Identification of key pathways and genes in TP53 mutation acute myeloid leukemia: evidence from bioinformatics analysis

Rui Huang1,*
Xiwen Liao2,*
Qiaochuan Li1

1Department of Hematology, 2Department of Hepatobiliary Surgery, The First Affiliated Hospital of Guangxi Medical University, Nanning, Guangxi, People’s Republic of China

*These authors contributed equally to this work

Background: Tumor protein p53 (TP53) mutations are not only a risk factor in acute myeloid leukemia (AML) but also a potential biomarker for individualized treatment options. This study aimed to investigate potential pathways and genes associated with TP53 mutations in adult de novo AML.

Methods: An RNA sequencing dataset of adult de novo AML was downloaded from The Cancer Genome Atlas database. Differentially expressed genes (DEGs) were identified by edgeR of the R platform. Key pathways and genes were identified using the following bioinformatics tools: gene set enrichment analysis (GSEA), gene ontology (GO), the Kyoto Encyclopedia of Genes and Genomes (KEGG), Search Tool for the Retrieval of Interacting Genes/Proteins, and Molecular Complex Detection.

Results: GSEA suggested that TP53 mutations were significantly associated with cell differentiation, proliferation, cell adhesion biological processes, and MAPK pathway. In total, 1,287 genes were identified as DEGs. GO and KEGG analysis suggested that upregulation of DEGs was significantly enriched in categories associated with cell adhesion biological processes, Ras-associated protein 1, PI3K–Akt pathway, and cell adhesion molecules. The top ten genes ranked by degree, CDH1, BMP2, KDR, LEP, CASR, ITGA2B, APOE, MNX1, NMU, and TRH, were identified as hub genes from the protein–protein interaction network. Survival analysis suggested that patients with TP53 mutations had a significantly increased risk of death, while the mRNA expression level in patients with TP53 mutation was similar to those carrying TP53 wild type.

Conclusion: Our findings have indicated that multiple genes and pathways may play a crucial role in TP53 mutation AML, offering candidate targets and strategies for TP53 mutation AML individualized treatment.

Keywords: bioinformatics analysis, acute myeloid leukemia, TCGA, RNA sequencing, TP53 mutation

Introduction

Tumor protein p53 ([TP53] p53, encoded by the human gene TP53) is a tumor suppressor protein, which is involved in transcriptional activation, DNA binding and oligomerization with other proteins. The functions of p53 are primarily in inducing cell cycle arrest, apoptosis, senescence, DNA repair, and changes in metabolism.1,2 Somatic mutation of the TP53 gene is one of the most frequent alterations in human cancers, and the majority of cancer-associated TP53 mutations are missense mutations in the DNA-binding domain.3,4 Missense mutations of the TP53 gene result in a mutant p53 protein, which loses wild-type p53 tumor suppressive function and leads to a deleterious gain of function (GOF).5 Morton et al’s research on pancreatic cancer suggest...
that mutant p53 GOF activities may play a role in increasing the metastatic potential and/or promoting metastatic tumor cells’ survival and proliferation. The GOF activity of mutant p53 may depend largely on multiple signals for mutant p53 protein stabilization, thereby, resulting in sustained activation of multiple biologic mechanisms in cancers.

Recent studies have shown that TP53 mutations are frequently detected in acute myeloid leukemia (AML) patients, with a mutation rate of 5%–8% in newly diagnosed patients and 30%–40% in therapy-related AML (t-AML). The TP53 pathway is frequently dysfunctional in AML via MDM2 and MDM4 overexpression, as well as by TP53 mutations/deletions. Overexpression of these proteins can regulate p53 stabilization and lead to poor outcomes, regardless of TP53 mutational status. A previous study by Wong et al has revealed that TP53 mutants were selectively enriched in t-AML or therapy-related myelodysplastic syndrome patients, but their results do not support the hypothesis that cytotoxic therapy is directly responsible for the induced mutation of TP53. In addition, they used next-generation sequencing techniques to detect TP53 mutations in blood or bone marrow before any chemotherapy and demonstrated a potential mechanism whereby hematopoietic stem/progenitor cells with heterozygous TP53 mutations may have a competitive advantage in the presence of cytotoxic therapy. TP53 alterations are significantly associated with multiple risk factors, including older age, complex karyotype, white blood counts, and specific DNA copy number alterations. In addition, multivariate analysis in these studies suggests that TP53 mutation was an important prognostic factor leading to a poor overall survival (OS) and disease-free survival (DFS) in AML patients. Elderly AML patients often have a TP53 mutation and genetic markers that are associated with reduced sensitivity of less responsiveness to standard cytotoxic therapy which leads to poorer outcome. Welch et al reported that approximately 67% of patients with an unfavorable-risk cytogenetic profile and 100% of TP53 mutation AML patients were responsive to decitabine treatment, which seems to narrow the prognosis gap between patients with TP53 mutations and TP53 wild type.

