Am J Transl Res 2021;13(12):13683-13696
www.ajtr.org /ISSN:1943-8141/AJTR0138108

Original Article

Identification of a prognostic signature based on copy number variations (CNVs) and CNV-modulated gene expression in acute myeloid leukemia

Changchun Niu[1,2], Di Wu[2], Alexander J Li[2], Kevin H Qin[2], Daniel A Hu[2], Eric J Wang[2], Andrew Blake Tucker[2], Fang He[2], Linjuan Huang[2], Hao Wang[2], Qing Liu[2], Na Ni[2], Deyao Shi[2], Xia Zhao[2], Yafang Wan[1], Tian Li[1], Tongchuan He[2], Pu Liao[1]

[1]Department of Laboratory Diagnostic Medicine, The Affiliated Chongqing Hospital of The University of Chinese Academy of Sciences, Chongqing General Hospital, Chongqing 400021, China; [2]Molecular Oncology Laboratory, Department of Orthopaedic Surgery and Rehabilitation Medicine, The University of Chicago Medical Center, Chicago, IL 60637, USA

Received August 7, 2021; Accepted November 9, 2021; Epub December 15, 2021; Published December 30, 2021

Abstract: Objectives: Acute myeloid leukemia (AML) is caused by multiple genetic alterations in hematopoietic progenitors, and molecular genetic analyses have provided useful information for AML diagnosis and prognostication. This study aimed to integratively understand the prognostic value of specific copy number variation (CNV) patterns and CNV-modulated gene expression in AML. Methods: We conducted integrative CNV profiling and gene expression analysis using data from the Therapeutically Applicable Research To Generate Effective Treatments (TARGET) and The Cancer Genome Atlas (TCGA) AML cohorts. CNV-related genes associated with survival were identified using the TARGET AML cohort and validated using the TCGA AML cohort. Genes whose CNV-modulated expression was associated with survival were also identified using the TARGET AML cohort and validated using the TCGA AML cohort, and patient bone marrow samples were then used to further validate the effects of CNV-modulated gene expression on survival. CNV and mRNA survival analyses were conducted using proportional hazards regression models (Cox regression) and the "survminer" and "survival" packages of the R Project for Statistical Computing. Genes belonging to the Kyoto Encyclopedia of Genes and Genomes (KEGG) cancer panel were extracted from KEGG cancer-related pathways. Results: One hundred two CNV-related genes (located at 7q31-34, 16q24) associated with patient survival were identified using the TARGET cohort and validated with the TCGA AML cohort. Among these 102 validated genes, three miRNA genes (*MIR29A, MIR183,* and *MIR335*) were included in the KEGG cancer panel. Five genes (*SEMA4D*, *CBFB*, *CHAF1B*, *SAE1*, and *DNMT1*) whose expression was modulated by CNVs and significantly associated with clinical outcomes were identified, and the deletion of *SEMA4D* and *CBFB* was found to potentially exert protective effects against AML. The results of these five genes were also validated using patient marrow samples. Additionally, the distribution of CNVs affecting these five CNV-modulated genes was independent of the risk group (favorable-, intermediate-, and adverse-risk groups). Conclusions: Overall, this study identified 102 CNV-related genes associated with patient survival and identified five genes whose expression was modulated by CNVs and associated with patient survival. Our findings are crucial for the development of new modes of prognosis evaluation and targeted therapy for AML.

Keywords: Acute myeloid leukemia, copy number variation, gene expression, prognosis

Introduction

In 2008, whole-genome sequencing (WGS) was first applied to a patient with acute myeloid leukemia (AML) [1], and since then, molecular analysis has provided abundant useful information for AML diagnosis, prognosis, and treatment [2]. Prognostic studies have provided information concerning potential outcomes and survival rates that is crucial to understanding the disease and improving its treatment [3]. Copy number variation (CNV), a potential valuable prognostic marker, is a type of structural variation larger than 1 kilobase pair and involves unbalanced rearrangements that result in gains or losses in the DNA content [4]. CNVs

drive genome evolution and alter gene expression [5, 6] and thereby influence the progression of diseases such as cancers [7]. For example, in acute lymphoblastic leukemia (ALL), *IKZF* deletion has been identified as a poor prognosis marker [8].

Although the prognostication and diagnosis of AML have shifted from histological analyses to more comprehensive analyses, such as analyses of genetic alterations [9], prognostication must be further improved because approximately 60% of patients with AML are at intermediate risk and respond differently to therapy [10], and patients with AML and rarer abnormalities must be evaluated even more specifically [11].

At present, cytomorphology, cytochemistry, immunophenotyping, cytogenetics and molecular genetics play crucial roles in AML diagnosis, prognostication and risk stratification [12]. Cytogenetic CNV abnormalities have been included in the World Health Organization (WHO) classification (2016) [13] and other risk stratification strategies [14-16] and constitute the single strongest prognostic factors for complete remission (CR) and overall survival (OS) in AML. According to their cytogenetic profile, patients with AML can be stratified into favorable, intermediate or adverse prognostic risk groups [17]; for example, the monosomy 5 or 7, t(6:9), and inv(3) profiles are associated with a higher prognostic risk [18]. CNV analysis has undergone a greater revolution than cytogenetics analysis [19]. Although many studies concerning CNVs in AML [10, 20-22] have been conducted, none have performed an integrative evaluation of the roles of gene expression derived from CNV. Additionally, an understanding of how CNVs can act as prognostic markers by modulating gene expression in AML remains elusive. Therefore, investigating the association between CNVs and AML prognosis is crucial to obtaining a better understanding of the genetic disorders of the disease. In turn, more novel, reliable predictions of the outcomes of patients with AML could arise.

To thoroughly understand the association between CNVs and AML, we collected CNV WGS data and gene expression data from the Therapeutically Applicable Research To Generate Effective Treatments (TARGET) AML cohort and CNV array analysis data and gene expression data from The Cancer Genome Atlas (TCGA) consortium AML cohort. Through a comprehensive analysis of CNVs and CNV-modulated gene expression alterations, we identified the essential CNV-related genes needed for the outcome and survival of AML patients.

Methods

Gene data and bone marrow samples

In the present study, the TARGET and TCGA AML cohorts were used to analyze the prognostic roles of CNVs and CNV-modulated gene expression. TARGET CNV (WGS data) and mRNA (mRNA and miRNA sequencing) data were downloaded from target-data.nci.nih.gov (https://target-data.nci.nih.gov/Public/AML/). The R project package "circlize" was used to map the CNV genes to chromosomes. The clinical data, including patient identification, sex, risk group, age, overall survival time, and vital status, were downloaded and preprocessed. The CNV data (Affymetrix Genome-Wide SNP array 6.0 data), mRNA data (mRNA and miRNA sequencing), and clinical data of the TCGA AML cohort were downloaded from gdac.broadinstitute.org. Overall survival time = "patient.days_to_death" + "patient.days_to_last_followup".

The CNV data from TCGA were divided into two groups (samples from primary blood-derived cancer and samples from solid normal tissue) and analyzed separately using the computational approach Genomic Identification of Significant Targets in Cancer (GISTIC, version 2.0.23) downloaded from ftp.broadinstitute.org, and the parameters were set up as suggested by TCGA (https://docs.gdc.cancer.gov/Data/Bioinformatics_Pipelines/CNV_Pipeline/).

The data in "all_thresholded.by_genes.txt" were analyzed as follows: a result greater than or equal to 1 was defined as amplification, and a result less than or equal to -1 was defined as deletion.

Bone marrow samples were collected from patients with AML (n=121; 53 female and 68 male patients; median age: 39 years) at Xinqiao Hospital and Chongqing General Hospital, Chongqing, China, since November 2016. The patients were followed up until death or the

end of the study in September 2020. The patients with AML were diagnosed according to the French-American-British (FAB) and WHO classifications and had not received bone marrow transplantation. All procedures performed in studies involving human participants were in accordance with the ethical standards of the institutional and/or national research committee and the Helsinki declaration. The study was approved by the Ethics Committee of Chongqing General Hospital.

Data analysis

KEGG cancer panel genes: Using an API with the Python project package "Bio. KEGG" (from https://www.genome.jp/kegg/), we obtained the pathways related to cancer from the Kyoto Encyclopedia of Genes and Genomes (KEGG) database. We identified a total of 483 genes involved in these cancer-related pathways, and this group of genes was considered the KEGG cancer panel gene.

Survival analysis of CNVs and cytogenetic changes: A multivariate proportional hazards regression model (Cox regression) was used to identify survival-related CNVs in patients with AML. From the TARGET cohort, 193 patients (96 female and 97 male; median age: 9 years) with integrated CNV data (WGS data) and clinical data were included in the survival analysis. From the TCGA cohort, 191 patients (87 female and 104 male; median age: 58 years) with integrated CNV data and clinical data were included to validate the survival analysis result from the TARGET cohort.

We removed the CNVs occurring in fewer than 4 samples. Based on each gene, the patients were divided into two groups denoted "normal" or "CNV" and were analyzed using the Cox regression model adjusted for sex and age. The *p*-values were adjusted with "HDR", and $P<0.05$ and adjusted $P<0.05$ were considered to indicate significance. Kaplan-Meier curves were used for the visualization of survival. We performed all analyses using the "survminer" and "survival" packages in the R project.

A survival analysis of cytogenetic changes in the samples from the TARGET and TCGA cohorts was also performed using a Cox regression model adjusted for sex and age.

Survival analysis of CNV-modulated gene expression: The mRNA expression data (RPKM) were log2-transformed. Patients with integrated mRNA sequencing data and clinical data were included and evaluated using multivariate Cox proportional hazards regression models adjusted for sex and age and univariate Cox proportional hazards regression models adjusted for sex or age. We used the R "survminer" and "survival" packages to perform Cox proportional hazards regression model analyses and "HDR" to adjust the *P* values afterward. In the statistical analysis, $P<0.05$ and adjusted $P<0.05$ were considered to indicate significance, and the upper and lower 95% confidence intervals are reported.

An integrative analysis of gene expression and CNVs was performed using 156 patients in the TARGET cohort and 171 patients in the TCGA cohort, and these patients were selected because they had integrated CNV data and RNA-sequencing data. The correlations between gene CNVs and expression were analyzed as follows: The patients were divided into groups with different CNV statuses ("Normal", "Duplication" and "Deletion"), the Kruskal test was used to compare more than two groups, and the Wilcoxon test was used to compare two groups. The *P* values were adjusted with "HDR", and $P<0.05$ and adjusted $P<0.2$ (adjusted with "HDR") were considered to indicate statistical significance.

