fortune Journals

Research Article

JOURNAL OF BIOINFORMATICS AND SYSTEMS BIOLOGY

ISSN: 2688-5107

Machine Learning Reveals Two Distinct Transcriptomics Subtypes of Acute Myeloid Leukaemia with Differences in Disease Outcomes and Genetic Landscape

Panji Nkhoma[1*], Doris Kafita[1], Kevin Dzobo[2], Sinkala Musalula[1,3,4]

Abstract

Background: Accurately subtyping diseases, particularly in cancer research, is crucial for enhancing the precision of treatment decisions and improving outcomes across various cancers, including hematological malignancies like acute myeloid leukaemia (AML).

Methods: Consequently, we utilised an unsupervised K-means clustering on transcriptomics data from 173 acute myeloid leukaemia samples profiled by The Cancer Genome Atlas (TCGA). In our analysis, we categorised patients into two distinct groups: Subtype-1, comprising 68 individuals, and subtype-2, encompassing 105 individuals.

Results: Analysis revealed that individuals within subtype-2 experienced a markedly prolonged period of disease-free survival compared to those in subtype-1, as evidenced by the Log-rank test ($p = 0.00273$). Furthermore, it was observed that patients in subtype-1 presented with elevated white blood cell counts, suggesting a potential biomarker of disease progression within this subgroup. We also identified differentially expressed genes linked to poor survival, prognosis, and chemoresistance, involving pathways like Aminoacyl-tRNA biosynthesis, apoptosis, NF-kappa B and HIF-1, through bioinformatics analysis of subtype-1. Our findings show that AML patients categorised within subtype-1 exhibit a more aggressive form of the disease compared to those allocated to subtype-2.

Conclusion: Consequently, these observations underscore the feasibility of employing subtype-specific precision treatments for AML patients, offering a tailored therapeutic approach based on the distinct disease characteristics of different patient subtypes.

Keywords: Acute myeloid leukaemia; Bioinformatics; Machine learning; The Cancer Genome Atlas; Cancer subtypes;

List of Abbreviations

AML	Acute Myeloid Leukaemia
DFS	Disease Free Survival
FDR	False Discovery Rate
GO	Gene Ontology
GSEA	Gene Set Enrichment Analysis
KEGG	Kyoto Encyclopaedia of Genes and Genomes

Affiliation:

[1]University of Zambia, School of Health Sciences, Department of Biomedical Sciences, Nationalist Road, Lusaka, Zambia

[2]Medical Research Council-SA Wound Healing Unit, Hair and Skin Research Laboratory, Division of Dermatology, Department of Medicine, Groote Schuur Hospital, Faculty of Health Sciences University of Cape Town, Anzio Road, Observatory, Cape Town, South Africa

[3]Division of Computational Biology, Department of Integrative Biomedical Sciences, Faculty of Health Sciences, University of Cape Town, Cape Town, South Africa

[4]Institute of Infectious Diseases and Molecular Medicine, University of Cape Town, Cape Town, South Africa

***Corresponding author:**
Panji Nkhoma, University of Zambia, School of Health Sciences, Department of Biomedical Sciences, Nationalist Road, Lusaka, Zambia

Citation: Panji Nkhoma, Doris Kafita, Kevin Dzobo, Sinkala Musalula. Machine Learning Reveals Two Distinct Transcriptomics Subtypes of Acute Myeloid Leukaemia with Differences in Disease Outcomes and Genetic Landscape. Journal of Bioinformatics and Systems Biology. 7 (2024): 227-236.

Received: August 12, 2024
Accepted: August 29, 2024
Published: December 23, 2024

Nkhoma P, et al., J Bioinform Syst Biol 2024
DOI:10.26502/jbsb.5107094

ML	Machine Learning
mRNA	messenger Ribonucleic Acid
OS	Overall Survival
PCA	Principal Component Analysis
TCGA	The Cancer Genome Atlas
WBC	White Blood Cell Count
X2K	Expression 2 Kinase

Introduction

Acute myeloid leukaemia (AML) represents a malignant haematological condition that impacts myeloid cells, a subset of white blood cells, within the blood and bone marrow [1]. This aggressive cancer is marked by a swift progression and is defined by the presence of more than 20% blast cells in the bone marrow or blood [1]. It arises from the malignant clonal expansion of myeloid progenitor cells, which is accompanied by a failure in the cells' ability to differentiate appropriately [2]. In 2023, the estimated incidence of AML was 20,330, accounting for 1% of all cancers, with an estimated death rate of 11,310 and a 5-year relative survival of 31.7% [3]. AML is characterised by numerous genetic aberrations, providing key insights into the mechanisms underlying its development and progression [4, 5]. Genetic aberrations in AML alter cellular transcriptional programs, allowing for the categorisation of patient's classification based on gene expression signatures, which offer valuable insight into disrupted signalling pathways and help identify potential targets for personalised therapeutic approaches [6, 7].

Machine learning (ML) methods offer significant opportunities for refining the definition of disease subtypes and enhancing risk prediction across various diseases [8]. ML can facilitate a more nuanced understanding of disease characteristics by leveraging complex datasets and identifying patterns that may not be evident to human observers. With the continuous refinement of classification schemes, further specific molecular markers correlated with patient survival and cancer aggressiveness are anticipated to be identified [9, 10]. As classification schemes improve, additional specific molecular correlates of patient survival and cancer aggressiveness are expected to be uncovered [11-13]. Unsupervised clustering algorithms, a cornerstone of machine learning methodologies, have been instrumental in elucidating cancer heterogeneity by discerning distinct subtypes across a spectrum of malignancies [14-16]. This is achieved through the meticulous analysis of gene expression data, where these algorithms identify natural groupings or clusters within the data that correspond to varying cancer subtypes [17]. Such advancements in computational biology have paved the way for more nuanced understandings of cancer biology and the potential for more personalised approaches to cancer treatment [18]

Numerous researchers have applied machine learning algorithms to the task of cancer subtyping across various cancer types, including but not limited to breast cancer [19, 20] and pancreatic cancer [21]. In our research, we utilise an unsupervised K-means clustering machine learning algorithm to delineate transcriptomic subtypes of AML by analysing gene expression data from The Cancer Genome Atlas (TCGA) [22]. After identifying AML subtypes, we identified differentially expressed genes across these groups. We then conducted bioinformatics analyses to uncover variations in the genetic landscape of these subtypes, aiming to reveal the molecular distinctions between them. This approach deepens our understanding of AML's heterogeneity and supports the development of targeted therapies.

