

Exploring key gene in Acute myeloid leukemia with weighted genes co-expression network analysis

Hang Zhang

School of Public Health, Nantong University

Yue Gao

Jiangsu Preventive Medicine Association

Yinnong Jia

Department of Pharmaceutical Sciences, School of Pharmaceutical Sciences and Yunnan Key Laboratory of Pharmacology for Natural Products, Kunming Medical University

Yang Shen

Nanjing Prevention and Treatment Center for Occupational Disease

Qing Zhang

Nanjing Prevention and Treatment Center for Occupational Disease

Lei Han

Department of Occupational Disease Prevention, Jiangsu Provincial Center for Disease Control and Prevention

Hengdong Zhang

School of Public Health, Nantong University

Baoli Zhu

Center for Global Health, School of Public Health, Nanjing Medical University

Tingting Yu

Department of Medical Genetics, School of Basic Medical Science Jiangsu Key Laboratory of Xenotransplantation, Nanjing Medical University

Ming Xu (✉ sosolou@jscdc.cn)

School of Public Health, Nantong University

Research Article

Keywords: WGCNA, Acute myeloid leukemia, PPM1J, survival time

Posted Date: January 17th, 2022

DOI: <https://doi.org/10.21203/rs.3.rs-1230880/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Purpose: This study applied the WGCNA method to calculate the data of AML acquired from the TCGA database to discover the genes that are highly related to AML, and figure out key genes from these highly expressed genes, providing novel potential targets and biomarkers for therapy.

Methods: The TCGA-AML was downloaded to construct gene co-expression module according to the correlation of genes, and the most significant module exhibiting the highest correlation coefficient with clinical characteristics was selected. The Cox proportional hazards model was selected for gene survival analysis to screen risk factors related to AML death. Time-dependent ROC was analyzed to screen the genes with the largest ROC curve area and identified as the key genes, and Cox model forest and Nomogram method were constructed to verify the prognosis model.

Results: According to the correlation of gene expression, 22 co-expression modules were found. The Grey60 module was significantly positively correlated with clinical characteristics (Cor=0.27, $P<0.001$). PTCRA, SLC4A3, SERPINF1, PPM1, and MB were screened by survival analysis. Furthermore, PPM1J with the largest AUC area was selected as the key gene. The death risk of the PPM1J high expression group was 2.83 times higher than that of the PPM1J low expression group. The survival rates of patients with different ages and cytogenetic risk types were statistically significantly different ($P<0.05$). Nomogram displayed that the expression of PPM1J in patients has a great impact on the survival time. The predicted 1-year, 3-year, and 5-year survival rates of patients in the low-expression group of PPM1J are >85%, >85%, and >70%, respectively, while the 1-year, 3-year, and 5-year survival rates of patients in the high-expression group are significantly decreased to 65%, 35%, and 20%, respectively.

Conclusion: In this study, the PPM1J gene was screened out by constructing a gene co-expression network, which may become a potential genetic biomarker, affecting the occurrence and prognosis of AML.

Introduction

Acute myeloid leukemia (AML) is a common adult hematologic malignancy, with typical characteristics such as high incidence rate, high mortality rate and rapid progression^[1,2]. In 2015, the total number of leukemia cases occurred in China was 75,300 and the number of deaths caused by this was 53,400. And the death rate was increased significantly with the growth of the age^[3]. Global cancer statistics in 2021 showed that the new cases and deaths of AML are higher than the previous period^[4]. With the development of urbanization and the severity of industrial pollution, the incidence rate of AML remains a growing trend year by year. It is obvious that AML has seriously threatened human health worldwide. However, the molecular mechanism of AML is still unrevealed because of the complicity. Till now, it is still a lack of specialists consensus to identify the exact factors for both AML occurrence and survival rate.

For its special derivation, AML is a malignant disease from bone marrow, characterized by aberrant clone expansion and stagnant differentiation of bone marrow progenitor cells^[5, 6]. In recent years, with the improvement of chemotherapy regimens, the application of various novel drugs and hematopoietic stem cell transplantation have gradually exhibited their curative effects on the therapy of AML patients^[7, 8]. Despite of the intervention by the advanced and modern technology, the prognosis results of AML are still unsatisfactory^[7, 8]. Therefore, it is crucial to investigate deeper to understand the molecular mechanism of AML and to find out the specific prognostic molecular markers of AML, which would not only help us to understand the pathogenesis of leukemia but also illustrate more important and useful information for the design of individualized treatment and targeted therapy.

With the continuous development of bioinformatics and traditional biological researches, the current studies for single genes have been unable to face the new challenges brought by the era of big data. Therefore, the interaction of biological network analysis with high-throughput data has attracted more attention. Weighted gene co-expression network analysis (WGCNA) is mainly used in large samples of gene expression data to mine genes with similar expression profiles, and then group these genes and store them in the same module. Considering the possibility of genes with the same expression trends may also have certain correlations in their biological functions, the WGCNA provides a strategy to investigate the momentous interaction of key genes and their potential relationship of regulation with each other in diseases^[9, 10]. This method can be used to screen candidate biological markers or latent therapeutic sites, but also has been successfully applied to various biological fields^[9, 10].

Based on the data in TCGA (The Cancer Genome Atlas), we proposed to distinguish the AML-related survival modules by eigengene or hub gene, and then calculated the correlation between the one module and the other module, as well as the correlation between the module and the sample clinical traits. Time-dependent ROC was also performed to screen the target genes with the largest ROC curve area and identified as the key genes in this study. Our study might provide a potential target of significant importance for early warning, prognostic analysis and targeted therapy of AML.

1 Material And Methods

1.1 Data sources and pre-processing

The data were downloaded from the TCGA data portal (<https://portal.gdc.cancer.gov>) by the access of TCGA-AML, which contained 200 clinical cases and 151 RNA-seq cases.

We used R software (R.64-bit.3.5.1) and RMA (robust multi-array average) package to standardize the data set. After the initial filtering, 44347 genes were obtained. By calculating the variance, the first 50% genes with the largest variance were selected as the analysis data set of WGCNA.

1.2 Construction of weighted gene co-expression modules

The method of clustering tree is applied to eliminate outlier samples according to the clustering graph to ensure the construction of a stable co-expression network. Scale-free network was established to ensure the co-expression network consistent with the scale-free phenomenon. The scale-free network index (R^2) > 0.85 was used as the criteria to satisfy the scale-free condition. At the same time, the threshold parameter β was determined according to the average connectivity^[11, 12]. The adjacent matrix and topological matrix were obtained according to the β -value, and the genes were clustered by the degree of dissimilarity. Finally, the clustering tree was cut into different modules by the dynamic cutting method^[13].

