

Analysis of the Differential Expression and Prognostic Relationship of DEGs in AML based on TCGA Database

Hang Zhang

Nantong University Medical School

Qian Li

Affiliated Kunshan Hospital of Jiangsu University

Lan Li

Guizhou Province People's Hospital

Yue Gao

Jiangsu Province Center for Disease Control and Prevention

Rong Rong

The First Hospital of China Medical University: The First Affiliated Hospital of China Medical University

Kai Sun

The first affiliated hospital of nanjing Medical university

Zhengmin Yu

Jiangsu Province Center for Disease Control and Prevention

Xinyu Ji

Jiangsu Province Center for Disease Control and Prevention

Xiaowen Liu

Jiangsu Province Center for Disease Control and Prevention

Lei Han

Jiangsu Province Center for Disease Control and Prevention

Hengdong Zhang

Jiangsu Province Center for Disease Control and Prevention

Baoli Zhu

Nanjing Medical University

Tingting Yu

Nanjing Medical University

MING XU (✉ sosolou@jscdc.cn)

Jiangsu Province Center for Disease Control and Prevention <https://orcid.org/0000-0001-8846-2880>

Primary research

Keywords: Acute myeloid leukemia, TCGA, DEGs, Survival Analysis

Posted Date: December 4th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-120259/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License. [Read Full License](#)

Abstract

Background: Acute myeloid leukemia (AML) is a common and lethal hematological malignant hyperplastic disease originating from hematopoietic stem cells. The purpose of this study is to obtain the key AML survival-related differentially expressed gene.

Methods: RNA sequencing (RNA-Seq) data and clinical information of patients were downloaded from the TCGA-LAML database. We focused the intersectional genes of survival-related DEGs, cytogenetics risk related-DEGs, and the top 10 pathways of both up- and down-regulated of Normalization Enrichment Score (NES). Multivariate Cox regression analyses were performed to analyze the independent factors for AML. The Kaplan–Meier and Nomogram analyses were plotted to predict and compare survival of AML patients. The validation of DEGs were performed by a clinical follow-up investigation.

Results: 151 RNA-Seq samples for 60,488 genes and 200 clinical samples were selected from the TCGA-database. After filtering with the conditions about survival-related DEGs, cytogenetics-risk associated, and predicted in top 10 pathways, IGHM was the only remaining gene. Cox analysis showed that IGHM expression and age displayed were independent factors to the patients' survival ($P < 0.05$). Higher IGHM expression was identified in poor survival group ($P < 0.05$), with 68%, 43%, and 30% in 1, 3, and 5-year survival investigation. GSEA analysis revealed that IGHM were mainly enriched in the immune response. In Chinese population, IGHM displayed its significance not in the childhood AML patients but in adult ones.

Conclusion: High expression of IGHM gene is an independent risky factor for the survival of patients with AML, which can be an important molecular marker for AML prognosis.

Introduction

Leukemia is a malignant disease of the hematopoietic system. Based on disease progression and cell types, leukemia can be classified as acute lymphocytic leukemia (ALL), chronic lymphocytic leukemia (CLL), acute myelogenous leukemia (AML) and chronic myelogenous leukemia (CML)[1]. AML is a group of malignant clonal diseases originating from bone marrow hematopoietic stem cells, characterized by abnormal differentiation of hematopoietic stem cells and excessive proliferation of myeloid progenitor cells[2]. The abnormal accumulation of leukemia cells leads to the inhibition of normal hematopoietic cell growth and the infiltration of leukemia cells in the bone marrow, resulting in multiple organ failures and such symptoms as anemia, bleeding and infection[3]. The onset of AML can cover all periods of lifetime, but it mostly occurs in the elder-age and male patients, the incidence of which usually increases following the age, with a median of 67 years at diagnosis[4, 5]. According to the previous study, approximate 80% of AML occurs in adults [6].

At present, with the development of new diagnosis and treatment technologies, the survival rate of AML patients has significantly improved, but the long-term survival rate of patients still remains poor. For patients < 60 years old, according to the related previous studies, the 5-year overall survival (OS) rate is less than 40%; for the majority of patients with AML (aged over 60 years old), the 5-year OS rate is only 10–20% [7, 8]. Meanwhile, AML patients mainly have accompanied with poor prognosis. Even though most patients achieve a complete remission (CR) with intensive induction chemotherapy, over half of young adult patients and about 90% of elderly patients still succumb to the disease [9]. The culprits of above phenomenon could be regarded as the limited current knowledge of the underlying molecular mechanisms and its progression of AML, and the low-effective early clinical diagnosis. Some studies show that patients with different subtypes, different karyotypes, different gene expressions and mutation types have different prognosis[10]. Therefore, finding out effective target genes and prognostic molecules is of great significance to the early diagnosis and prognosis of patients with AML.

With the developments of gene detecting and computing sciences, the high-throughput sequencing technologies and bioinformatics analyses have been widely used in clinical research to screen meaningful oncogene and epigenetic changes. It is helpful to identify the differentially expressed genes (DEGs), functional pathways involved in AML carcinogenesis, and the biomarkers for diagnosis or prognosis. In order to conduct a comprehensive study of the human cancer genome, the United States initiated the establishment of The Cancer Genome Atlas (TCGA) database in 2006. The TCGA is a human cancer gene information database established and collaborated by the National Cancer Institute (NCI) and the National Human Genome Research Institute (NHGRI), owning comprehensive and well-curated genomic data of over 11,000 tumors across 33 major cancer types. This database mainly include the repository containing genomic, transcriptomic, epigenetic, proteomics and clinical information of various kinds of cancer[11].

We combined the RNA sequencing (RNA-Seq) data and clinical information of AML patients downloaded from TCGA database to perform our present study. By screen and analyzing the DEGs related to prognosis of AML patients, we tended to identify some novel AML-related genes. For further determination of the association between their aberrant expression levels and prognosis, the nomogram prognostic model was constructed to predict the long-term survival rate of AML patients. Moreover, we enriched DEGs-related signaling pathways in different cytogenetic groups using gene ontology (GO) enrichment analysis. Finally, 87 Chinese adult AML patients and 43 children AML patients were recruited to validate the clinical value for diagnosis of our findings.

Materials And Methods

Ethic principle

The present study was approved by the ethics committee of Jiangsu Provincial Center for Disease Control and Prevention (Approval number: 2012025), according to the principles of Helsinki Declaration.

Data preprocessing

The RNA-Seq data and clinical information of TCGA-LAML were downloaded from the TCGA database (<https://portal.gdc.cancer.gov/>), and the raw data included 151 RNA-Seq samples for 60,488 genes and 200 clinical samples. Firstly, we filtered the genes with low expression (counts per million (CPM)<1 in all samples), which reduced our RNA-Seq data to 56,822 genes. After excluding Asian race (2 cases), missing clinical information (2 cases), missing of Cytogenetics risk information (3 cases), 193 clinical samples were left. For matching the information from different databases, only 146 samples were selected because of the explicit intersection of RNA-Seq and Clinical data. Considering 9 patients were excluded because of the lack of follow-up information, we selected only 137 samples for the further investigation. On the other hand of genes filtration, RNA-Seq data was reduced to 37,362 genes after removing genes with null expression in over 50% samples. We also excluded the genes with their high or low z-score covering 95% samples, and the RNA-Seq data decreased from 37,297 to 16,226 genes. Finally, we merged RNA-Seq data with clinical information, then obtained 137 samples with 16,226 genes expression profiles for final analysis.