Results

Transcriptomics subtypes of acute myeloid leukaemia

We applied an unsupervised K-mean clustering machine learning algorithm with a squared Euclidian distance metric to the transcriptomics data on acute myeloid leukaemia samples from the TCGA. We came up with two consistent subtypes of cancer patients. One subtype, which we named subtype-1, consisted of 68 samples, whereas the second subtype, named subtype-2, consisted of 105 samples (Figure 1).

Clinical characteristics and survival outcomes of the transcriptomics subtypes of acute myeloid leukaemia

We used the Kaplan Meier [23] test to compare the survival outcomes between patients with the two transcriptomics subtypes of AML. We found a similar overall survival (OS) duration (Log-rank test; p = 0.109) between patients with subtype-1 (13.6 months) and subtype-2 (26.3 months);

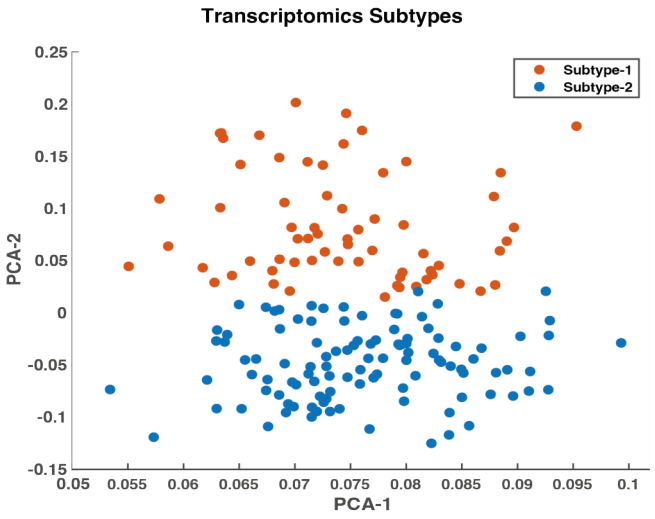

Figure 1: Clustering of acute myeloid leukaemia; the first and second principal components of the PCA are the plot on the X axis and the response on the Y axis. Points are coloured according to subtypes defined by K-means clustering.

Citation: Panji Nkhoma, Doris Kafita, Kevin Dzobo, Sinkala Musalula. Machine Learning Reveals Two Distinct Transcriptomics Subtypes of Acute Myeloid Leukaemia with Differences in Disease Outcomes and Genetic Landscape. Journal of Bioinformatics and Systems Biology. 7 (2024): 227-237.

Nkhoma P, et al., J Bioinform Syst Biol 2024
DOI:10.26502/jbsb.5107094

Figure 2a. However, the Disease-free survival (DFS) status for patients with subtype-2 acute myeloid leukaemia was significantly longer (26.2 months) than that for patients with subtype-1 AML (12 months), Log-rank test; p = 0.0273; Figure 2b. We further evaluated differences in other disease outcomes, including disease progression or recurrence at the end of follow-up between the two AML transcriptomics subtypes. Our findings showed that 45% of subtype-1 AML patients were disease-free, while 55% experienced disease progression or recurrence, with only 26% surviving. In comparison, 57% of subtype-2 AML patients were disease-free, while 43% experienced disease progression or recurrence, with 39% surviving; Figures 2c and 2e. A comparison of the mean ages of the two disease subtypes revealed a close distribution of 57 years for subtype-1 and 54 years for subtype-2; Figure 2d. We also performed the Wilcoxon rank sum test to assess differences in white blood cell count (WBC) between the two AML subtypes. This analysis revealed a significant discrepancy in WBC values (Z = 3.0552, p = 0.0022), with patients in Subtype-1 exhibiting markedly higher WBC counts than those in Subtype-2, as illustrated in Figure 2f.

Differentially expressed genes between subtype-1 and subtype-2

To understand the biological differences between the two transcriptomics Subtypes of acute myeloid leukaemia, we applied the Negative Binomial test [24] to determine the differentially expressed genes [25] between the two Subtypes. We found 310 mRNA transcripts that were significantly upregulated in Subtype-1 than in Subtype-2 AML, whereas 248 were significantly upregulated in Subtype-2 (Figure 3a and b; Supplementary file 1).

Altered signalling pathways and molecular processes distinguish disease Subtypes

Transcription factors and kinases are critical in the progression of cancer. In order to understand the genetic landscape and signalling networks of the two transcriptomics Subtypes of AML, we utilised the expression-2-kinase [26] software to extract the transcription factors and kinases associated with the two AML Subtypes.

Our findings showed considerable enrichment for transcription factors in the two Subtypes of AML. Several