Patient sample validation analyses: The expression of five genes (*SEMA4D, CBFB, CHAF1B, SAE1*, and *DNMT1*) in the patient samples was assessed. DNA and RNA from patient bone marrow samples were extracted using the Tiangen DNA/RNA kit (Beijing, China) according to the manufacturer's instructions. The extracted DNA was used for CNV analysis using the AccuCopy™ method developed by Genesky Biotechnologies Inc. (Shanghai, China) as described previously [23]. The primers used were as follows: *SEMA4D*, 5'-GGATGAAACTTGCCACGTGAA-3', 5'-GGAAATGCCTTGCCCTAAACC-3'; *DNMT1*, 5'-GATCAGGCAGCTCAATAATTTGTGT-3', 5'-TGACCTCAAATATGGGCAGCA-3'; *CBFB,* 5'-GTCATTGCAGGCAAGAAGACAAC-3', 5'-GAGAACAGCGACAAACACCTA-3'; *CHAF1B*, 5'-TAAATGGCTCCTGGCCCCTAT-3', 5'-TCTTCCACGGACGGTTACTGCT-3'; and *SAE1*, 5'-TGATTCTGCAAGCTCACTGTTCTGT-3', 5'-CTATCTGA-

ATCGGGCCCTCCT-3'. For each gene, two primers were used to increase the accuracy of the CNV analysis.

A ReverTra Ace-α First-Strand cDNA Synthesis Kit (Toyobo) was used for the generation of cDNA using the extracted RNA, and real-time quantitative PCR with SYBR Green using a 7500 PCR System (ABI) was performed to determine gene expression. The primers used for real-time PCR were as follows: *CBFB*, forward 5'-ACTGGATGGTATGGGCTGTC-3', reverse 5'-AAGGCCTGTTGTGCTAATGC-3'; *CHAF1B,* forward 5'-CTGGGCAACTGATGGGAATT-3', reverse 5'-GCAGCACCCTGTCACAGCT-3'; *DNMT1*, forward 5'-GTTCTTCCTCCTGGAGAATGTCA-3', reverse 5'-GGGCCACGCCGTACTG-3'; *SAE1*, forward 5'-AGGACTGACCATGCTGGATCAC-3', reverse 5'-CTCAGTGTCCACCTTCACATCC-3'; *SEMA4D*, forward 5'-GTCTTCAAAGAAGGGCAACAGG-3', reverse 5'-GAGCATTTCAGTTCCGCTGTG-3'; and *β-actin* (internal control), forward 5'-AGTTGCGTTACACCCTTTC-3', reverse 5'-CCTTCACCGTTCCAGTTT-3'. PCR was conducted in triplicate for each sample. All gene expression levels were normalized to that of *β-actin* using the $2^{-\Delta\Delta Ct}$ method.

The patient samples were divided into groups based on their CNV statuses ("Normal", "Duplication" and "Deletion"). Using R, the Kruskal test was used to compare more than two groups, and the Wilcox test was used to compare two groups. $P<0.05$ was considered to indicate statistical significance.

The gene expression levels were normalized and multivariate Cox proportional hazards regression models adjusted for sex and age were used for survival analysis using the R "survminer" and "survival" packages.

VENN diagram generation

VENN diagrams were plotted using the R package "Venn Diagram".

Results

CNV distribution

CNV data (WGS) from 193 patients in the TARGET AML cohort (96 female and 97 male; median age: 9 years; median follow-up time: 1652 days) and CNV data (SNP array) from 191 patients in the TCGA cohort (87 female and 104 male; median age: 58 years; median follow-up time of 304 days) were used to investigate the distribution of CNVs in AML.

In the TARGET AML cohort, the distribution of CNVs in 99.6% (17,507/17,586) of genes was less than 6%. However, 35 CNV-affected genes were present in more than 90% of patients. In contrast, the prevalence of CNVs in the TCGA cohort did not exceed 14.1% (**Figure 1A**).

Consistent with a previous report [10], our study showed that CNVs were not randomly distributed across the chromosomes. Amplifications were more frequent on chromosomes 1, 8, 19, 21, and 22 in the TARGET and TCGA cohorts, and deletions were more frequent on chromosomes 7, 16, and X. Fewer CNVs were found in normal tissues than in the AML samples (**Figure 1B**).

Distinct CNV distribution patterns in the three risk groups were observed between the TARGET and TCGA cohorts (**Figure 1C**, **1D**). Different patterns were found in the two cohorts: CNVs occurred more frequently in the intermediate-risk group of the TARGET cohort and more frequently in the adverse-risk group of the TCGA cohort. In the adverse-risk groups, more frequent amplifications in chromosome 7 and more frequent deletions in chromosome 19 occurred in the TARGET cohort, and deletions in chromosome 7 frequently occurred in the intermediate- and favorable-risk groups. In the TCGA cohort, CNVs occurred less frequently in the intermediate- and favorable-risk groups than in the adverse-risk group. We also compared the distributions of CNVs and cytogenetic changes in the TCGA and TARGET cohorts and found that del(7q)/7q- was related to deletions of the genes in the TCGA cohort but not in the TARGET cohort (Figures S1, S2).

Prognostic value of CNVs and cytogenetic changes in AML

To explore the prognostic value of CNVs, multivariate Cox proportional hazards regression model analysis adjusted for age and sex was performed using the TARGET cohort (193 patients). Seven hundred fifty-eight CNV-related genes were significantly associated with patient survival ($P<0.05$; adjusted $P<0.05$). We validated these 758 CNV genes in the TCGA

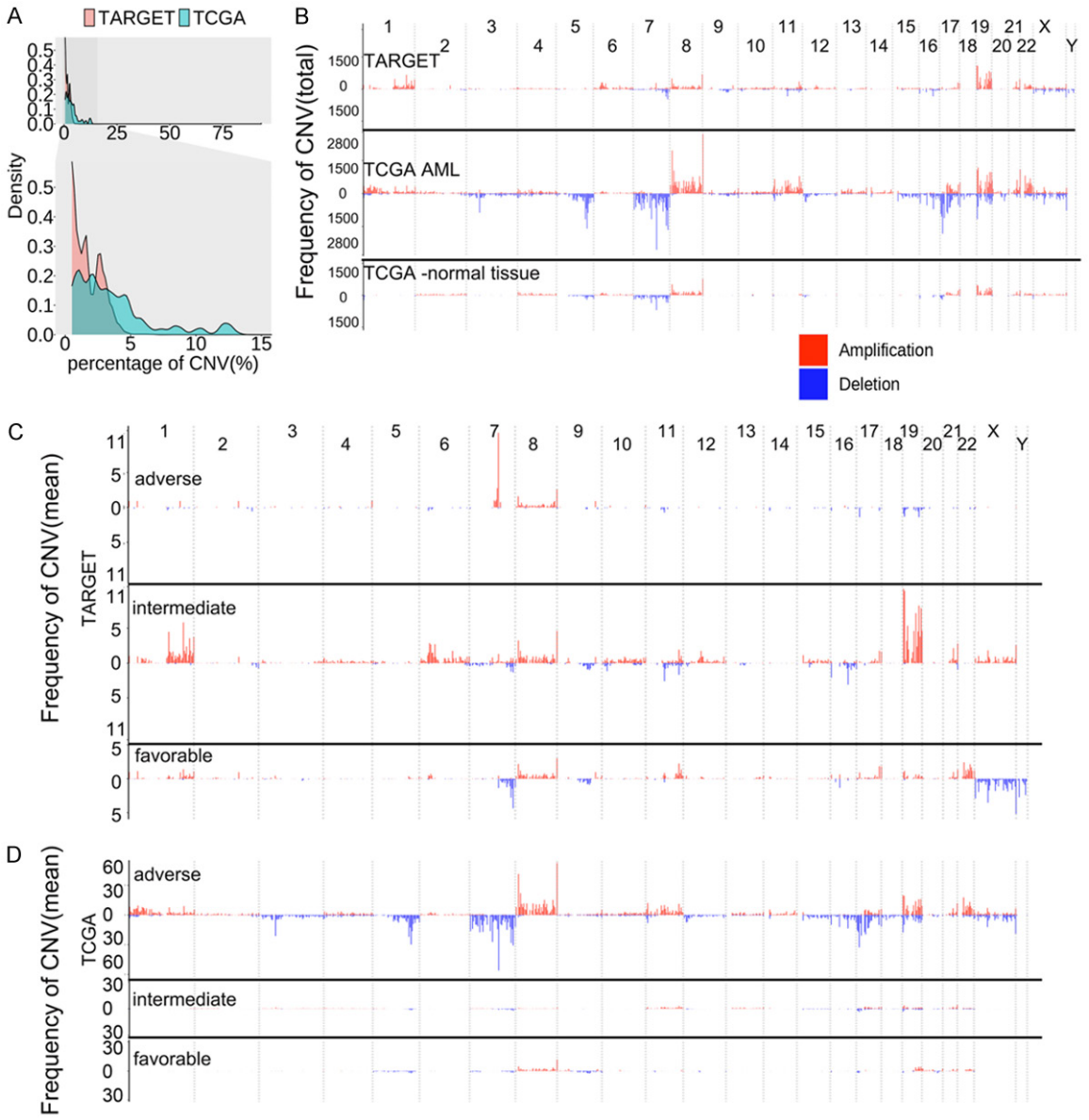

Figure 1. Genome-wide distribution of CNV-related genes in two AML cohorts. A. Distribution of CNV percentages in the TARGET AML cohort (n=193), TCGA AML cohort (n=191) and TCGA normal tissues (n=189). B. Frequency of cumulative CNV counts (total CNV counts of all patients) in chromosome map locations. C, D. Frequency of cumulative CNV counts (mean CNV counts of each patient) in chromosome map locations of the TARGET and TCGA cohorts.

AML cohort (191 patients), and 102 of the CNV genes were associated with patient survival (P<0.05) (Table S1). Therefore, these 102 CNV genes showing the greatest association with high risk were found in both the TARGET and TCGA cohorts.

These 102 CNV genes were located on chromosomes 7 and 16 (7q31, 7q32, 7q33, 7q34, and 16q24.1) (**Figure 2A**) and included 7 lncRNA and 10 miRNA genes, which implies that they might have gene-regulating capabilities. Additionally, 7.8% (8/102) of the CNVs modulated gene expression in the TARGET cohort, and 52.0% (53/102) of the CNVs modulated gene expression in the TCGA cohort (Table S1), which suggested that the modulation of gene expression affects patient prognosis.