Screen differentially expressed genes

Cox proportional hazards model was used to evaluate the effects of gene expression level on the survival time of the patients with AML (P -value < 0.05, exp (coef) >1). The DESeq2 R-package (v3.5.1) was introduced to perform differential gene expression analysis between the low-risk and high-risk cytogenetics groups. The enrolling conditions for the differentially expressed gene were \log_2 Fold Change > 1 and adjusted P -value < 0.05. The Gene Set Enrichment Analysis (GSEA) of gene functional was performed by using ClusterProfiler R-package, and the meaningfully enriched pathways were screened with abs (NES) ≥ 1 (abs refers to the absolute value), NOM p-value ≤ 0.05 , and FDR q-value ≤ 0.25 as threshold. Venn diagrams were performed by using the limma R-package to determine the intersection of survival related DEGs, cytogenetics risk related DEGs, and the genes in the top 10 pathways of up-regulation. By the result of Venn diagram, we further screened out the key genes of AML as our target for the present investigation.

Survival analysis of DEGs

Multivariate Cox regression analyses were performed to investigate the independent factors that affect patient prognosis. Proportional hazards assumption was evaluated to test Cox regression model by Schoenfeld residuals, and to analyze the impact of clinical characteristics and IGHM gene expression on the survival time of AML patients. The Kaplan–Meier survival curve and Nomogram were plotted to compare and predict the influence of IGHM on the 1-year, 3-year, and 5-year survival rates of AML patients. The concordance index (C-index) and calibration curve were used to evaluate the prediction performance of nomogram. All statistical tests were two-sided and $P < 0.05$ was considered to indicate a statistically significant difference.

Samples collection and validation

Total 24 childhood AML patients and 64 adult AML patients without history of blood-related diseases were recruited from the first affiliated hospital with Nanjing Medical University, Children's Hospital of Nanjing Medical University, Guizhou Provincial People' Hospital, and Affiliated Kunshan Hospital of Jiangsu University during Jan 2012 to Dec 2016. The corresponding demographic and clinical outcomes were recorded. All participants were genetically unrelated Han Chinese. The participants or their legal guardians were signed a written informed consent after understanding the content of entire research and potential side effects. 2 ml marrow samples and the corresponding clinical information of AML patients were enrolled for further analyses.

The RNA was extracted by Qiagen RNeasy Plus Universal Kit (Qiagen Dusseldorf, Germany), following the manufacturer's protocol. And A260/280 of the purified RNA were tested by the Nanodrop OneC Ultramicro ultraviolet spectrophotometer (ThermoFisher Scientific, Waltham, MA, USA) to avoid the DNA contamination. The RNA expression of IGHM was performed by ABI7900HT real-time PCR System (Applied Biosystems, Foster City, CA, USA).

Results

Samples clinical characteristics

In the present study, a total of 137 AML patients were included in the analysis, including 77 males and 60 females, 126 whites and 11 blacks, with mean age of diagnosis was 54.44 ± 15.85 years. Using current SWOG criteria for cytogenetics risk category, the sample can be divided roughly into three categories: 28 patient had favorable cytogenetics, 74 had intermediate-risk/normal cytogenetics, and 35 had poor cytogenetics. 48 cases survival and 89 cases death. The clinical characteristics of included AML patients are shown in Table 1.

Table 1
Gene Set Enrichment Analysis (GSEA) of IGHM

| pathway | ES | NES | NOM p-value | FDR q-value | size |
|---|----------|----------|-------------|-------------|------|
| GO_HUMORAL_IMMUNE_RESPONSE_MEDIATED_BY_CIRCULATING_IMMUNOGLOBULIN | 0.749178 | 2.288786 | 0.000106 | 0.009878 | 138 |
| GO_COMPLEMENT_ACTIVATION | 0.712199 | 2.196018 | 0.000105 | 0.009878 | 157 |
| GO_PHAGOCYTOSIS_RECOGNITION | 0.7644 | 2.194633 | 0.000113 | 0.009938 | 72 |
| GO_B_CELL_MEDIATED_IMMUNITY | 0.669893 | 2.099131 | 0.000103 | 0.009878 | 206 |
| GO_B_CELL_RECEPTOR_SIGNALING_PATHWAY | 0.657912 | 1.977392 | 0.000108 | 0.009878 | 112 |
| GO_MEMBRANE_INVAGINATION | 0.644349 | 1.951241 | 0.000107 | 0.009878 | 123 |
| GO_POSITIVE_REGULATION_OF_B_CELL_ACTIVATION | 0.637783 | 1.942857 | 0.000106 | 0.009878 | 131 |

Identification and analysis of Survival- and cytogenetics- dual related DEGs

For survival-related DEG analysis, the data was adjusted for gender (male, female (reference group)), age and race (white, Black or African American (reference group)). After screening by Cox proportional hazard regression analyses (P -value < 0.05, exp (coef) > 1), and 1022 survival related DEGs were obtained. 17 genes were lost by conversion of gene names, and finally 1005 survival-related DEGs were obtained.

Considering the cytogenetics risk as an imperative clinical factor in AML, we further re-filtered the survival-related DEGs in different cytogenetics risk in order to find out the target genes which also associated with AML cytogenetics risk. The cytogenetics risk category usually can be divided into two groups: favorable intermediate/normal risk (104 cases) and poor risk (33 cases). Genes were considered up-regulated (down-regulated) if \log_2 Fold Change in expression was higher (or lower) than 1 (abscissa), and adjusted P -value < 0.05 (ordinate), a total of 2291 cytogenetics risk category related DEGs were identified using DESeq2 (as shown in Fig. 1).

Because the large number of survival-related DEGs might cost a huge calculation power to analysis, we performed the functional enrichment analysis to enrich the genes in DESeq2 with high relevancy biological functions. A total of 790 meaningful pathways of cytogenetics risk category were listed by GSEA. The top 10 pathways of up-regulated and down-regulated gene sets of NES were marked and listed in Fig. 2.

Finally, a Venn diagram was drawn to reflect the intersection of survival related DEGs (Overall Survival), cytogenetics risk related DEGs (CR DESeq2), and functional enrichment analysis (Fig. 3). A total of 42 intersecting DEGs were identified in the two sets, and there is only one gene in the CR GSEA top10 pathways of up-regulated among the 42 intersecting DEGs, namely IGHM.

Survival Analysis Of Ighm

Multivariate Cox regression analysis was used to analyze the influence of clinical characteristics and IGHM gene expression on the survival time of AML patients. Age, gender, and race were included as covariates, the survival rate of patients with different IGHM amounts of expression was analyzed by Cox proportional hazards model. The proportional hazards (PH) assumption of cox regress model reveal its $P = 0.21$ (Fig. 4), and the Wald test value of the overall Cox regression model is 31.54, and its P -value < 0.001, suggesting the availability of Cox regression in our present study. The average survival time of patients based on all factors was 19 months (95%CI: 12 ~ 27 months), which is about 1.5 years. The results of multivariate analysis showed that the two factors, IGHM expression and age, have their statistical significance with the survival time of AML patients, all P -values were < 0.05 (Table 2).