Non-mutational inactivation of p53 is frequent in AML. This mechanism can be used for wild-type TP53 AML treatment and occurs via overexpression of MDM2 and MDM4. However, AML patients with mutated p53 were resistant to MDM2 inhibitor MI219. The treatment strategy of TP53 mutation AML was consequently different from TP53 wild-type AML. It is therefore necessary to investigate the difference between the TP53 mutation AML and the associated wild type. According to previous studies, TP53 high-frequency mutations reported in AML patients can be used as a prediction factor for prognosis and individualized treatment options. However, the TP53 mutation’s occurrence and the changes in biological processes they cause, remain unclear. In the present study, an RNA sequencing (RNA-Seq) dataset of adult de novo AML was employed to identify the key pathways and genes associated with TP53 mutations using bioinformatics analysis approaches.

Materials and methods

RNA-Seq data
An RNA-Seq dataset of adult AML, which included human AML patients’ whole transcriptome sequencing dataset and corresponding survival profiles, was downloaded from The Cancer Genome Atlas (TCGA) database (https://gdc-portal.nci.nih.gov). All of the AML patients in TCGA were adult de novo AML patients and the sequencing dataset came from bone marrow tissue analyses at diagnosis. The corresponding information related to patients with TP53 mutation was obtained from the cBioPortal for Cancer Genomics website (http://www.cbioportal.org/index.do).

Gene set enrichment analysis (GSEA)
To investigate the effect of TP53 mutations on various biological function gene sets in adult AML patients, differences in gene mRNA expression levels of biological functional annotation and pathways between TP53 mutation and wild-type patients were analyzed by GSEA v2.2.3 (http://software.broadinstitute.org/gsea/downloads.jsp). Reference gene sets from the Molecular Signatures Database (MSigDB) of c2 (c2.cp.kegg.v5.2 symbols.gmt) and c5 (c5.bp.v5.2 symbols.gmt; c5.mf.v5.2 symbols.gmt; c5.cc.v5.2 symbols.gmt; consist of genes annotated by the same GO terms), respectively. The MSigDB of c2 is a pathway gene set, which was curated from publications and extracted from canonical pathways and experimental signatures, whereas the MSigDB of c5 was constructed on genes annotated by the same GO terms. The number of permutations was set at 1,000. Enrichment results satisfying a nominal P-value cutoff of <0.05 with a false discovery rate (FDR) <0.25 were considered statistically significant.

Identification of differentially expressed genes (DEGs)
EdgeR, an R package for examining differential expression of RNA-Seq count data, was used according to the user’s guide.
for screening differential expression of genes at gene levels between TP53 mutation and wild-type AML patients.\textsuperscript{27,28} DEGs were identified with the following criterion: $|\log_2$ fold change (FC) $| \geq 2$; both the $P$-value and FDR $< 0.05$. The DEGs were used for further bioinformatics analysis. A heat map and volcano plot of the DEGs were drawn via the ggplots package in the R platform.

**Functional annotation and pathway enrichment analysis of DEGs**

Database for Annotation, Visualization, and Integrated Discovery (DAVID) v.6.8 (https://david.ncifcrf.gov/tools.jsp) was used to annotate input genes, classify gene functions, identify gene conversions, and carry out gene ontology (GO) term analysis.\textsuperscript{29} To identify the DEGs' functional annotation, we analyzed GO terms and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment with DAVID, while specifying a $P$-value $< 0.05$ for statistical significance.

