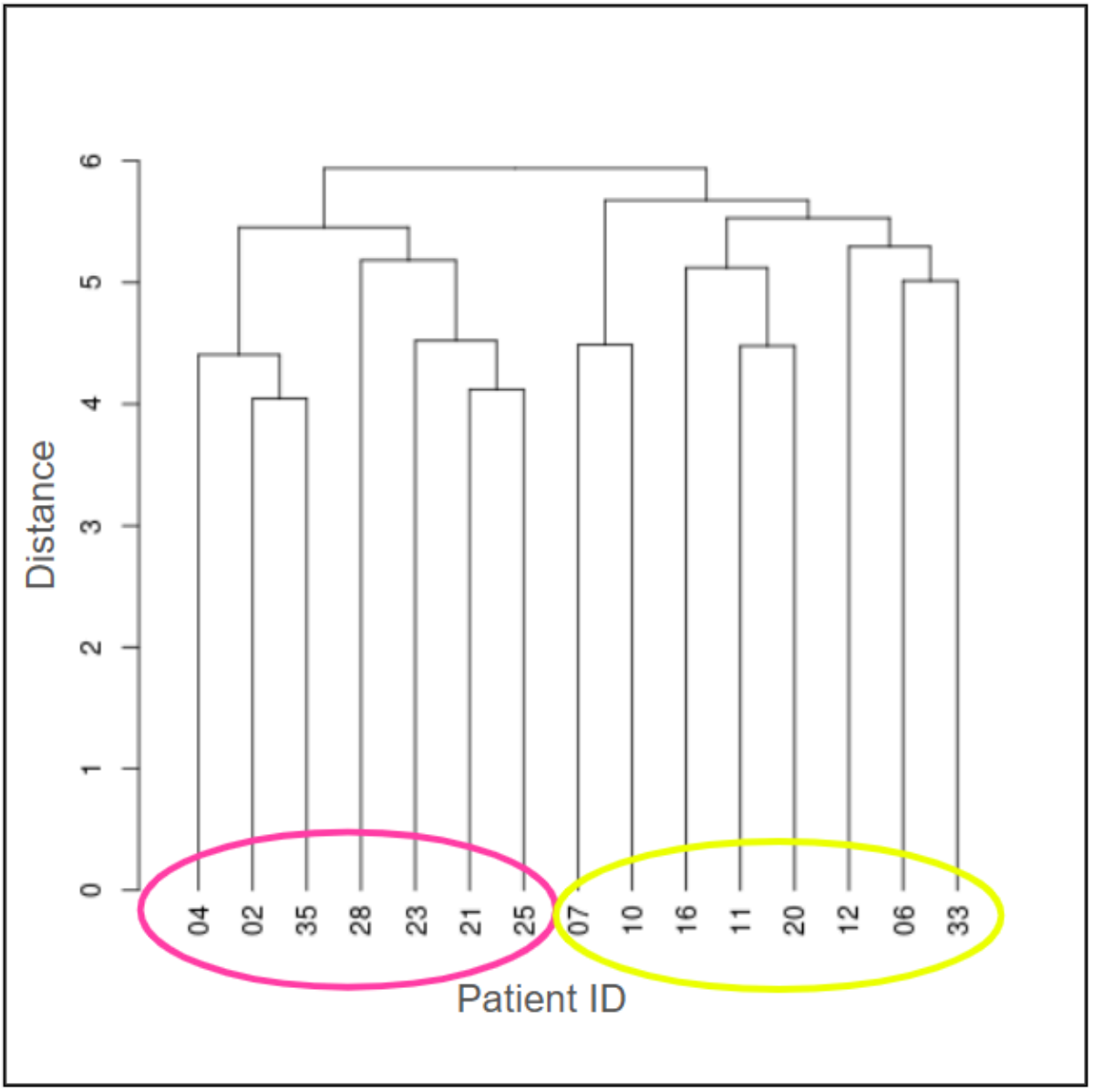

Figure 4: Hierarchical clustering dendrogram. This figures illustrates the relationships among samples based on their similarity. The vertical axis represents the distance at which clusters merge, with shorter links indicating greater similarity between samples. This dendrogram provides insights into the clustering structure, highlighting how samples group together, which is essential for understanding their underlying biological relationships. Samples of cluster A are circled in fuchsiam while samples of cluster B in yellow

To further explore the biological significance of the gene expression changes, GSEA was performed using BMMC t0 samples

matched with PBMC t0-t48 samples on the Reactome 2022 database. A total of 12 BMMC samples were analyzed in the GSEA, consisting of 6 samples from Cluster A and 6 from Cluster B, finding 36 significant gene sets identified with a p-value < 0.05 and an absolute normalized enrichment score (NES) > 1.5, summarized in Table 2.

Table 2. GSEA significant gene sets. The following table summarizes the significant gene sets identified through GSEA using data from the Reactome 2022 database. Each gene set is associated with its NES and p-value, indicating the strength and significance of the association.

Pathway Name	NES	p-value
TP53 Regulates Transcription Of Genes Involved In G2 Cell Cycle Arrest R-HSA-6804114	-1.85	8.2E-3
XBP1(S) Activates Chaperone Genes R-HSA-381038	-1.84	6.1E-3
TYSND1 Cleaves Peroxisomal Proteins R-HSA-9033500	-1.83	6.2E-3
Translation Of Structural Proteins R-HSA-9683701	-1.81	1.6E-2

NOTCH3 Activation And Transmission Of Signal To Nucleus R-HSA-9013507	-1.79	1.7E-2
IRE1alpha Activates Chaperones R-HSA-381070	-1.79	1.0E-2
Deactivation Of Beta-Catenin Transactivating Complex R-HSA-3769402	-1.76	1.0E-2
Maturation Of Spike Protein R-HSA-9683686	-1.72	8.0E-3
Inhibition Of DNA Recombination At Telomere R-HSA-9670095	-1.68	2.8E-2
Regulation Of TP53 Activity Thru Association With Co-factors R-HSA-6804759	-1.67	2.9E-2
Synthesis Of PG R-HSA-1483148	-1.67	2.7E-2
Sulfur Amino Acid Metabolism R-HSA-1614635	-1.65	3.1E-2
Late SARS-CoV-2 Infection Events R-HSA-9772573	-1.65	4.4E-2
Alpha-Linolenic (Omega3) And Linoleic (Omega6) Acid Metabolism R-HSA-2046104	-1.64	2.8E-2

Interleukin-9 Signaling R-HSA-8985947	-1.64	4.7E-2
Signaling By NTRK3 (TRKC) R-HSA-9034015	-1.60	1.7E-2
Formation Of Senescence-Associated Heterochromatin Foci (SAHF) R-HSA-2559584	-1.60	3.5E-2
Role Of Phospholipids In Phagocytosis R-HSA-2029485	-1.58	1.7E-2
Regulation Of PTEN Gene Transcription R-HSA-8943724	-1.57	2.8E-2
FGFR1b Ligand Binding And Activation R-HSA-190370	-1.57	2.7E-2
Signal Attenuation R-HSA-74749	-1.56	4.2E-2
Eicosanoid Ligand-Binding Receptors R-HSA-391903	-1.56	2.3E-2
Release Of Apoptotic Factors From Mitochondria R-HSA-111457	-1.56	4.9E-2

AKT Phosphorylates Targets In Cytosol R-HSA-198323	-1.55	4.3E-2
Transport Of RCbl Within Body R-HSA-9758890	-1.52	3.5E-2
RUNX2 Regulates Osteoblast Differentiation R-HSA-8940973	-1.52	4.9E-2
Prolactin Receptor Signaling R-HSA-1170546	-1.51	5.0E-2
Synthesis Of PA R-HSA-1483166	-1.50	2.3E-2
Free Fatty Acids Regulate Insulin Secretion R-HSA-400451	1.50	3.9E-2
Defective C1GALT1C1 Causes TNPS R-HSA-5083632	1.53	2.0E-2
SMAC (DIABLO) Binds To IAPs R-HSA-111463	1.57	3.4E-2
SMAC, XIAP-regulated Apoptotic Response R-HSA-111469	1.57	3.4E-2

RUNX1 Regulates Transcription Of Genes Involved In Differentiation Of Myeloid Cells R-HSA-8939246	1.61	4.4E-2
Fertilization R-HSA-1187000	1.63	2.1E-2
Activated NTRK2 Signals Thru PI3K R-HSA-9028335	1.70	2.1E-3
MET Activates PI3K/AKT Signaling R-HSA-8851907	1.84	6.5E-2

Among them, the most relevant are summarized in Figure 5. The TP53 pathway is crucial for regulating cell cycle progression and apoptosis (R-HSA-6804114). In AML, TP53 mutations are prevalent and associated with poor outcomes and treatment resistance. This pathway can induce G1 and G2 cell cycle arrest through various mechanisms, including the activation of CDKN1A (p21), which inhibits cyclin-dependent kinases [88]. The XBP1 pathway (R-HSA-381038) is involved in the unfolded protein response (UPR), which helps cells to manage endoplasmic reticulum stress. This pathway is activated in response to 5-Azacytidine treatment and may influence both survival and apoptosis in AML cells [89,90]. IRE1alpha (R-HSA-381070), also part of the UPR, helps maintain cellular homeostasis during stress by activating chaperone proteins that assist in proper protein folding. In AML, the activation of IRE1alpha has been shown to support cell survival under stress conditions, which is particularly relevant when

considering treatment with 5-Azacytidine. This drug can induce ER stress, leading to the activation of IRE1α pathways that promote cell survival through the upregulation of chaperone proteins [91]. Research indicates that 5-Azacytidine treatment can activate IRE1alpha signaling pathways in AML cells, contributing to their adaptive response to the drug. Inhibition of IRE1alpha has been proposed as a therapeutic strategy in AML, as targeting this pathway may enhance the sensitivity of leukemic cells to 5-Azacytidine. Studies have shown that inhibiting IRE1alpha can lead to increased apoptosis in AML cells treated with hypomethylating agents, suggesting a potential for combination therapies that exploit this vulnerability [92].