Table 2
Multivariate Cox regression analysis of factors for AML patients survival

| Factor | coef | exp(coef) [95%CI] | SE(coef) | Z | P-value |
|--------|-------|-------------------|----------|-------|----------|
| IGHM | 0.73 | 2.07 [1.03, 4.15] | 0.36 | 2.04 | 0.041 |
| Age | 0.04 | 1.04 [1.02, 1.06] | 0.01 | 4.74 | 2.17E-06 |
| Sex | -0.15 | 0.86 [0.55, 1.34] | 0.23 | -0.68 | 0.499 |
| Race | -0.09 | 0.91 [0.33, 2.55] | 0.52 | -0.17 | 0.863 |

After adjusting for other variables, the Kaplan-Meier survival curve for each factor in different conditions is demonstrated in Fig. 5. The survival rate of patients in the IGHM high expression group was statistically lower than that in low expression group ($P = 0.041$). The risk of death in IGHM high expression patients was 2.07 times higher than that in low expression ones ($P < 0.05$). The mean value 55 years of age was used as a cut-off point because of age is a continuous variable, these variables were converted into categorical variables by the cut function and the survival curve was

plotted. Kaplan-Meier survival analysis further showed that the survival rate of old group is lower than that of young group, and the difference was statistically significant ($P < 0.001$). The risk of death in the patients with old group having 3.00 times the risk compared to young patients ($P < 0.001$).

Establish And Evaluate Nomogram Prognostic Model

The nomogram, which is also called alignment diagram, the advantage of application in medical research is to estimate the survival of individual patient by incorporating multiple clinical variables and their interdependent relationships [12]. Its essence is the visualization of the built model, the closer the C-index of the nomogram is to 1, the better the accuracy of the model. In this study, the C-index of fitting the prediction model of Cox regression is 0.69, this result shows that the model has a good accuracy. The calibration curve is shown in Fig. 6, and the predicted calibration curve is closer to the standard curve, it also shows that the prediction ability of nomogram is better.

Nomogram results showed that the expression of IGHM in LAML patients has a great influence on their survival time. It is predicted that the 1-year, 3-year and 5-year survival of the patients with low expression group of IGHM are $> 85\%$, $> 70\%$ and $> 60\%$, and 68%, 43%, and 30% in the high expression group, respectively (in Fig. 7).

Analysis Of Ighm-related Signaling Pathways

After focusing IGHM as our target gene, we reviewed the results of GSEA analysis to evaluate the potential function of IGHM (in Table 3). Among the significant top 10 up-regulated pathways with statistical significance ($abs(NES) \geq 1$, $NOM\ p\text{-value} \leq 0.05$ and $FDR\ q\text{-value} \leq 0.25$), IGHM participated in 7 pathways, including humoral immune response by circulating immunoglobulin, complement activation, phagocytosis recognition, B cell mediated immunity, B cell receptor signaling pathway, membrane invagination, and positive regulation of B cell activation.

Identification of IGHM in marrow samples of Chinese AML patients

To identify the findings in our study, we tested the IGHM mRNA expression in marrow samples of AML patients in childhood- and adult- groups. As the Fig. 8 described, there was no difference between non end-point and end-point childhood patients, while the IGHM expressed significantly higher in end-point adult patients than the non end-point ones. These results suggested the IGHM might perform its impact specifically on adult AML survival.

Discussion

The present study analyzed the RNA-seq data and clinical information of 146 AML patients from TCGA database through bioinformatics analysis. In this study, we performed multivariate Cox regression analysis on prognostic factors, the findings suggested that age and IGHM expression are independent prognostic factors for patients with AML ($P < 0.05$), which means IGHM could be a key gene which is statistically related to the AML patients' survival. To our knowledge, it was the first study to confirm IGHM as a potential target gene to AML by using bioinformatics analysis, accompanied with further clinical sample verification.

Previous studies have shown that IGHM (Immunoglobulin Heavy Constant Mu) is a gene marker for the short transition period in the differentiation and development of myeloid and lymphoid progenitor cells [13, 14]. Normally, IGHM maintained in an inactive status, without antibody response. But in both B lymphocytic leukemia and myeloid leukemia, with the process of the cellular malignant transformation and abnormal clone amplification, the expression of IGHM was abnormally highly increased [15]. In our study, the high expression of IGHM was identified as an independent risky factor for AML patients' survival, which was highly consistent with the conclusion that high expressions of IGHM are more common in AML patients [15, 16]. For the function of IGHM in coding protein, the IGHM defines the IgM isotype in B cells; aberrant expression of IGHM is closely associated with the malfunction of mu-chain of immunoglobulin, including mutations and rearrangements. A recent research suggests that patients with autosomal recessive agammaglobulinemia have IGHM gene mutations [17]. Also, the mu-chain of immunoglobulin gene rearrangement was detected in myeloid leukemia cells from AML patients, and the survival rate of AML patients was significantly lower [13, 14, 16, 18, 19]. Neeraj et al. also found that gene rearrangement of IGHM in the diffuse large B cell lymphoma (DLBCL) [20]. Therefore, we believed the IGHM could be a functional and reasonable marker to evaluate the prognosis of AML.

Our present study also analyses differences between the expression of IGHM in different cytogenetic risk groups, the results show that the expression of IGHM is higher in patients with high-risk group. Unlike most other solid tumors, many hematological malignancies strongly associated with single characteristic cytogenetic abnormalities. A previous study had shown that patients, younger than 65 years treated with standard chemotherapy with a favorable karyotype, have CR rates in the 85–90% range and a 5-year OS of 50–60% [21]. Also, patients with intermediate-risk cytogenetics have CR rates of 65–75%, and a 5-year OS of 35–45%, while patients with poor cytogenetics have CR rates of 45–55% and a 5-year OS of only 10–20% [21]. Therefore, the problem of AML with unfavorable-risk cytogenetics deserves special attention. The GSEA of our study showed that DEGs mainly enriched in immune response (BP) regulated the occurrence and development of AML and influence the prognosis of AML, which was consistent with the contributes of immune response to tumor progression and drug resistance in various cancers [22, 23], including lung cancer [24], breast cancer [25] and bladder cancer [26]. We also found that IGHM genes were primarily enriched in the B cell mediated immunity and the B-cell receptor signaling pathway, these results indicate that the development and prognosis of AML may be related to these biological processes.

There are several limitations existing in our study. Firstly, this study concerns only the TCGA database, and only the RNA-seq data and clinical information, we hope that future research can combine multiple databases and include data on DNA sequencing, and methylation profiling, exome sequencing, miRNA expression, etc., in order to gain a more comprehensive understanding of AML pathogenesis. Secondly, the sample size for both TCGA database analysis and our validation in marrow were small, it was hard to avoid the sample selecting bias. Finally, we did not check the status of mu-chain of immunoglobulin in our AML sample to evaluate the relationship between IGHM high expression and the rearrangement, because of the rare bone marrow samples of AML patients. Further investigations are expected to reveal the mechanism of how high IGHM expression leads to poor prognosis of AML.

In conclusion, we confirmed that IGHM is an independent risky factor for the prognosis in AML patients. Besides, we also performed a nomogram model for predicting the long-term survival rate of AML patients, patients with high expression of IGHM showed lower survival, and the expression of IGHM is higher in patients with high-risk cytogenetic group. The further molecular biology experiments and clinical studies are needed to verify the possibility of IGHM as a prognostic molecular marker for AML.

Abbreviations

IGHM: immunoglobulin heavy constant mu; AML: acute myeloid leukemia; NES: Normalization Enrichment Score; DEGs: different expressed genes; OS: overall survival; TCGA: the Cancer Genome Atlas program; GO: gene ontology; RNA-Seq: RNA sequencing; GSEA: Gene Set Enrichment Analysis

Declarations

Ethics approval and consent to participate

The present study was approved by the ethics committee of Jiangsu Provincial Center for Disease Control and Prevention (Approval number: 2012025), according to the principles of Helsinki Declaration.