**Protein–protein interactions (PPI) network and module analysis**

The Search Tool for the Retrieval of Interacting Genes/Proteins (STRING) database (http://string.embl.de/), which provides a critical assessment and integration of PPI,\textsuperscript{30} was used to assess the direct (physical) and indirect (functional) associations of DEGs. Interactive relationships among DEGs with an interaction score $> 0.4$ was considered statistically significant. The PPI network was then used for module screening by Molecular Complex Detection (MCODE) (scores $> 3$ and nodes $> 4$) in Cytoscape, a bioinformatics integration platform.\textsuperscript{31} Furthermore, we also analyzed the GO terms and KEGG pathway enrichment for DEGs in the top three modules, respectively.

**Statistical analysis**

The Student’s $t$-test was used to compare the TP53 mRNA expression level between TP53 mutation and wild-type AML bone marrow tissue. The Kaplan–Meier method with log-rank test was used for calculating the clinical outcome between different gene expression groups. HR and the 95% CIs were calculated from the Cox proportional hazards regression model. FDR in edgeR and GSEA were adjusted for multiple testing with the Benjamini–Hochberg procedure to control FDR, respectively.\textsuperscript{32–34} A value of $P<0.05$ was considered statistically significant. All the statistical analyses were conducted with SPSS version 20.0 (IBM Corporation, Armonk, NY, USA) and R 3.3.0.

**Results**

**Data source**

Information for 151 patients with adult de novo AML and corresponding bone marrow RNA-Seq dataset were obtained from TCGA database. A total of 140 patients with complete follow-up profiles were available. There were eleven AML patients (7%) with TP53 mutations and the rest were TP53 wild type, these data were obtained from the cBioPortal for Cancer Genomics website.

**GSEA**

To investigate the effect of TP53 mutations on the prognosis of AML patients, the effects of TP53 mutations on various biological functional gene sets were analyzed by the GSEA approach. The GSEA results are shown in Figure 1 and Table S1. In the GSEA analysis of GO enrichment (Figure 1A–J), biological processes pertaining to positive regulation of cell differentiation, regulation of myeloid cell differentiation, negative regulation of cell proliferation, leukocyte differentiation and activation, and cell adhesion, were significantly enriched. This suggests that TP53 mutations may contribute to disease progression and affect prognosis by influencing cell differentiation, proliferation, and cell adhesion in AML patients. However, the GO enrichment analysis of molecular function was significantly enriched in receptor signaling proteins, serine-threonine kinase activity, and protein tyrosine kinase activity. Furthermore, the cellular component was enriched for adherens junction and cell fraction. In the GSEA analysis of KEGG pathways, the TP53 mutation group was associated with the MAPK signaling pathway and pathways in cancer (Figure 1K–L).

**Identification of DEGs**

Overall, RNA-Seq datasets from eleven TP53 mutation-bearing and 140 TP53 wild-type AML patients were used for DEG screening. The criteria of the DEGs were set as follows: i) $|\log_2$ FC $| \geq 2$; ii) both the $P$-value and FDR value of edgeR $< 0.05$. A total of 1,287 DEGs met the criteria, of which, 1,067 were upregulated and 220 were downregulated (Table S2). The volcano plot of the DEGs is shown in Figure 2, while the heat map is shown in Figure S1.

**GO and KEGG of DEGs**

Altogether, 220 downregulated and 1,067 upregulated DEGs were submitted for further GO and KEGG pathway analyses with DAVID, respectively. The GO analysis of downregulated DEGs (Figure 3) suggested significant enrichment in skeletal system development, cartilage development involved in endochondral bone morphogenesis, central nervous
Figure 1 GSEA results of TP53 mutations in AML patients.

Note: Results are shown for c5 (A-J) and c2 (K, L) reference gene sets.

Abbreviations: GSEA, gene set enrichment analysis; AML, acute myeloid leukemia; FDR, false discovery rate; GO, gene ontology; KEGG, Kyoto Encyclopedia of Genes and Genomes.
system neuron differentiation, negative regulation of neuron differentiation, cell–cell signaling, multicellular organism development, and positive regulation of developmental growth. However, with regard to molecular function and cellular components, DEGs were significantly enriched in extracellular space, sequence-specific DNA binding, transcriptional activator activity, RNA polymerase II distal enhancer sequence-specific binding, and calcium ion binding. No pathways were significantly enriched in the KEGG enrichment analysis of downregulated DEGs.