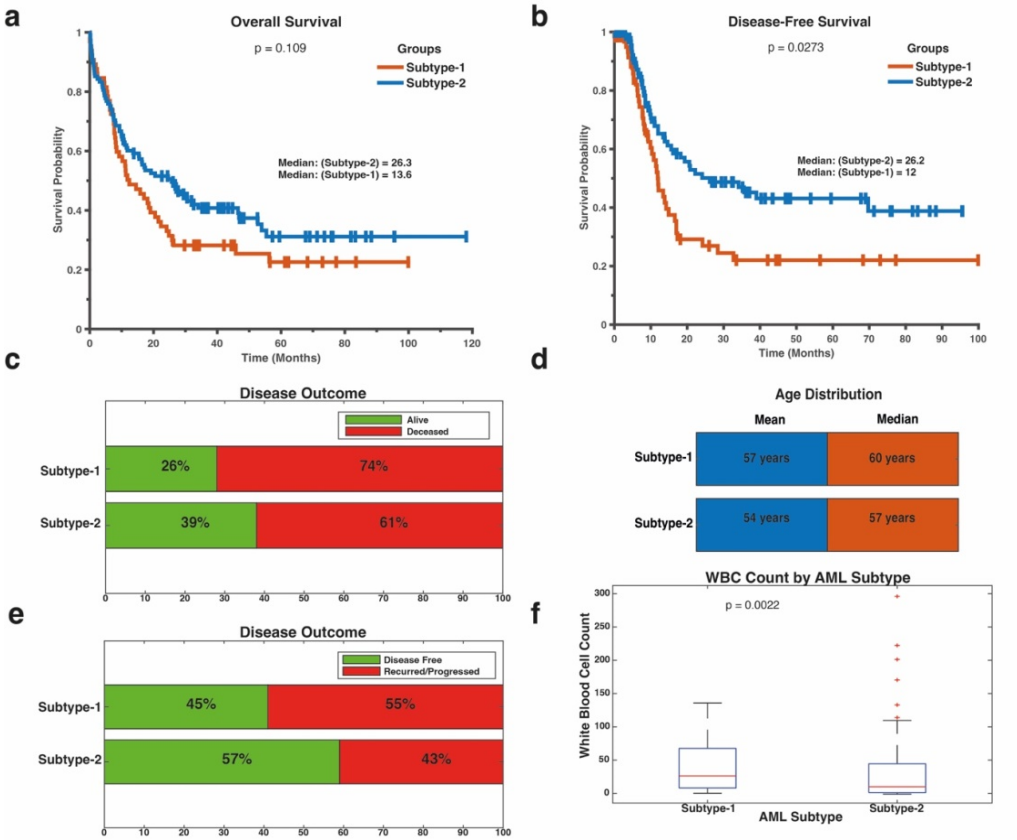

Figure 2: (a)Kaplan-Meier curve for overall survival months for patients with AML across the two AML transcriptomics subtypes; (b) Kaplan-Meier curve for Disease-Free Survival months for patients with AML across the two transcriptomics subtypes; (c) bar chart showing vital statistics after follow-up across the two AML subtypes; (d) bar chart showing the distribution of age of AML patients across the two cancer subtypes; (e) bar chart showing disease outcome after treatment across the two AML subtypes; (f) Box plot showing the median values of WBC counts for the two AML subtypes

Citation: Panji Nkhoma, Doris Kafita, Kevin Dzobo, Sinkala Musalula. Machine Learning Reveals Two Distinct Transcriptomics Subtypes of Acute Myeloid Leukaemia with Differences in Disease Outcomes and Genetic Landscape. Journal of Bioinformatics and Systems Biology. 7 (2024): 227-237.

Nkhoma P, et al., J Bioinform Syst Biol 2024
DOI:10.26502/jbsb.5107094

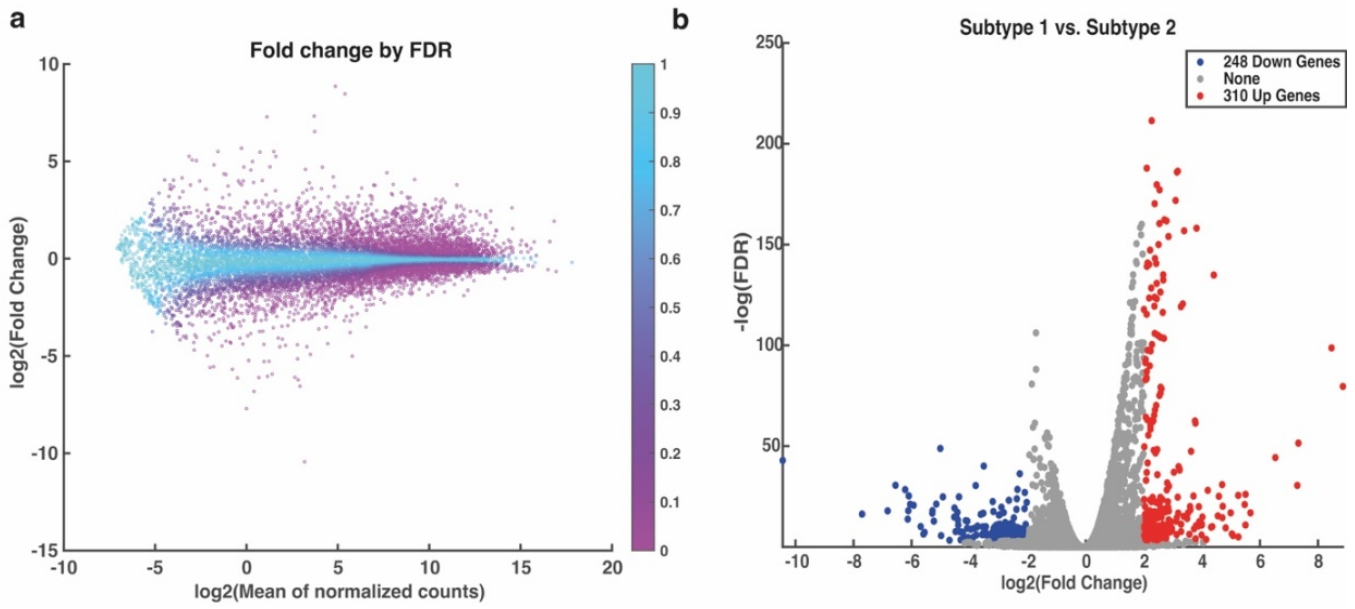

Figure 3: (a) Scatter plot of gene expressions and their significance showing fold change vs mean normalised counts in log scale. Data points are coloured according to adjusted p values; (b) Volcano plot showing 310 significantly upregulated genes and downregulated genes in subtype 1 AML

were shared, and others were unique to either Subtype (Figure 4a; Supplementary file 2). The shared transcription factors include RNF2 (subtype-1 p = 1.40 X 10-9 and subtype-2 p = 4.74 X 10-11), JARID2 (p = 1.82 X 10-10 and p = 1.63 X 10-22), TP53 (p = 4.90 X 10-4 and p = 1.90 X 10-10), EZH2 (p = 1.62 X 10-9 and p = 1.33 X 10-18), FOXA2 (p = 2.89 X 10-6 and p = 0.006) respectively among others.