We also compared these 102 genes with the KEGG cancer panel and found that three miRNA genes (*MIR29A, MIR183,* and *MIR335*)

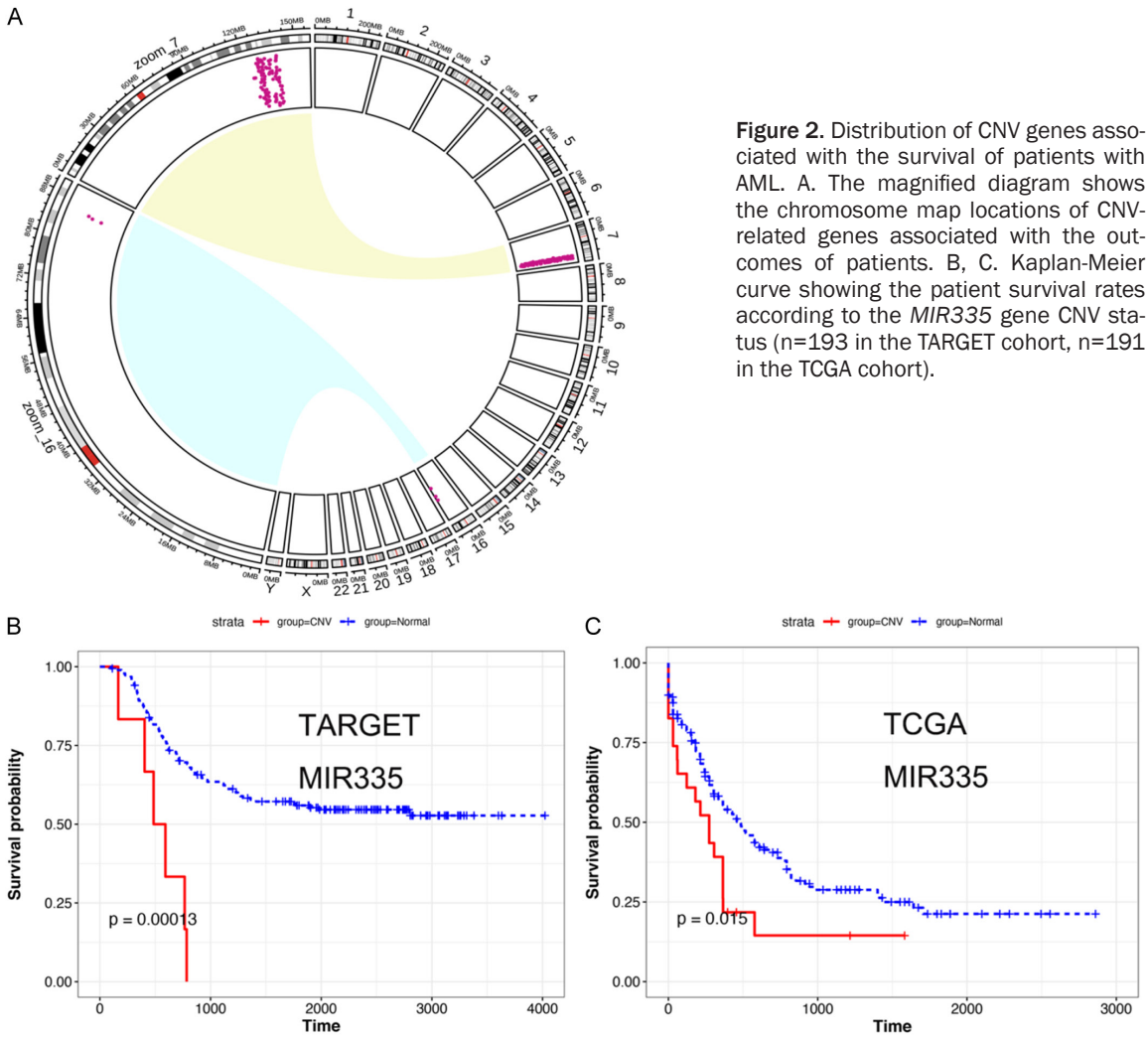

Figure 2. Distribution of CNV genes associated with the survival of patients with AML. A. The magnified diagram shows the chromosome map locations of CNV-related genes associated with the outcomes of patients. B, C. Kaplan-Meier curve showing the patient survival rates according to the *MIR335* gene CNV status (n=193 in the TARGET cohort, n=191 in the TCGA cohort).

were included in the KEGG cancer panel. A Cox model analysis adjusted for age and sex additionally showed that *MIR335* expression (miRNA sequencing data) was associated with survival in the TARGET cohort (n=300; *P*= 0.015; HR: 0.911; 95% confidence interval: 0.844-0.982), and the association of MIR335 with the survival of patients with AML is shown in **Figure 2B**, **2C**.

We also evaluated the prognostic value of cytogenetic changes. In the TARGET cohort (96 female and 97 male patients; median age: 9 years), t(10;11)(p11.2;q23), MLL and trisomy 8 were associated with high-risk prognosis, and inv(16) was associated with low-risk prognosis (**Table 1**). In the TCGA cohort (87 female and 104 male patients; median age: 58 years), del(5q)/5q- was associated with high-risk prognosis, and t(15;17) and variants were associated with low-risk prognosis (**Table 2**).

Identification of CNV-modulated gene expression associated with survival

We analyzed the relationship between CNVs and gene expression in both the TARGET (156 patients) and TCGA cohorts (171 patients). In the TCGA cohort, the expression of 5,022 genes was driven by CNVs (pink circle in **Figure 3A**), and that of 577 genes was driven by CNVs in the TARGET cohort (purple circle in **Figure 3A**). Two hundred fifty-one genes whose expression was driven by CNVs in both cohorts were identified (overlap between the pink and purple circles in **Figure 3A**).

Additionally, a survival analysis of gene expression adjusted for sex and age in the TARGET

Table 1. Prognostic value of cytogenetic changes in the TARGET cohort (n=193)

Cytogenetic Change	*P* value	Hazard ratios (HR)	HR.confint.lower (95%)	HR.confint.upper (95%)	Frequency
del5q	0.99	4.36E-07	Not available	Not available	1
del7q	0.16	2.31	0.727	7.35	4
del9q	0.95	1.05	0.254	4.31	5
inv(16)	0.025	0.434	0.21	0.898	30
Minus.X	0.24	1.84	0.659	5.13	7
Minus.Y	0.7	0.797	0.246	2.58	8
MLL	0.017	1.92	1.12	3.27	30
t(10;11)(p11.2;q23)	0.0047	3.71	1.5	9.21	5
t(11:19)(q23:p13.1)	0.084	2.46	0.887	6.85	6
t(3;5)(q25;q34)	0.072	6.4	0.848	48.3	1
t(6;11)(q27;q23)	0.57	0.56	0.0771	4.07	3
t(6;9)	0.23	2.38	0.577	9.85	2
t(8;21)	0.37	0.727	0.363	1.46	25
t(9;11)(p22;q23)	0.55	1.27	0.587	2.73	18
trisomy.21	0.31	1.71	0.601	4.87	7
trisomy.8	0.025	2.16	1.1	4.24	15

Table 2. Prognostic value of cytogenetic changes in the TCGA cohort (n=191)

Cytogenetic Change	*P* value	Hazard ratios (HR)	HR.confint.lower (95%)	HR.confint.upper (95%)	Frequency
+8	0.63	0.872	0.498	1.53	21
del(5q)/5q-	0.029	1.88	1.07	3.33	15
del(7q)/7q-	0.11	1.53	0.912	2.56	21
inv(16)	0.11	0.386	0.122	1.22	9
t(15;17) and variants	0.027	0.359	0.145	0.891	17
t(9;11)	0.086	3.46	0.84	14.2	2

AML cohort (156 patients) was performed, and the results showed that the expression of 2,058 genes was associated with survival. We then validated these findings with the TCGA cohort (171 patients) and found that 268 of the genes were associated with survival in this cohort (green circle in **Figure 3A**). Furthermore, among these 268 genes, 5 were driven by CNVs in both the TCGA and TARGET cohorts (overlap of the green, pink and purple circles in **Figure 3A**). In summary, the expression of these 5 genes (*CBFB* (16q22), *CHAF1B* (21q22), *DNMT1* (19p13), *SAE1* (19q13), and *SEMA4D* (9q22)) was associated with survival and was modulated by CNVs in both cohorts (**Figure 3B**). The upregulation of these five genes was associated with an adverse prognosis in AML (**Figure 3B**). We also showed that the deletion of *CBFB* and *SEMA4D* downregulated gene expression, and the amplification of *CHAF1B*, *SAE1,* and *DNMT1* upregulated gene expression (**Figure 3C-L**). These results implied that the deletion of *CBFB* and *SEMA4D* may be protective and that the amplification of *CHAF1B, SAE1*, and *DNMT1* has adverse effects on prognosis.

Additionally, these CNVs were not affected by the risk groups. In the TARGET cohort, CNVs occurred more frequently in intermediate-risk patients, whereas in the TCGA cohort, they occurred more frequently in the patients belonging to the adverse-risk group (**Figure 3M**).