GO and KEGG analysis was also carried out for upregulated DEGs. The upregulated DEGs were significantly enriched in proteolysis, complement activation, cell adhesion, and Fc-gamma receptor signaling pathway involved in phagocytosis biological process (Figure 4A). Cellular component enrichment suggests that upregulated DEGs were significantly enriched in extracellular region, proteinaceous extracellular matrix, immunoglobulin complex, and circulating and blood microparticle. Whereas, molecular function was enriched in antigen binding and serine-type endopeptidase activity (Figure 4A). In KEGG pathway analysis, upregulated DEGs were significantly enriched in
extracellular matrix (ECM)-receptor interaction, Rap1 signaling pathway, PI3K-AKT signaling pathway, hematopoietic cell lineage, calcium signaling pathway, cell adhesion molecules, and chemical carcinogenesis (Figure 4B).

PPI network and module analysis
To investigate the interaction and hub genes of DEGs, protein–protein interactions were constructed using STRING (Figure S2). The top ten genes ranked by degree were identified as hub genes. These hub genes included CDH1, also known as E-cadherin, BMP2, KDR, LEP, CASR, ITGA2B, APOE, MXN1, NMU, and TRH. CDH1 had the highest degree of nodes among the hub genes with 34. Modules of genes in PPI network were identified by the MCODE plugin in Cytoscape. The top three significant modules were selected for GO term and KEGG pathway enrichment analysis (Table S3 and Figure 5A–D, respectively). KEGG pathway enrichment by DAVID suggested that the top three module genes were mainly related to neuroactive ligand-receptor interactions, calcium signaling pathway, cAMP signaling pathway, and olfactory transduction pathways. Functional annotation in module 1 genes was enriched in the integral component of plasma membrane, phospholipase C-activating G-protein coupled receptor (PLC-activating GPCR) signaling pathway, positive regulation of cytosolic calcium ion concentration, and G-protein coupled receptor (GPCR) signaling pathway. The genes of module 2 were mainly enriched in the integral component of plasma membrane, GPCR signaling pathway, coupled to cyclic nucleotide second messenger, adenylate cyclase-inhibiting G-protein coupled receptor signaling pathway, somatostatin signaling pathway, and cell–cell signaling. In addition, GO term enrichment of module 3 genes was also associated with olfactory receptor activity, G-protein coupled receptor activity, and the G-protein coupled receptor signaling pathway (Table S3).

Survival analysis
In this study, we observed that TP53 mRNA expression was similar between TP53 mutation and wild-type patients’ bone marrow tissue (Figure 6A). Overall, 151 patients with a 7.3% lost follow-up rate were included in the survival analysis. A total of 87 (57.6%) patients died before the last follow-up date with a median survival time (MST) of 19 months. OS analysis for AML patients grouped by TP53 mutation revealed that AML patients with TP53 mutations had a shorter MST than those without TP53 mutations (TP53 mutations vs TP53 wild type, 31 days vs 608 days, log-rank $P<0.0001$). Univariate Cox proportional hazards regression analysis also suggested that TP53 mutations were significantly associated with a poor clinical outcome and an increased risk of death, compared to those patients without TP53 mutations ($P<0.0001$, HR $=4.510$, 95% CI $=2.266–8.979$, Figure 6B).

Discussion
AML is characterized by proliferative, clonal, abnormally differentiated, and occasionally poorly differentiated cells of the hematopoietic system, which infiltrate the bone marrow,
Key pathways and genes in TP53 mutation AML

Blood, and other tissues. AML differentiation can be induced by ATP depletion and the subsequent activation of DNA damage signaling via the ATR/Chk1-dependent and p53-independent pathway. p53 tumor suppressive function limits proliferation in response to cellular stress, whereas p53 inactivation enhances proliferation of myeloid progenitor cells in combination with aberrant self-renewal to promote AML. TP53 mutations, which cause p53 suppressive function and p53 pathway dysfunction, affect AML cell differentiation and proliferation. Hepatocyte growth factor promotes proliferation, invasion, and metastasis of AML cells by regulating the PI3K-Akt and MAPK/ERK signaling pathway. In addition, drug resistance of AML stem cells can be modulated by survivin expression via the MAPK signaling

![Figure 5](image)

Top three modules from the PPI network.