In addition, our study indicates unique expression of VDR (p = 0.025) in subtype 1 AML. Increased VDR expression is associated with good prognosis in acute myeloid leukaemia [27, 28]. Transcription factors uniquely expressed in Subtype-2 AML include GLI1 (p = 0.008) and DMRT1 (0.04). High GLI1 expression has be shown to be an indicator of poor prognosis in acute leukaemia patients [29] and also reduces drug sensitivity by regulating the cell cycle in AML [30]. However, studies have shown that GLI1 is a strong target to treat AML patients and is also an excellent approach for developing novel therapies [31, 32].

Our kinase enrichment also showed an overlap for many kinases between the two transcriptomics Subtypes of AML. However, some kinases were still unique to each of the two Subtypes (Figure 4b; Supplementary file 3). The kinases overlapping between the two groups include; CSNK2A1 (subtype -1 p = 0.002 and subtype -2 p = 1.54 X 10-5), AKT1 (p = 0.002 and p = 0.001), TAF1 (p = 0.005 and p = 0.001), HIPK2 (p = 1.67 X 10-4 and p = 0.01), ABL1 (p = 0.01 and p = 0.02), and CDK2 (p = 0.013 and p = 0.001), respectively among others.

Kinases uniquely expressed in subtype 1 AML include MAPKAPK3 (p = 0.003), PKN2 (p = 0.03) and PTK6 (p = 0.04). Expression of high levels of PTK6 promotes tumour development and is associated with poor prognosis in breast cancer [33, 34]. However, among the kinases exclusively expressed in subtype 2 AML patients, only STK17A (p = 0.03) was significantly expressed.

We further applied Gene Set Enrichment Analysis (GSEA) [35] to extract knowledge of the KEGG signalling pathways enriched in subtype-1 AML compared to subtype-2 AML. Among the signalling pathways more enriched or upregulated in subtype-1 AML are those involved in Aminoacyl-tRNA biosynthesis, apoptosis, NF-kappa B and HIF-1 signalling pathways with Normalised Enrichment Score p values of 0.001, 0.03, 0.03, and 0.03, respectively. NF-kB transcription factors are critical regulators of immunity, stress response, apoptosis, and differentiation. These transcription factors have recently become potential targets for cancer treatment [36]. However, the diversity in the defects that lead to abnormal NF-κB activation makes the possibility of finding a universal target difficult [37] (Figure 4c, 4d, 4e and 4f; Supplementary file 4).

The mutational landscape of acute myeloid leukaemia transcriptomics subtypes

With a focus only on the consensus cancer genes [38], we evaluated the extent to which copy number variations and gene mutations affect patients with AML. We found no statistical difference in the mutations and copy number variations between the two AML subtypes (Supplementary file 5). However, across the two transcriptomics subtypes, we found that the gene mutations were similar to those reported in other studies [39-41], mainly affecting FLT3