Consent for publication

Written consent for publication was obtained from all patients involved.

Availability of data and material

The datasets analyzed during the testing step in the present study are available in the TCGA repository, <https://portal.gdc.cancer.gov/>; the data analyzed during the current validation step in this study available from the corresponding author on reasonable request.

Competing Interests

The authors declare no competing financial interest about the present study.

Authors' contributions

TTY and MX conceived and designed the project. HZ, QL, LL, and YG conducted the database analysis. LL, RR, KS, and QL collected the samples of leukemia. HZ and LH performed the experiments of mRNA expression. ZMY, XYJ, and XLW wrote the article. MX, HDZ, and BLZ supported the corresponding funding. All authors read and approved the manuscript and agree to be accountable for all aspects of the research in ensuring that the accuracy or integrity of any part of the work are appropriately investigated and resolved. All authors read and approved the manuscript.

Funding

This study was financially supported by the National Natural Science Foundation of China (81703201 and 81602431), the Nature Science Foundation of Jiangsu Province (BK20171076 and BK20181488), the Jiangsu Provincial Medical Innovation Team (CXTDA2017029), the Jiangsu Provincial Medical Youth Talent (QNRC2016548), the Program of Jiangsu Prevention Medicine Association (Y2018086) and the Lifting Program of Jiangsu Provincial Scientific and Technological Association.

Acknowledgements

We appreciated for the help of Yuanfang Chen in Jiangsu Center for Disease Control and prevention. And we thank for the technical support from the TargetGene Biotechnology Co.,LTD in Nanjing.

References

1. Szczepanski T, van der Velden VH, van Dongen JJ. Classification systems for acute and chronic leukaemias. *Best Pract Res Clin Haematol.* 2003;16(4):561–82.
2. Tamamyan G, Kadia T, Ravandi F, Borthakur G, Cortes J, Jabbour E, Daver N, Ohanian M, Kantarjian H, Konopleva M. Frontline treatment of acute myeloid leukemia in adults. *Crit Rev Oncol Hematol.* 2017;110:20–34.

3. Stefanidakis M, Karjalainen K, Jaalouk DE, Gahmberg CG, O'Brien S, Pasqualini R, Arap W, Koivunen E. Role of leukemia cell invadosome in extramedullary infiltration. *Blood*. 2009;114(14):3008–17.
4. Siegel R, Ma J, Zou Z, Jemal A. Cancer statistics, 2014. *CA Cancer J Clin*. 2014;64(1):9–29.
5. Gyurkocza B, Storb R, Storer BE, Chauncey TR, Lange T, Shizuru JA, Langston AA, Pulsipher MA, Bredeson CN, Maziarz RT, et al. Nonmyeloablative allogeneic hematopoietic cell transplantation in patients with acute myeloid leukemia. *J Clin Oncol*. 2010;28(17):2859–67.
6. Estey EH. Acute myeloid leukemia: 2013 update on risk-stratification and management. *Am J Hematol*. 2013;88(4):318–27.
7. Dohner H, Estey EH, Amadori S, Appelbaum FR, Buchner T, Burnett AK, Dombret H, Fenaux P, Grimwade D, Larson RA, et al. Diagnosis and management of acute myeloid leukemia in adults: recommendations from an international expert panel, on behalf of the European LeukemiaNet. *Blood*. 2010;115(3):453–74.
8. Dohner H, Weisdorf DJ, Bloomfield CD. Acute Myeloid Leukemia. *N Engl J Med*. 2015;373(12):1136–52.
9. Ferrara F, Schiffer CA. Acute myeloid leukaemia in adults. *Lancet*. 2013;381(9865):484–95.
10. Takahashi K, Wang F, Kantarjian H, Doss D, Khanna K, Thompson E, Zhao L, Patel K, Neelapu S, Gumbs C, et al. Preleukaemic clonal haemopoiesis and risk of therapy-related myeloid neoplasms: a case-control study. *Lancet Oncol*. 2017;18(1):100–11.
11. Tomczak K, Czerwinski P, Wiznerowicz M. The Cancer Genome Atlas (TCGA): an immeasurable source of knowledge. *Contemporary oncology*. 2015;19(1A):A68–77.
12. Balachandran VP, Gonen M, Smith JJ, DeMatteo RP. Nomograms in oncology: more than meets the eye. *Lancet Oncol*. 2015;16(4):e173–80.
13. Palumbo A, Minowada J, Erikson J, Croce CM, Rovera G. Lineage infidelity of a human myelogenous leukemia cell line. *Blood*. 1984;64(5):1059–63.
14. Ha K, Minden M, Hozumi N, Gelfand EW. Immunoglobulin gene rearrangement in acute myelogenous leukemia. *Cancer Res*. 1984;44(10):4658–60.
15. Williams L, Moscinski LC, Medveczky PG. Immunoglobulin germline mu transcripts in acute myelogenous leukemia cells vary in splicing pattern and are heterogeneous. *Leukemia*. 1995;9(12):2016–22.
16. Guo M, Dong L, Huang S. [Aberrant expression of immunoglobulin germline gene C mu in leukemias]. *Zhonghua Xue Ye Xue Za Zhi*. 1998;19(7):359–62.
17. Silva P, Justicia A, Regueiro A, Fariña S, Couselo JM, Loidi L. Autosomal recessive agammaglobulinemia due to defect in μ heavy chain caused by a novel mutation in the IGHM gene. *Genes Immun*. 2017;18(3):197–9.
18. Dong L, Guo M, Huang SM, Jia SQ, Wang H. Transcripts of immunoglobulin germline mu: an amplified myeloid and B-lymphoid common gene program in various leukemias. *Acta Haematol*. 1999;101(3):119–23.
19. Zhong L, Chen J, Huang X, Li Y, Jiang T. Monitoring immunoglobulin heavy chain and T-cell receptor gene rearrangement in cfDNA as minimal residual disease detection for patients with acute myeloid leukemia. *Oncol Lett*. 2018;16(2):2279–88.
20. Jain N, Hartert K, Tadros S, Fiskus W, Havranek O, Ma MCJ, Bouska A, Heavican T, Kumar D, Deng Q, et al: **Targetable genetic alterations of TCF4 (E2-2) drive immunoglobulin expression in diffuse large B cell lymphoma**. *Sci Transl Med* 2019, 11(497).
21. Orozco JJ, Appelbaum FR. Unfavorable, complex, and monosomal karyotypes: the most challenging forms of acute myeloid leukemia. *Oncology (Williston Park)*. 2012;26(8):706–12.
22. Miranda A, Hamilton PT, Zhang AW, Pattnaik S, Becht E, Mezheyeuski A, Bruun J, Micke P, de Reynies A, Nelson BH. Cancer stemness, intratumoral heterogeneity, and immune response across cancers. *Proc Natl Acad Sci U S A*. 2019;116(18):9020–9.
23. Astaneh M, Dashti S, Esfahani ZT. Humoral immune responses against cancer-testis antigens in human malignancies. *Hum Antibodies*. 2019;27(4):237–40.
24. Sorich MJ, Rowland A, Karapetis CS, Hopkins AM. Evaluation of the Lung Immune Prognostic Index for Prediction of Survival and Response in Patients Treated With Atezolizumab for NSCLC: Pooled Analysis of Clinical Trials. *J Thorac Oncol*. 2019;14(8):1440–6.
25. Wagner J, Rapsomaniki MA, Chevrier S, Anzeneder T, Langwieder C, Dykgers A, Rees M, Ramaswamy A, Muenst S, Soysal SD, et al. A Single-Cell Atlas of the Tumor and Immune Ecosystem of Human Breast Cancer. *Cell*. 2019;177(5):1330–45 e1318.
26. Flores-Martin JF, Perea F, Exposito-Ruiz M, Carretero FJ, Rodriguez T, Villamediana M, Ruiz-Cabello F, Garrido F, Cozar-Olmo JM, Aptsiauri N. A Combination of Positive Tumor HLA-I and Negative PD-L1 Expression Provides an Immune Rejection Mechanism in Bladder Cancer. *Ann Surg Oncol*. 2019;26(8):2631–9.