We also validated the analysis using patient bone marrow samples (121 patients with a median follow-up time of 555 days). The expression of the five genes (*SEMA4D, CBFB, CHAF1B, SAE1,* and *DNMT1*) was significantly associated with the survival of patients with AML (**Figure 4A**), and the CNVs of four genes (*SEMA4D, CBFB, CHAF1B,* and *DNMT1*) regulated the expression of the genes (**Figure 4B-E**). The CNVs of *SAE1* affected SAE1 expres-

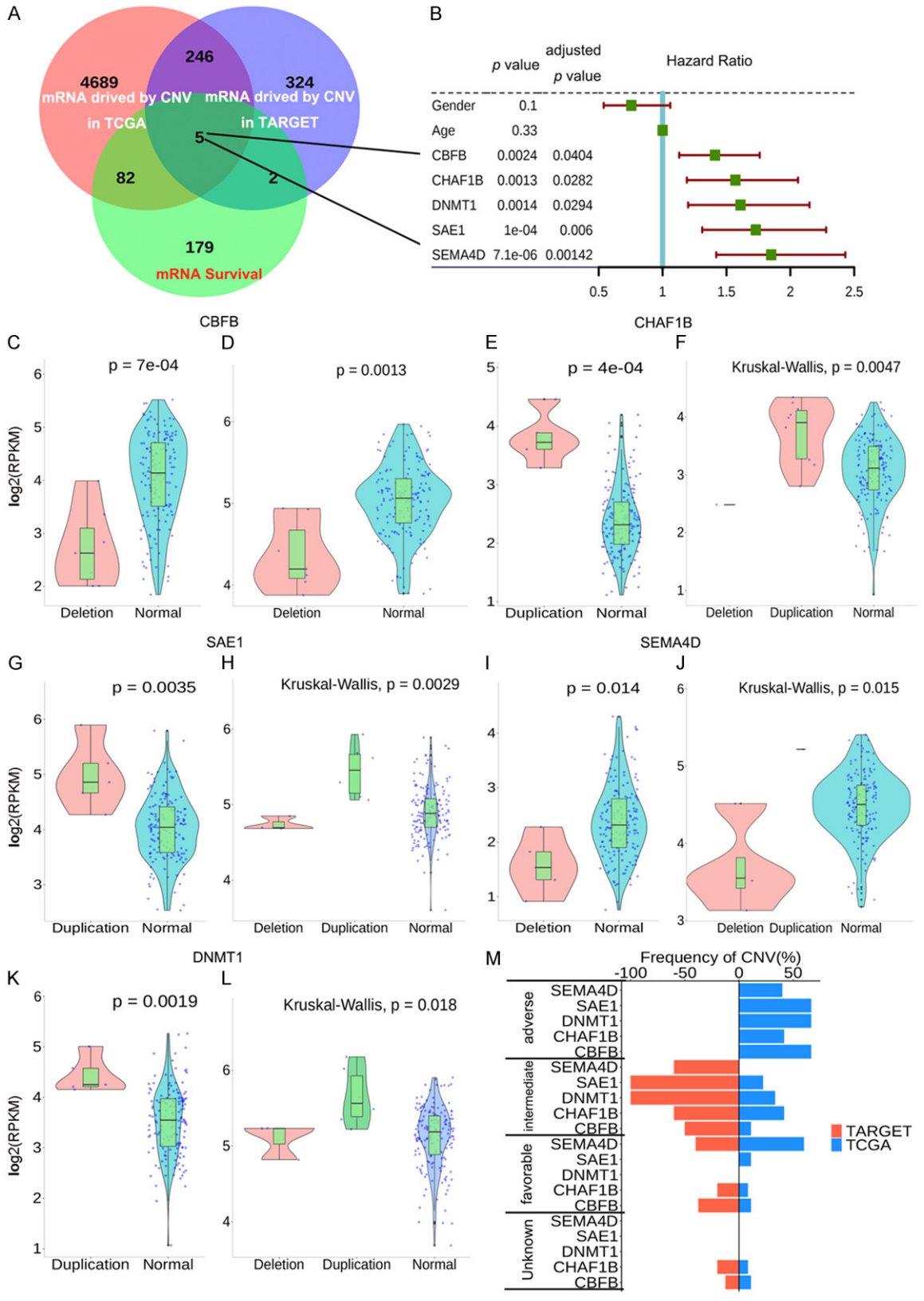

Figure 3. CNV-modulated genes associated with survival in AML. (A) Venn diagram of the relationship of genes whose expression was associated with clinical outcomes in AML and genes whose expression was modulated by CNVs. (B) Forest plot of the results from Cox proportional hazards regression models with the TARGET AML cohort

(n=156) and univariate Cox proportional hazards regression models corrected for sex and age. (C-L) Gene expression changes induced by CNVs: (C, E, G, I, and K) show the results from the TARGET cohort (n=156); (D, F, H, J, and L) show the results from the TCGA cohort (n=171). (M) Distribution of 5 CNV-related genes in the risk groups in the two AML cohorts.

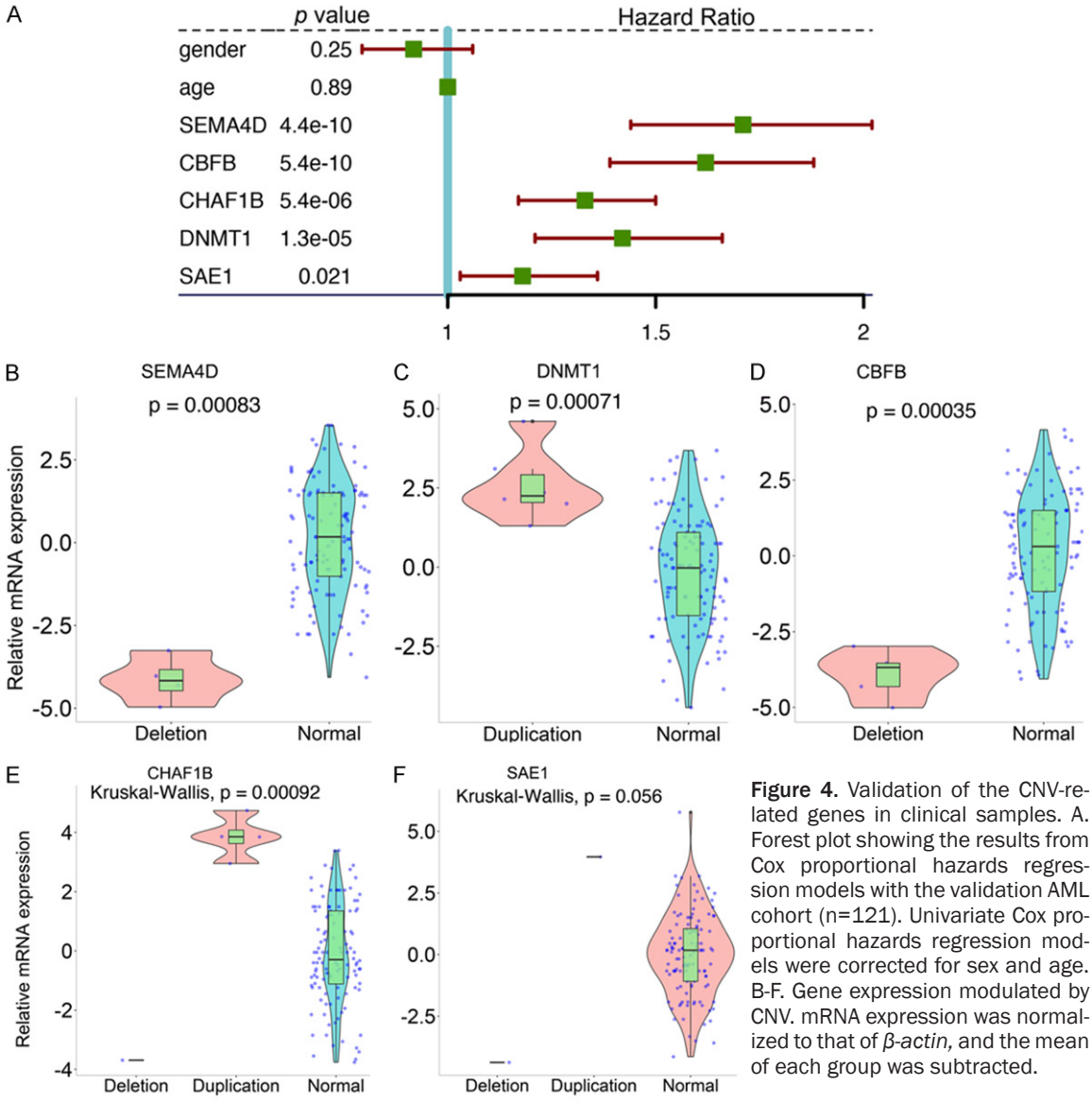

Figure 4. Validation of the CNV-related genes in clinical samples. A. Forest plot showing the results from Cox proportional hazards regression models with the validation AML cohort (n=121). Univariate Cox proportional hazards regression models were corrected for sex and age. B-F. Gene expression modulated by CNV. mRNA expression was normalized to that of *β-actin,* and the mean of each group was subtracted.

sion, but the effect was not significant because fewer CNVs (1 duplication and 1 deletion) were detected in the patients (**Figure 4F**).

Discussion

CNVs influence the progression of diseases by driving genome evolution and altering gene expression [5, 6]. Therefore, the present study aimed to explore the potential prognostic roles of CNVs and CNV-modulated gene expression in AML.

Using many subjects from the TARGET and TCGA AML cohorts, we identified 102 CNV-related genes directly associated with the survival of patients with AML. We also identified five genes modulated by CNVs that have vital effects on the survival of subjects.

First, we investigated the distribution of CNVs on chromosomes using both the TARGET and TCGA cohorts. The CNVs in the TARGET cohort showed a different distribution pattern than those in the TCGA cohort, and these differenc-

es are likely due to the different approaches used in the studies. Array data have been used in CNV analyses performed in previous AML studies [10, 20, 24, 25]. In the present study, we used WGS data from TARGET to identify valuable CNV markers and validated these markers using array data from TCGA. A previous study indicated that GWS is the most powerful approach due to its high sensitivity and ability to provide accurate breakpoint information [26], whereas the resolution capacity of arrays is limited [4]. Concerning the distribution of the CNV frequencies across the chromosomes, our study showed more frequent amplifications on chromosomes 8, 11, and 21 and more frequent deletions on chromosome 7, and these findings are consistent with those reported by Nibourel et al., which showed more frequent amplifications on chromosomes 8, 11 and 21 and more frequent deletions on chromosomes 7, 12, 17 and 21 in AML samples [10]. The comparison of the TCGA cohort data with data from normal tissues of patients with AML revealed that CNVs occurred more frequently in AML tissues, supporting the crucial role of CNVs in cancer [6]. The distinct distribution patterns of CNVs based on risk group suggest that CNVs are potential risk stratification markers that must be further investigated and validated.