1.3 Correlation analysis between co-expression modules and clinical phenotypes

When constructing the network of modules, the algorithm would automatically generate the correlation curve of the clinical phenotype for each gene. And correlation coefficients related to the phenotypes would be calculated, companies with correlation curves. Regarding to the coefficient values, if modules were negatively correlated with the clinical characteristics, the core module would be marked in green color. The positive correlation was represented by the red color. And the darkness of both colors suggested the intensity of coefficients correlations. Accordingly, the relationship between the genes of each module was analyzed according to the clinical characteristics in each sample, and the modules were selected as the largest coefficient correlation with survival as the pivot core module for the follow-up studies.

1.4 Validation of key genes

1.4.1 Multivariate regression analysis to screen genes

The related information in the AML database was extracted from the core module. We merged the survival files and RS files into a matrix, including gender, age, race and cytogenetic risks. Further survival analysis was performed to screen out the risk genes affecting the survival outcome of AML.

1.4.2 Time-dependent ROC to further screen genes

The R software "survival ROC" package was used to draw a time-dependent ROC curve. The risk scoring model was chosen to verify the accuracy of patient survival. We also calculated the parameter-like area under the curve (AUC) to evaluate the efficiency of the model for AML survival, and the AUC value was usually enrolled between 0.5 and 1.

1.4.3 Establishment and evaluation of prediction model

To further verify the model as an independent predictor of the survival outcome of AML or not, we adopted the consistency index (C-index) to predict the performance based on the clinical parameters of the patient. Also, a relative R software package was used to construct a nomogram for our findings. As its advantage in medical research, nomograms can calculate the survival rate of patients in a personalized method and combine a variety of important factors to predict specific end-point events.

2 Results

2.1 Pre-processing of TCGA RNA sequencing and clinical data

We obtained 60488 genes corresponding to 151 samples from the download package of the TCGA database. After preprocessing chip data by R software and filtering the genes with incomplete and duplicate probe information, we obtained a total of 146 samples and 44347 genes. The top 50% of genes with the largest variance were further screened as the WGCNA analysis data set. The hierarchical clustering was performed according to the expression levels of the selected genes to obtain a sample cluster diagram, as shown in Fig. 1A, which suggested that all samples are located in the clusters and pass the cutoff thresholds. Additionally, basic patient information was added below the resulting tree (Fig. 1B). The clustered samples exhibited similar clinical characteristics.

2.2 Analysis of weighted gene co-expression network

2.2.1 Cluster analysis of genes

Pearson correlation matrix of paired genes was established, and it was weighted by soft threshold β to ensure the Pearson correlation matrix consistent with the scale-free network distribution. The soft threshold $\beta = 4$ was determined by calculation (Fig. 2A, B). The linear regression results showed that the correlation coefficient was $R^2 = 0.87$ (Fig. 2C, D), which was consistent with the characteristics of a scale-free network.

The Adjacency function was adopted to generate the adjacency matrix according to the β value, and then the TOM similarity function was used to convert the adjacency matrix into a topological matrix. The topological matrix was utilized to cluster genes by their dissimilarity. The minimum number of genes in the module was defined as 100, and a total of 22 gene modules was obtained (one among which was the grey module representing genes that were not divided into any module) (Fig. 3). The correlation coefficients of all modules were shown in Fig. 4 ($r^2 \leq 0.8$).

2.2.2 Correlation analysis between co-expression module and clinical phenotypes

In order to determine the relationship between relative gene modules and AML survival status, we calculated the correlation between clinical information and gene modules. The results were described in Fig. 5 and Fig. 6. The horizontal axis represented different clinical features, and the vertical axis represented different gene modules. Different colors in each box indicated the size of correlation; red indicated positive correlation and green indicated negative correlation, following a gradient ramp according to their different intensities of expression. The number in each square represented the Pearson correlation coefficient, which was the R-value, and the value in the brackets represented the P-value.

Among all selected modules, the module with the highest absolute value was considered to be the most clinically relevant module, which was the grey60 model in our findings. Moreover, it was revealed from Fig. 5 that the grey60 module displayed the strongest correlation with mortality ($r = 0.27$, $P < 0.001$). Therefore, the grey60 module was identified as the key research object for further investigation.

2.3 Verification of related genes

2.3.1 Multivariate regression analysis to screen genes

Additionally, the 296 genes remaining in the grey60 module were screened by using a multivariate Cox regression model to analyze the influences of each gene's expression level in the grey60 module on AML survival. Four associations of gender, age, ethnicity and cytogenetic risk were adjusted to estimate the influence of genes. The grey60 module was positively correlated with death ($p < 0.01$ was considered as the selection criterion). In addition, 5 genes in the grey60 module were marked as risk factors according to the results from survival analysis, as shown in Table 1 (the data are reserved to three decimal places).

Table 1
Multivariate Cox regression analysis results of 5 genes as risk factor

Symbol	coefficient	exp(coef)	se(coef)	z	P-value
PTCRA	1.52	4.59	0.45	3.42	0.001
SLC4A3	1.37	3.92	0.46	2.98	0.003
SERPINF1	1.25	3.50	0.46	2.70	0.007
PPM1J	1.04	2.83	0.45	2.29	0.022
MB	1.03	2.81	0.41	2.50	0.012

2.3.2 Time-dependent ROC to further screen genes

Based on the results of multivariate analysis, the ROC curve for the overall survival rate of AML patients was constructed using R software to evaluate the prognosis efficiency of the above-selected genes. The prediction model results showed that the AUC values were 0.55 for PPM1J, 0.51 for MB, 0.48 for SLC4A3, 0.46 for PTCRA, and 0.41 for SERPINF1. Among these findings, PPM1J showed the largest AUC area, suggesting that it may be the optimal prognosis biomarker for AML (as shown in Fig. 7). Therefore, the subsequent analysis was performed with a focus on the influence of PPM1J in AML patients' survival.

2.3.3 Construction and verification of clinical prediction models

The closer the model C-index approach the value of 1, the better accuracy the model could achieve. The results showed that the accuracy of the model was good (c-index = 0.71, $P < 0.001$). As shown in Fig. 8, the risk of death in the PPM1J high expression group was 2.83 times higher than that in the low expression group. The survival rates of patients with different ages and cytogenetic risk types were

statistically different ($P < 0.05$). For the different cytogenetic risk categories, the risk of death in the intermediate group was 2.44 times higher than that in the favorable group, and the risk of death in the poor group was 3.23 times higher than that of the favorable group.