Nkhoma P, et al., J Bioinform Syst Biol 2024
DOI:10.26502/jbsb.5107094

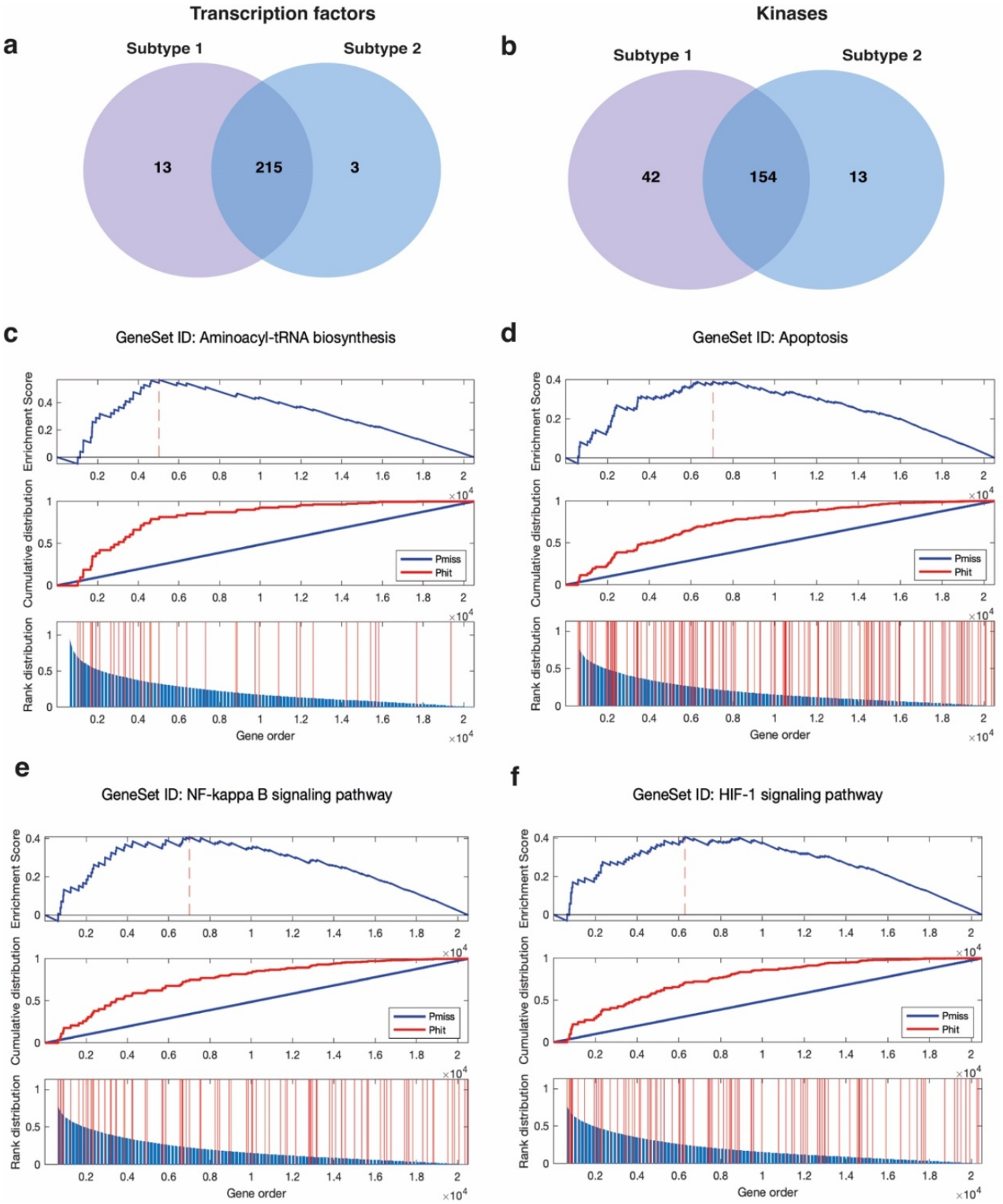

Figure 4: (a) Venn diagram showing overlap of 215 transcription factors across the two AML subtypes, with 13 expressed only in subtype-1 and 3 only in subtype-2; (b) Venn diagram showing overlap of 154 kinases across the two AML subtypes with 42 expressed only in subtype-1 and 13 only in subtype-2; GSEA plots showing significant enrichment for (C) Aminoacyl-tRNA biosynthesis pathway; (d) apoptosis signalling pathway; (e) NF-kappa B signalling pathway and (f) HIF-1 signalling pathway in subtype-1 AML

Citation: Panji Nkhoma, Doris Kafita, Kevin Dzobo, Sinkala Musalula. Machine Learning Reveals Two Distinct Transcriptomics Subtypes of Acute Myeloid Leukaemia with Differences in Disease Outcomes and Genetic Landscape. Journal of Bioinformatics and Systems Biology. 7 (2024): 227-237.

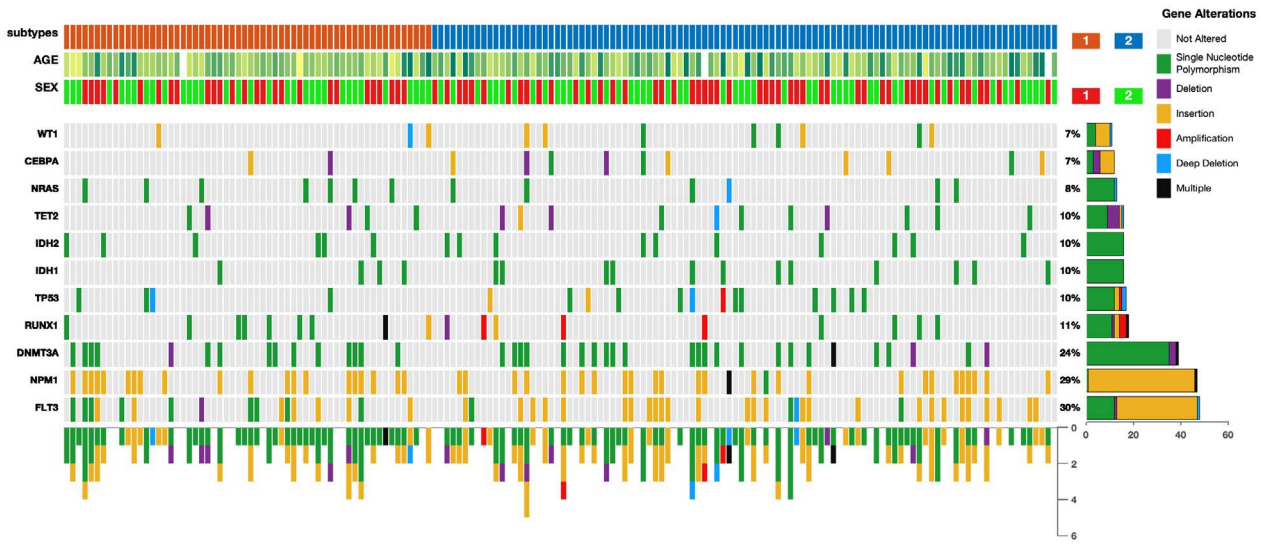

Figure 5: (a) Integrated plot of gene mutations, copy number alterations and clinical features of AML patients. From top to bottom panels indicate the transcriptomic subtypes of AML; the patients' age; patient's gender; mutation and copy number frequency across all AML patients

(31%), NPM1 (29%), DNMT3A (26%), IDH2, RUNX1, IDH1, TET2 and TP53 all at 10% (Figure 5). We further revealed that the frequent mutations in FLT3 and NPM1 were insertions, whereas single nucleotide polymorphisms (SNPs) were predominantly observed in DNMT3A, IDH2, RUNX1, IDH1, TET2 and TP53.

Methods

We downloaded the TCGA [22] dataset of 200 AML patients from cBioPortal (http://www.cbioportal.org) [42]. We only returned and analysed 173 AML patient samples with whole genome transcriptomics data. We further utilised DNA copy number alterations, mutation data, and comprehensively de-identified clinical and sample information.

Transcriptomics classification of acute myeloid leukaemia

We used the 1000 most variable genes for the K-means clustering machine learning method and for principal component analysis (PCA) to classify AML samples according to their gene expression levels. We used the Calinski-Harabasz clustering evaluation criterion [43] to determine the ideal number of subtypes, and the results indicated that 2 was the optimum number of subtypes (Figure 1). After that, to define the transcriptomics data and determine the leukaemia subtypes, we used unsupervised k means across 1000 iterations with random initialisation to minimize the likelihood of the algorithm converging to local minima. This was done using the squared Euclidian distance metric and then selected the clustering solution with the highest average Silhouette score [44]. Next, we used Principal Component Analysis [45, 46] to minimize the dimensionality of the transcriptomics measured data, which allowed us to visualise the clustering of the AML clusters. Lastly, we plotted the first two dimensions of the principal components with points coloured based on the K-means clustering group assignment. We utilized a dataset of pre-processed and normalized read count mRNA expression data measured in Fragments Per Kilobase of transcript per Million mapped reads (FPKM).

Survival analysis

We compared the overall survival and disease-free survival of subtype-1 and subtype-2 patients using the Kaplan-Meier technique [23]. We further applied descriptive statistics to determine percentages of patients who were alive, deceased, disease-free, and those with disease progression or recurrence at the end of follow-up for each of the two AML subtypes.