The present study identified 102 CNV genes through a survival analysis using the Cox model, and these genes were associated with an adverse survival risk. Among these 102 genes, 26 (*AGBL3*, *AKR1B15*, *FAM180A*, *FLJ40288, KCP*, *KLHDC10*, *LINC00311*, *LOC100129148*, *LOC100130705*, *LOC100506860*, *LOC154872*, *LOC400548*, *LOC407835*, *LUZP6*, *MESTIT1*, *MGC27345*, *MIR182*, *MIR29A*, *MIR4468*, *MIR96*, *PRRT4*, *TMEM213*, *TMEM229A*, *TPI1-P2*, *UBN2*, and *ZDHHC7*) were first identified in the present study [10, 20, 21, 24, 25, 27]. These genes are located at 7q31, 7q32, 7q33, 7q34, and 16q24.1, and we showed that some cytogenetic changes in chromosomes 7 or 16 were significantly associated with survival of the patients in both the TARGET and TCGA cohorts. Deletion in the 7q33-34 area has been previously reported [24]. Additionally, these 102 CNV genes include 7 lncRNA and 10 miRNA genes, which implies that they may have gene-regulating capability because lncRNAs and miRNAs function as master regulators of gene expression [28, 29]. Furthermore, three miRNA genes (*MIR29A*, *MIR183*, and *MIR335*) overlapped with the KEGG cancer gene panel. miR-29a is a key epigenetic regulator and plays crucial roles in cancers including AML [30], and miR-335 is involved in diverse cancers including AML [31]. Our results indicate that high miR-335 expression is a protective factor.

The present study also identified 5 CNV genes that indirectly affected the outcomes of patients by regulating gene expression; deletions of *CBFB* and *SEMA4D* induced potentially protective downregulated expression, and *DNMT1* expression had adverse effects on prognosis. Among these 5 genes, 4 (*SEMA4D*, *CHAF1B*, *SAE1*, and *DNMT1*) were first identified in the present study [10, 20, 21, 24, 25, 27]. Additionally, the distribution of CNVs was independent of the risk group, which suggested that they may be independent prognostic markers. In a previous study, gene expression levels were found to be strongly correlated with CNVs in tumor samples [32]. Furthermore, these genes and related pathways are crucial for leukemia cell survival, and CNV-modulated expression occurs in genomically unstable leukemia, which is the mechanism through which CNVs modulate these five genes in AML. Previous studies have shown that some CNVs are associated with AML. *CBFB* forms the fusion gene *CBFB/MYH11,* which blocks differentiation in AML [33], but *CBFB* deletion is usually accompanied by *MYH11* deletion*,* and *MYH11/CBFB* deletion has no negative effect on prognosis [34]. Furthermore, the CBFB protein expression levels were higher in AML cells than in corresponding normal tissues [35], and the CBFB protein expression levels in AML cells from relapsed cases were higher than those in primary AML cells at diagnosis [36]. Our findings suggest that the deletion-induced downregulation of *CBFB* expression has positive effects on prognosis, and the mechanism of these effects may be associated with the RUNX transcription factor family, which is involved in AML pathogenesis [37]. Elevated *CHAF1B* expression has been observed in leukemia and is critical for leukemogenesis and maintenance of the undifferentiated state of leukemia cells; additionally, heterozygous deletions of *Chaf1b* in mice block leukemogenesis [38]. We showed that in both cohorts, the CNV-

based amplification of *CHAF1B* led to expression upregulation and a negative prognosis. *DNMT1*, a DNA methyltransferase, is upregulated and considered a potential oncogene in AML because it downregulates p15 expression. We found that CNV-modulated upregulation of *DNMT1* expression may lead to poor outcomes in patients with AML. *SAE1* is a key molecule in the small ubiquitin (Ub)-like modifier (SUMO) pathway, and the promyelocytic leukemia protein (PML) and fusion protein PML-RARα (retinoic acid receptor-α) oncoproteins are substrates of the SUMO pathway [39]. Upregulated *SAE1* expression derived from CNV amplifications leads to enhanced activation of mature SUMO (small ubiquitinlike modifier) [40]. Subsequently, through the SUMOylation of proteins (such as sPRDM16 and Ubc9), the proliferation of human AML cells is promoted, the differentiation of AML cells is inhibited, and the antileukemic activity of ATRA is also inhibited [41], which likely represents the mechanism through which upregulated *SAE1* expression derived from CNV amplifications is associated with poor outcomes. *SEMA4D* expression is considered a strong predictor of worse clinical outcomes [42], and our findings showed that downregulated *SEMA4D* expression due to deletion is protective. Additionally, it has been reported that SEMA4D, CHAF1B, and DNMT1 are associated with tumor metastasis [43-49], which may be one of the reasons that these genes showed a prognostic signature in AML. Moreover, some drugs targeting the genes SEMA4D and DNMT1 (5-Aza-2'deoxycytidine (5-AzadC)), NSC-319745, and Harmine have been used as special drugs targeting DNMT1 [50, 51]; while anti-SEMA4D monoclonal antibodies [52] and pepinemab (anti-SEMA4D) have been used to target SEMA4D, providing novel approaches for AML treatment.

However, the present study has some limitations. First, the median ages of the two AML cohorts were different (9 years in TARGET vs. 58 years in TCGA), but our sample set showed that age was not associated with the outcomes of patients with AML when the survival analysis was adjusted for age (**Figures 3B**, **4B**). Second, the distinct CNV distribution patterns between the TARGET and TCGA cohorts in the three risk groups may have resulted from the different approaches used for CNV analysis and the different patient populations. Therefore, for the potential widespread clinical use of these prognostic signatures, more investigations of the biological roles of these CNV-related genes in AML are needed.

In conclusion, we identified 102 CNV-related genes (related to 7q31-34 and 16q24) associated with a higher risk of worse outcomes and 5 genes whose expression was modulated by CNVs and associated with the prognostic risk of patients. Additionally, the deletion of *SEMA4D* and *CBFB* may be protective.

Acknowledgements

The authors thank Therapeutically Applicable Research To Generate Effective Treatments (TARGET) and The Cancer Genome Atlas (TCGA) for providing the data. The reported work was supported in part by research grants from the National Natural Science Foundation of China (81702075) and the Natural Science Foundation of Chongqing, China (cstc2019jcyj-msxmX0276).

Disclosure of conflict of interest

None.

Address correspondence to: Drs. Changchun Niu and Pu Liao, Department of Laboratory Diagnostic Medicine, The Affiliated Chongqing Hospital of The University of Chinese Academy of Sciences, Chongqing General Hospital, No. 1 Zhulin County, Jiangbei District, Chongqing 400021, China. Tel: +86-023-63510942; E-mail: bright_star2000@163.com (CCN); Tel: +86-023-63516362; E-mail: liaopu2015@163.com (PL)

References

[1] Ley TJ, Mardis ER, Ding L, Fulton B, McLellan MD, Chen K, Dooling D, Dunford-Shore BH, McGrath S, Hickenbotham M, Cook L, Abbott R, Larson DE, Koboldt DC, Pohl C, Smith S, Hawkins A, Abbott S, Locke D, Hillier LW, Miner T, Fulton L, Magrini V, Wylie T, Glasscock J, Conyers J, Sander N, Shi X, Osborne JR, Minx P, Gordon D, Chinwalla A, Zhao Y, Ries RE, Payton JE, Westervelt P, Tomasson MH, Watson M, Baty J, Ivanovich J, Heath S, Shannon WD, Nagarajan R, Walter MJ, Link DC, Graubert TA, DiPersio JF and Wilson RK. DNA sequencing of a cytogenetically normal acute myeloid leukaemia genome. Nature 2008; 456: 66-72.

[2] Cai SF and Levine RL. Genetic and epigenetic determinants of AML pathogenesis. Semin Hematol 2019; 56: 84-89.
[3] Wee Y, Liu Y, Lu J, Li X and Zhao M. Identification of novel prognosis-related genes associated with cancer using integrative network analysis. Sci Rep 2018; 8: 3233.
[4] Zarrei M, MacDonald JR, Merico D and Scherer SW. A copy number variation map of the human genome. Nat Rev Genet 2015; 16: 172-183.
[5] Liang L, Fang JY and Xu J. Gastric cancer and gene copy number variation: emerging cancer drivers for targeted therapy. Oncogene 2016; 35: 1475-1482.
[6] Shlien A and Malkin D. Copy number variations and cancer. Genome Med 2009; 1: 62.
[7] Lauer S and Gresham D. An evolving view of copy number variants. Curr Genet 2019; 65: 1287-1295.
[8] Medeiros BC. Deletion of IKZF1 and prognosis in acute lymphoblastic leukemia. N Engl J Med 2009; 360: 1787.
[9] Arber DA. The 2016 WHO classification of acute myeloid leukemia: what the practicing clinician needs to know. Semin Hematol 2019; 56: 90-95.
[10] Nibourel O, Guihard S, Roumier C, Pottier N, Terre C, Paquet A, Peyrouze P, Geffroy S, Quentin S, Alberdi A, Abdelali RB, Renneville A, Demay C, Celli-Lebras K, Barbry P, Quesnel B, Castaigne S, Dombret H, Soulier J, Preudhomme C and Cheok MH. Copy-number analysis identified new prognostic marker in acute myeloid leukemia. Leukemia 2017; 31: 555-564.
[11] Grimwade D, Hills RK, Moorman AV, Walker H, Chatters S, Goldstone AH, Wheatley K, Harrison CJ and Burnett AK. Refinement of cytogenetic classification in acute myeloid leukemia: determination of prognostic significance of rare recurring chromosomal abnormalities among 5876 younger adult patients treated in the United Kingdom Medical Research Council trials. Blood 2010; 116: 354-365.
[12] Haferlach T and Schmidts I. The power and potential of integrated diagnostics in acute myeloid leukaemia. Br J Haematol 2020; 188: 36-48.
[13] Arber DA, Orazi A, Hasserjian R, Thiele J, Borowitz MJ, Le Beau MM, Bloomfield CD, Cazzola M and Vardiman JW. The 2016 revision to the World Health Organization classification of myeloid neoplasms and acute leukemia. Blood 2016; 127: 2391-2405.
[14] Döhner H, Estey E, Grimwade D, Amadori S, Appelbaum FR, Büchner T, Dombret H, Ebert BL, Fenaux P, Larson RA, Levine RL, Lo-Coco F, Naoe T, Niederwieser D, Ossenkoppele GJ, Sanz M, Sierra J, Tallman MS, Tien HF, Wei AH, Löwenberg B and Bloomfield CD. Diagnosis and management of AML in adults: 2017 ELN recommendations from an international expert panel. Blood 2017; 129: 424-447.
[15] Papaemmanuil E, Gerstung M, Bullinger L, Gaidzik VI, Paschka P, Roberts ND, Potter NE, Heuser M, Thol F, Bolli N, Gundem G, Van Loo P, Martincorena I, Ganly P, Mudie L, McLaren S, O'Meara S, Raine K, Jones DR, Teague JW, Butler AP, Greaves MF, Ganser A, Döhner K, Schlenk RF, Döhner H and Campbell PJ. Genomic classification and prognosis in acute myeloid leukemia. N Engl J Med 2016; 374: 2209-2221.
[16] Grimwade D, Ivey A and Huntly BJ. Molecular landscape of acute myeloid leukemia in younger adults and its clinical relevance. Blood 2016; 127: 29-41.
[17] De Kouchkovsky I and Abdul-Hay M. 'Acute myeloid leukemia: a comprehensive review and 2016 update'. Blood Cancer J 2016; 6: e441.
[18] Mrózek K, Marcucci G, Nicolet D, Maharry KS, Becker H, Whitman SP, Metzeler KH, Schwind S, Wu YZ, Kohlschmidt J, Pettenati MJ, Heerema NA, Block AW, Patil SR, Baer MR, Kolitz JE, Moore JO, Carroll AJ, Stone RM, Larson RA and Bloomfield CD. Prognostic significance of the European LeukemiaNet standardized system for reporting cytogenetic and molecular alterations in adults with acute myeloid leukemia. J Clin Oncol 2012; 30: 4515-4523.
[19] Mehrotra M, Luthra R, Ravandi F, Sargent RL, Barkoh BA, Abraham R, Mishra BM, Medeiros LJ and Patel KP. Identification of clinically important chromosomal aberrations in acute myeloid leukemia by array-based comparative genomic hybridization. Leuk Lymphoma 2014; 55: 2538-2548.
[20] Radtke I, Mullighan CG, Ishii M, Su X, Cheng J, Ma J, Ganti R, Cai Z, Goorha S, Pounds SB, Cao X, Obert C, Armstrong J, Zhang J, Song G, Ribeiro RC, Rubnitz JE, Raimondi SC, Shurtleff SA and Downing JR. Genomic analysis reveals few genetic alterations in pediatric acute myeloid leukemia. Proc Natl Acad Sci U S A 2009; 106: 12944-12949.
[21] Vosberg S, Herold T, Hartmann L, Neumann M, Opatz S, Metzeler KH, Schneider S, Graf A, Krebs S, Blum H, Baldus CD, Hiddemann W, Spiekermann K, Bohlander SK, Mansmann U and Greif PA. Close correlation of copy number aberrations detected by next-generation sequencing with results from routine cytogenetics in acute myeloid leukemia. Genes Chromosomes Cancer 2016; 55: 553-567.
[22] Bolouri H, Farrar JE, Triche T Jr, Ries RE, Lim EL, Alonzo TA, Ma Y, Moore R, Mungall AJ, Marra MA, Zhang J, Ma X, Liu Y, Liu Y, Auvil JMG, Davidsen TM, Gesuwan P, Hermida LC, Salhia