Nomogram showed that the expression of PPM1J in AML patients exhibited a negative impact on their survival time. The 1-year, 3-year, and 5-year survival rates of patients in the PPM1J low expression group were predicted as >85%, >85%, >70%, while that in the PPM1J high expression group were 65%, 35%, and 20%, respectively. In addition, the cytogenetic risk was significantly different between the favorable group and the poor group. The 1-year, 3-year, and 5-year survival rates of the favorable group was > 85%, > 85%, > 70%, respectively, while the poor group was 58%, 25%, 13% (Fig. 9).

4 Discussion

In this study, we analyzed the AML samples provided by the TCGA database using the WGCNA method to obtain 22 AML gene modules. The clinical characteristics of the sample were combined to further analyze its correlation with age, gender, race and so on. The most positive correlation module with clinical characteristics is the grey60 model, and the correlation between the grey60 module and death is the strongest. The selected modular genes were further screened by the Cox risk proportional regression model to meet our requirements to achieve the most optimized prediction effect with fewer genetic variables. To our knowledge, this effective combination of the two methods to construct a model was the first study to result in a more comprehensive understanding of the prognosis of AML. The following screening of prognostic genes provides a more convenient and optimal method, which suggested five potential prognostic genes named PTCRA, SLC4A3, SERPINF1, PPM1J, and MB, respectively.

The relevant literature showed only a few studies have investigated the above-mentioned genes in AML, which are in fact worthy of more interest and further evaluation. According to the time-dependent ROC of these genes, PPM1J displayed the largest area under the curve as well as the significant association with the survival time of AML patients in nomogram, indicating that PPM1J may be a valuable tumor biomarker for AML and benefit the prognosis of patients. The PPM1J gene encodes the serine/threonine protein phosphatase and belongs to the protein phosphatase 2C gene family^[14, 15]. Li *et al.* reported the single nucleotide polymorphism (SNP) of the PPM1J locus as a new locus related to the development of renal function^[16]. Whisenant *et al.* have developed a new method combining professional knowledge and machine learning to find a new 'D-site' in the genome, and realized that PPM1J can be used as a new c-Jun N-terminal protein kinase (JNK), which is of great significance for the study of cell biology and cancer protease network^[15]. Yan *et al.* utilized comprehensive bioinformatics approaches such as ESTIMATE algorithm and PPI to mine the AML TCGA database and found that the expression of PPM1J is closely related to the prognosis of AML^[17], which is consistent with the results of our research.

The other co-expression genes also demonstrated their potential importance for the occurrence and survival of cancers. PTCRA is a single-channel membrane protein that expresses in immature T cells and regulates T cell development, which is essential for the initiation of methylation during embryonic development^[18, 19]. By performing the whole-exome sequencing of patients with chronic myelogenous

leukemia, Russian researchers found that different genotyping variants of PTCRA may be a leading cause for the failure of targeted tumor therapy^[20]. SLC4A3 gene family proteins contain bicarbonate transporters, which are involved in ion transportation in most cells^[21, 22]. Kimi *et al.* screened out SLC4A3 as a significant radiation-sensitive gene from the database of breast cancer as well as head and neck cancer^[23]. As a member of the serine protease inhibitor family, PEDF has a high affinity to type I collagen in bone tissue. The high expression of PEDF in osteoblasts and active areas of bone formation suggested that it played an important role in the process of bone angiogenesis and matrix reconstruction^[24]. For SERPINF1 (serpin peptidase inhibitor, clade F, member 1), located on chromosome 17p13.3, it encodes pigment-epithelium-derived factor (PEDF)^[25]. Meanwhile, the SERPINF1 is usually reported to associate with PEDF gene, which is universally involved in antitumor-related cellular processes, such as antiangiogenic^[26], tumor growth inhibition and anti-metastasis^[27, 28]. Regarding to the antitumor effect, it is rational to detect their aberrant expression in the endpoints, including recurrence, metastasis, or even death. The myoglobin (MB) is a monomeric blood protein present in striated muscle cells, which accounts for 1%-2% of the total weight of skeletal muscle. It promotes the oxygen transportation to mitochondria from the total oxygen stored in the human body^[29, 30], which is an essential cellular process leading to the chemotherapy-resistance in AML^[31, 32]. With the development of high-throughput screening, various genes related to human diseases are discovered. It is of high biological significance as providing a new direction for the investigation of clinical problems for various cancers, including pathological mechanisms and the discovery of effective molecular targeted therapies.

Although we screened out the target genes that were significantly related to the survival of AML, this study still has certain limitations. First of all, the sample source of this study is based on the data of AML samples in the TCGA database. The results might be affected by ethnic differences in the source. In future studies, AML samples from China and the rest of Asia need to be collected for further verification. Secondly, additional rigorous experiments are needed to be performed on the key genes revealed in this study. As a follow-up, further *in vitro* and *in vivo* experiments such as RT-PCR, western blot, and animal experiments are desired to be conducted as the biological verification for the key genes screened.

5 Conclusions

In summary, based on the TCGA database screening and the WGCNA method applied, we have discovered key genes related to the prognosis of AML, which would be valuable to provide more reference information for the ongoing studies of the regulation mechanism, target selection, diagnosis and treatment of AML. The biological function of these co-expression genes and their specific regulatory mechanism in AML have not been thoroughly studied, which is probably an important direction of future research.

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

Not applicable.

Availability of data and materials

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 is available from the corresponding author on reasonable request.

Competing interests

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

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, BK20181488, and BK20201485), the Jiangsu Provincial Medical Innovation Team (CXTDA2017029), the Jiangsu Provincial Medical Youth Talent (QNRC2016548), Jiangsu Provincial Health Development Research Center (JSHD2021013), the Program of Jiangsu Prevention Medicine Association (Y2018086) and the Lifting Program of Jiangsu Provincial Scientific and Technological Association.

Authors' contributions

H Zhang: Data analysis and manuscript writing; Y Gao: Drafting and revision; YN Jia: Data analysis and interpretation; Y Shen, Q Zhang: data collection; L Han, HD Zhang, BL Zhu: Funding and administrative support; TT Yu: Conception and design; M Xu: Sequencing data analysis and data analysis; All authors read and approved the final manuscript.

Acknowledgements

Thanks to all authors for their contributions to this article.