Identification of the differentially expressed genes

We examined mRNA expression data to determine which genes were expressed differently in transcriptomics subtype-1 and subtype-2 AML patients. Using the Negative Binomial Model [24], we performed statistical analysis on the mRNA transcripts of both groups. The False Discovery Rate (FDR) [47] approach was utilised to correct the p-values acquired from the analysis. When mRNA expression showed an adjusted p-value of less than 0.05 and a fold-change of more than 4, it was deemed statistically significantly differentiable between subtype-1 and subtype-2 AML.

Functional enrichment analyses

We used the Expression 2 Kinase (X2K) software [48] to infer transcription factors and kinases from the lists of differentially expressed genes for the two disease subtypes. We independently ran the lists of significantly upregulated

genes in subtype-1 and subtype-2 AML through the X2K software and extracted statistically significant transcription factors and kinases from the .csv file outputs.

We then downloaded the Kyoto Encyclopaedia of Genes and Genomes (KEGG) 2019 human database and Gene Ontology (GO) molecular function 2021. Then, for each gene set within each database, we modified the gene sets by returning only the genes present in our gene expression dataset, thus limiting the gene background to genes only present in the mRNA expression data. Finally, we used Gene set enrichment analysis (GSEA) to determine the KEGG pathways that are enriched for in subtype-1 AML compared to subtype-2 AML

We used mutation data (single nucleotide polymorphisms and indels) and copy number alteration data to assess the extent of genetic changes in AML subtypes. First, we combined these two genetic modification data. We then used information from the Sanger Consensus Cancer Gene Database [38] to return only genes associated with human cancer. Additionally, oncogenes and tumour suppressor genes in the gene alteration dataset were annotated using information from UniProt Knowledgebase, TSGene database, and ONGene database [38, 49, 50]. Finally, using the chi-square test, We compared genetic changes between disease subtypes.

Statistical analysis

All analyses described herein were conducted using MATLAB version 2023a. Fisher's exact test was employed to discern associations among categorical variables, while the Welch and Wilcoxon rank sum tests were utilised to evaluate differences in continuous variables across AML subtypes within various categories. Significance was established when the p-value was less than 0.05 for individual tests and when the Benjamini-Hochberg adjusted p-value was below 0.05 for multiple comparisons.

Discussion

Our comprehensive analysis of gene expression, clinical data, mutations, and copy number alterations in AML leveraged machine learning to discern two distinct AML subtypes: subtype-1 and subtype-2. This classification underscores the potential for personalized treatment strategies, reflecting the heterogeneity within AML and highlighting the importance of integrating diverse data types for a holistic understanding of the disease's molecular underpinnings [51, 52]. The demographic characteristics of the patients in the two disease subtypes were relatively similar, showing that we were dealing with adult-type acute myeloid leukaemia.

Our findings indicate that while overall survival rates were similar across both AML subtypes, subtype-2 patients experienced significantly better disease-free survival than those in subtype-1. This observation suggests that subtype-2 patients, post-treatment, had a prolonged period without disease relapse or progression compared to subtype-1, which also exhibited higher white blood cell counts found to be associated with poor prognosis in AML [53, 54], hinting at a more aggressive disease form potentially driven by an adverse genetic landscape. This differentiation in survival outcomes emphasises the critical need for subtype-specific therapeutic strategies in AML management. Some of the genetic alterations may include significant aberrations in the genes involved in signalling pathways associated with poor survival, prognosis and chemoresistance, such as the Aminoacyl-tRNA biosynthesis, NF-kappa B and the P53 [55-58] signalling pathways, which we found to be more enhanced in subtype-1 AML after carrying out GSEA. Activation of the NF-kB pathway has been reported to regulate the transcription of target genes that promote cell survival and proliferation, inhibit apoptosis, and mediate invasion and metastasis [59]. Constitutive activation of NF-kB in AML has been associated with enhanced proliferation and survival of cancer cells [60], likely contributing to the notably higher white blood cell (WBC) count observed in subtype-1 compared to subtype-2 AML.

Hyperleukocytosis in AML is linked to increased mortality and a higher incidence of severe complications, including leukostasis, disseminated intravascular coagulation, and tumour lysis syndrome, compared to AML patients without hyperleukocytosis [61]. Furthermore, hyperleukocytosis is associated with a significantly lower event-free survival and complete remission rate in AML patients [62].

Conclusion

Our investigation has effectively delineated AML patients into two distinct subtypes based on gene expression profiles, revealing notable differences in clinical characteristics and genetic backgrounds. Subtype-1 patients exhibited markedly reduced disease-free survival and were characterised by a more aggressive disease course underpinned by a detrimental genetic landscape. Notably, genes overexpressed in Subtype-1 are implicated in signalling pathways linked to adverse survival outcomes, poor prognosis, and chemoresistance, including Aminoacyl-tRNA biosynthesis, P53, and NF-kB, alongside the occurrence of leucocytosis. These insights underscore the viability of developing subtype-specific precision treatments for AML, promising more targeted and effective therapeutic approaches.

Ethics Approval

The protocol for this study was approved by the University of Zambia Biomedical Research Ethics Committee IRB00001131 of IORG0000774, approval number 3062-2022. The publicly available datasets were collected by the cBioPortal and TCGA projects and made available through their respective databases. All methods used were performed

within the stipulated guidelines provided by the cBioPortal and TCGA projects.

Data Availability

The data that support our results are available from the following repositories: cBioPortal; https://www.cbioportal.org/

Conflicts of Interest

Authors declare no conflicts of interest.

