Analyzing the key gene expression and prognostics values for acute myeloid leukemia

Lingling Shi¹#, Yan Huang²#, Xunjun Huang², Weijie Zhou², Jie Wei², Donghong Deng¹, Yongrong Lai¹

¹Department of Hematology, the First Affiliated Hospital of Guangxi Medical University, Nanning, China; ²Department of Hematology and Rheumatology, People’s Hospital of Baise, Baise, China

Contributions: (I) Conception and design: L Shi, Y Huang; (II) Administrative support: X Huang, W Zhou; (III) Provision of study materials or patients: J Wei, D Deng; (IV) Collection and assembly of data: J Wei; (V) Data analysis and interpretation: L Shi, Y Lai; (VI) Manuscript writing: All authors; (VII) Final approval of manuscript: All authors.

#These authors contributed equally to this work.

Correspondence to: Jie Wei. Department of Hematology and Rheumatology, People’s Hospital of Baise, Baise, China. Email: 654680956@qq.com; Donghong Deng. Department of Hematology, the First Affiliated Hospital of Guangxi Medical University, Nanning, China. Email: ddh_gx@163.com; Yongrong Lai. Department of Hematology, the First Affiliated Hospital of Guangxi Medical University, Nanning, China. Email: laiyongrong@hotmail.com.

Background: Acute myeloid leukemia (AML) is one of the first tumor types sequenced at the whole genome level. However, numbers of the mutated genes expression levels, functions, and prognostics values still unclear.

Methods: To most ordinary mutated genes were analyzed via cancer virtual cohort discovery analysis platform (CVCDAP), and further investigated the mutational conversions, variant allele frequencies (VAF), driver genes, and potential druggable mutated genes in AML. The top mutated gene mRNA expression levels and the relationship between gene expression levels and prognosis for AML patients were performed by Gene Expression Profiling Interactive Analysis (GEPIA). Moreover, we used the UALCAN dataset to confirm the association between gene expression levels and prognosis for AML patients. Enrichment functions of the top mutated genes of AML were analyzed through Metascape. Finally, the role of these defined genes in cancer pathways and potential drug targets were analyzed by gene set cancer analysis (GSCALite).

Results: The top 20 mutated genes for AML included FLT3, HPS3, ABCA6, PCLO, SLIT2, and other ones. Compared to normal control samples, NPM1 and GABRB3 were significantly downregulated in AML samples, but TP53, DNMT3A, HPS3, FLT3, SENP6, and RUNX1 were significantly overexpressed (all these genes P value <0.01). Overexpression of FLT3 and PCLO indicated a poor prognosis, but the overexpression of SLIT3 functioned as a protector for AML via GEPIA. HSP3 indicates the favorable factor for AML, but overexpression of ABCA6 (P=0.066) may act as the adverse factor by UALCAN analysis. Enrichment function analysis shows the functions of defining genes, including negative regulation of cell differentiation, small GTPase mediated signal transduction, and immune system process. Finally, these genes participate in apoptosis, cell cycle, PI3K/AKT, and RAS/MAPK signaling pathway, and FLT3 is sensitive to 5-Fluorouracil, Methotrexate, ATRA. DNMT3A and IDH2 are resistant to Trametinib. RUNX1 and TP53 were sensitive to I-BET-762 and Tubastatin A.

Conclusions: Present study showed overexpression of FLT3, ABCA6, and PCLO indicated the poor prognosis of AML, but overexpression of SLIT3 and HSP3 functioned as an AML protector. There are several drugs and small molecules that target the top 20 mutated genes in AML.