References

1. Medeiros, B.C., Is there a standard of care for relapsed AML? *Best Pract Res Clin Haematol*, 2018. 31(4): p. 384–386.
2. Deschler, B. and M. Lubbert, Acute myeloid leukemia: epidemiology and etiology. *Cancer*, 2006. 107(9): p. 2099–107.
3. Chen, W., et al., Cancer statistics in China, 2015. *CA Cancer J Clin*, 2016. 66(2): p. 115–32.

4. Siegel, R.L., et al., *Cancer Statistics, 2021*. *CA Cancer J Clin*, 2021. 71(1): p. 7-33.
5. De Kouchkovsky, I. and M. Abdul-Hay, 'Acute myeloid leukemia: a comprehensive review and 2016 update'. *Blood Cancer J*, 2016. 6(7): p. e441.
6. Shallis, R.M., et al., *Epidemiology of acute myeloid leukemia: Recent progress and enduring challenges*. *Blood Rev*, 2019. 36: p. 70–87.
7. Langfelder, P. and S. Horvath, *WGCNA: an R package for weighted correlation network analysis*. *BMC Bioinformatics*, 2008. 9: p. 559.
8. Xu, P., et al., *Identification of glioblastoma gene prognosis modules based on weighted gene co-expression network analysis*. *BMC Med Genomics*, 2018. 11(1): p. 96.
9. Liu, W., et al., [Weighted gene co-expression network analysis in biomedicine research]. *Sheng Wu Gong Cheng Xue Bao*, 2017. 33(11): p. 1791–1801.
10. Pei, G., L. Chen, and W. Zhang, *WGCNA Application to Proteomic and Metabolomic Data Analysis*. *Methods Enzymol*, 2017. 585: p. 135–158.
11. Bhonsle, S.P., et al., *Mitigation of impedance changes due to electroporation therapy using bursts of high-frequency bipolar pulses*. *Biomed Eng Online*, 2015. 14 Suppl 3: p. S3.
12. Zhang, B. and S. Horvath, *A general framework for weighted gene co-expression network analysis*. *Stat Appl Genet Mol Biol*, 2005. 4: p. Article17.
13. Horvath, S. and J. Dong, *Geometric interpretation of gene coexpression network analysis*. *PLoS Comput Biol*, 2008. 4(8): p. e1000117.
14. Kamada, R., et al., *Metal-dependent Ser/Thr protein phosphatase PPM family: Evolution, structures, diseases and inhibitors*. *Pharmacol Ther*, 2020: p. 107622.
15. Whisenant, T.C., et al., *Computational prediction and experimental verification of new MAP kinase docking sites and substrates including Gli transcription factors*. *PLoS Comput Biol*, 2010. 6(8).
16. Li, M., et al., *SOS2 and ACP1 Loci Identified through Large-Scale Exome Chip Analysis Regulate Kidney Development and Function*. *J Am Soc Nephrol*, 2017. 28(3): p. 981–994.
17. Yan, H., et al., *Identification of prognostic genes in the acute myeloid leukemia immune microenvironment based on TCGA data analysis*. *Cancer Immunol Immunother*, 2019. 68(12): p. 1971–1978.
18. Shi, J.Y., et al., *Genetic variations of DNA repair genes and their prognostic significance in patients with acute myeloid leukemia*. *Int J Cancer*, 2011. 128(1): p. 233–8.
19. Rodriguez, R.M., et al., *Regulation of the transcriptional program by DNA methylation during human alphabeta T-cell development*. *Nucleic Acids Res*, 2015. 43(2): p. 760–74.
20. Lavrov, A.V., et al., *Frequent variations in cancer-related genes may play prognostic role in treatment of patients with chronic myeloid leukemia*. *BMC Genet*, 2016. 17 Suppl 1: p. 14.
21. Linser, P.J., et al., *Carbonic anhydrases and anion transport in mosquito midgut pH regulation*. *J Exp Biol*, 2009. 212(Pt 11): p. 1662–71.

22. He, B.J., et al., Molecular Genetic Mechanisms of Hereditary Spherocytosis: Current Perspectives. *Acta Haematol*, 2018. 139(1): p. 60–66.
23. Drobin, K., et al., Molecular Profiling for Predictors of Radiosensitivity in Patients with Breast or Head-and-Neck Cancer. *Cancers (Basel)*, 2020. 12(3).
24. Carnagarin, R., A.M. Dharmarajan, and C.R. Dass, PEDF attenuates insulin-dependent molecular pathways of glucose homeostasis in skeletal myocytes. *Mol Cell Endocrinol*, 2016. 422: p. 115–124.
25. Becker, J., et al., Exome sequencing identifies truncating mutations in human SERPINF1 in autosomal-recessive osteogenesis imperfecta. *Am J Hum Genet*, 2011. 88(3): p. 362–71.
26. Eslani, M., et al., Corneal Mesenchymal Stromal Cells Are Directly Antiangiogenic via PEDF and sFLT-1. *Invest Ophthalmol Vis Sci*, 2017. 58(12): p. 5507–5517.
27. Wu, Q.J., et al., AAV-mediated human PEDF inhibits tumor growth and metastasis in murine colorectal peritoneal carcinomatosis model. *BMC Cancer*, 2012. 12: p. 129.
28. Abe, R., Y. Fujita, and S. Yamagishi, Angiogenesis and metastasis inhibitors for the treatment of malignant melanoma. *Mini Rev Med Chem*, 2007. 7(6): p. 649–61.
29. Benhamou, E., N. Sultan, and J.D. Warren, Biblical myoglobinuria. *Intern Med J*, 2018. 48(8): p. 1010.
30. David, W.S., Myoglobinuria. *Neurol Clin*, 2000. 18(1): p. 215–43.
31. Farge, T., et al., Chemotherapy-Resistant Human Acute Myeloid Leukemia Cells Are Not Enriched for Leukemic Stem Cells but Require Oxidative Metabolism. *Cancer Discov*, 2017. 7(7): p. 716–735.
32. Testa, U., et al., Oxidative stress and hypoxia in normal and leukemic stem cells. *Exp Hematol*, 2016. 44(7): p. 540–60.