Figures

DESeq2 Volcano

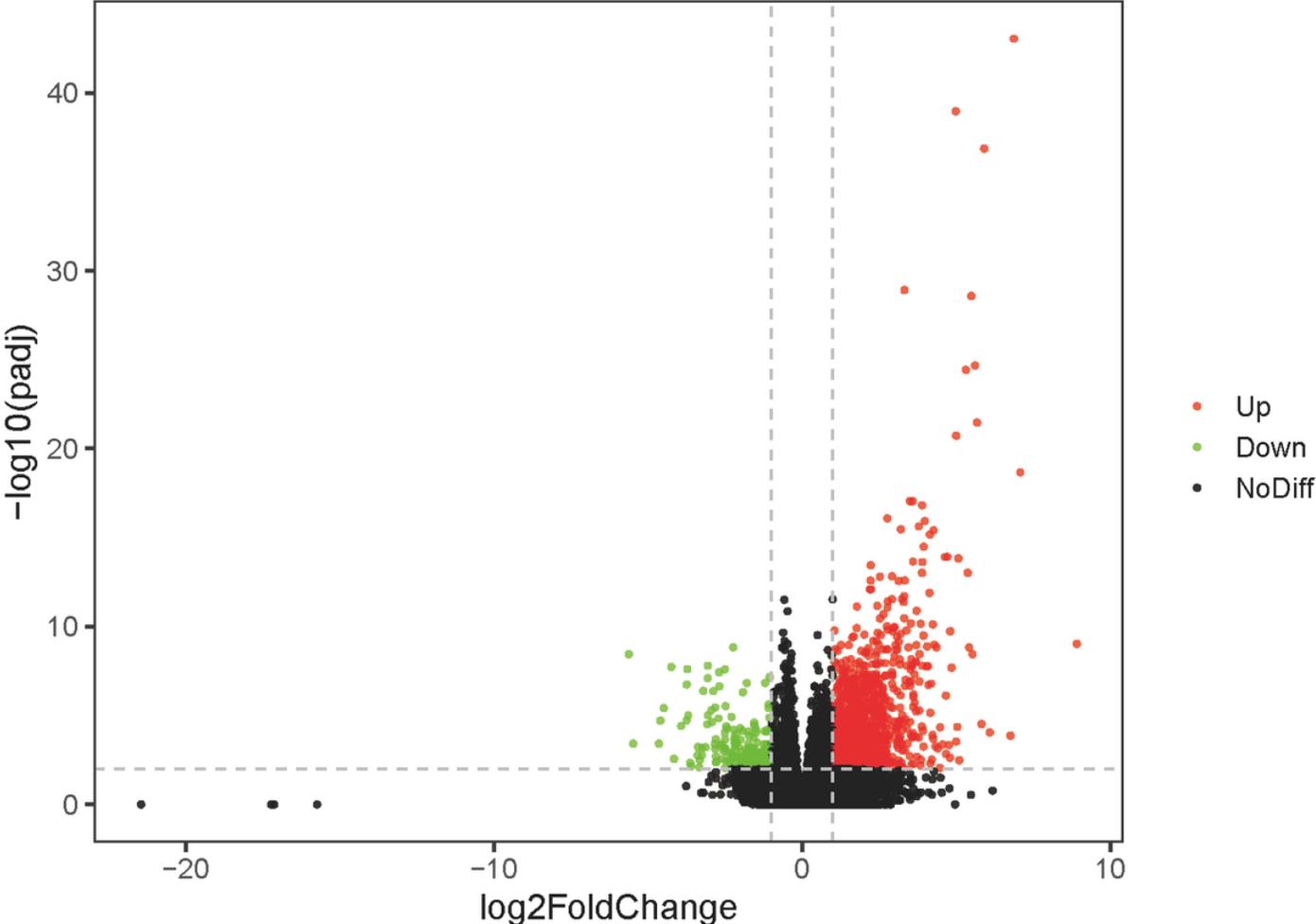


Figure 1

Volcano plot of DEGs in AML with different cytogenetics risks.