Keywords: Acute myeloid leukemia (AML); bioinformatics analysis; top mutated genes

doi: 10.21037/tcr-20-3177

View this article at: http://dx.doi.org/10.21037/tcr-20-3177
Introduction

Acute myeloid leukemia (AML) is a heterogeneous hematologic cancer characterized by the proliferation and differentiation arrest of immature clonal myeloid cells. The older, the higher the incidence of AML, and the older, the more mortality (1). With the aging society worldwide, the number of AML patients will increase by times, but it is a significant challenge for clinicians to manage the older AML patients. Since intensive chemotherapy is not a suitable option for many older patients with significant comorbidities, baseline organ dysfunction, or low-performance status, the risk of complications and treatment-related mortality is unacceptably high. Herein, we urgently need to understand further tumorigenesis’s underlying mechanisms, treatment resistance, and progression.

Although the advance of cell biology and comprehensive genomic analysis has shown the possible leukemogenesis mechanisms, it is still incompletely understood. The gene aberrations have been confirmed, a vital part of leukemogenesis; these genetics showed a favorable or adverse outcome, and even as the therapy targets of AML patients. For the past decades, many studies have improved some cytogenetic changes that play a crucial role in tumorigenesis and prognosis of AML. For example, the mutations of TP53, WT1, double CEBPA mutation, RUNX1, DNMT3A, and FLT3 were observed in AML (2-7). The genes mutated types, including amplifications, deletions, rearrangements, and point mutations, the different mutated types may lead to a different outcome of AML. For FLT3, the most typical mutated type is internal tandem duplication (ITD) in AML, and FLT3-ITD mutations seem like an adverse biomarker for AML (8,9). Döhner et al. study showed that RUNX1 mutation led to poor prognosis of AML patients. The wild-type NPM1 and the mutated TP53 is the adverse genetic change of AML as well (10). With understanding the fundamental role of these mutations in AML. Some of these mutated genes have functioned as the therapy targets of AML. Enasidenib had been introduced into inhibiting the IDH2 pathway in AML (11). Several regimens that include also have been used to target the FIL3-ITD mutation AML (12-14). Despite some therapies, target regimens have put into practice, the functions and the prognostic role of several mutated genes of AML remain obscure. Herein, we used bioinformatics to analyze the most typical mutated genes for AML and further explore the mutated types, mutational conversions, variant allele frequencies (VAF), potential driver genes, and potential druggable mutated genes. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013).

Methods

CVCDAP analysis

CVCDAP a web-based platform to deliver an interactive and customizable toolbox off the shelf for cohort-level analysis of TCGA and CPTAC public datasets that allows flexible selection of patients sharing common molecular and clinical characteristics across multiple studies as a virtual cohort and provides dozens of built-in customizable tools for seamless, genomic, transcriptomic, proteomic and clinical analysis of a single virtual cohort, as well as, to compare two virtual cohorts with relevance (16). We used the CVCDAP to analyze the top 20 mutated genes of AML and further explore the mutated types, mutational conversions, variant allele frequencies (VAF), potential driver genes, and potential druggable mutated genes. The study was conducted in accordance with the REMARK reporting checklist (available at http://dx.doi.org/10.21037/tcr-20-3177).

Gene expression profiling interactive analysis (GEPIA)

GEPIA is a web-based tool to deliver fast and customizable functionalities for TCGA and GTEx data (17). We analyze the top 20 mutated gene expression levels and prognostics role for AML via GEPIA.

UALCAN analysis

The UALCAN dataset is user-friendly, which allows us to perform: (I) analyze relative expression of a query gene(s) across tumor and normal samples, as well as in various tumor sub-groups for individual cancer stages, tumor grade, race, body weight, or other clinicopathologic features; (II) estimate the effect of gene expression level and clinicopathologic features on patient survival (18). To
further investigate the top 20 mutated gene expression levels and prognostics role, we used the UALCAN dataset to perform the mutated gene expression levels and prognostics values for AML.

**Metascape analysis**

Metascape is a web-based portal designed to supply a comprehensive gene list annotation and analysis resource for experimental biologists. In terms of design features, Metascape combines functional enrichment, interactome analysis, gene annotation, and membership search to use over 40 independent knowledge bases within one integrated portal. We used it to analyze the potential biological function (biological process) of the top 20 mutated genes, the protein-protein network (PPI), and gene-related diseases.