Figures

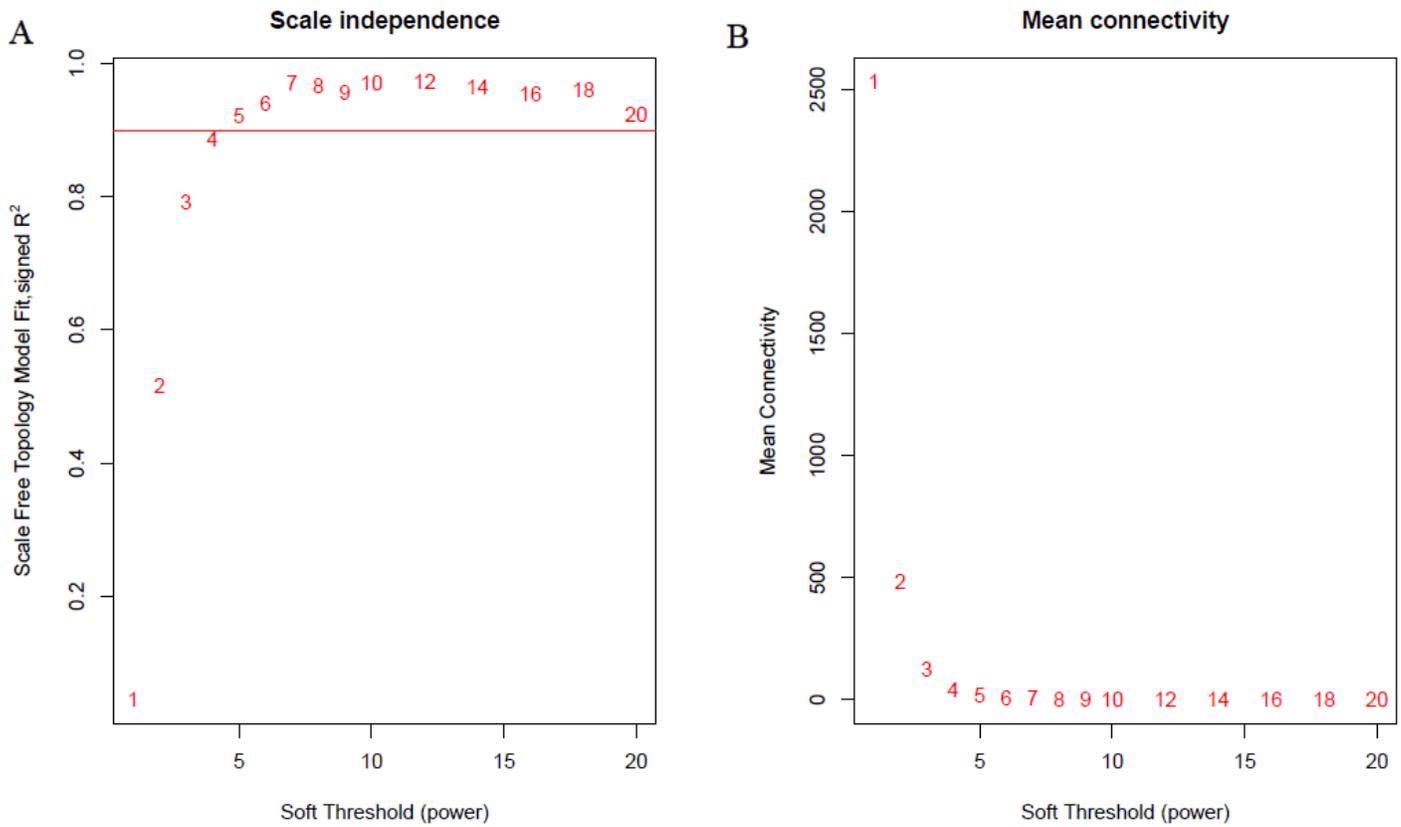


Figure 2

Soft-thresholding powers selection

(A) The scale-free fit index. (B) The mean connectivity. The red line in panel A corresponds to 0.87. Red numbers in both panels A and B denote soft-thresholding powers that are associated with the correlation coefficient square values and the mean connectivity. (C) and (D) Scale-free topology when soft-thresholding power $\beta=4$.