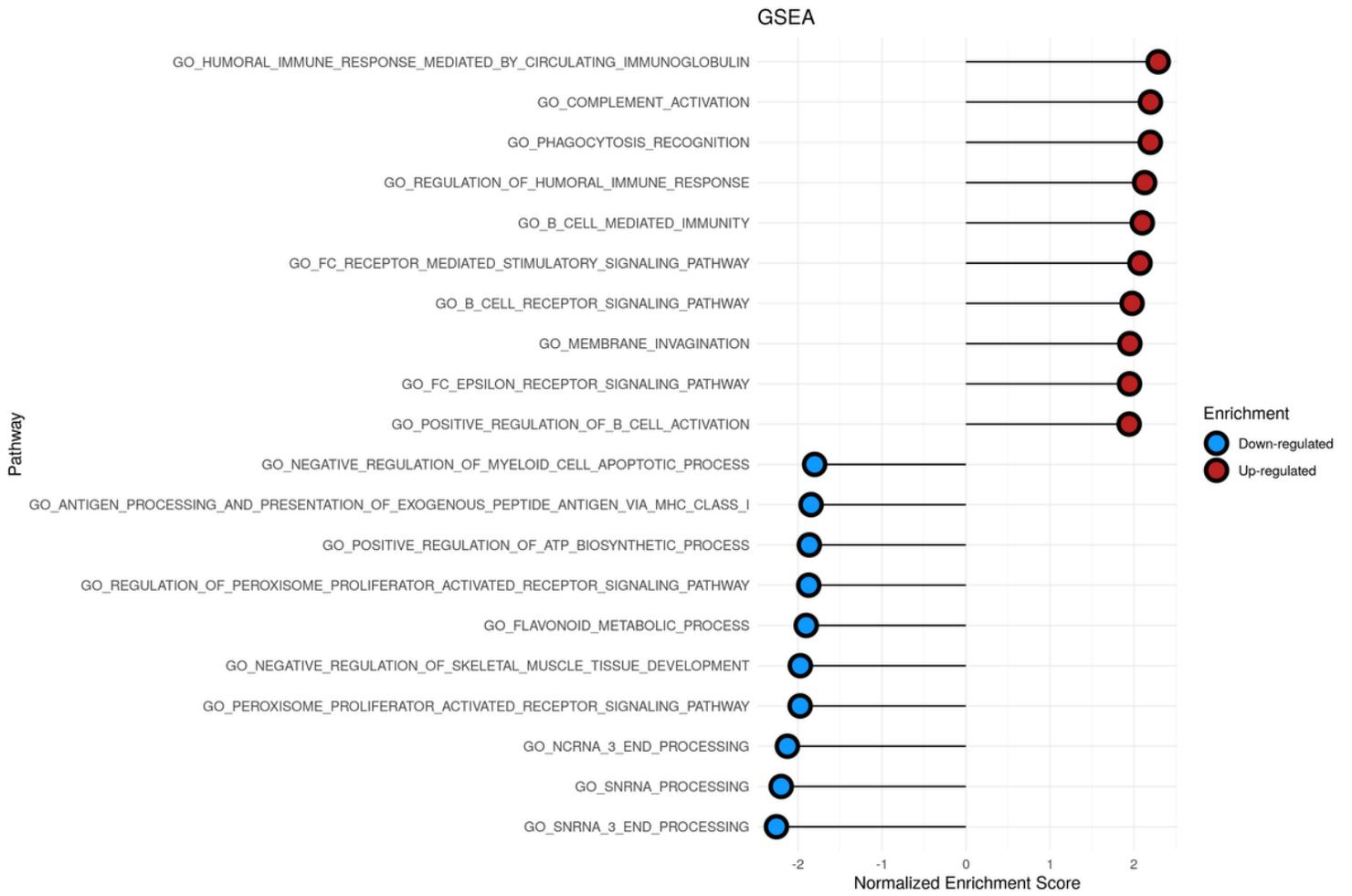


Figure 2

Top10 significant pathways of cyto genetics risk category among LAML by GSEA.

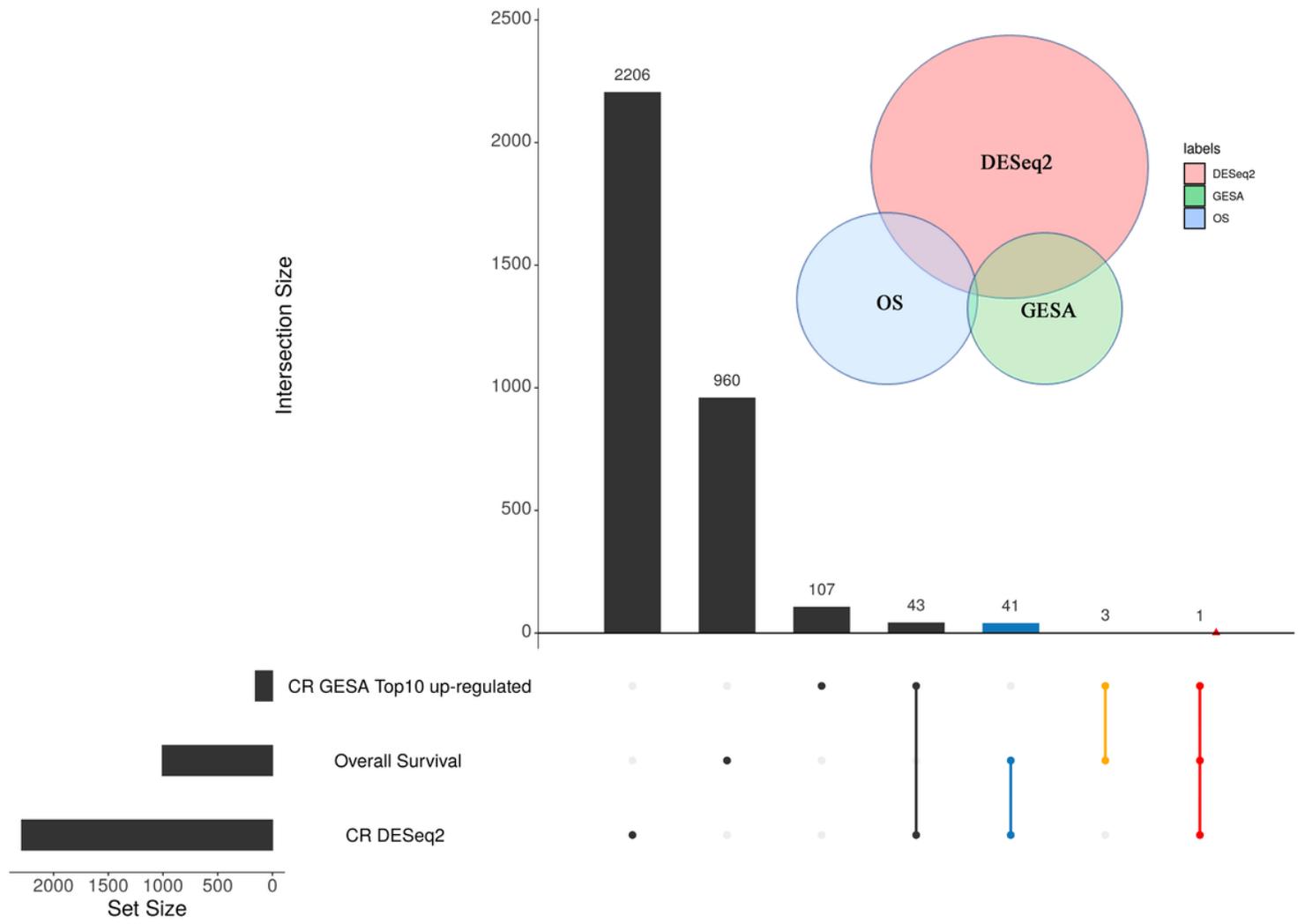


Figure 3

Venn diagram of DEGs for the AML-target gene selection.