B, Capone S, Ramsingh G, Zwaan CM, Noort S, Piccolo SR, Kolb EA, Gamis AS, Smith MA, Gerhard DS and Meshinchi S. The molecular landscape of pediatric acute myeloid leukemia reveals recurrent structural alterations and age-specific mutational interactions. Nat Med 2018; 24: 103-112.

[23] Du R, Lu C, Jiang Z, Li S, Ma R, An H, Xu M, An Y, Xia Y, Jin L, Wang X and Zhang F. Efficient typing of copy number variations in a segmental duplication-mediated rearrangement hotspot using multiplex competitive amplification. J Hum Genet 2012; 57: 545-551.

[24] Kühn MW, Radtke I, Bullinger L, Goorha S, Cheng J, Edelmann J, Gohlke J, Su X, Paschka P, Pounds S, Krauter J, Ganser A, Quessar A, Ribeiro R, Gaidzik VI, Shurtleff S, Krönke J, Holzmann K, Ma J, Schlenk RF, Rubnitz JE, Döhner K, Döhner H and Downing JR. High-resolution genomic profiling of adult and pediatric core-binding factor acute myeloid leukemia reveals new recurrent genomic alterations. Blood 2012; 119: e67-75.

[25] Vujkovic M, Attiyeh EF, Ries RE, Goodman EK, Ding Y, Kavcic M, Alonzo TA, Wang YC, Gerbing RB, Sung L, Hirsch B, Raimondi S, Gamis AS, Meshinchi S and Aplenc R. Genomic architecture and treatment outcome in pediatric acute myeloid leukemia: a Children's Oncology Group report. Blood 2017; 129: 3051-3058.

[26] Liu B, Morrison CD, Johnson CS, Trump DL, Qin M, Conroy JC, Wang J and Liu S. Computational methods for detecting copy number variations in cancer genome using next generation sequencing: principles and challenges. Oncotarget 2013; 4: 1868-1881.

[27] Walter MJ, Payton JE, Ries RE, Shannon WD, Deshmukh H, Zhao Y, Baty J, Heath S, Westervelt P, Watson MA, Tomasson MH, Nagarajan R, O'Gara BP, Bloomfield CD, Mrózek K, Selzer RR, Richmond TA, Kitzman J, Geoghegan J, Eis PS, Maupin R, Fulton RS, McLellan M, Wilson RK, Mardis ER, Link DC, Graubert TA, DiPersio JF and Ley TJ. Acquired copy number alterations in adult acute myeloid leukemia genomes. Proc Natl Acad Sci U S A 2009; 106: 12950-12955.

[28] Peng WX, Koirala P and Mo YY. LncRNA-mediated regulation of cell signaling in cancer. Oncogene 2017; 36: 5661-5667.

[29] Bartel DP. MicroRNAs: target recognition and regulatory functions. Cell 2009; 136: 215-233.

[30] Gado MM, Mousa NO, Badawy MA, El Taweel MA and Osman A. Assessment of the diagnostic potential of miR-29a-3p and miR-92a-3p as circulatory biomarkers in acute myeloid leukemia. Asian Pac J Cancer Prev 2019; 20: 3625-3633.

[31] Zhou JD, Li XX, Zhang TJ, Xu ZJ, Zhang ZH, Gu Y, Wen XM, Zhang W, Ji RB, Deng ZQ, Lin J and Qian J. MicroRNA-335/ID4 dysregulation predicts clinical outcome and facilitates leukemogenesis by activating PI3K/Akt signaling pathway in acute myeloid leukemia. Aging (Albany NY) 2019; 11: 3376-3391.

[32] Fehrmann RS, Karjalainen JM, Krajewska M, Westra HJ, Maloney D, Simeonov A, Pers TH, Hirschhorn JN, Jansen RC, Schultes EA, van Haagen HH, de Vries EG, te Meerman GJ, Wijmenga C, van Vugt MA and Franke L. Gene expression analysis identifies global gene dosage sensitivity in cancer. Nat Genet 2015; 47: 115-125.

[33] Opatz S, Bamopoulos SA, Metzeler KH, Herold T, Ksienzyk B, Bräundl K, Tschuri S, Vosberg S, Konstandin NP, Wang C, Hartmann L, Graf A, Krebs S, Blum H, Schneider S, Thiede C, Middeke JM, Stölzel F, Röllig C, Schetelig J, Ehninger G, Krämer A, Braess J, Görlich D, Sauerland MC, Berdel WE, Wörmann BJ, Hiddemann W, Spiekermann K, Bohlander SK and Greif PA. The clinical mutatome of core binding factor leukemia. Leukemia 2020; 34: 1553-1562.

[34] Lv L, Yu J and Qi Z. Acute myeloid leukemia with inv(16)(p13.1q22) and deletion of the 5'MYH11/3'CBFB gene fusion: a report of two cases and literature review. Mol Cytogenet 2020; 13: 4.

[35] Morita K, Suzuki K, Maeda S, Matsuo A, Mitsuda Y, Tokushige C, Kashiwazaki G, Taniguchi J, Maeda R, Noura M, Hirata M, Kataoka T, Yano A, Yamada Y, Kiyose H, Tokumasu M, Matsuo H, Tanaka S, Okuno Y, Muto M, Naka K, Ito K, Kitamura T, Kaneda Y, Liu PP, Bando T, Adachi S, Sugiyama H and Kamikubo Y. Genetic regulation of the RUNX transcription factor family has antitumor effects. J Clin Invest 2017; 127: 2815-2828.

[36] Morita K, Noura M, Tokushige C, Maeda S, Kiyose H, Kashiwazaki G, Taniguchi J, Bando T, Yoshida K, Ozaki T, Matsuo H, Ogawa S, Liu PP, Nakahata T, Sugiyama H, Adachi S and Kamikubo Y. Autonomous feedback loop of RUNX1-p53-CBFB in acute myeloid leukemia cells. Sci Rep 2017; 7: 16604.

[37] Krygier A, Szmajda-Krygier D, Sałagacka-Kubiak A, Jamroziak K, Żebrowska-Nawrocka M and Balcerczak E. Association between the CEBPA and c-MYC genes expression levels and acute myeloid leukemia pathogenesis and development. Med Oncol 2020; 37: 109.

[38] Volk A, Liang K, Suraneni P, Li X, Zhao J, Bulic M, Marshall S, Pulakanti K, Malinge S, Taub J, Ge Y, Rao S, Bartom E, Shilatifard A and Crispino JD. A CHAF1B-dependent molecular switch in hematopoiesis and leukemia pathogenesis. Cancer Cell 2018; 34: 707-723, e707.