Regulation of PTEN gene transcription (R-HSA-8943724) is significant as PTEN is a tumor suppressor that negatively regulates the PI3K/AKT signaling pathway. Its regulation influences cell growth and survival pathways, potentially affecting responses to treatments like 5-Azacytidine [93]. The deactivation of the beta-catenin transactivating complex (R-HSA-3769402) is important in cancer biology, including AML, as this pathway is frequently implicated in various cancers, including AML. When the beta-catenin destruction complex is deactivated, beta-catenin accumulates and translocates to the nucleus, where it regulates gene expression related to cell proliferation and survival. This accumulation can significantly impact how cancer cells respond to therapies, including hypomethylating agents like 5-Azacytidine, which can lead to changes in the expression levels of beta-catenin and its target genes, thereby influencing cell survival and proliferation. This modulation may enhance the therapeutic effects of 5-Azacytidine by altering the cellular response to treatment [94,95]. The formation of senescence-associated

heterochromatin foci (SAHF) (R-HSA-2559584) is involved in cellular senescence, which can be induced by DNA damage and stress. 5-Azacytidine has been shown to induce cellular senescence in AML cells. It acts as a hypomethylating agent that can lead to the reactivation of silenced tumor suppressor genes, promoting senescence and apoptosis in leukemic cells. The induction of SAHF in response to 5-Azacytidine treatment suggests that this drug may enhance the senescence program, contributing to its therapeutic efficacy [96-98].

The AKT signaling pathway (R-HSA-198323) is often activated in cancers, including AML, and plays a significant role in cell survival and proliferation. This activation can result from various genetic alterations, such as mutations in the PI3K pathway or loss of tumor suppressor genes like PTEN. In the context of AML, studies have shown that the activation of the AKT signaling pathway can influence the effectiveness of treatments like 5-Azacytidine. For instance, AKT signaling has been implicated in mediating cellular responses to hypomethylating agents, where its inhibition may enhance the sensitivity of leukemic cells to these therapies. Its role in promoting cell survival and growth makes it critical when evaluating responses to hypomethylating agents like 5-Azacytidine [99-101].

SMAC proteins (R-HSA-111463 and R-HSA-111469) promotes apoptosis by binding to IAPs, thereby relieving their inhibition on caspases. This interaction is crucial for the apoptotic process, particularly in cancer cells where IAPs are often overexpressed. The combination of SMAC mimetics with 5-Azacytidine has been shown to enhance apoptotic signaling in AML cells. By antagonizing IAPs, SMAC proteins can potentiate the effects of 5-Azacytidine, leading to increased cell death in leukemic cells promote apoptosis by antagonizing inhibitors

of apoptosis proteins (IAPs), enhancing the effects of 5-Azacytidine treatment [102,103].

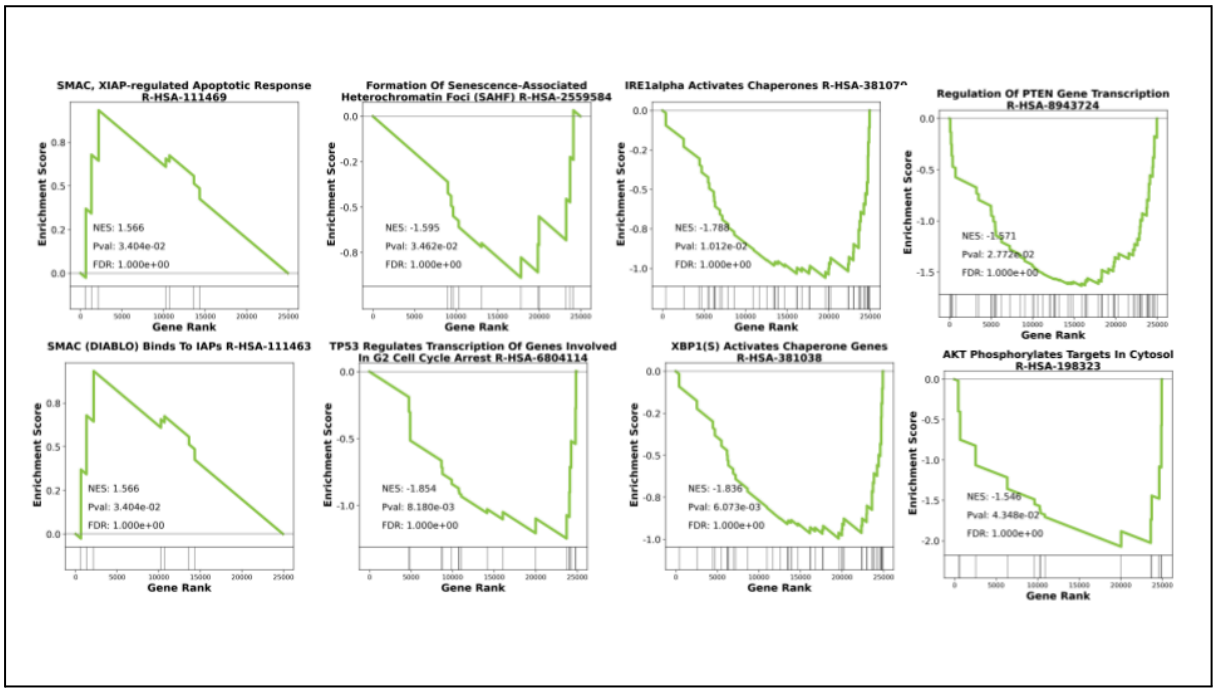

Figure 5. Key GSEA plots. This figure illustrates the most relevant GSEA plots derived from RNA-seq data, highlighting gene sets notably associated with the therapeutic response to 5-Azacytidine in AML patients.

5.2 Genomic Profiling

After applying quality filters, variants were further analyzed based on depth threshold and allele frequency. A minimum depth of 100X coverage was required for SNVs and INDELs, with a minimum VAF of 5% in tumor samples. Synonymous or intronic variants were excluded, retaining only missense, nonsense, frameshift, in-frame, and splicing variants for further analysis. Variants were classified following ACMG guidelines, considering only VUS, Likely Pathogenic, and Pathogenic variants for further investigation.

After applying the filtering criteria, we identified a total of 257 variants: 94 INDELs, 90 missense variants, 42 splicing

variants, and 31 stop-gain mutations across the samples. These variants were classified into categories based on ACMG classification guidelines: 87 were deemed as VUS, 120 were classified as Likely Pathogenic, and 50 as Pathogenic. Additionally, variants were stratified by clinical relevance into Tier I (18 variants), Tier II (106 variants), Tier III (92 variants), and Tier IV (41 variants), according to established criteria for prioritizing actionable mutations. For CNVs, filtering was conducted with specific criteria. Only CNVs with a copy number ratio between 0.5 and 2 were retained for analysis. After filtering, a total of 40 CNVs were kept for further investigation.

Subsequently, we generated an oncoprint that illustrates the most frequently mutated genes along with the specific types of variants identified, categorized by cluster (Figure 6). This visualization effectively conveys the distribution of mutations across the different clusters, providing valuable insights into the molecular landscape of the samples.

The WES samples were stratified into two clusters based on gene expression changes observed between t48 and t0 RNA-seq samples. Cluster A included 5 samples while Cluster B included 8 samples. Notably, 12 samples did not match either cluster due to the absence of corresponding RNA-seq data from t0 or t48 time points.