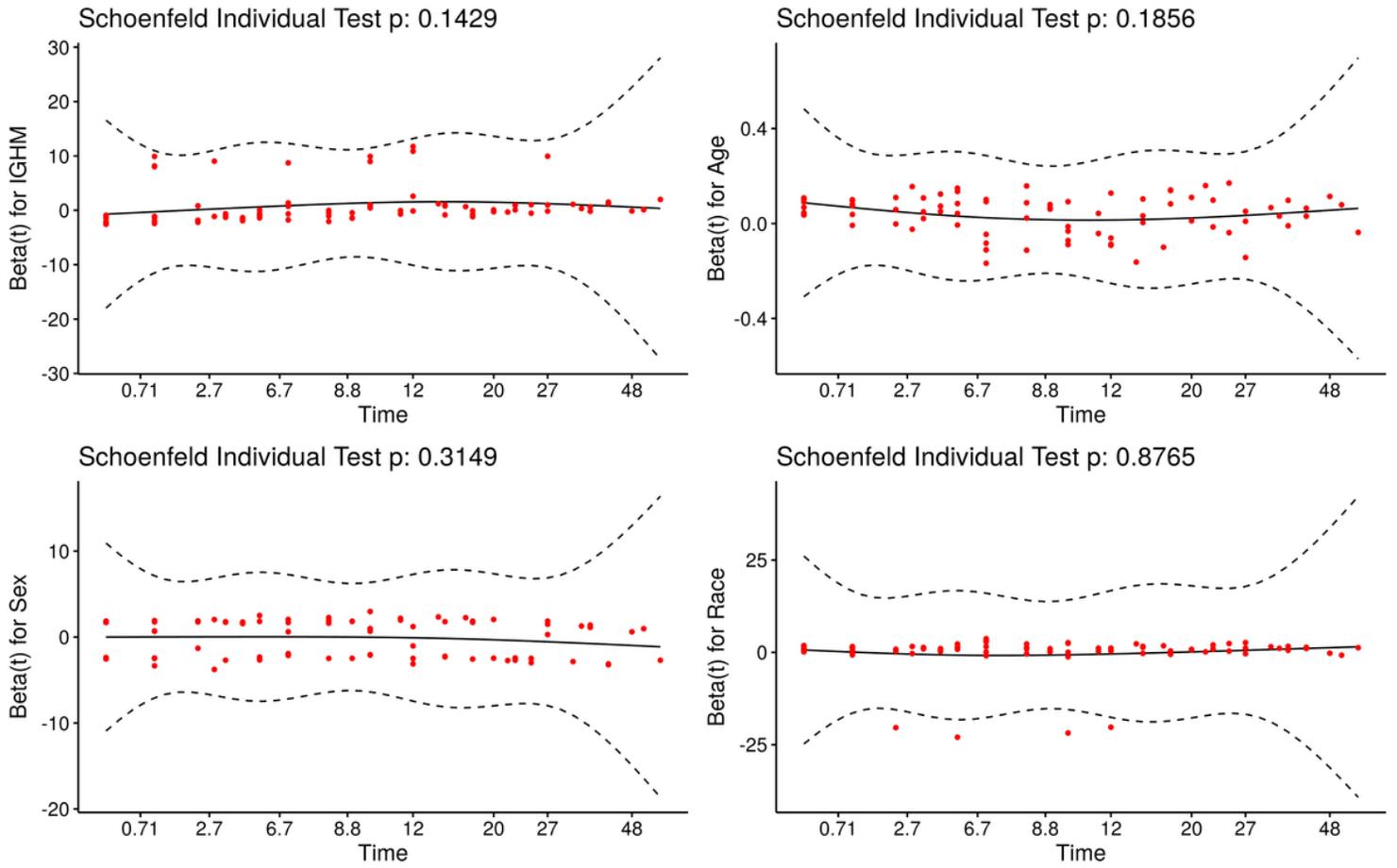


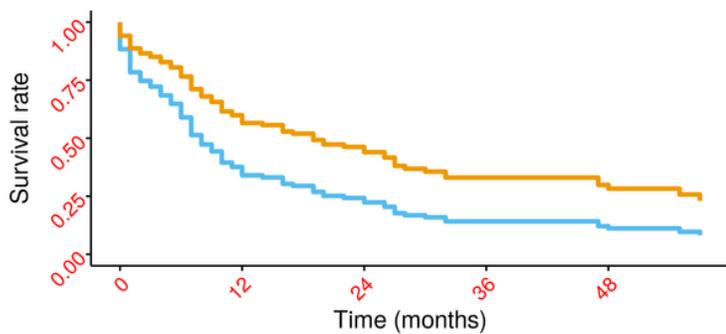
Figure 4

Proportional Hazards Assumption test for Cox model of TCGA AML survival

A. IGHM

$\exp(\text{coef}) = 2.07, z = 2.04, P = 0.041$

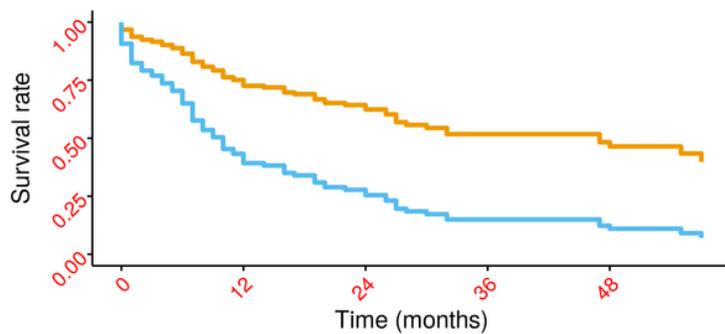
Strata — High expression — Low expression *



B. Age

$\exp(\text{coef}) = 3.00, z = 4.57, P < 0.001$

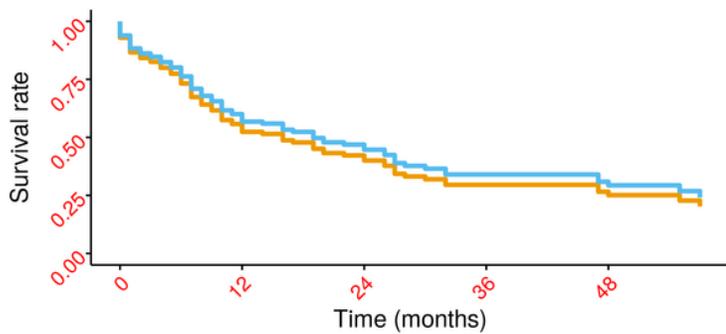
Strata — Young (age <= 55) * — Old (age > 55)



C. Sex

$\exp(\text{coef}) = 0.86, z = -0.68, P = 0.499$

Strata — Female * — Male



D. Race

$\exp(\text{coef}) = 0.91, z = -0.17, P = 0.863$

Strata — Black or African American * — White

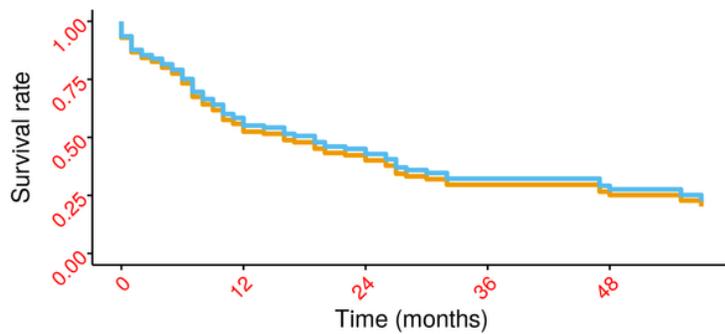


Figure 5

Adjusted Kaplan–Meier survival curves for IGHM expression in TCGA AML patients.