Funding

The authors did not receive any funding to conduct the research

Author Contribution

Conceptualisation: Panji Nkhoma, Sinkala Musalula, Kevin Dzobo, Doris Kafita

Formal Analysis: Panji Nkhoma, Sinkala Musalula

Methodology: Panji Nkhoma, Sinkala Musalula, Kevin Dzobo, Doris Kafita

Visualisation: Panji Nkhoma, Sinkala Musalula

Writing – original draft: Panji Nkhoma, Sinkala Musalula

Manuscript – review & editing: Panji Nkhoma, Sinkala Musalula, Kevin Dzobo, Doris Kafita

References

1. What is acute myeloid leukaemia (AML)? https://www.cancerresearchuk.org/about-cancer/acute-myeloid-leukaemia-aml/about-acute-myeloid-leukaemia.
2. De Kouchkovsky I, Abdul-Hay M. Acute myeloid leukemia: a comprehensive review and 2016 update. Blood Cancer J 6 (2016): e441-e441.
3. Cancer Stat Facts: Leukemia — Acute Myeloid Leukemia (AML) https://seer.cancer.gov/statfacts/html/amyl.html
4. Bullinger L, Döhner K, Döhner H. Genomics of acute myeloid leukemia diagnosis and pathways. J of Clin Oncol 35 (2017): 934-946.
5. Döhner H, Weisdorf DJ, Bloomfield CD. Acute myeloid leukemia. New England J Med 373 (2015): 1136-1152.
6. Lilljebjörn H, Orsmark-Pietras C, Mitelman F, et al. Transcriptomics paving the way for improved diagnostics and precision medicine of acute leukemia. Seminars in Cancer Biol 84 (2022): 40-49.
7. Nkhoma P, Dzobo K, Kafita D, et al. Racial Disparities in the Genetic Landscape of Acute Myeloid Leukaemia from The Cancer Genome Atlas: Insights from a Bioinformatics Analysis. bioRxiv (2023).
8. Banerjee A, Chen S, Fatemifar G, et al. Machine learning for subtype definition and risk prediction in heart failure, acute coronary syndromes and atrial fibrillation: systematic review of validity and clinical utility. BMC Med 19 (2021): 85.
9. Khaliq AM, Erdogan C, Kurt Z, et al. Refining colorectal cancer classification and clinical stratification through a single-cell atlas. Genome Biol 23 (2022): 113.
10. Søkilde R, Persson H, Ehinger A, et al. Refinement of breast cancer molecular classification by miRNA expression profiles. BMC Genomics 20 (2012): 1-12.
11. Sinkala M, Naran K, Ramamurthy D, et al. Machine learning and bioinformatic analyses link the cell surface receptor transcript levels to the drug response of breast cancer cells and drug off-target effects. PLoS One 19 (2021): e0296511.
12. Sinkala M, Mulder N, Martin D. Machine Learning and Network Analyses Reveal Disease Subtypes of Pancreatic Cancer and their Molecular Characteristics. Sci Rep 10 (2020): 1212.
13. Ye H, Sowalsky AG. Molecular correlates of intermediate- and high-risk localized prostate cancer. Urol Oncol 36 (2018): 368-374.
14. Singh MP, Rai S, Gupta SK, et al. Unsupervised machine learning-based clustering identifies unique molecular signatures of colorectal cancer with distinct clinical outcomes. Genes & Diseases 10 (2023): 2270-2273.
15. Eckardt J-N, Röllig C, Metzeler K, et al. Unsupervised meta-clustering identifies risk clusters in acute myeloid leukemia based on clinical and genetic profiles. Commun Med 3 (2023): 68.
16. Xu H, Mohamed M, Flannery M, et al. An Unsupervised Machine Learning Approach to Evaluating the Association of Symptom Clusters With Adverse Outcomes Among Older Adults With Advanced Cancer: A Secondary Analysis of a Randomized Clinical Trial. JAMA Network Open 6 (2023): e234198-e234198.
17. Štancl P, Karlic R. Machine learning for pan-cancer classification based on RNA sequencing data. Front Mol Biosci 10 (2023): 1285795.
18. Ke J, Shen Y, Lu Y, et al. Mine local homogeneous representation by interaction information clustering with unsupervised learning in histopathology images. Comput Methods and Programs Biomed 235 (2023): 107520.
19. Cascianelli S, Molineris I, Isella C, et al. Machine learning for RNA sequencing-based intrinsic subtyping of breast cancer. Sci Rep 10(2020): 14071.
20. List M, Hauschild A-C, Tan Q, et al. Classification of breast cancer subtypes by combining gene expression and

DNA methylation data. J Integrative Bioinformatics 11 (2014): 1-14.

21. Sinkala M, Mulder N, Martin D. Machine Learning and Network Analyses Reveal Disease Subtypes of Pancreatic Cancer and their Molecular Characteristics. Sci Rep 10 (2022): 1212.

22. Weinstein JN, Collisson EA, Mills GB, et al. The cancer genome atlas pan-cancer analysis project. Nat Genetics 45 (2013): 1113-1120.

23. Goel MK, Khanna P, Kishore J. Understanding survival analysis: Kaplan-Meier estimate. Int J Ayurveda Res 1 (2014): 274.

24. Anders S, Huber W. Differential expression analysis for sequence count data. Genome Biol 11 (2010): R106.

25. Anders S, Huber W. Differential expression analysis for sequence count data. Nat Preced (2010): 1-1.

26. Clarke DJB, Kuleshov MV, Schilder BM, et al. eXpression2Kinases (X2K) Web: linking expression signatures to upstream cell signaling networks. Nucleic Acid Res 46 (2018): W171-W179.

27. Paubelle E, Zylbersztejn F, Maciel TT, et al. Vitamin D Receptor Controls Cell Stemness in Acute Myeloid Leukemia and in Normal Bone Marrow. Cell Reports 30 (2020): 739-754.e734.

28. Pezeshki SMS, Asnafi AA, Khosravi A, et al. Vitamin D and its receptor polymorphisms: New possible prognostic biomarkers in leukemias. Oncol Rev 12 (2018): 366.

29. El Zaiat RS, Nabil R, Khalifa KA, et al. High GLI-1 Expression is a Reliable Indicator of Bad Prognosis in Newly Diagnosed Acute Leukemia Patients. Indian J Hematol Blood Transfusion 39 (2023): 376-382.

30. Zhou C, Du J, Zhao L, et al. GLI1 reduces drug sensitivity by regulating cell cycle through PI3K/AKT/GSK3/CDK pathway in acute myeloid leukemia. Cell Death Dis 12 (2021): 231.