Gene dendrogram and module colors

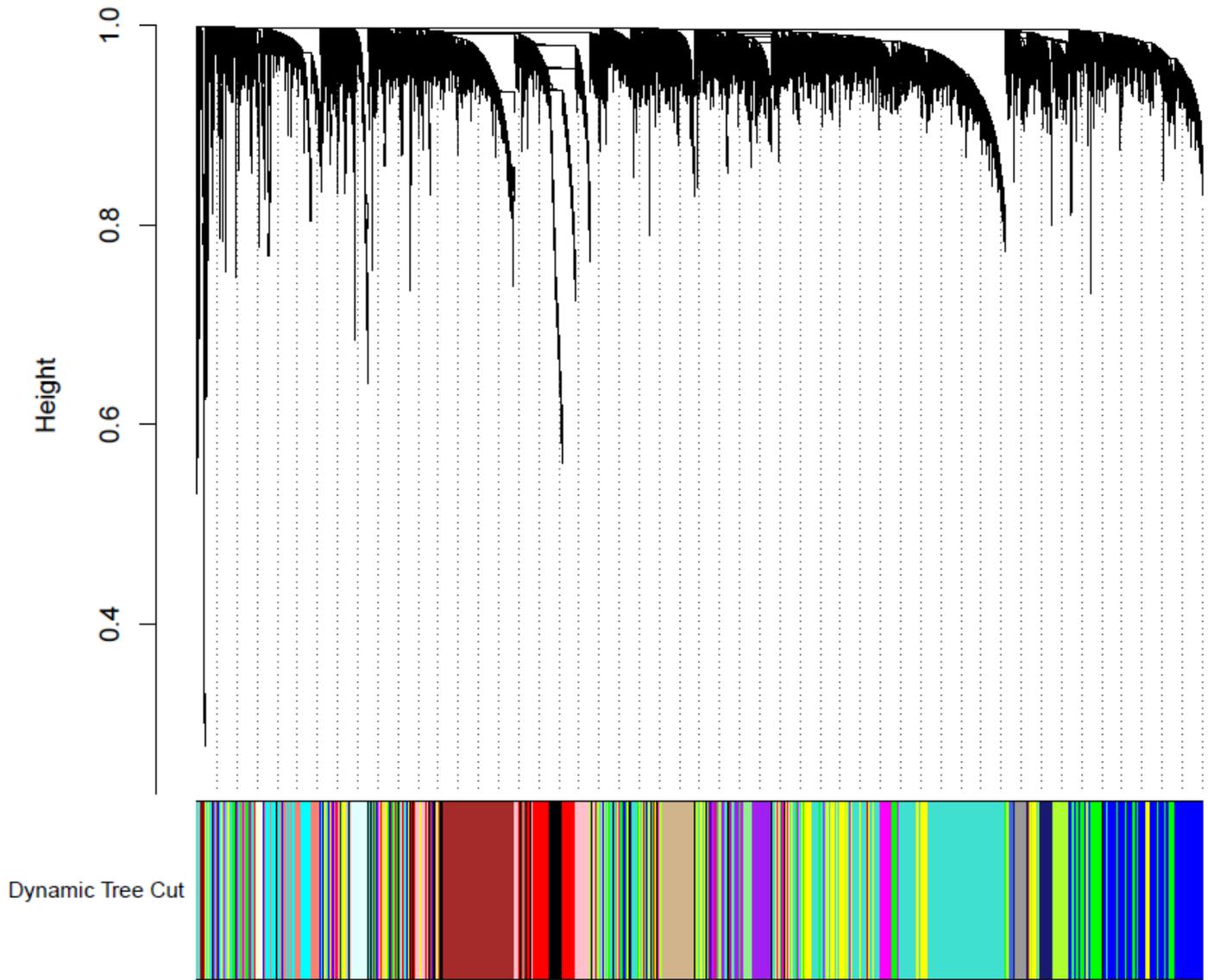


Figure 3

The model established by dynamic tree cutting

The clustering dendrogram of all differentially expressed genes was constructed according to the dissimilarity matrices, and the color band shows the module membership acquired using the dynamic tree cut algorithm.