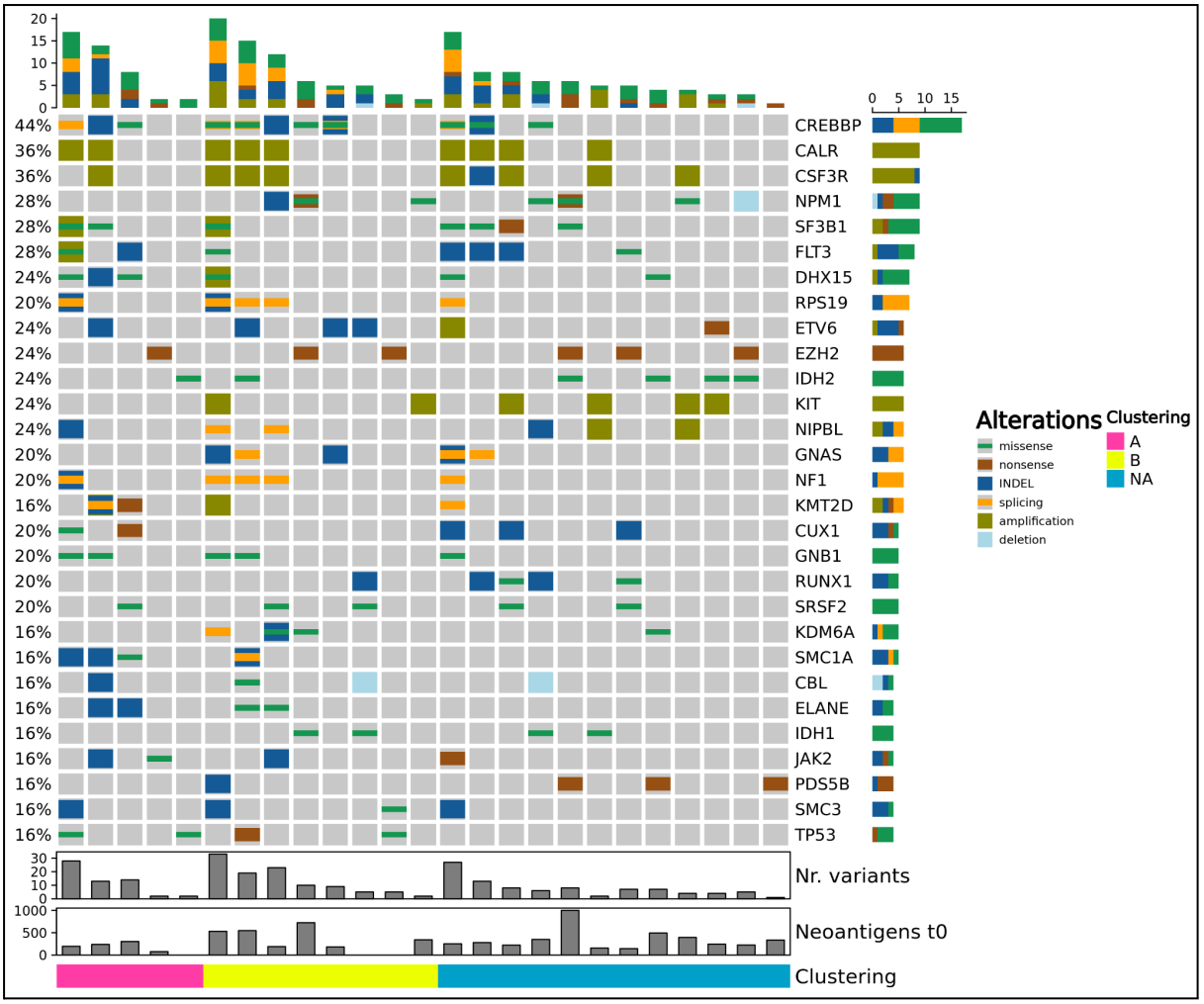

Figure 6. Oncoprint of most frequently altered genes. It shows genomic alterations across samples from patients with AML. Each row represents a specific gene, while each column corresponds to an individual patient sample. Color-coded boxes indicate the types of mutations for each gene, including missense variants, nonsense variants, INDELs, splicing variants, and CNVs. Additionally, the bottom of the plot provides information on the number of variants per sample, the count of neoantigens at time point t0 for each sample, and sample clusters.

The analysis of mutations across the clusters revealed varying frequencies of specific gene mutations in patients (Table 3). Notably, mutations in CUX1 and KMT2D were exclusive to Cluster A and not found in Cluster B. Conversely, mutations in ATM, GNAS, KDM6A, and NPM1 were absent in Cluster A but present in Cluster B.

Table 3. Most altered genes in each cluster. This table shows the percentages of mutated samples for the most altered genes in each cluster.

Gene	% Mutated Cluster 1	% Mutated Cluster 2
ATM	0	38
CREBBP	60	62
CUX1	40	0
DHX15	60	12
ELANE	40	25
ETV6	20	38
FLT3	40	12
GNAS	0	38
GNB1	40	25
JAK2	40	12
KDM6A	0	38
KMT2D	40	0
NF1	20	38
NPM1	0	38
RPS19	20	38
SF3B1	40	12
SMC1A	60	12
TP53	40	25

In the CNV analysis, the most frequently altered genes across all samples were CALR (69%), CSF3R (62%), and KIT (46%). These alterations were similarly distributed between the two clusters for CALR and CSF3R, with CALR showing alterations in 40% of Cluster A samples and 38% in Cluster B, while CSF3R was altered in 20% of Cluster A and 38% of Cluster B. However, alterations in KIT were notably different between the clusters: it was absent in Cluster A but altered in 25% of Cluster B, contrasting with its global alteration frequency of 46% across all samples.

The consistent alterations in CALR and CSF3R suggest these genes play a role in AML across both patient groups, potentially influencing common disease mechanisms. These mutations can contribute to aberrant signaling pathways that promote cell survival and proliferation, thereby affecting how patients respond to therapies. The mechanism of 5-Azacytidine involves inducing DNA demethylation, which can reactivate silenced tumor suppressor genes. The presence of CALR and CSF3R mutations may influence the effectiveness of 5-Azacytidine by modulating the cellular response to treatment. For instance, studies have shown that these mutations can lead to altered expression of genes involved in apoptosis and cell cycle regulation, potentially affecting how leukemic cells respond to 5-Azacytidine treatment [104-106].

Conversely, the absence of KIT alterations in Cluster A compared to their presence in Cluster B suggests a distinct role for KIT CNVs in shaping the oncogenic pathways and possibly the immune environment specific to Cluster B in AML. KIT alterations have been linked to poorer survival outcomes in AML, as shown in studies with AML1-ETO patients, underscoring KIT's impact on prognosis and cellular signaling

in AML subsets [107]. Studies have shown that specific AML mutations, including those in genes related to splicing factors and metabolic pathways, can influence therapeutic sensitivity to 5-Azacytidine and other epigenetic drugs. For example, a synergistic effect with 5-Azacytidine was observed in AML cells harboring SF3B1 mutations, underscoring the impact of mutation profiles on treatment outcomes with azacytidine-based regimens [108]. The differential presence of KIT mutations between clusters may reflect unique molecular characteristics, possibly pointing to alternative signaling cascades or immune modulations unique to each cluster. This supports the hypothesis that KIT alterations contribute to specific molecular signatures that define AML subtypes and influence patient response to treatment [109].

5.3 Immune Profiling

The integration of 10 matched WES and RNA-seq samples enabled a comprehensive analysis of neoantigens, which are critical for understanding the immune response to tumors. Notably, none of the neoantigen sequences were shared across patients. Furthermore, each patient was evaluated for the number of somatic variants and neoantigens (Table 4 and Figure 7).

Patient	Cluster	nr. Variants	nr. Neoantigens
02	A	21	302
04	A	108	196
06	B	25	342
07	B	32	724

10	B	42	181
11	B	93	188
12	B	66	547
16	B	150	531
21	A	12	75
35	A	34	237

Table 4. Summary of patient multi-omics approach data. This table includes cluster assignments, number of somatic variants, and counts of neoantigens identified in patients with AML. Each patient is categorized into either Cluster A or Cluster B, highlighting the diversity of genomic alterations and potential immunogenic profiles across the cohort.

The distinct neoantigen profiles observed in Clusters A and B not only point to differences in immunogenicity but may also be linked to varied responses to treatments like 5-Azacytidine. In Cluster B, the higher neoantigen counts—peaking at 724—suggest an environment potentially more responsive to immune-modulatory effects, which could enhance the efficacy of 5-Azacytidine. As a DNA methyltransferase inhibitor, 5-Azacytidine acts by reactivating silenced tumor suppressor genes through demethylation, and this epigenetic reprogramming can increase the expression of neoantigens on tumor cells. In tumors with a higher neoantigen load, such as those in Cluster B, the reactivation of additional neoantigens may intensify immune recognition, potentially leading to enhanced responses to immunotherapies when combined with

5-Azacytidine.