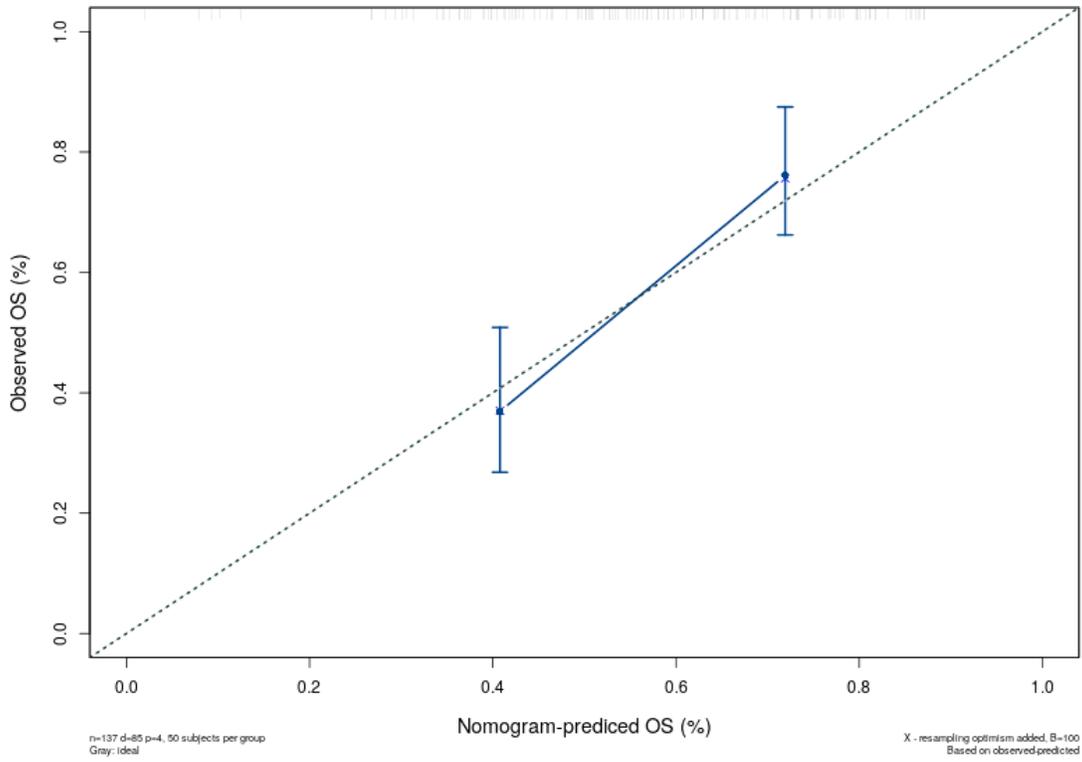


Figure 6

Calibration curve of Cox regression model of IGHM in TCGA AML patients

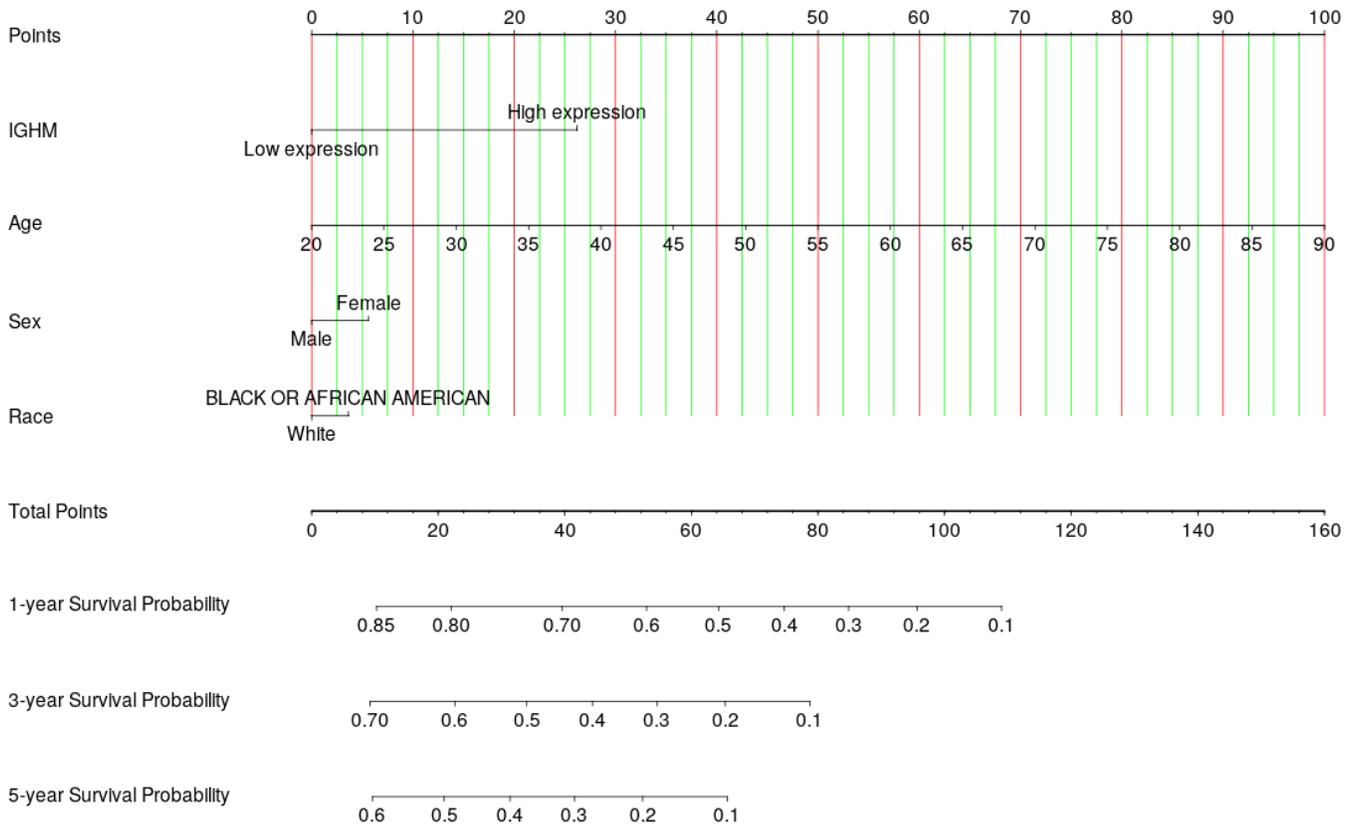


Figure 7

Nomogram fitting the Cox proportional risk model in TCGA AML patients.

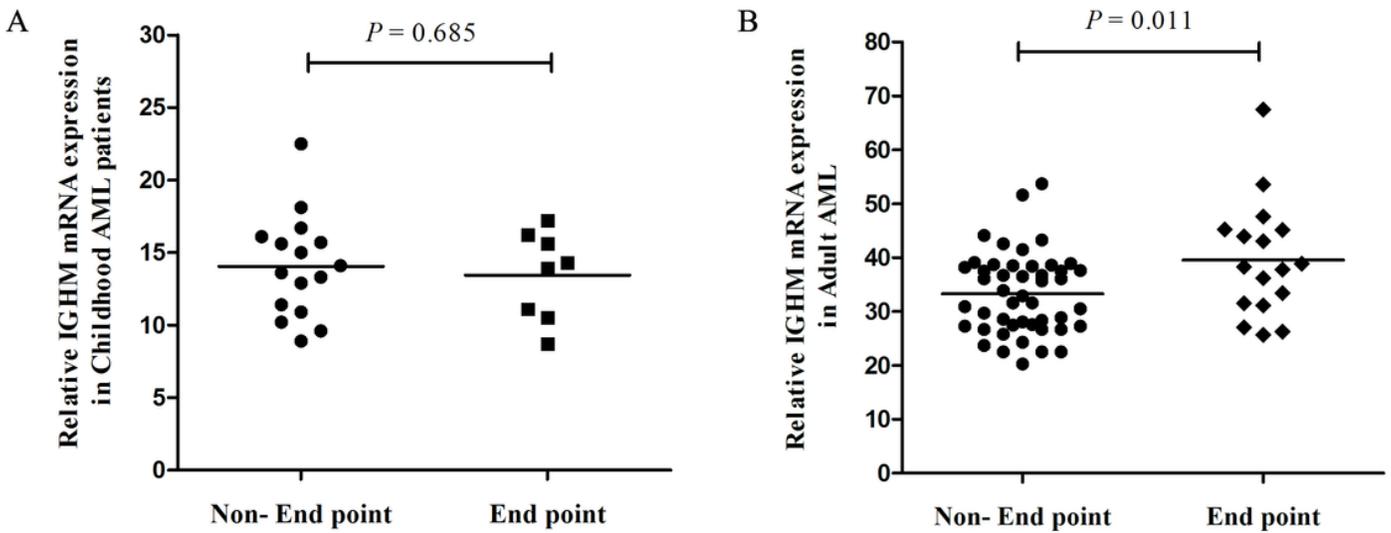


Figure 8

Validation of IGHM expression in Chinese childhood and adult AML patients with different survival end-points. A. Childhood AML patients; B. Adult AML patients.