[39] Seeler JS and Dejean A. SUMO and the robustness of cancer. Nat Rev Cancer 2017; 17: 184-197.
[40] Wang L, Qian J, Yang Y and Gu C. Novel insights into the impact of the SUMOylation pathway in hematological malignancies (review). Int J Oncol 2021; 59: 73.
[41] Boulanger M, Paolillo R, Piechaczyk M and Bossis G. The SUMO pathway in hematomalignancies and their response to therapies. Int J Mol Sci 2019; 20: 3895.
[42] de Boer B, Prick J, Pruis MG, Keane P, Imperato MR, Jaques J, Brouwers-Vos AZ, Hogeling SM, Woolthuis CM, Nijk MT, Diepstra A, Wandinger S, Versele M, Attar RM, Cockerill PN, Huls G, Vellenga E, Mulder AB, Bonifer C and Schuringa JJ. Prospective isolation and characterization of genetically and functionally distinct AML subclones. Cancer Cell 2018; 34: 674-689, e678.
[43] Zuazo-Gaztelu I, Pàez-Ribes M, Carrasco P, Martín L, Soler A, Martínez-Lozano M, Pons R, Llena J, Palomero L, Graupera M and Casanovas O. Antitumor effects of anti-semaphorin 4D antibody unravel a novel proinvasive mechanism of vascular-targeting agents. Cancer Res 2019; 79: 5328-5341.
[44] Chen WG, Sun J, Shen WW, Yang SZ, Zhang Y, Hu X, Qiu H, Xu SC and Chu TW. Sema4D expression and secretion are increased by HIF-1α and inhibit osteogenesis in bone metastases of lung cancer. Clin Exp Metastasis 2019; 36: 39-56.
[45] Lu JJ, Su YW, Wang CJ, Li DF and Zhou L. Semaphorin 4D promotes the proliferation and metastasis of bladder cancer by activating the PI3K/AKT pathway. Tumori 2019; 105: 231-242.
[46] Gomes AP, Ilter D, Low V, Rosenzweig A, Shen ZJ, Schild T, Rivas MA, Er EE, McNally DR, Mutvei AP, Han J, Ou YH, Cavaliere P, Mullarky E, Nagiec M, Shin S, Yoon SO, Dephoure N, Massagué J, Melnick AM, Cantley LC, Tyler JK and Blenis J. Dynamic incorporation of histone H3 variants into chromatin is essential for acquisition of aggressive traits and metastatic colonization. Cancer Cell 2019; 36: 402-417, e413.
[47] Jiang H, Cao HJ, Ma N, Bao WD, Wang JJ, Chen TW, Zhang EB, Yuan YM, Ni QZ, Zhang FK, Ding XF, Zheng QW, Wang YK, Zhu M, Wang X, Feng J, Zhang XL, Cheng SQ, Ma DJ, Qiu L, Li JJ and Xie D. Chromatin remodeling factor ARID2 suppresses hepatocellular carcinoma metastasis via DNMT1-Snail axis. Proc Natl Acad Sci U S A 2020; 117: 4770-4780.
[48] Lee E, Wang J, Yumoto K, Jung Y, Cackowski FC, Decker AM, Li Y, Franceschi RT, Pienta KJ and Taichman RS. DNMT1 regulates epithelial-mesenchymal transition and cancer stem cells, which promotes prostate cancer metastasis. Neoplasia 2016; 18: 553-566.
[49] Huang L, Hu B, Ni J, Wu J, Jiang W, Chen C, Yang L, Zeng Y, Wan R, Hu G and Wang X. Transcriptional repression of SOCS3 mediated by IL-6/STAT3 signaling via DNMT1 promotes pancreatic cancer growth and metastasis. J Exp Clin Cancer Res 2016; 35: 27.
[50] Wong KK, Lawrie CH and Green TM. Oncogenic roles and inhibitors of DNMT1, DNMT3A, and DNMT3B in acute myeloid leukaemia. Biomark Insights 2019; 14: 1177271919846454.
[51] Sharma V and Montano MM. Non-epigenetic induction of HEXIM1 by DNMT1 inhibitors and functional relevance. Sci Rep 2020; 10: 21015.
[52] Mao Y, Evans EE, Mishra V, Balch L, Eberhardt A, Zauderer M and Gold WA. Anti-semaphorin 4D rescues motor, cognitive, and respiratory phenotypes in a rett syndrome mouse model. Int J Mol Sci 2021; 22: 9465.

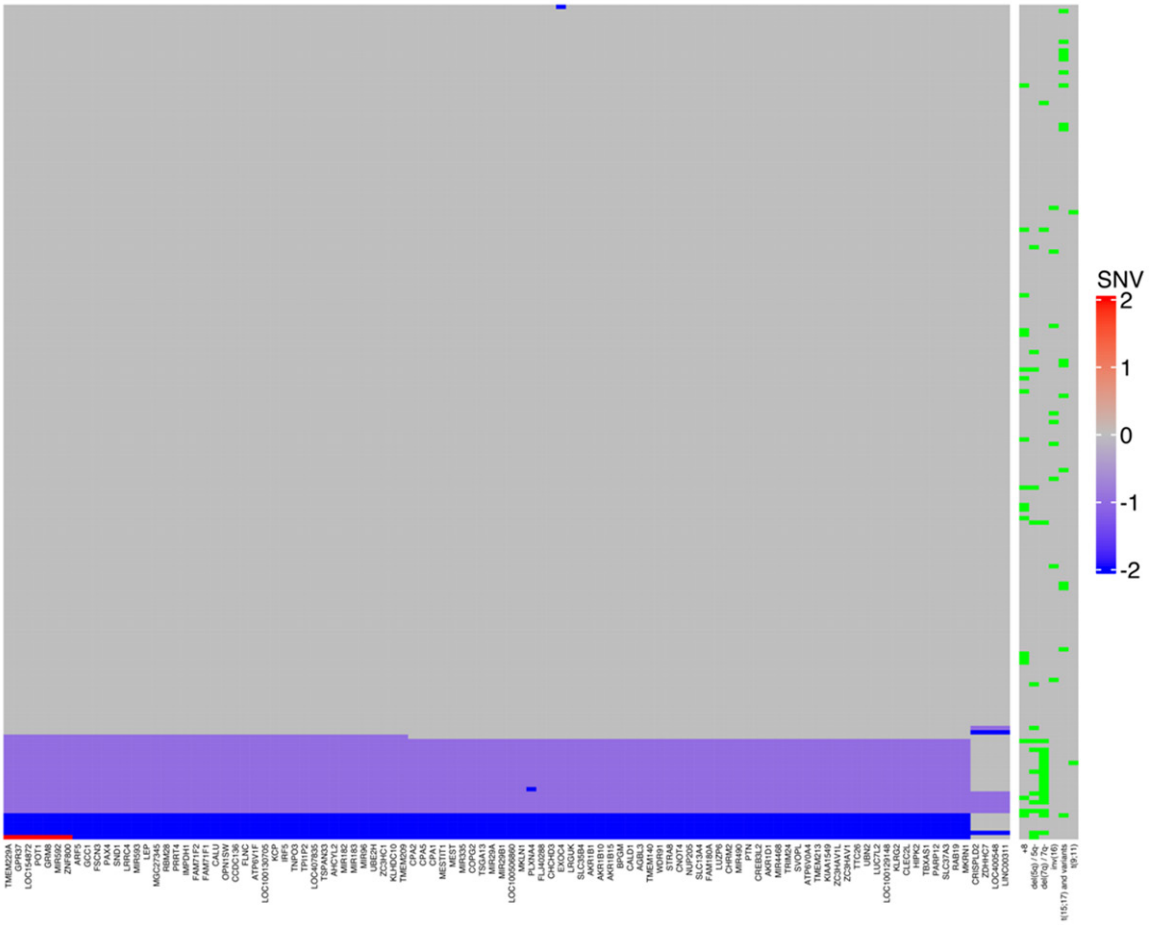

Figure S1. Distributions of CNVs and cytogenetic changes in TCGA cohort.

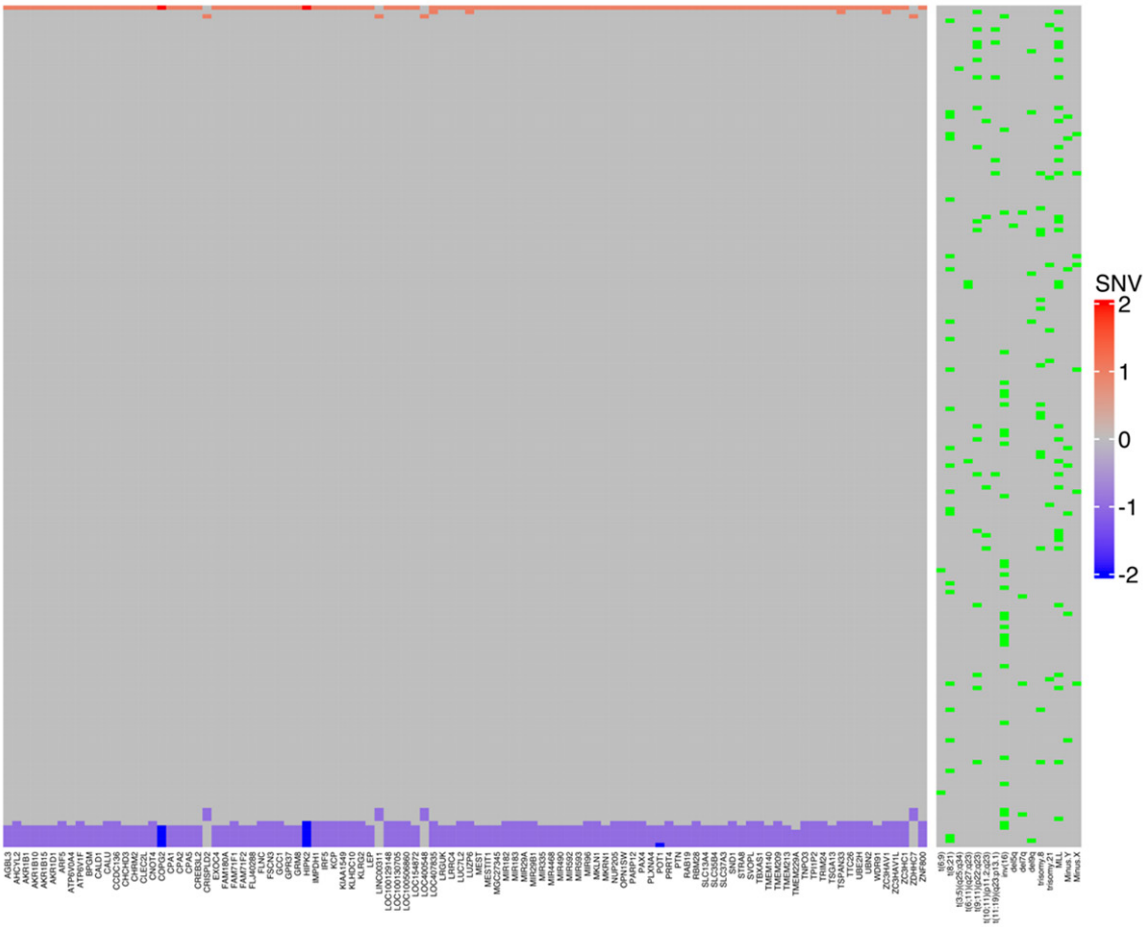

Figure S2. Distributions of CNVs and cytogenetic changes in TARGET cohort.