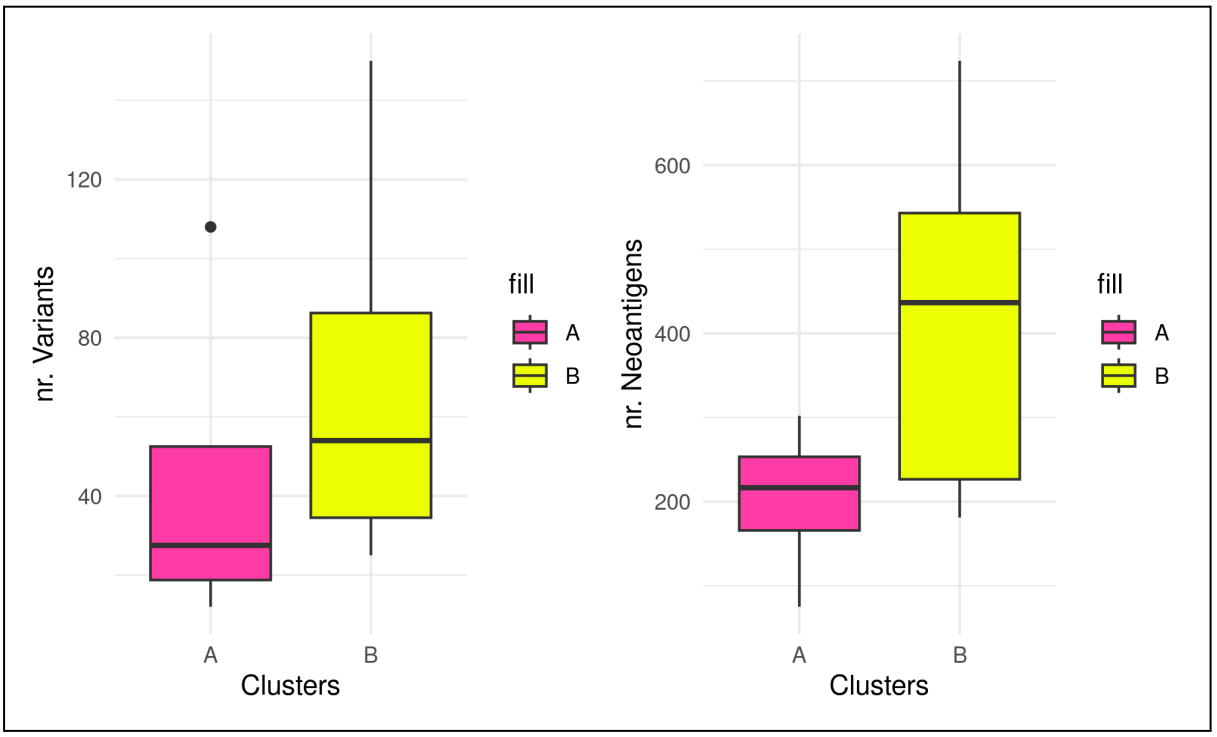

Figure 7. Variants and Neoantigen Boxplots. This figure illustrates the distribution of data among patients in Cluster A and Cluster B. The boxplot on the left shows the distribution of the number of somatic variants, while the boxplot on the right depicts the distribution of neoantigen counts. The differences in these distributions provide insights into the mutational landscape and potential immunogenic profiles associated with each cluster.

Moreover, the variability in neoantigen burden within Cluster A, with counts ranging widely from 75 to 302, could indicate differential sensitivity to 5-Azacytidine within this cluster. Tumors with lower neoantigen loads may respond less robustly to the immune-stimulatory effects of 5-Azacytidine, as fewer neoantigens are available to elicit a significant immune response. For these patients, the therapeutic benefits of 5-Azacytidine might be limited to its direct anti-leukemic effects without the added benefit of heightened immune recognition. Conversely, those within Cluster A with relatively higher neoantigen counts might experience a

stronger immune-mediated response to 5-Azacytidine, though still potentially less pronounced than in Cluster B.

These findings imply that neoantigen profiling could serve as a predictive biomarker for 5-Azacytidine response. Patients in Cluster B, with their inherently higher neoantigen counts, might benefit more from combination strategies that incorporate 5-Azacytidine with immune checkpoint inhibitors or neoantigen-targeted vaccines, taking advantage of the immune-stimulating properties of 5-Azacytidine. In Cluster A, patients with lower neoantigen levels may require additional adjunctive therapies beyond 5-Azacytidine to enhance immune recognition or to directly target leukemia cell survival pathways.

6. Discussion

The WES analysis revealed significant mutational differences between the two primary AML clusters. The exclusive presence of mutations in CUX1 and KMT2D in Cluster A, alongside their absence in Cluster B, points to unique oncogenic pathways potentially contributing to disease progression and therapeutic response within this cluster. CUX1, a known transcription factor, has been linked to tumor suppressive functions, with mutations contributing to genomic instability, an aspect that could influence the response to hypomethylating agents like 5-Azacytidine [110]. Similarly, KMT2D is involved in histone methylation and chromatin remodeling, and its mutations may lead to transcriptional dysregulation, suggesting that patients in Cluster A could experience differential transcriptional activity that impacts treatment outcomes [111,112]. Conversely, Cluster B's unique mutation profile in ATM, GNAS, KDM6A, and NPM1 provides a contrasting molecular

landscape. The absence of ATM and KDM6A mutations in Cluster A may indicate lower genomic instability and a potentially less responsive immune microenvironment, given ATM's role in DNA repair and immune response [113,114]. The presence of NPM1 mutations in Cluster B, however, is clinically significant, as NPM1 mutations are frequently associated with AML and linked to specific responses to chemotherapy and hypomethylating agents [117]. This cluster-specific mutation profile may help explain differences in immune-related pathways and neoantigen presentation seen in the neoantigen analysis, with higher neoantigen counts in Cluster B potentially attributable to these distinctive mutations. The CNV analysis across both clusters further highlighted key genes with consistent alterations, such as CALR and CSF3R, known to drive myeloid proliferation through JAK-STAT pathway activation [104-106]. Interestingly, the presence of KIT alterations in Cluster B and their absence in Cluster A could suggest divergent oncogenic pathways or distinct immunogenic characteristics influencing response patterns. KIT mutations, known for their association with poor prognosis and chemoresistance in AML, underscore the potential need for combination therapies in Cluster B to enhance treatment efficacy [107].

RNA-seq analysis allowed us to explore transcriptional responses to 5-Azacytidine, specifically through clustering based on gene expression shifts between t0 and t48. The expression profiles revealed divergent transcriptional programs associated with immune responses, metabolic pathways, and stress responses, contributing to an understanding of response heterogeneity in AML. Differentially expressed genes in Cluster A were primarily enriched in pathways such as caffeine metabolism, alpha-linolenic acid metabolism, and pyruvate metabolism, which have been linked to cellular stress

and metabolic reprogramming in cancer. These pathways are known to influence cancer cell proliferation and survival, potentially affecting sensitivity to 5-Azacytidine [117,118]. Conversely, Cluster B exhibited significant enrichment in pathways involved in cell signaling and immune regulation, such as the JAK-STAT and MAPK pathways. These findings are in line with recent studies that underscore the role of immune signaling in AML progression and treatment response, particularly in relation to 5-Azacytidine, which has been shown to influence immune-modulating pathways [119,120]. The enhanced expression of genes within these immune-regulatory pathways may contribute to improved antigen presentation and immune recognition, suggesting that Cluster B patients could exhibit distinct immunogenic profiles potentially responsive to combination therapies involving immunotherapy.

Neoantigen analysis revealed differences in the immune response potential between clusters, with Cluster B exhibiting a higher median neoantigen count, suggesting a more immunogenic tumor environment. Although not statistically significant, this increased neoantigen load might enhance recognition by T cells, aligning with the hypothesis that a greater neoantigen burden could stimulate anti-tumor immunity, as described in various cancers, including AML [52,120,121]. The findings support an immunologically distinct profile for Cluster B, which may influence the effectiveness of immune-based therapies. Interestingly, the discrepancy between neoantigen load and mutational burden indicates that neoantigen formation involves factors beyond mutation rate alone, such as variant expression levels and peptide-MHC binding affinities, both crucial for effective immune recognition [83,124,125]. This insight is relevant for immunotherapy approaches, particularly in Cluster B, where higher neoantigen

loads might improve the engagement of immune responses. This cluster-specific neoantigen profile suggests that the efficacy of therapies combining immune modulation with agents like 5-Azacytidine may differ by cluster, potentially enhancing outcomes in patients with a higher neoantigen load.