31. Hasipek M, Al-Harbi S, Phillips JG, et al. Therapeutic Targeting of GLI1 in Acute Myeloid Leukemia. Blood 130 (2017): 2654-2654.

32. Tesanovic S, Krenn PW, Aberger F. Hedgehog/GLI signaling in hematopoietic development and acute myeloid leukemia—From bench to bedside. Front Cell Developmental Biol 10 (2022): 944760.

33. Chen Y, Qu W, Tu J, et al. Prognostic impact of PTK6 expression in triple negative breast cancer. BMC Women's Health 23 (2023): 575.

34. Wang L, Luo S, Wang Z, et al. Comprehensive Analysis Reveals PTK6 as a Prognostic Biomarker Involved in the Immunosuppressive Microenvironment in Breast Cancer. J Immunol Res 2022 (2022): 5160705.

35. Subramanian A, Tamayo P, Mootha VK, et al. Gene set enrichment analysis: a knowledge-based approach for interpreting genome-wide expression profiles. Proceed Nat Acad Sci 102 (2005): 15545-15550.

36. Burley TA, Kennedy E, Broad G, et al. Targeting the Non-Canonical NF-κB Pathway in Chronic Lymphocytic Leukemia and Multiple Myeloma. Cancers 14 (2022):1489.

37. Imbert V, Peyron JF. NF-κB in Hematological Malignancies. Biomedicines 5 (2017).

38. Forbes SA, Beare D, Gunasekaran P, et al. COSMIC: exploring the world's knowledge of somatic mutations in human cancer. Nucleic Acid Res 43 (2015): D805-D811.

39. Medinger M, Passweg JR. Acute myeloid leukaemia genomics. British J Haematol 179 (2017): 530-542.

40. Naoe T, Kiyoi H. Gene mutations of acute myeloid leukemia in the genome era. Int J Hematol 97 (2013): 165-174.

41. Network CGAR. Genomic and epigenomic landscapes of adult de novo acute myeloid leukemia. New England J Med 368 (2013): 2059-2074.

42. Cerami E, Gao J, Dogrusoz U, et al. The cBio cancer genomics portal: an open platform for exploring multidimensional cancer genomics data. Cancer Discovery 2 (2012): 401-404.

43. Caliński T, Harabasz J. A dendrite method for cluster analysis. Commun Stat Simul Comput 3 (1974): 1-27.

44. Wu Y, Ianakiev K, Govindaraju V. Improved k-nearest neighbor classification. Pattern recognition 35 (2002): 2311-2318.

45. Abdi H, Williams LJ. Principal component analysis. WIREs Computational Statistics 2 (2010): 433-459.

46. Jolliffe I. Principal Component Analysis. In: International Encyclopedia of Statistical Science. Edited by Lovric M. Berlin, Heidelberg: Springer Berlin Heidelberg; 2011: 1094-1096.

47. Jung K, Friede T, Beißbarth T. Reporting FDR analogous confidence intervals for the log fold change of differentially expressed genes. BMC Bioinformatics 12 (2011): 1-9.

48. Clarke DJB, Kuleshov MV, Schilder BM, et al. eXpression2Kinases (X2K) Web: linking expression signatures to upstream cell signaling networks. Nucleic Acid Res 46 (2018): W171-W179.

49. UniProt. the universal protein knowledgebase. Nucleic Acid Res 45 (2017): D158-D169.

50. Zhao M, Kim P, Mitra R, et al. TSGene 2.0: an updated literature-based knowledgebase for tumor suppressor genes. Nucleic Acid Res 44 (2016): D1023-1031.

51. Dozzo A, Galvin A, Shin JW, et al. Modelling acute myeloid leukemia (AML): What's new? A transition from the classical to the modern. Drug Deliv Transl Res 13 (2023): 2110-2141.

52. Bhat GR, Sethi I, Sadida HQ, et al. Cancer cell plasticity: from cellular, molecular, and genetic mechanisms to tumor heterogeneity and drug resistance. Cancer Metastasis Rev (2024): 1-32.

53. Creutzig U, Rössig C, Dworzak M, et al. Exchange Transfusion and Leukapheresis in Pediatric Patients with AML With High Risk of Early Death by Bleeding and Leukostasis. Pediatr Blood Cancer 63 (2016): 640-645.

54. Kuo KH, Callum JL, Panzarella T, et al. A retrospective observational study of leucoreductive strategies to manage patients with acute myeloid leukaemia presenting with hyperleucocytosis. Br J Haematol 168 (2015): 384-394.

55. Shin DY. TP53 Mutation in Acute Myeloid Leukemia: An Old Foe Revisited. Cancers (Basel) 15 (2023).

56. Wang H, Guo M, Wei H, et al. Targeting p53 pathways: mechanisms, structures, and advances in therapy. Signal Transduct Target Ther 8 (2023): 92.

57. Hernández Borrero LJ, El-Deiry WS. Tumor suppressor p53: Biology, signaling pathways, and therapeutic targeting. Biochimica et Biophysica Acta (BBA) - Rev Cancer 1876 (2021): 188556.

58. Gao X, Guo R, Li Y, et al. Contribution of upregulated aminoacyl-tRNA biosynthesis to metabolic dysregulation in gastric cancer. J Gastroenterol Hepatol 36 (2021): 3113-3126.

59. Rinkenbaugh AL, Baldwin AS. The NF-κB Pathway and Cancer Stem Cells. Cells 5 (2016): 16.

60. Bosman MCJ, Schuringa JJ, Vellenga E. Constitutive NF-κB activation in AML: Causes and treatment strategies. Critical Rev Oncol 98 (2016): 35-44.

61. Bewersdorf JP, Zeidan AM. Hyperleukocytosis and Leukostasis in Acute Myeloid Leukemia: Can a Better Understanding of the Underlying Molecular Pathophysiology Lead to Novel Treatments? Cells 9 (2020).

62. de Jonge HJ, Valk PJ, de Bont ES, et al. Prognostic impact of white blood cell count in intermediate risk acute myeloid leukemia: relevance of mutated NPM1 and FLT3-ITD. Haematologica 96 (2011): 1310-1317.