**Gene Set Cancer Analysis (GSCAlite)**

GSCAlite is a web that can analyze a set of genes in cancers with the following functional modules: (I) differential expression in tumor versus normal, and the survival analysis; (II) genomic variations and their survival analysis; (III) gene expression associated cancer pathway activity; (IV) miRNA regulatory network for genes; (V) drug sensitivity for genes; (VI) normal tissue expression and eQTL for genes. We used GSCAlite to investigate the relations between top mutated gene methylation, copy number variation (CNV), single-nucleotide variant (SNV), gene expression level, and the prognosis of AML patients. We used it to analyze the top 20 mutated genes’ role in several cancer pathways and the potential small molecules and drugs that target the mutated genes.

**Statistical analysis**

The gene expression and survival analysis P value <0.05 is considered a significant difference in our study.

**Results**

**The results of CVCDAP analysis**

The analysis of top mutated genes of AML via CVCDAP dataset and the results showed top 20 mutated genes for AML including *BPIFC*, *DNMT3A*, *TTN*, *NPM1*, *RUNX1*, *TP53*, *IDH2*, *FLT3*, *HPS3*, *SENP6*, *ABCa6*, *ASXL1*, *BRWD1*, *DNAH11*, *GABRG3*, *GDI2*, *KRAS*, *MUC16*, *PCLO*, and *SLIT2*. The altered types of these genes included missense mutation, nonsense mutation, and others listed in *Figure 1A*. The mutational conversions showed that C>T was the most typical type (*Figure 1B*). The results of the VAF showed the top 10 genes (*Figure 1C*). The driver genes showed in *Figure 1D*. The potential drug targets showed that ABCA6, BPIFC, DNMT3A, FLT3, and GABRG3 are the candidate target therapy of AML (*Figure 1E*).

**Transcriptional levels of top 20 mutated genes for AML**

We used the GEIPA database to analyze the top 20 mutated genes’ expression levels between the tumor and the normal control samples. Compared to normal samples, NPM1 and GABRB3 were significantly downregulated in AML samples, but TP53, DNMT3A, HPS3, FLT3, SENP6, and RUNX1 were significantly overexpressed (all these genes P value <0.01), but the lower gene expression levels did not have a significant difference between AML and normal samples (*Figure 2*).

**The prognostic role of top 20 mutated genes**

We used the GEPIA dataset to investigate the prognostics values of the top 20 mutated genes. The results showed that the expression levels of SLIT3, FLT3, and PCLO had a considerable influence on the overall survival (OS) of AML patients. Overexpression of FLT3 and PCLO indicated the poor prognosis of AML, but the overexpression of SLIT3 functioned as a protector for AML (*Figure 3*). To confirm the top 20 mutated genes’ prognostics values, we used the UALCAN dataset to reanalyze the relation between gene expression levels and OS of AML. The results showed that overexpression of HSP3 indicated the favorable factor for AML, but overexpression of ABCA6 (P=0.066) may act as an adverse factor (*Figure 4*).