This multi-omic approach which integrates data from WES, RNA-seq, and neoantigen analyses provides a comprehensive view of the molecular and immunological landscapes within AML clusters. The data suggest that Cluster A and Cluster B exhibit distinct genetic and transcriptional signatures, alongside varied neoantigen profiles, which collectively influence their response to the hypomethylating agent 5-Azacytidine. Cluster A, with its transcriptional activation of apoptotic pathways and lack of survival-associated alterations (e.g., KIT), indicates a predisposition to undergo cell death in response to 5-Azacytidine, aligning with prior evidence that suggests hypomethylating agents are effective in AML subtypes marked by pro-apoptotic gene expression profiles [125-127]. Conversely, Cluster B's unique mutational landscape, featuring mutations in genes such as KIT and pathways promoting survival and immune evasion, implies a more resistant phenotype. Previous research has indicated that KIT mutations in AML may contribute to survival and therapy resistance, emphasizing the need for combined treatment strategies to target these pathways effectively [128,129]. Neoantigen analysis further differentiates the clusters, with Cluster B demonstrating a higher median neoantigen load. This immunologically distinct profile aligns with findings in solid tumors, where increased neoantigen loads correlate with enhanced T-cell response and immunotherapy sensitivity, suggesting a potential benefit of checkpoint inhibitors or neoantigen-based vaccines for Cluster B patients [53,130]. Thus,

multi-omic integration enables the identification of actionable targets in Cluster B, where resistance to 5-Azacytidine could be mitigated through combined targeted and immunotherapy approaches.

7. Conclusion and Future Perspectives

This study emphasizes the potential of a multi-omic approach to understand the molecular mechanisms driving response to 5-Azacytidine in AML. The integration of WES, RNA-seq, and neoantigen analysis offers a more comprehensive view of the genetic, transcriptional, and immunogenic variations in AML, allowing for a stratified analysis of cluster-specific behaviors. These insights suggest the need for developing personalized therapeutic strategies, potentially incorporating targeted therapy, immune modulation, and hypomethylating agents in accordance with the unique molecular profiles identified in each AML cluster. Further studies leveraging expanded patient datasets can build on these findings to validate biomarkers that may predict response to 5-Azacytidine, as well as inform novel combination strategies that can enhance treatment efficacy and overcome resistance.

Despite the robust multi-omic data analysis, the study has several limitations. Foremost is the absence of clinical information, which has not been provided by all the participant-enrolling centers. This lack of clinical context limits our ability to correlate molecular findings with patient outcomes, treatment histories, or clinical characteristics, which would offer invaluable validation of the predictive biomarkers identified. Future studies incorporating comprehensive clinical data will be critical for translating these findings into clinical practice, enabling

correlations that link specific molecular profiles with therapeutic outcomes, disease progression, and overall survival.

References

1. Döhner, H., Wei, A. H., Appelbaum, F. R., Craddock, C., DiNardo, C. D., Dombret, H., ... & Löwenberg, B. (2022). Diagnosis and management of AML in adults: 2022 recommendations from an international expert panel on behalf of the ELN. *Blood, The Journal of the American Society of Hematology, 140(12)*, 1345-1377.
2. De Kouchkovsky, I., & Abdul-Hay, M. (2016). Acute myeloid leukemia: a comprehensive review and 2016 update. *Blood cancer journal, 6(7)*, e441-e441.
3. Emadi, A. (2019). Acute myeloid leukemia: Overview and current management strategies. MSD Manuals. Retrieved from https://www.msdmanuals.com/home/blood-disorders/leukemias/acute-myeloid-leukemia-aml
4. Döhner, H., Weisdorf, D. J., & Bloomfield, C. D. (2015). Acute myeloid leukemia. *New England Journal of Medicine*, 373(12), 1136-1152.
5. Vakiti A, Reynolds SB, & Mewawalla P. (2024). Acute Myeloid Leukemia. Retrieved from: https://www.ncbi.nlm.nih.gov/books/NBK507875/
6. Blood Cancer UK. (2024). Symptoms of acute myeloid leukaemia (AML). Retrieved from https://bloodcancer.org.uk/understanding-blood-cancer/leukaemia/acute-myeloid-leukaemia/aml-symptoms-diagnosis/symptoms/
7. Snaith, O., Poveda-Rogers, C., Laczko, D., Yang, G., & Morrissette, J. J. (2023). Cytogenetics and genomics of acute myeloid leukemia. *Best Practice & Research Clinical Haematology*, 101533.
8. Shallis, R. M., Wang, R., Davidoff, A., Ma, X., & Zeidan, A. M. (2019). Epidemiology of acute myeloid leukemia:

Recent progress and enduring challenges. *Blood reviews, 36*, 70-87.

9. Labrador, J., Martinez-Cuadron, D., Boluda, B., Serrano, J., Gil, C., Perez-Simon, J. A., ... & Alonso-Dominguez, J. M. (2022). Evolution of the Genetic and Biological Studies Performed at Diagnosis in Patients with Acute Myeloid Leukemia Included in the Pethema Epidemiological Registry (REALMOL Study). *Blood, 140*(Supplement 1), 9100-9102.
10. Jani, C. T., Ahmed, A., Singh, H., Mouchati, C., Al Omari, O., Bhatt, P. S., ... & Lam, P. (2023). Burden of AML, 1990-2019: Estimates from the global burden of disease study. *JCO Global Oncology, 9*, e2300229.
11. Eacker, S., Muratov, A., Malig, M., Nelson, B., Wood, M., VanDyke, M., ... & Radich, J. (2023). 4. Comprehensive next generation cytogenomics improves risk stratification of acute myeloid leukemia. *Cancer Genetics, 278*, 2.
12. Weinberg, O. K., Porwit, A., Orazi, A., Hasserjian, R. P., Foucar, K., Duncavage, E. J., & Arber, D. A. (2023). The International Consensus Classification of acute myeloid leukemia. *Virchows Archiv, 482*(1), 27-37.
13. Yokota, T., & Kanakura, Y. (2016). Genetic abnormalities associated with acute lymphoblastic leukemia. *Cancer science, 107*(6), 721-725.
14. Hamed, N. A. What is New in the 5th Edition of the World Health Organization Classification of Acute Myeloid Leukemia.
15. Fu, C., Kou, R., Meng, J., Jiang, D., Zhong, R., & Dong, M. (2023). m6A genotypes and prognostic signature for assessing the prognosis of patients with acute myeloid leukemia. *BMC Medical Genomics, 16*(1), 191.

16. Tosi, J. F. D. S., Severino, A. R., Bertolucci, C. M., Barbosa, L. G. R., Higashi, M., Mattos, E. R., & Ikom-Colturato, M. R. V. (2023). ACUTE MYELOID LEUKEMIA WITH A RARE IMMUNOPHENOTYPE AND COMPLEX GENETIC PROFILE. *Hematology, Transfusion and Cell Therapy, 45*, S141.
17. Park, H. S. (2024). What is new in acute myeloid leukemia classification?. *Blood Research, 59*(1), 15.
18. Daver, N., Schlenk, R. F., Russell, N. H., & Levis, M. J. (2019). Targeting FLT3 mutations in AML: review of current knowledge and evidence. *Leukemia, 33*(2), 299-312.
19. Falini, B. (2023). NPM1-mutated acute myeloid leukemia: New pathogenetic and therapeutic insights and open questions. *American journal of hematology, 98*(9), 1452-1464.
20. Stengel, A., Kern, W., Haferlach, T., Meggendorfer, M., Fasan, A., & Haferlach, C. (2017). The impact of TP53 mutations and TP53 deletions on survival varies between AML, ALL, MDS and CLL: an analysis of 3307 cases. *Leukemia, 31*(3), 705-711.
21. Kudalkar, E. M., Pang, C., Haag, M. M., Pollyea, D. A., Kamdar, M., Xu, G., ... & Bao, L. (2022). 21q22 amplification detection in three patients with acute myeloid leukemia: cytogenomic profiling and literature review. *Molecular cytogenetics, 15*(1), 30.
22. Mondesir, J., Willekens, C., Touat, M., & de Botton, S. (2016). IDH1 and IDH2 mutations as novel therapeutic targets: current perspectives. *Journal of blood medicine*, 171-180.
23. Fasan, A., Haferlach, C., Alpermann, T., Jeromin, S., Grossmann, V., Eder, C., ... & Schnittger, S. (2014). The role of different genetic subtypes of CEBPA mutated AML. *Leukemia, 28*(4), 794-803.