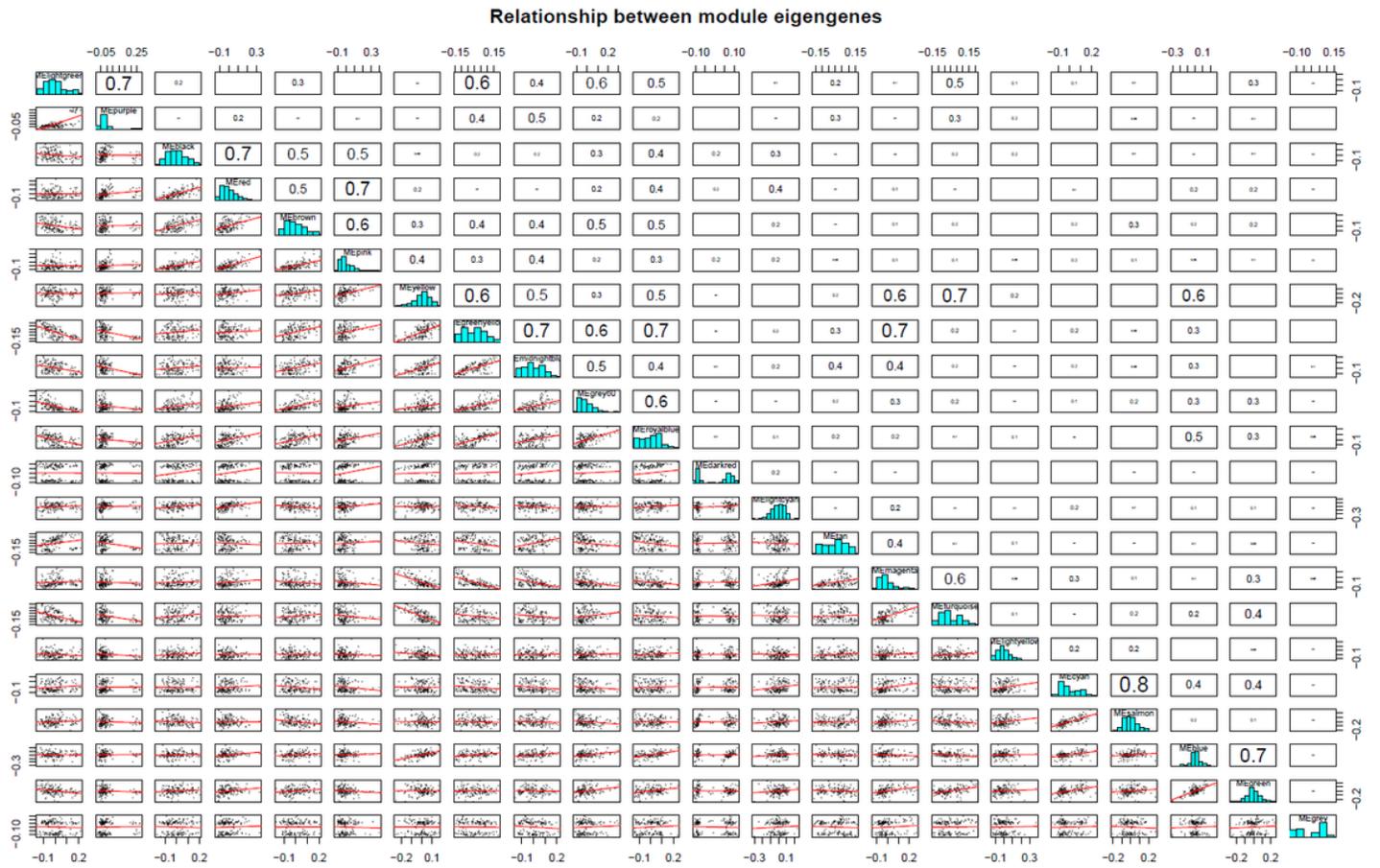


Figure 4

Relationship between module eigengenes

Module-trait relationships



Figure 5

Module-trait relationships

Module Significance p-value=0

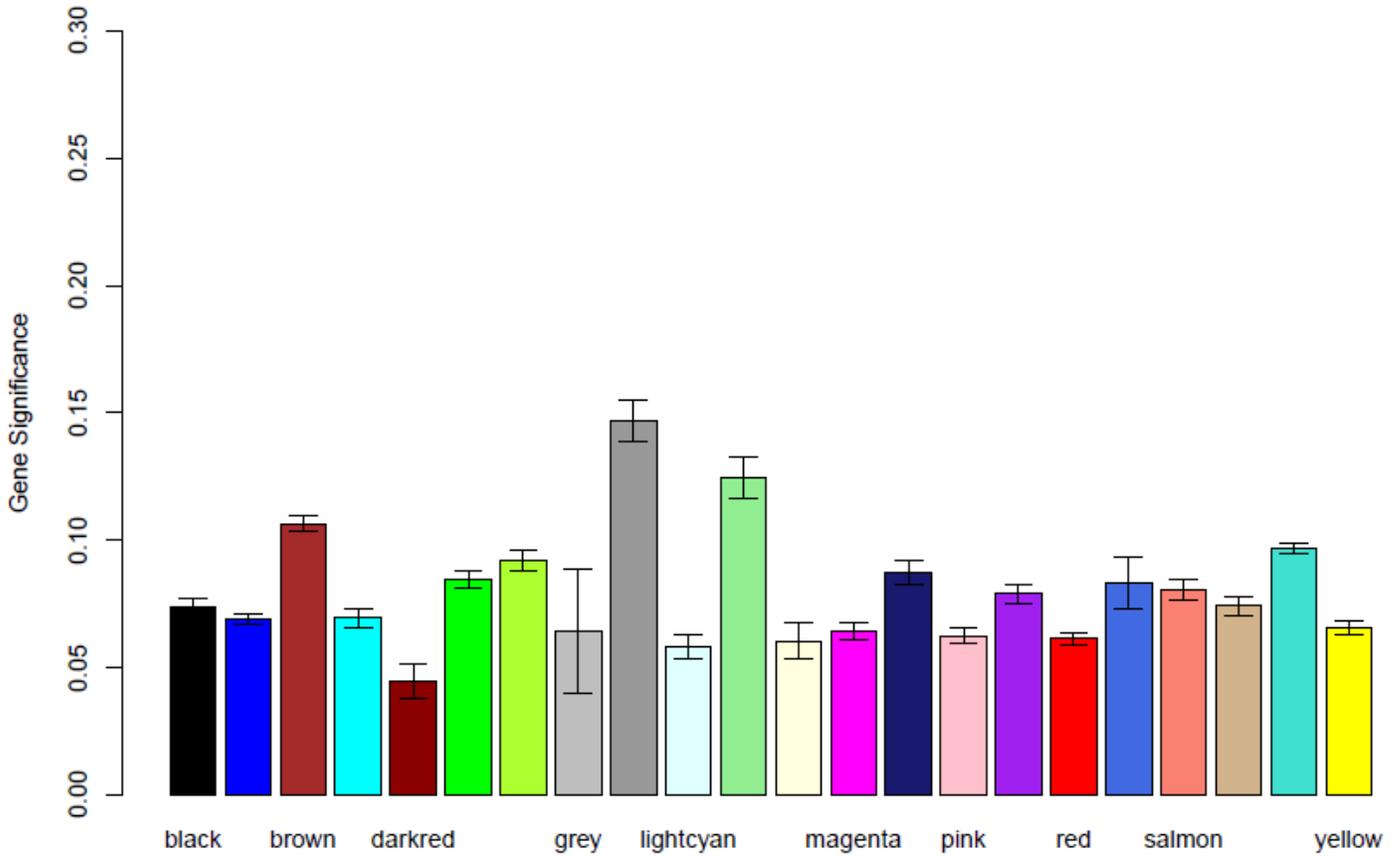


Figure 6

Relationships among model membership and gene significance

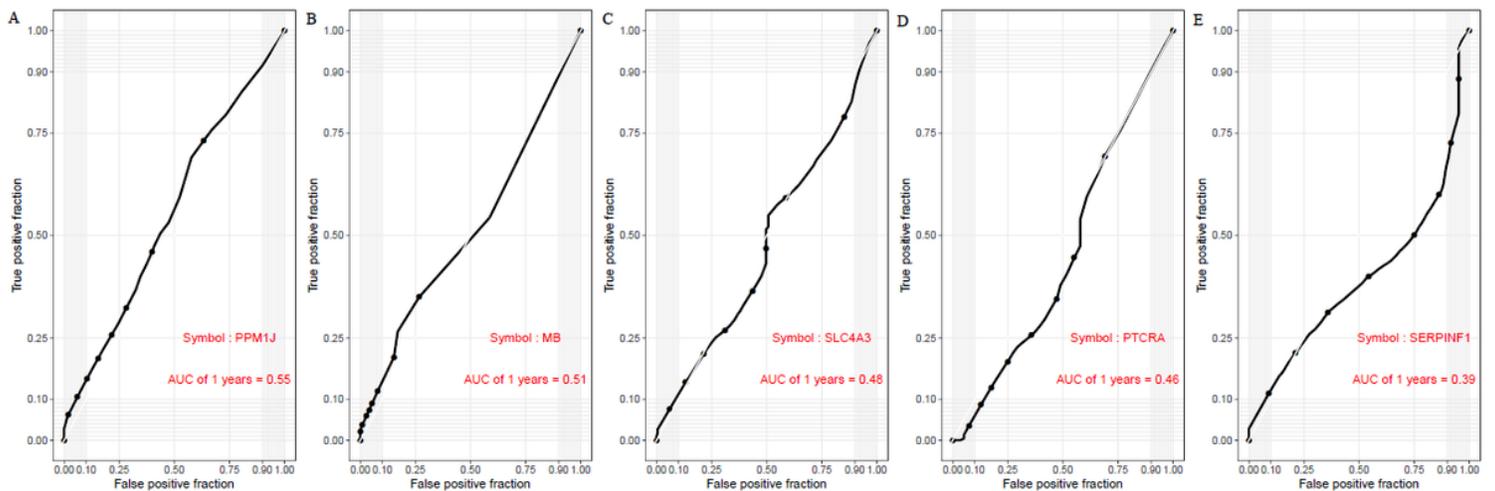


Figure 7

Time-dependent ROC

ROC curves of the five hub genes. The patients were stratified into high- and low- expression groups according to the median expression. (A) PPM1J. (B) MB. (C) SLC4A3. (D) PTCRA. (E) SERPINF1.