**Enrichment function analysis**

To explore the functions of the top 20 mutated genes for AML, we used the Metascape dataset to perform the analysis. The results showed that the top 20 mutated genes’ biological functions were cellular component organization or biogenesis, immune system process, response to a stimulus, negative regulation of the biological process, metabolic process, and developmental process (*Figure 5*).
Figure 1 The mutational conversions, variant allele frequencies, driver genes, and potential druggable mutated genes of top 20 mutated genes. (A) The results of the top 20 mutated genes for AML (CVCDAP); (B) the overall distribution of six types of mutational conversions for the top 20 mutated genes; (C) the variant allele frequencies of the top 10 mutated genes in AML; (D) the size of a point represents the number of clusters found in a given gene. Genes satisfying FDR thresholds are colored in red, with others in blue; (E) the potential druggable mutated genes in categories of the top 20 mutated genes for AML. AML, acute myeloid leukemia.
Figure 2 The mRNA expression levels of the top 20 mutated genes for AML (GEPIA). AML, acute myeloid leukemia; GEPIA, Gene Expression Profiling Interactive Analysis.
Figure 3 The relations between the top 20 mutated gene expression levels and the prognostic role for AML (GEPIA). AML, acute myeloid leukemia; GEPIA, Gene Expression Profiling Interactive Analysis.
Figure 4 The relations between the top 20 mutated gene expression levels and the prognostic role for AML (UALCAN). AML, acute myeloid leukemia.
The main biological functions include regulation of the microtubule-based process, negative regulation of cell differentiation, small GTPase mediated signal transduction, protein secretion, heart morphogenesis, and the Kyoto Encyclopedia of Genes and Genomes (KEGG) which is AML (Figure 5B). These genes were significantly related to multiple hematological malignancies, including myeloid leukemia and myeloproliferative disease (Figure 5C). The target transcriptional factors that regulated these genes were IGLV5 37, TOP2B, GTF2A2, BANP, and ZNF140 (Figure 5D). The PPI network showed that the top 20 mutated genes’ primary biological functions include microtubule-based processes, negative regulation of cell differentiation, small GTPase mediated signal transduction, protein secretion, and heart morphogenesis as well (Figure 6A,B,C).

GSCAlite analysis

We used GSCAlite to explore the SNV, CNV, the methylation of the top 20 mutated genes, and the relation
Figure 6 The protein-protein network of the top 20 mutated genes in AML (Matascape). (A) Colored by cluster ID; (B) colored by P value, where terms containing more genes tend to have a more significant P value; (C) protein-protein interaction network and MCODE components identified in the gene list. AML, acute myeloid leukemia.

Figure 7 The methylation status of the top 20 mutated genes affects mRNA expression and prognosis (GSCAlite). (A) The relationship between the top 20 mutated genes, methylation and gene expression; (B) the relationship between top 20 mutated genes methylation and OS of AML. AML, acute myeloid leukemia; GSCAlite, Gene Set Cancer Analysis; OS, overall survival.

between them and the prognosis for AML. The results showed that DNMT3A, with the most frequency of SNV, further analysis showed that the SNV of TP53, RUNX1, and DNMT3A indicated the poor prognosis of AML (Figure 7A,B). The top 20 mutated genes CNV status and the relation of CNV and genes mRNA expression showed in Figure 8A,B. The relations between methylation and OS showed that DNMT3A hypermethylation indicated the adverse biomarker, but PCLO hypermethylation led to a satisfactory prognosis (Figure 9A,B). All these genes play a
Figure 8 The status of CNV and the association between CNV and top 20 mutated gene expression (GSCAlite). (A) CNV pie distribution; (B) the association between CNV and top 20 mutated gene expression. CNV, copy number variations.

Figure 9 The status of the top 20 mutated genes mutations frequency and the effect of mutations of survival risk for AML (GSCAlite). (A) Top 20 mutated gene mutation frequency in AML; (B) the effect of mutations of survival risk for AML. AML, acute myeloid leukemia.
crucial role in several cancer pathways, including apoptosis, cell cycle, DNA damage response.

EMT, hormone AR, hormone ER, PI3K/AKT, RAS/MAPK, RTK, and TSC/mTOR, and they served as the inhibitors or activators for these cancer pathways (Figure 10A,B). The drug sensitivity is shown in Figure 11.

Discussion

Despite the development of therapy regimens, the AML still has the highest incidence and cancer-related death for hematological malignancies. There is remaining a long way to fully understand the AML tumorigenesis, therapy resistance, and progress. Many studies’ results showed that genetic changes functioned as one of the most critical promotors for tumorigenesis and therapy resistance from the past decades. With the advance of sequence methods, increasingly mutated genes have been observed in AML, and the biological functions of mutated genes have preliminarily learned. Herein, we used bioinformatics to analyze the top mutated genes in AML. And further investigate the prognostics value, functions, and potential target therapy.