24. DiNardo, C. D., & Cortes, J. E. (2016). Mutations in AML: prognostic and therapeutic implications. *Hematology 2014, the American Society of Hematology Education Program Book, 2016*(1), 348-355.
25. Wouters, B. J. (2021). Targeting IDH1 and IDH2 mutations in acute myeloid leukemia: Emerging options and pending questions. *Hemasphere, 5*(6), e583.
26. Tazi, Y., Arango-Ossa, J. E., Zhou, Y., Bernard, E., Thomas, I., Gilkes, A., ... & Papaemmanuil, E. (2022). Unified classification and risk-stratification in acute myeloid leukemia. *Nature Communications, 13*(1), 1-16.
27. Lee, S. J., Kwag, D., Lee, J. M., Kim, H. S., Cho, B. S., Kim, M., ... & Jung, J. (2023). 331MO Comparing data distribution and prognosis of acute myeloid leukemia according to 5th WHO classification and international consensus. *Annals of Oncology, 34*, S1600.
28. Moffitt, A. B., Alexander, J., Stepansky, A., Famulare, C., Lopez, C. A., Hakim, N., ... & Wigler, M. (2022). Whole Genome Sequencing Comparison of Acute Myeloid Leukemia at Presentation and Remission Predicts Patient Outcome. *Blood, 140*(Supplement 1), 9167-9169.
29. Bohl, S. R., Bullinger, L., & Rücker, F. G. (2018). Epigenetic therapy: azacytidine and decitabine in acute myeloid leukemia. *Expert review of hematology, 11*(5), 361-371.
30. DiNardo, C. D., Jonas, B. A., Pullarkat, V., Thirman, M. J., Garcia, J. S., Wei, A. H., ... & Pratz, K. W. (2020). Azacitidine and venetoclax in previously untreated acute myeloid leukemia. *New England Journal of Medicine, 383*(7), 617-629.
31. Wang, H., Li, Y., Lv, N., Li, Y., Wang, L., & Yu, L. (2018). Predictors of clinical responses to

hypomethylating agents in acute myeloid leukemia or myelodysplastic syndromes. *Annals of hematology, 97*, 2025-2038.

32. Schmutz, M., Zucknick, M., Schlenk, R. F., Mertens, D., Benner, A., Weichenhan, D., ... & Claus, R. (2023). Predictive value of DNA methylation patterns in AML patients treated with an azacytidine containing induction regimen. *Clinical Epigenetics, 15*(1), 171.
33. Bhansali, R. S., Pratz, K. W., & Lai, C. (2023). Recent advances in targeted therapies in acute myeloid leukemia. *Journal of Hematology & Oncology, 16*(1), 29.
34. Kayser, S., & Levis, M. J. (2023). The clinical impact of the molecular landscape of acute myeloid leukemia. *Haematologica, 108*(2), 308.
35. Bhatia, K., Sandhu, V., Wong, M. H., Iyer, P., & Bhatt, S. (2024). Therapeutic biomarkers in acute myeloid leukemia: functional and genomic approaches. *Frontiers in Oncology, 14*, 1275251.
36. Demir, D. (2023). Insights into the New Molecular Updates in Acute Myeloid Leukemia Pathogenesis. *Genes, 14*(7), 1424.
37. Makkar, H., Majhi, R. K., Goel, H., Gupta, A. K., Chopra, A., Tanwar, P., & Seth, R. (2023). Acute myeloid leukemia: novel mutations and their clinical implications. *American journal of blood research, 13*(1), 12.
38. Jakobsen, I., Sundkvist, M., Björn, N., Gréen, H., & Lotfi, K. (2022). Early changes in gene expression profiles in AML patients during induction chemotherapy. *BMC genomics, 23*(1), 752.
39. Handschuh, L., Wojciechowski, P., Kazmierczak, M., & Lewandowski, K. (2021). Transcript-level dysregulation of

BCL2 family genes in acute myeloblastic leukemia. *Cancers, 13*(13), 3175.

40. Petiti, J., Pignochino, Y., Schiavon, A., Giugliano, E., Berrino, E., Giordano, G., ... & Lo Iacono, M. (2024). Comprehensive Molecular Profiling of NPM1-Mutated Acute Myeloid Leukemia Using RNAseq Approach. *International Journal of Molecular Sciences, 25*(7), 3631.
41. Liu, P., Liu, J. P., Sun, S. J., Gao, Y., Ai, Y., Chen, X., ... & Yuan, H. X. (2021). CBFB-MYH11 fusion sequesters RUNX1 in cytoplasm to prevent DNMT3A recruitment to target genes in AML. *Frontiers in Cell and Developmental Biology, 9*, 675424.
42. Jakobsen, I., Sundkvist, M., Björn, N., Gréen, H., & Lotfi, K. (2022). Early changes in gene expression profiles in AML patients during induction chemotherapy. *BMC genomics, 23*(1), 752.
43. Yu, J., Li, Y., Zhang, D., Wan, D., & Jiang, Z. (2020). Clinical implications of recurrent gene mutations in acute myeloid leukemia. *Experimental hematology & oncology, 9*(1), 4.
44. Hu, Z., Yang, Y., Li, J., & Hu, Z. (2024). Genetic mutations and immune microenvironment: unveiling the connection to AML prognosis. *Hematology, 29*(1), 2346965.
45. Isidori, A., Daver, N., & Curti, A. (2021). the biological landscape of immunotherapy in AML. *Frontiers in Oncology, 11*, 671252.
46. Khaldoyanidi, S., Nagorsen, D., Stein, A., Ossenkoppele, G., & Subklewe, M. (2021). Immune biology of acute myeloid leukemia: implications for immunotherapy. *Journal of Clinical Oncology, 39*(5), 419.

47. Sim, M. J., & Sun, P. D. (2022). T cell recognition of tumor neoantigens and insights into T cell immunotherapy. *Frontiers in Immunology, 13*, 833017.
48. Okada, M., Shimizu, K., & Fujii, S. I. (2022). Identification of neoantigens in cancer cells as targets for immunotherapy. *International Journal of Molecular Sciences, 23*(5), 2594.
49. Chergui, A., & Reagan, J. L. (2023). Immunotherapy in acute leukemias: past success paves the way for future progress. *Cancers, 15*(16), 4137.
50. Saini, S. K., Holmberg-Thydén, S., Bjerregaard, A. M., Unnikrishnan, A., Dorfmüller, S., Platzbecker, U., ... & Hadrup, S. R. (2022). Neoantigen reactive T cells correlate with the low mutational burden in hematological malignancies. *Leukemia, 36*(11), 2734-2738.
51. Vadakekolathu, J., Minden, M. D., Hood, T., Church, S. E., Reeder, S., Altmann, H., ... & Rutella, S. (2020). Immune landscapes predict chemotherapy resistance and immunotherapy response in acute myeloid leukemia. *Science translational medicine, 12*(546), eaaz0463.
52. Pastorczak, A., Domka, K., Fidyt, K., Poprzeczko, M., & Firczuk, M. (2021). Mechanisms of immune evasion in acute lymphoblastic leukemia. *Cancers, 13*(07), 1536.
53. Zhou, W., Yu, J., Li, Y., & Wang, K. (2022). Neoantigen-specific TCR-T cell-based immunotherapy for acute myeloid leukemia. *Experimental Hematology & Oncology, 11*(1), 100.
54. Rutella, S., Vadakekolathu, J., Mazziotta, F., Reeder, S., Yau, T. O., Mukhopadhyay, R., ... & Luznik, L. (2022). Immune dysfunction signatures predict outcomes and define checkpoint blockade-unresponsive

microenvironments in acute myeloid leukemia. *The Journal of Clinical Investigation, 132*(21).

55. Gómez-Llobell, M., Peleteiro Raíndo, A., Climent Medina, J., Gómez Centurión, I., & Mosquera Orgueira, A. (2022). Immune checkpoint inhibitors in acute myeloid leukemia: A meta-analysis. *Frontiers in Oncology, 12*, 882531.
56. Subklewe, M., Bücklein, V., Sallman, D., & Daver, N. (2023). Novel immunotherapies in the treatment of AML: is there hope?. *Hematology, 2023*(1), 691-701.
57. Sargas, C., Ayala, R., Larráyoz, M. J., Chillón, M. C., Carrillo-Cruz, E., Bilbao-Sieyro, C., ... & PETHEMA group. (2023). Molecular landscape and validation of new genomic classification in 2668 adult AML patients: real life data from the PETHEMA registry. *Cancers, 15*(2), 438.
58. Unger, J. M., Hershman, D. L., Till, C., Minasian, L. M., Osarogiagbon, R. U., Fleury, M. E., & Vaidya, R. (2021). "When offered to participate": a systematic review and meta-analysis of patient agreement to participate in cancer clinical trials. *JNCI: Journal of the National Cancer Institute, 113*(3), 244-257.
59. Forsat, N. D., Palmowski, A., Palmowski, Y., Boers, M., & Buttgereit, F. (2020). Recruitment and retention of older people in clinical research: a systematic literature review. *Journal of the American Geriatrics Society, 68*(12), 2955-2963.
60. Zymo Research. Direct-Zol™ mini/micro DNA/RNA kit. Retrieved from https://www.eadscience.com/product/r2080t-r2080-r2081-direct-zol-dna-rna-miniprep-zymo-research/
61. Twist Bioscience. Twist total nucleic acids library preparation EF kit 2.0. Retrieved from

https://www.twistbioscience.com/sites/default/files/resources/2022-08/Protocol_LP2_0_EnzFragCDI-REV2.final_.pdf