Notes: (A) PPI network of module 1; (B) PPI network of module 2; (C) PPI network of module 3; (D) enrichment pathways of the top three modules. The different protein sizes in the network imply the degree of specific gene in the PPI network.

Abbreviation: PPI, protein–protein interactions.

![Figure 6](image)

The comparison of mRNA expression and survival between AML patients with TP53 mutation and wild type.

Notes: (A) The mRNA expression of the TP53 gene in AML patients’ bone marrow tissue between TP53 with mutations and the wild type; (B) Kaplan–Meier survival curves for AML patients stratified by TP53 mutation.

Abbreviation: AML, acute myeloid leukemia.
pathway. GSEA analysis in the present study suggests that TP53 mutations were significantly associated with the regulation of cell differentiation, especially for myeloid cell differentiation, negative regulation of cell proliferation, cell adhesion, and leukocyte differentiation and activation.

The GO term analysis indicated that downregulated DEGs in TP53 mutation AML patients were related to biological processes of cell differentiation and development. The GO term enrichment analysis suggests that upregulated DEGs were related to complement activation, and cell adhesion biological processes. The KEGG enrichment analysis implies that ECM-receptor interaction, the Rap1 signaling pathway, PI3K-Akt signaling pathway, calcium signaling pathway, and cell adhesion molecules were significantly enriched in upregulated DEGs’ pathways.

ECM-receptor interaction and cell adhesion molecule signaling pathways were mainly involved in cell adhesion, and the genes in those enrichments included genes involved with the vascular cell adhesion molecule, cadherin and integrin family (VCAM1, ITGB3, ITGB1, ITGA2B, CDH1). Cell-cell and cell-matrix interactions are mainly mediated by adhesion molecules including those of the selectin, integrin, and cadherin families. In addition, the integrin family proteins mainly mediate the interaction between cells and the matrix, whereas, the calcium family proteins are mainly involved in the adhesion between homogeneous cells. These adhesion molecules have been demonstrated to be involved in intercellular and intracellular signaling, and through the regulation of the cell cycle and apoptosis, affect response to antitumor therapy and result in drug resistance. Evidence shows that anti-adhesion therapy can improve tumor cell resistance to some extent, restore the sensitivity of tumor cells to chemotherapeutic drugs, and have synergistic and sensitization effects in traditional chemotherapy. Combretastatins cause the downregulation of bone marrow endothelial cell (BMEC) adhesion molecules VCAM-1 and VE-cadherin, then promote BMECs to enter G2/M of the cell cycle. Furthermore, when combined with combretastatin, cytotoxic chemotherapy can significantly induce AML cell death. AML cells carrying TP53 mutations are highly resistant to drugs, which subsequently lead to recurrence. Our results suggest that the upregulation of cell adhesion molecules and pathways may be the cause of the high drug resistance found in TP53 mutation AML patients.

Rap1 is a small GTPase that controls multiple processes, including cell adhesion, cell–cell junction formation, and cell polarity. Rap1 plays a dominant role in the control of cell–cell and cell–matrix interactions by regulating the function of integrins and other cell adhesion molecules in various cell types. Our present study shows that cell adhesion (ITGA2B, ITGB3, CDH1) and KDR genes were enriched in this pathway and, hence, implies that this pathway mainly involves cell adhesion in AML. Rap1GAP expression was lower in AML patients compared to normal subjects, upregulated Rap1GAP promotes leukemia cell differentiation and apoptosis, but increases leukemia cell invasion due to MMP9 overexpression.