In the present study, the results showed the top 20
Translational Cancer Research, Vol 9, No 11 November 2020 7295

Figure 11 The analysis of drug sensitivity top mutated genes for AML (GSCAlite). AML, acute myeloid leukemia.

Among the mutated genes of AML, including BP1FC, DNMT3A, TTN, NPM1, RUNX1, TP53, IDH2, FLT3, HPS3, SENP6, ABCA6, ASXL1, BRWD1, DANA11, GABRG3, GDI2, KRAS, MUC16, PCLO, and SLIT2. Among these mutated genes, compared to normal samples, the results showed NPM1 and GABRB3 were significantly downregulated, but TP53, DNMT3A, HPS3, FLT3, SENP6, and RUNX1 were significantly overexpressed in AML samples. Referring to the relationship between gene expression levels and prognosis of AML patients, the results showed overexpression of FLT3, ABCA6, and PCLO indicated the poor prognosis of AML, but the overexpression of HSP3 and SLIT3 acted as a protector for AML. Qiu et al. study showed that PCLO might lead to etoposide resistance for small-cell lung cancer (19). Lohr et al. revealed PCLO could be detectable in diffuse large B-cell lymphoma (DLBCL) as well, but the function of PCLO in DLBCL unknown (20). According to ABCA6, Heddit et al. research showed overexpression of ABCA6 led to the poor outcome for epithelial ovarian cancer (EOC) (21). The upregulation of ABCA6 was found in mantle cell lymphoma (MCL), but the role of ABCA6 is still obscure (22). Gołos et al. study revealed that down-expression of SLIT in the bone marrow of AML patients might indicate its expression alterations in AML (23). There were one-third of AML patients with mutated FLT3 (24). In-frame duplications of 3 to >400 base pairs (bp), known as internal tandem duplications (ITDs), are the most common mutated types for FLT3 in AML (25). And the FLT3-ITD acted as the adverse biomarker for AML (26). Among the top 20 mutated genes, the FLT3 mutations is the one that attracted more and more researcher to learn the role, functions, and as the target therapy for AML. The mutated types of FLT3 include tyrosine kinase domain (TKD) mutations and ITDs. Both FLT3-ITD and FLT3-TKD mutations constitutively activate FLT3 kinase activity, resulting in proliferation and survival of AML (27). Herein, patients with FLT3-ITD mutations indicated with a shorter OS and high-risk relapse compared to without mutation (28). Interestingly, the role of FLT3-TKD mutations is still unclear. To date, there were several multitargeted tyrosine kinase inhibitors (TKIs) had been put into treating the FLT3-ITD-mutated AML (29-31). Early-phase studies showed that first-generation FLT3 inhibitors as monotherapy did not achieve the satisfy results but increased the toxicity. Conversely, sunitinib in combination with standard induction and cytarabine/daunorubicin consolidation therapy, there were 50% of patients with FLT3-ITD mutations and 38% with FLT3-
TKD mutations achieved CR (31). These results indicated that TKIs combination with standard chemotherapy may improve the response rates of FLT3-mutated patients. Regrettably, the research of PCLO, HPS3, and ABCA6 in cancer still less; it urgently needs to be further explored.

The enrichment functions of the top mutated genes of AML show that the significant biological functions are cellular component organization or biogenesis, immune system process, response to a stimulus, metabolic process, regulation of the microtubule-based process, negative regulation of cell differentiation, small GTPase mediated signal transduction, and protein secretion. Among these functions, the immune system process, metabolic process, negative regulation of cell differentiation, and small GTPase mediated signal transduction (32-35). And the immune therapy and target small GTPase pathway to treat cancer have been put into practice (35,36).

Tumorigenesis is a complicated process. Several important regulators, including TFs, also take part in tumorigenesis via regulation of gene expression. From our study results, there were five TFs that, including IGLV537, TOP2B, GTF2A2, BANP, and ZNF140, significantly targeted the top 20 mutated genes. Girardot et al. research showed that GTF2A2 was detectable in platelets of myeloproliferative neoplasm (MPN) patients, not in platelets of normal control samples (37). However, no study investigates the functions of IGLV537, TOP2B, BANP, and ZNF140 in cancer.