62. Promega. Maxwell RSC Blood DNA Kit. Retrieved from https://ita.promega.com/-/media/files/resources/protocols/technical-manuals/101/maxwell-rsc-blood-dna-kit-protocol.pdf?rev=2af291fc365349e9a7a854930e1d0bf1&sc_lang=en

63. QIagen. QIAamp Micro DNA Kit. Retrieved from https://www.qiagen.com/us/products/discovery-and-translational-research/dna-rna-purification/dna-purification/genomic-dna/qiaamp-dna-kits

64. Illumina.. Illumina Stranded Total RNA Prep with Ribo-Zero Plus Kit. Retrieved from https://support-docs.illumina.com/LP/IlluminaStrandedTotalRNA/Content/LP/Illumina_RNA/Protocol_SM_ST.htm

65. Hamilton Company.. Hamilton Microlab STARlet. Retrieved from https://www.hamiltoncompany.com/automated-liquid-handling/platforms/microlab-star

66. Illumina. Illumina DNA Prep with Exome 2.0 Plus Enrichment Kit. Retrieved from https://support.illumina.com/content/dam/illumina-support/documents/documentation/chemistry_documentation/illumina_prep/1000000157112_03_illumina-dna-prep-exome-v2-enrichment-reference-guide.pdf

67. Illumina. bcl2fastq Conversion Software. Retrieved from https://support.illumina.com/downloads/bcl2fastq-conversion-software-v2-20.html

68. Bray, N. L., Pimentel, H., Melsted, P., & Pachter, L. (2016). Near-optimal probabilistic RNA-seq quantification. *Nature biotechnology, 34*(5), 525-527.

69. Soneson, C., Love, M. I., & Robinson, M. D. (2016). Differential analyses for RNA-seq: transcript-level

estimates improve gene-level inferences. *F1000Research, 4*, 1521.

70. Love, M. I., Huber, W., & Anders, S. (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. *Genome biology, 15*, 1-21.
71. Ward Jr, J. H. (1963). Hierarchical grouping to optimize an objective function. *Journal of the American statistical association, 58*(301), 236-244.
72. Fang, Z., Liu, X., & Peltz, G. (2023). GSEApy: a comprehensive package for performing gene set enrichment analysis in Python. *Bioinformatics, 39*(1), btac757.
73. Fabregat, A., Sidiropoulos, K., Garapati, P., Gillespie, M., Hausmann, K., Haw, R., ... & D'Eustachio, P. (2016). The reactome pathway knowledgebase. *Nucleic acids research, 44*(D1), D481-D487.
74. Subramanian, A., Tamayo, P., Mootha, V. K., Mukherjee, S., Ebert, B. L., Gillette, M. A., ... & Mesirov, J. P. (2005). Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. *Proceedings of the National Academy of Sciences, 102*(43), 15545-15550.
75. Illumina. Illumina DRAGEN DNA Pipeline. Retrieved from https://help.dragen.illumina.com/product-guides/dragen-v4.3/dragen-dna-pipeline
76. Wang, K., Li, M., & Hakonarson, H. (2010). ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. *Nucleic acids research, 38*(16), e164-e164.
77. Kopanos, C., Tsiolkas, V., Kouris, A., Chapple, C. E., Albarca Aguilera, M., Meyer, R., & Massouras, A. (2019). VarSome: the human genomic variant search engine. *Bioinformatics, 35*(11), 1978-1980.

78. Richards, S., Aziz, N., Bale, S., Bick, D., Das, S., Gastier-Foster, J., ... & Rehm, H. L. (2015). Standards and guidelines for the interpretation of sequence variants: a joint consensus recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology. *Genetics in medicine, 17*(5), 405-423.
79. Li, M. M., Datto, M., Duncavage, E. J., Kulkarni, S., Lindeman, N. I., Roy, S., ... & Nikiforova, M. N. (2017). Standards and guidelines for the interpretation and reporting of sequence variants in cancer: a joint consensus recommendation of the Association for Molecular Pathology, American Society of Clinical Oncology, and College of American Pathologists. *The Journal of molecular diagnostics, 19*(1), 4-23.
80. Gu, Z. (2022). Complex heatmap visualization. *Imeta, 1*(3), e43.
81. Vago, L., & Gojo, I. (2020). Immune escape and immunotherapy of acute myeloid leukemia. *The Journal of clinical investigation, 130*(4), 1552-1564.
82. Gupta, R. G., Li, F., Roszik, J., & Lizée, G. (2021). Exploiting tumor neoantigens to target cancer evolution: current challenges and promising therapeutic approaches. *Cancer discovery, 11*(5), 1024-1039.
83. Biernacki, M. A., & Bleakley, M. (2020). Neoantigens in hematologic malignancies. *Frontiers in immunology, 11*, 121.
84. Zhou, W., Yu, J., Li, Y., & Wang, K. (2022). Neoantigen-specific TCR-T cell-based immunotherapy for acute myeloid leukemia. *Experimental Hematology & Oncology, 11*(1), 100.

85. Reynisson, B., Alvarez, B., Paul, S., Peters, B., & Nielsen, M. (2020). NetMHCpan-4.1 and NetMHCIIpan-4.0: improved predictions of MHC antigen presentation by concurrent motif deconvolution and integration of MS MHC eluted ligand data. *Nucleic acids research, 48*(W1), W449-W454.
86. Atkins, T. K., Solanki, A., Vasmatzis, G., Cornette, J., & Riedel, M. (2024). Evaluating NetMHCpan performance on non-European HLA alleles not present in training data. *Frontiers in immunology, 14*, 1288105.
87. Aptsiauri, N., & Garrido, F. (2022). The challenges of HLA class I loss in cancer immunotherapy: facts and hopes. *Clinical Cancer Research, 28*(23), 5021-5029.
88. Engeland, K. (2022). Cell cycle regulation: p53-p21-RB signaling. *Cell Death & Differentiation, 29*(5), 946-960.
89. Liang, D., Khoonkari, M., Avril, T., Chevet, E., & Kruyt, F. A. (2021). The unfolded protein response as regulator of cancer stemness and differentiation: Mechanisms and implications for cancer therapy. *Biochemical Pharmacology, 192*, 114737.
90. Paradzik, T., Bandini, C., Mereu, E., Labrador, M., Taiana, E., Amodio, N., ... & Piva, R. (2021). The landscape of signaling pathways and proteasome inhibitors combinations in multiple myeloma. *Cancers, 13*(6), 1235.
91. Mnasri, N., Mamarbachi, M., Allen, B. G., & Mayer, G. (2018). 5-Azacytidine engages an IRE1α-EGFR-ERK1/2 signaling pathway that stabilizes the LDL receptor mRNA. *Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms, 1861*(1), 29-40.
92. Śniegocka, M., Liccardo, F., Fazi, F., & Masciarelli, S. (2022). Understanding ER homeostasis and the UPR to

enhance treatment efficacy of acute myeloid leukemia. *Drug Resistance Updates, 64,* 100853.

93. Lee, P., Yim, R., Miu, K. K., Fung, S. H., Liao, J. J., Wang, Z., ... & Gill, H. (2022). Epigenetic Silencing of PTEN and Epi-Transcriptional Silencing of MDM2 Underlied Progression to Secondary Acute Myeloid Leukemia in Myelodysplastic Syndrome Treated with Hypomethylating Agents. *International Journal of Molecular Sciences, 23*(10), 5670.
94. Cardona-Echeverry, A., & Prada-Arismendy, J. (2020). Deciphering the role of Wnt signaling in acute myeloid leukemia prognosis: how alterations in DNA methylation come into play in patients' prognosis. J*ournal of Cancer Research and Clinical Oncology, 146*(12), 3097-3109.
95. Zhang, Y., & Wang, X. (2020). Targeting the Wnt/β-catenin signaling pathway in cancer. *Journal of hematology & oncology, 13*(1), 165.
96. Crouch, J., Shvedova, M., Thanapaul, R. J. R. S., Botchkarev, V., & Roh, D. (2022). Epigenetic regulation of cellular senescence. *Cells, 11*(4), 672.
97. Olan, I., Handa, T., & Narita, M. (2023). Beyond SAHF: An integrative view of chromatin compartmentalization during senescence. *Current Opinion in Cell Biology*, 83, 102206.
98. Hao, X., Wang, C., & Zhang, R. (2022). Chromatin basis of the senescence-associated secretory phenotype. T*rends in cell biology, 32*(6), 513-526.
99. DiNardo, C. D., Jonas, B. A., Pullarkat, V., Thirman, M. J., Garcia, J. S., Wei, A. H., ... & Pratz, K. W. (2020). Azacitidine and venetoclax in previously untreated acute myeloid leukemia. *New England Journal of Medicine, 383*(7), 617-629.

100. Darici, S., Alkhaldi, H., Horne, G., Jørgensen, H. G., Marmiroli, S., & Huang, X. (2020). Targeting PI3K/Akt/mTOR in AML: rationale and clinical evidence. *Journal of Clinical Medicine, 9*(9), 2934.