The PI3K-Akt signaling pathway can be activated by many types of cellular stimuli or toxic insults and regulates cancer cell transcription, translation, proliferation, growth, and survival. Uncontrolled activation of the PI3K/Akt pathway confers resistance to chemotherapy in vitro and in vivo, thus, inhibition of Akt activation is an attractive strategy for cancer treatment. The PI3K/Akt/mTOR signaling pathway is actively sustained in AML patients, and subsequently affects survival, proliferation, and the drug-resistance of leukemic cells. PI3Kδ and PI3Kγ inhibitor and knockdown experiments suggest that inhibition of PI3Kδ and PI3Kγ has an anti-adhesion, anti-proliferative, and anti-migration effect on AML cells. A recent study by Liang et al also demonstrated that PI3K/AKT pathways were upregulated in AML refractory/relapse patients, and inhibition of the PI3K/AKT pathway via GLI1 led to enhanced AML drug sensitivity. PI3K/Akt inhibition by wortmannin upregulated the p53 protein in AML patients, whereas, p53 inactivation was associated with PI3K/Akt activation in MDM2 phosphorylation, and subsequently increased MRPI expression, which is an important drug resistance mechanism via MRPI overexpression. Wang et al revealed that adhesion-dependent activation via PI3K/Akt pathway upregulation of an X-linked inhibitor of the apoptosis protein may be involved in the cell adhesion-mediated drug resistance of U937 cells, a human AML cell line. It is worth noting that PI3K/Akt was activated in AML patients and that it regulates drug resistance in a variety of ways. However, cell adhesion-mediated drug resistance also plays an important role in the drug resistance of AML. Based on the annotation of the KEGG website (http://www.kegg.jp/), the Rap1 signaling pathway (hsa04015) participated in PI3K/AKT pathway regulation and correlated to cell adhesion, survival, and proliferation. However, the PI3K/Akt pathway (hsa04151) was also associated with the adhesion, survival, cycle, and proliferation of cells. Increasing evidence suggests that the PI3K/Akt pathway affects the adhesion, cycle, proliferation, and migration of leukemic cells in AML, especially in connection with drug resistance. In our study, we found that upregulated genes were significantly enriched in the PI3K/Akt pathway in TP53 mutation AML,
which suggests that PI3K/Akt pathway activation may play a crucial role in TP53 mutation AML and can be used as a potential therapeutic target for multiple biological processes in TP53 mutation AML.

In PPI network analysis, we identified hub genes which consisted of the ten DEGs with the highest degree of interaction. The CDH1 gene, which encodes a classical cadherin of the cadherin superfamily and is a calcium-dependent cell-cell adhesion protein, had the highest degree of interaction among these hub genes. Evidence has shown that methylation of CDH1 was associated with unfavorable prognosis in AML, whereas, CDH1 downregulation led to markedly shorter OS and acted as an independent prognostic biomarker in cytogenetic normal AML. The KDR gene was upregulated in refractory leukemia cells and initiates a PI3K-dependent clonogenic response in AML. The combination of systemic LEP level and local bone marrow stroma LEP production may affect leukemic hematopoiesis; furthermore, LEP acts as a growth factor and promotes cellular proliferation in AML. A study by Ferguson et al has shown that MNX1 is frequently inactivated by epigenetic mechanisms and may play a dual role in the development of leukemia, as a tumor suppressor gene in acute lymphoid leukemia, but as an oncogene in infant AML. The NMU gene was highly expressed in AML cells, and its silencing inhibits cell proliferation. In previous studies, there were no reports that the other five hub genes were correlated with AML. However, ITGA2B and BMP2 were associated with cancer cell adhesion and migration, as well as PI3K/AKT signaling.

The significant GO term enrichments of the top three modules consisted of the ten DEGs with the highest degree of interaction. The key pathways and genes in TP53 mutation AML.

**Conclusion**

The TP53 mutation frequency in the current study fits well with the previous studies. Our findings suggest that TP53 mutations promote poor OS in adult de novo AML patients, consistent with previous studies. Multiple bioinformatics analysis results revealed that cell adhesion, differentiation, and proliferation may play a crucial role in TP53 mutation AML. Cell adhesion-related genes and pathways, including CDH1 gene, Rap1, and the PI3K-Akt pathway, which are associated with a drug resistance mechanism, may be the potential mechanism of drug resistance in TP53 mutation AML and will provide new targets and strategies for individualized treatment. However, further molecular biological experiments are required to confirm our findings. Additional findings obtained in our study are also worthy of further investigation.

**Acknowledgments**

This work was supported in part by the National Natural Science Foundation of China (81160075); Natural Science Foundation of Guangxi (0728124); Self-raised Scientific Research Fund of the Ministry of Health of Guangxi Province (Z2014035). The authors thank the contributors of The Cancer Genome Atlas (https://portal.gdc.cancer.gov/) and the cBioPortal for Cancer Genomics (http://www.cbioportal.org/index.do) for their contribution by sharing the AML dataset through open access. In addition, we would also like to acknowledge the helpful comments on this paper received from our reviewers.

**Disclosure**

The authors report no conflicts of interest in this work.

**References**


29. Huang et al