Table S1. A total of 102 CNV genes that were significantly associated with patient survival in both the TARGET and TCGA cohorts

102 CNV genes that were significantly associated with patient survival								
symbol of gene	Hazard ratios (HR) in TARGET cohort	HR confint Lower (95%) in TARGET cohort	HR confint upper (95%) in TARGET cohort	*P* value for CNV associating with patient survival in TARGET cohort	adjusted *P* value for CNV associating with patient survival in TARGET cohort	*P* value for CNV associating with patient survival in TCGA cohort	*P* value for CNV modulating gene expression in TCGA cohort	*P* value for CNV modulating gene expression in TARGET cohort
AGBL3	4.5	1.9	10	0.00048	0.00985389	0.047		
AHCYL2	4.3	1.9	9.4	0.00032	0.00770642	0.044	1.9573E-05	0.00057098
AKR1B1	4.5	1.9	10	0.00048	0.00985389	0.047	0.00611843	0.04634702
AKR1B10	4.5	1.9	10	0.00048	0.00985389	0.047		
AKR1B15	4.5	1.9	10	0.00048	0.00985389	0.047	0.02874165	0.12487514
AKR1D1	4.5	1.9	10	0.00048	0.00985389	0.047		
ARF5	4.3	1.9	9.4	0.00032	0.00770642	0.044	1.2459E-11	1.3726E-08
ATP6V0A4	4.5	1.9	10	0.00048	0.00985389	0.047		
ATP6V1F	4.3	1.9	9.4	0.00032	0.00770642	0.044	4.6522E-10	1.4644E-07
BPGM	4.5	1.9	10	0.00048	0.00985389	0.047		
CALD1	4.5	1.9	10	0.00048	0.00985389	0.047		
CALU	4.3	1.9	9.4	0.00032	0.00770642	0.044	1.9756E-07	1.445E-05
CCDC136	4.3	1.9	9.4	0.00032	0.00770642	0.044	0.00710411	0.05140568
CHCHD3	4.5	1.9	10	0.00048	0.00985389	0.047	3.1658E-09	5.7894E-07
CHRM2	4.5	1.9	10	0.00048	0.00985389	0.047		
CLEC2L	4.5	1.9	10	0.00048	0.00985389	0.047		
CNOT4	4.3	1.9	9.4	0.00032	0.00770642	0.047	5.5881E-09	8.7948E-07
COPG2	4.5	1.9	10	0.00048	0.00985389	0.047	2.9049E-09	5.5658E-07
CPA1	4.5	1.9	10	0.00048	0.00985389	0.047	0.00511289	0.04104045
CPA2	4.5	1.9	10	0.00048	0.00985389	0.047		
CPA5	4.5	1.9	10	0.00048	0.00985389	0.047		
CREB3L2	4.5	1.9	10	0.00048	0.00985389	0.047	6.2642E-05	0.00136827
CRISPLD2	5.8	2.1	16	0.00071	0.01428372	0.0097		
EXOC4	4.5	1.9	10	0.00048	0.00985389	0.022	6.3993E-09	9.479E-07
FAM180A	4.5	1.9	10	0.00048	0.00985389	0.047		
FAM71F1	4.3	1.9	9.4	0.00032	0.00770642	0.044		
FAM71F2	4.5	1.9	10	0.00048	0.00985389	0.044		
FLJ40288	4.5	1.9	10	0.00048	0.00985389	0.047		
FLNC	4.3	1.9	9.4	0.00032	0.00770642	0.044		
FSCN3	4.3	1.9	9.4	0.00032	0.00770642	0.044	1.4609E-11	1.5148E-08

GCC1	4.3	1.9	9.4	0.00032	0.00770642	0.044	2.1367E-11	1.9063E-08
GPR37	4.5	1.9	10	0.00048	0.00985389	0.044		
GRM8	4.5	1.9	10	0.00048	0.00985389	0.044	0.00252367	0.02450953
HIPK2	4.3	1.9	9.4	0.00032	0.00770642	0.047	4.2758E-05	0.00100606
IMPDH1	4.3	1.9	9.4	0.00032	0.00770642	0.044	8.3049E-09	1.1415E-06
IRF5	4.3	1.9	9.4	0.00032	0.00770642	0.044	2.0392E-06	9.1695E-05
KCP	4.3	1.9	9.4	0.00032	0.00770642	0.044	6.9497E-09	1.0124E-06
KIAA1549	4.3	1.9	9.4	0.00032	0.00770642	0.047	0.02462633	0.11417369
KLHDC10	4.3	1.9	9.4	0.00032	0.00770642	0.044	1.4644E-07	1.1575E-05
KLRG2	4.3	1.9	9.4	0.00032	0.00770642	0.047		
LEP	4.5	1.9	10	0.00048	0.00985389	0.044		
LINC00311	5.8	2.1	16	0.00071	0.01428372	0.0097		
LOC100129148	4.3	1.9	9.4	0.00032	0.00770642	0.047		
LOC100130705	4.3	1.9	9.4	0.00032	0.00770642	0.044		
LOC100506860	4.3	1.9	9.4	0.00032	0.00770642	0.047		
LOC154872	4.5	1.9	10	0.00048	0.00985389	0.044		
LOC400548	5.8	2.1	16	0.00071	0.01428372	0.0097		
LOC407835	3.6	1.7	7.4	0.00071	0.01428372	0.044		
LRGUK	4.5	1.9	10	0.00048	0.00985389	0.047	0.00036067	0.00567432
LRRC4	4.5	1.9	10	0.00048	0.00985389	0.044	0.00408081	0.03480037
LUC7L2	4.3	1.9	9.4	0.00032	0.00770642	0.047	2.0843E-12	3.674E-09
LUZP6	3.6	1.7	7.4	0.00071	0.01428372	0.047	8.2753E-09	1.1415E-06
MEST	4.5	1.9	10	0.00048	0.00985389	0.047		
MESTIT1	4.5	1.9	10	0.00048	0.00985389	0.047		
MGC27345	4.5	1.9	10	0.00048	0.00985389	0.044		
MIR182	4.3	1.9	9.4	0.00032	0.00770642	0.044		
MIR183	4.3	1.9	9.4	0.00032	0.00770642	0.044		
MIR29A	4.3	1.9	9.4	0.00032	0.00770642	0.047		
MIR29B1	4.3	1.9	9.4	0.00032	0.00770642	0.047		
MIR335	4.5	1.9	10	0.00048	0.00985389	0.047		
MIR4468	4.5	1.9	10	0.00048	0.00985389	0.047		
MIR490	4.5	1.9	10	0.00048	0.00985389	0.047		
MIR592	4.5	1.9	10	0.00048	0.00985389	0.044		
MIR593	4.5	1.9	10	0.00048	0.00985389	0.044		
MIR96	4.3	1.9	9.4	0.00032	0.00770642	0.044		
MKLN1	4.3	1.9	9.4	0.00032	0.00770642	0.047	1.2659E-06	6.357E-05

MKRN1	4.5	1.9	10	0.00048	0.00985389	0.047	2.1629E-11	1.9063E-08
NUP205	4.3	1.9	9.4	0.00032	0.00770642	0.047	2.4036E-07	1.6746E-05
OPN1SW	4.3	1.9	9.4	0.00032	0.00770642	0.044	2.1281E-07	1.5273E-05
PARP12	4.3	1.9	9.4	0.00032	0.00770642	0.047		
PAX4	4.3	1.9	9.4	0.00032	0.00770642	0.044		
PLXNA4	4.5	1.9	10	0.00048	0.00985389	0.047		
POT1	4.5	1.9	10	0.00048	0.00985389	0.044	8.3608E-06	0.00029011
PRRT4	4.3	1.9	9.4	0.00032	0.00770642	0.044		
PTN	4.5	1.9	10	0.00048	0.00985389	0.047		
RAB19	4.5	1.9	10	0.00048	0.00985389	0.047		
RBM28	4.3	1.9	9.4	0.00032	0.00770642	0.044	3.6686E-08	3.7597E-06
SLC13A4	4.5	1.9	10	0.00048	0.00985389	0.047	0.00012173	0.00236049
SLC35B4	4.5	1.9	10	0.00048	0.00985389	0.047	3.48E-10	1.278E-07
SLC37A3	4.5	1.9	10	0.00048	0.00985389	0.047		
SND1	4.3	1.9	9.4	0.00032	0.00770642	0.044	3.0195E-11	2.2177E-08
STRA8	4.5	1.9	10	0.00048	0.00985389	0.047		
SVOPL	4.3	1.9	9.4	0.00032	0.00770642	0.047	0.00339381	0.03053508
TBXAS1	4.3	1.9	9.4	0.00032	0.00770642	0.047	1.8415E-08	2.1078E-06
TMEM140	4.5	1.9	10	0.00048	0.00985389	0.047	0.00036392	0.00570716
TMEM209	4.3	1.9	9.4	0.00032	0.00770642	0.044	2.17E-05	0.00061694
TMEM213	4.5	1.9	10	0.00048	0.00985389	0.047		
TMEM229A	3.7	1.5	9.3	0.005	0.04368599	0.044		
TNPO3	4.3	1.9	9.4	0.00032	0.00770642	0.044	4.5869E-12	6.5345E-09
TPI1P2	4.3	1.9	9.4	0.00032	0.00770642	0.044	1.5589E-07	1.2213E-05
TRIM24	4.3	1.9	9.4	0.00032	0.00770642	0.047	0.00011941	0.00232075
TSGA13	4.5	1.9	10	0.00048	0.00985389	0.047	9.8026E-06	0.00032912
TSPAN33	3.6	1.7	7.4	0.00071	0.01428372	0.044	0.00536612	0.04241644
TTC26	4.3	1.9	9.4	0.00032	0.00770642	0.047	3.5588E-05	0.00088853
UBE2H	4.3	1.9	9.4	0.00032	0.00770642	0.044	1.2889E-05	0.00041287
UBN2	4.3	1.9	9.4	0.00032	0.00770642	0.047	2.1181E-05	0.00060809
WDR91	4.5	1.9	10	0.00048	0.00985389	0.047	1.0928E-05	0.00036276
ZC3HAV1	3.6	1.7	7.4	0.00071	0.01428372	0.047	1.5091E-05	0.00046668
ZC3HAV1L	4.3	1.9	9.4	0.00032	0.00770642	0.047	0.00431297	0.03615988
ZC3HC1	4.3	1.9	9.4	0.00032	0.00770642	0.044	6.5183E-10	1.8238E-07
ZDHHC7	5.8	2.1	16	0.00071	0.01428372	0.0097	0.00308482	0.02855581
ZNF800	4.3	1.9	9.4	0.00032	0.00770642	0.044	7.6367E-07	4.2622E-05