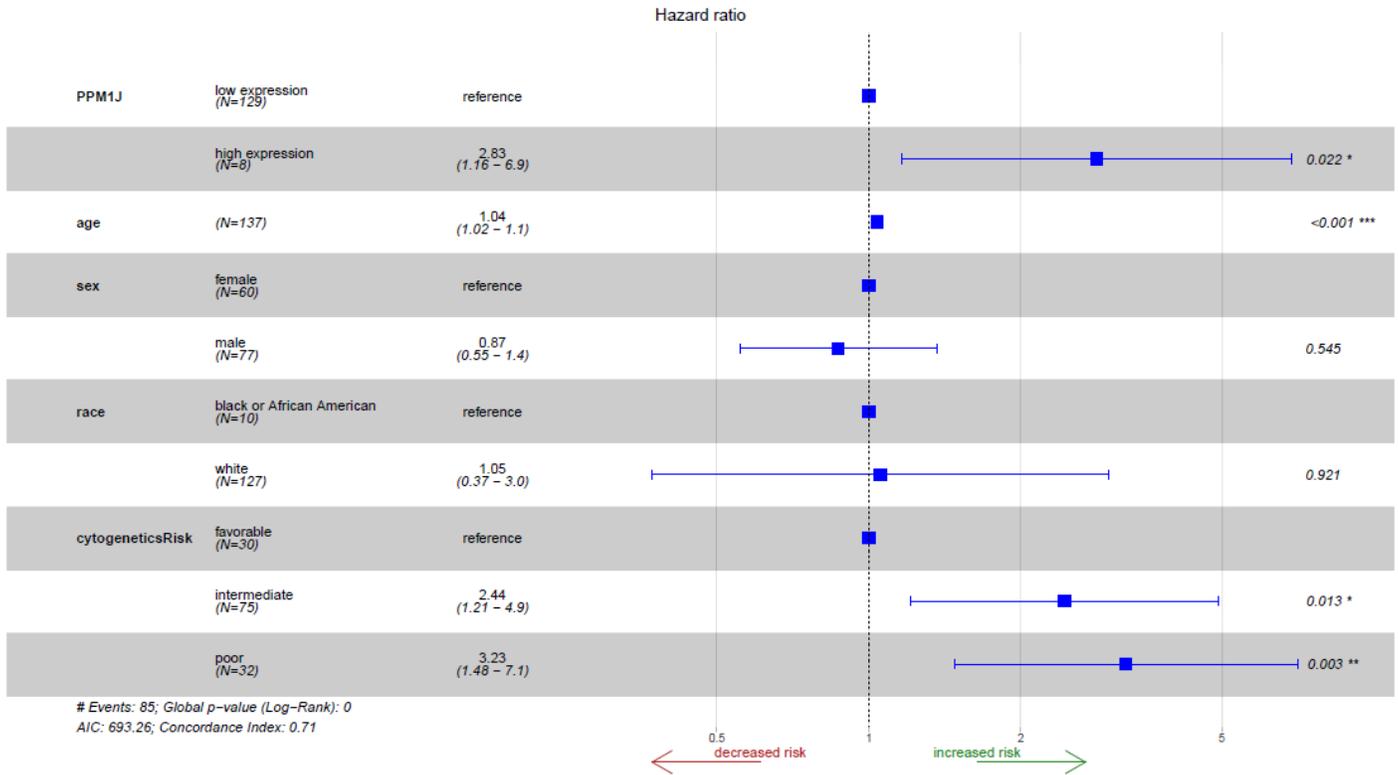


Figure 8

PPM1J forest plot based on Cox proportional hazards model

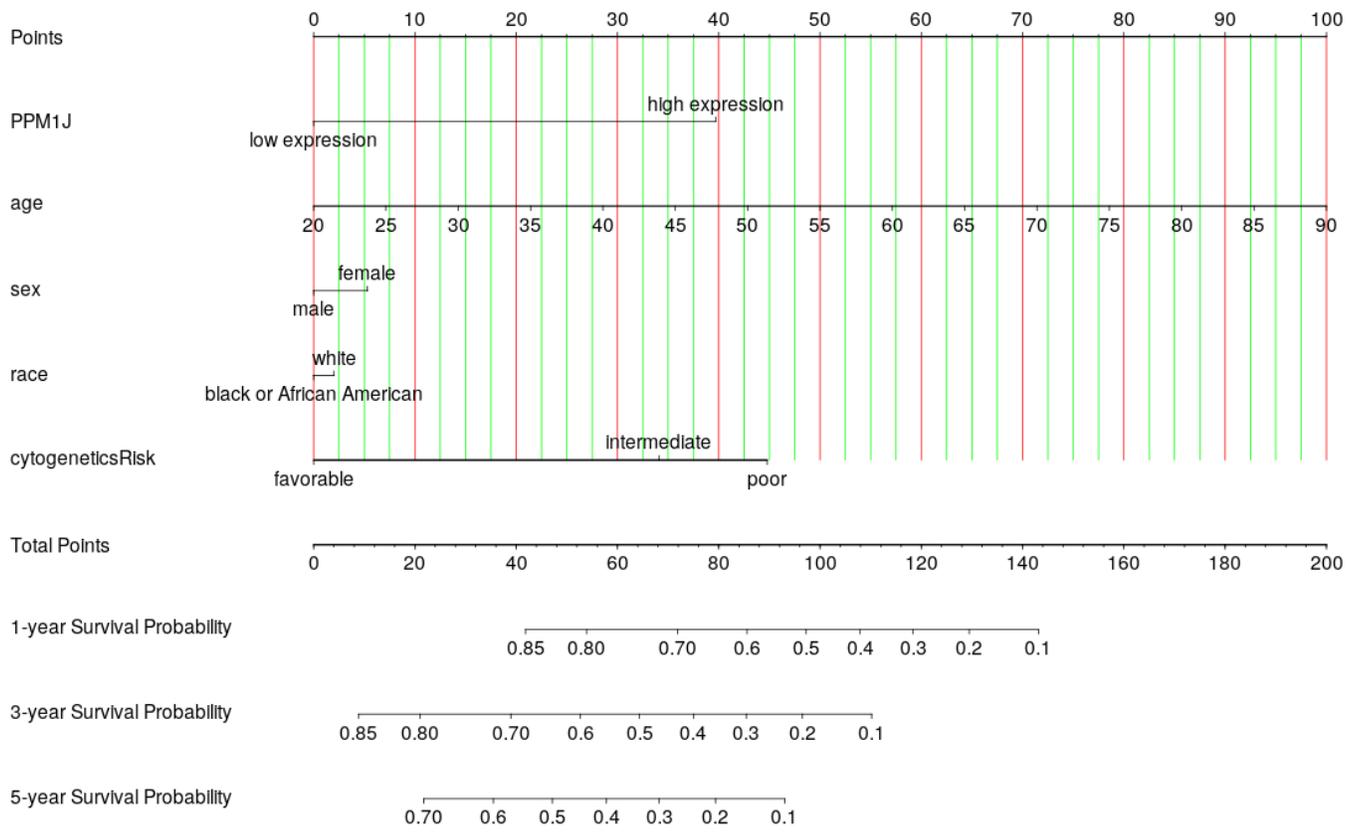


Figure 9

Nomogram fitting the Cox proportional risk model