101. Rascio, F., Spadaccino, F., Rocchetti, M. T., Castellano, G., Stallone, G., Netti, G. S., & Ranieri, E. (2021). The pathogenic role of PI3K/AKT pathway in cancer onset and drug resistance: an updated review. *Cancers, 13*(16), 3949.

102. Cetraro, P., Plaza-Diaz, J., MacKenzie, A., & Abadía-Molina, F. (2022). A review of the current impact of inhibitors of apoptosis proteins and their repression in cancer. *Cancers, 14*(7), 1671.

103. Dittmann, J., Haydn, T., Metzger, P., Ward, G. A., Boerries, M., Vogler, M., & Fulda, S. (2020). Next-generation hypomethylating agent SGI-110 primes acute myeloid leukemia cells to IAP antagonist by activating extrinsic and intrinsic apoptosis pathways. *Cell Death & Differentiation, 27*(6), 1878-1895.

104. Kim, S. Y., Song, I. C., Kim, J., & Kwon, G. C. (2024). Analysis of CSF3R mutations in atypical chronic myeloid leukemia and other myeloid malignancies. *Annals of Diagnostic Pathology, 71*, 152317.

105. Guastafierro, V., Ubezio, M., Manes, N., Milanesi, C., Della Porta, M., & Bonometti, A. (2023). CSF3R-mutant chronic myelomonocytic leukemia is a distinct clinically subset with abysmal prognosis: a case report and systematic review of the literature. *Leukemia & Lymphoma, 64*(9), 1566-1573.

106. Orlova, A., Wingelhofer, B., Neubauer, H. A., Maurer, B., Berger-Becvar, A., Keserű, G. M., ... & Moriggl, R. (2018). Emerging therapeutic targets in

myeloproliferative neoplasms and peripheral T-cell leukemia and lymphomas. *Expert opinion on therapeutic targets, 22*(1), 45-57.

107. Yang, M., Zhao, B., Wang, J., Zhang, Y., Hu, C., Liu, L., ... & Jin, J. (2022). A predictor combining clinical and genetic factors for AML1-ETO leukemia patients. *Frontiers in Oncology, 11*, 783114.

108. Bernardo, P., Pereira, M., Galvão, B., Desterro, J., Nunes, A., Frade, M. D. J., ... & Carmo-Fonseca, M. (2022). P388: PREFERENTIAL SYNERGISTIC EFFECT OF INDISULAM IN COMBINATION WITH VENETOCLAX AND AZACYTIDINE IN AML CELLS WITH SF3B1 MUTATIONS. *HemaSphere, 6*, 288-289.

109. Sun, R., Sun, L., Xie, X., Li, X., Wu, P., Wang, L., & Zhu, P. (2022). Single-cell analysis of transcription factor regulatory networks reveals molecular basis for subtype-specific dysregulation in acute myeloid leukemia. *Blood Science, 4*(02), 65-75.

110. Ramdzan, Z. M., & Nepveu, A. (2014). CUX1, a haploinsufficient tumour suppressor gene overexpressed in advanced cancers. *Nature Reviews Cancer, 14*(10), 673-682.

111. Yang, W., & Ernst, P. (2017). SET/MLL family proteins in hematopoiesis and leukemia. *International journal of hematology, 105*, 7-16.

112. Antunes, E. T., & Ottersbach, K. (2020). The MLL/SET family and haematopoiesis. *Biochimica et Biophysica Acta (BBA)-Gene Regulatory Mechanisms, 1863*(8), 194579.

113. Quintás-Cardama, A., Hu, C., Qutub, A., Qiu, Y. H., Zhang, X., Post, S. M., ... & Kornblau, S. M. (2017). p53 pathway dysfunction is highly prevalent in acute myeloid leukemia independent of TP53 mutational status. *Leukemia, 31*(6), 1296-1305.

114. George, B., Kantarjian, H., Baran, N., Krocker, J. D., & Rios, A. (2021). TP53 in acute myeloid leukemia: molecular aspects and patterns of mutation. *International journal of molecular sciences, 22*(19), 10782.
115. Wertheim, G. B., Hexner, E., & Bagg, A. (2012). Molecular-based classification of acute myeloid leukemia and its role in directing rational therapy: personalized medicine for profoundly promiscuous proliferations. *Molecular diagnosis & therapy*, 16, 357-369.
116. Dermawan, J. K., Wensel, C., Visconte, V., Maciejewski, J. P., Cook, J. R., & Bosler, D. S. (2022). Clinically Significant CUX1 Mutations Are Frequently Subclonal and Common in Myeloid Disorders With a High Number of Co-mutated Genes and Dysplastic Features. *American Journal of Clinical Pathology, 157*(4), 586-594.
117. Soltani, M., Zhao, Y., Xia, Z., Ganjalikhani Hakemi, M., & Bazhin, A. V. (2021). The importance of cellular metabolic pathways in pathogenesis and selective treatments of hematological malignancies. *Frontiers in Oncology, 11*, 767026.
118. Nair, R., Salinas-Illarena, A., & Baldauf, H. M. (2021). New strategies to treat AML: novel insights into AML survival pathways and combination therapies. *Leukemia, 35*(2), 299-311.
119. Xu, Q. Y., & Yu, L. (2020). Epigenetic therapies in acute myeloid leukemia: the role of hypomethylating agents, histone deacetylase inhibitors and the combination of hypomethylating agents with histone deacetylase inhibitors. *Chinese Medical Journal, 133*(06), 699-715.
120. Babar, Q., Saeed, A., Tabish, T. A., Pricl, S., Townley, H., & Thorat, N. (2022). Novel epigenetic

therapeutic strategies and targets in cancer. *Biochimica et Biophysica Acta (BBA)-Molecular Basis of Disease, 1868*(12), 166552.

121. Cai, Y., Li, D., Lv, D., Yu, J., Ma, Y., Jiang, T., ... & Xu, J. (2024). MHC-I-presented non-canonical antigens expand the cancer immunotherapy targets in acute myeloid leukemia. *Scientific Data, 11*(1), 831.

122. Wen, X. M., Xu, Z. J., Jin, Y., Xia, P. H., Ma, J. C., Qian, W., ... & Qian, J. (2021). Association analyses of TP53 mutation with prognosis, tumor mutational burden, and immunological features in acute myeloid leukemia. *Frontiers in Immunology, 12*, 717527.

123. Łuksza, M., Sethna, Z. M., Rojas, L. A., Lihm, J., Bravi, B., Elhanati, Y., ... & Balachandran, V. P. (2022). Neoantigen quality predicts immunoediting in survivors of pancreatic cancer. *Nature, 606*(7913), 389-395.

124. Penter, L., & Wu, C. J. (2020). Personal tumor antigens in blood malignancies: genomics-directed identification and targeting. *The Journal of Clinical Investigation, 130*(4), 1595-1607.

125. Wei, A. H., Roboz, G. J., Dombret, H., Dohner, H., Schuh, A. C., Montesinos, P., ... & Ravandi, F. (2023). Survival outcomes with oral azacitidine maintenance in patients with acute myeloid leukemia in remission by receipt of initial chemotherapy: subgroup analyses from the phase III QUAZAR AML-001 trial. *Haematologica, 108*(10), 2820-2825.

126. Roboz, G. J., Ravandi, F., Wei, A. H., Dombret, H., Thol, F., Voso, M. T., ... & Döhner, H. (2022). Oral azacitidine prolongs survival of patients with AML in remission independently of measurable residual disease

status. Blood, *The Journal of the American Society of Hematology, 139*(14), 2145-2155.

127. Stomper, J., Rotondo, J. C., Greve, G., & Lübbert, M. (2021). Hypomethylating agents (HMA) for the treatment of acute myeloid leukemia and myelodysplastic syndromes: mechanisms of resistance and novel HMA-based therapies. *Leukemia, 35*(7), 1873-1889.

128. Katagiri, S., Chi, S., Minami, Y., Fukushima, K., Shibayama, H., Hosono, N., ... & Gotoh, A. (2022). Mutated KIT tyrosine kinase as a novel molecular target in acute myeloid leukemia. *International Journal of Molecular Sciences, 23*(9), 4694.

129. Estruch, M., Reckzeh, K., Vittori, C., Centio, A., Ali, M., Engelhard, S., ... & Theilgaard-Mönch, K. (2021). Targeted inhibition of cooperative mutation-and therapy-induced AKT activation in AML effectively enhances response to chemotherapy. *Leukemia, 35*(7), 2030-2042.

130. Jeong, A. R., Trando, A. H., Thomas, S. D., Riviere, P., Sakowski, P. J., Sokol, E. S., ... & Kurzrock, R. (2024). Higher tumor mutational burden and PD-L1 expression correlate with shorter survival in hematologic malignancies. *Therapeutic Advances in Medical Oncology, 16*.