The final goal of cancer research is to explore the more efficient and safe therapy regimens for cancer patients. So, we used the GSCAlite to explore the role of the top 20 mutated genes in the define cancer pathway and the Sensitivity to drugs or/and small molecules. The results showed that ABCA6, FLT3, HPS3, RUNX1, and SLIT2 had a double role for DNA damage response. To date, several studies explored therapy cancer via targeting DNA damage response (38-40). Apoptosis, cell cycle, PI3K/AKT, and RAS/MAPK also significant pathways for multiple cancer types tumorigenesis, including AML (41-44). Finally, we investigated the potential small molecules or/drugs that targeted these genes. The results showed that FLT3 was sensitive to Navitoclax, 5-Fluorouracil, Methotrexate, AZD7762, ATRA. DNMT3A and IDH2 are resistance to Docetaxel, 17-AAG, and Trametinib. RUNX1, TP53, SENP6, GDI2, TTN, BRWD1 is sensitive to I-BET-762, and Tubastatin A. These clues indicated the top 20 mutated genes sever as an essential role in AML, but there was a limitation of our study, lacking the validation of these mutated gene expression levels in AML and normal control samples.

Conclusions

Our study preliminary explored the top 20 mutated genes of AML and further investigated the top mutated gene expression levels and the prognostics value for AML. BPIFC, DNMT3A, TTN, NPM1, RUNX1, TP53, IDH2, FLT3, HPS3, SENP6, ABCA6, ASXL1, BRWD1, DNAH11, GABRG3, GDI2, KRAS, MUC16, PCLO, and SLIT2 were the top 20 mutated genes for AML via CVCDAP analysis. Overexpression of FLT3, ABCA6, and PCLO indicated the poor prognosis of AML, but overexpression of SLIT3 and HSP3 functioned as a protector for AML. The enrichment function analysis of the top 20 mutated genes indicates significant parts of the immune system process, metabolic process, negative regulation of cell differentiation, small GTPase mediated signal transduction, and protein secretion. All the top 20 mutated genes also take part in several essential cancer pathways, including DNA damage response, apoptosis, cell cycle, PI3K/AKT, and RAS/MAPK pathway. The drug sensitivity analysis showed that FLT3 was sensitive to Navitoclax, 5-Fluorouracil, Methotrexate, AZD7762, ATRA. DNMT3A and IDH2 are resistance to Docetaxel, 17-AAG, and Trametinib. RUNX1, TP53, SENP6, GDI2, TTN, BRWD1 is sensitive to I-BET-762, and Tubastatin A. These clues indicated the top 20 mutated genes sever as an essential role in AML, but there was a limitation of our study, lacking the validation of these mutated gene expression levels in AML and normal control samples.

Acknowledgments

Funding: None.

Footnote

Reporting Checklist: The authors have completed the REMARK reporting checklist. Available at http://dx.doi.org/10.21037/tcr-20-3177

Conflicts of Interest: All authors have completed the ICMJE uniform disclosure form (available at http://dx.doi.org/10.21037/tcr-20-3177). The authors have no conflicts of interest to declare.

Ethical Statement: The authors are accountable for all treating the patients with these genes’ mutations.
aspects of the work in ensuring that questions related to the accuracy or integrity of any part of the work are appropriately investigated and resolved. The study was conducted in accordance with the Declaration of Helsinki (as revised in 2013).

Open Access Statement: This is an Open Access article distributed in accordance with the Creative Commons Attribution-NonCommercial-NoDerivs 4.0 International License (CC BY-NC-ND 4.0), which permits the non-commercial replication and distribution of the article with the strict proviso that no changes or edits are made and the original work is properly cited (including links to both the formal publication through the relevant DOI and the license). See: https://creativecommons.org/licenses/by-nc-nd/4.0/.

References


