Int J Clin Exp Med 2020;13(8):5702-5716
www.ijcem.com /ISSN:1940-5901/IJCEM0108497

Original Article

Identification of upregulated DOCK1 in leukemia stem cells associated with risk stratification in acute myeloid leukemia through bioinformatics analysis

Yeqi Sun[1], Peng Chen[2,3], Daiqiang Li[1]

[1]Department of Pathology, Xiangya Second Hospital, Central South University, No. 139 Middle Renmin Road, Changsha 410011, Hunan, P. R. China; [2]Department of Clinical Pharmacology, Xiangya Hospital, Central South University, No. 87 Xiangya Road, Changsha 410008, Hunan, P. R. China; [3]Institute of Clinical Pharmacology, Central South University, Hunan Key Laboratory of Pharmacogenetics, No. 110 Xiangya Road, Changsha 410078, Hunan, P. R. China

Received January 30, 2020; Accepted June 3, 2020; Epub August 15, 2020; Published August 30, 2020

Abstract: Xenotransplantation studies showed that acute myeloid leukemia (AML) is organized as a hierarchy sustained by leukemia stem cells (LSCs), and the LSCs persistence influences AML clinical outcomes because of the high relapse rate and therapy resistance. This study aimed to identify differentially expressed genes between LSCs and non-LSCs, from which prognosis-related genes were extracted and confirmed in AML cohorts. Meanwhile, involved alterations of methylation pattern were analyzed. Differentially expressed genes between LSC and non-LSC fractions were screened in GSE30377 and GSE76008 datasets from the Gene Expression Omnibus (GEO). AML prognosis-related genes were identified and verified in TCGA, GEO, TARGET databases and Beat AML programme. To understand its molecular mechanisms, the altered degrees of methylation were investigated in GEO, TCGA and TARGET databases. The expression level of *DOCK1* was higher in LSCs than in non-LSCs, and was also implicated with LSC-related gene through the gene set enrichment analysis. Besides, the significant differences in gene expression between AML CD34+ and CD34- cells, normal CD34+ and mononuclear cells, as well as hematopoietic stem cells and granulocytes/monocytes were observed. The significantly lower methylation levels were found in the *DOCK1* promoter of LSCs, and the mechanism of *DOCK1* upregulation in LSCs could be mediated by hypomethylation. Furthermore, the high expression of *DOCK1* was associated with a worse prognosis in six independent cohorts of *de novo* AML patients. Our results indicated that the high expression of *DOCK1* regulated by promoter hypomethylation was a potential biomarker of LSCs and, more importantly, a predicted adverse prognosis factor for AML.

Keywords: AML, DOCK1, prognosis, leukemia stem cell, expression

Introduction

Acute myeloid leukemia (AML) is a heterogeneous malignancy characterized by differentiation block, abnormal proliferation and malignant clone of hematopoietic stem and progenitor cells (HSPCs) [1]. In recent decades, advances in understanding the pathogenesis and supportive care have led to improved prognosis of AML, but the overall 5-year survival rate in AML patients remains relatively low (about 15%), especially for elderly patients [2]. Apart from the age at diagnosis, white blood cell (WBC) counts, antecedent hematologic diseases and therapy-related leukemia, the factors of AML prognosis include cytogenetic aberrations and somatic mutations. Based on genomic analysis, AML cases can be compartmentalized into 11 classes with distinct diagnostic features and clinical outcomes [3]. Recent studies have also reported that some highly expressed genes were associated with adverse prognosis of AML patients, such as *ERG* [4], *BAALC* [4], *WT1* [5], *DNMT3B* [6], and *GLI1* [7]. However, existing prognostic factors fail to fully account for AML prognosis, especially in patients with intermediate-risk AML. The identification of effective prognostic biomarkers is one of the most urgent needs in AML clinical practice.

The cellular hierarchy and functional heterogeneity exist in leukemia cell populations. A cell subpopulation was found to possess properties similar to normal hematopoietic stem cells (HSCs), known as "leukemia initiating cells" or "leukemia stem cells" (LSCs) [8]. The LSCs were found to reside at the top of a cellular hierarchy that gives rise to and maintains the process of self-renewal and differentiation leukemia. Due to the characteristics of enhanced drug efflux, relative quiescence, and apoptosis inhibition, these LSCs are highly resistant to the majority of chemotherapy drugs aimed at cell cycle [9]. In addition, this subpopulation usually brings about minimal residual disease (MRD) after induction therapy in AML, which ultimately causes disease relapse. According to the cancer stem cell model, there is need for adopting better therapy strategies to eliminate LSCs for cure and novel biomarkers to identify these high-risk AML patients [10].

Recent high-throughput studies have increased our understanding of LSCs and revealed LSC-specific expression gene and epigenetic signatures, for example the extracted AML prognosis-related genes [11-13]. However, an independent large-scaled study often produced weak, inconsistent, and even contradictory results. Here, we showed that highly expressed dedicator of cytokinesis 1 (*DOCK1*) in LSCs was associated with a poor prognosis in multiple independent AML cohort. Besides, the distinctive epigenomic pattern differences between LSC and non-LSC populations were in-depth explored. This study demonstrated that the *DOCK1* expression level could be used as a poor prognostic biomarker for AML, which could gain new sights into risk stratification for patients with AML through genetic/epigenetic changes.

Materials and methods

Differential expression analysis

SOFT files of GSE30377 and GSE76008 were downloaded from Gene Expression Omnibus (GEO, http://www.ncbi.nlm.nih.gov/geo). The expression profile data of 25 LSC and 29 non-LSC, 138 LSC and 89 non-LSC fractions were respectively extracted from GSE30377 and GSE76008. The comparison of gene expression levels in LSC and non-LSC fractions used Smyth's moderated t-test with Benjamini-Hochberg multiple testing correction by the limma 3.30.13 package in R. Unadjusted *P*-value < 0.001 and absolute value of logFC ≥ 0.5 were selected as the cut-off criteria in the differential gene expression analyses. The heatmap was performed by the pheatmap 1.0.10 R package.

Screening of AML prognostic genes in TCGA database

The preprocessed data of level 3 RNA-seq of TCGA AML cohort were downloaded from the UCSC Xena Hubs (https://xenabrowser.net/), and the clinical information was downloaded from the cBioPortal (http://www.cbioportal.org/). A total of 173 AML cases with gene expression profile and detailed follow-up time were included for analysis.

Gene set enrichment analysis

The association of *DOCK1* gene with LSC property was investigated using the Gene Set Enrichment Analysis (GSEA, version 3.0). The global gene list was ranked by Pearson's correlation coefficients with *DOCK1* expression level, and the statistical significance of enrichment score was determined by 1000-time phenotype permutations. The LSC signature gene sets were generated by studies of GSE30377 and GSE76008 [11, 12].

Verification of the AML prognostic gene

GSE12417 SOFT file containing processed data of expression profiles and overall survival (OS) for two independent cohorts of normal karyotype AML cases were downloaded from GEO database [14]. We removed one case diagnosed with myelodysplastic syndromes (MDS) in each cohort, leaving 162 and 78 cases for analysis respectively. The exemplar sequence probe of 203187_at was selected for *DOCK1* gene in HG-U133 A (GPL96) or HG-U133_Plus_2 (GPL570) arrays. The cohort of GSE71014 included 104 cytogenetically normal AML patients [15], whose data was also downloaded from GEO. For the therapeutically applicable research to generate effective treatments (TARGET) AML project [16], clinical annotation and 358 open access RNA sequencing data were downloaded from National Cancer Institute (NCI) (https://ocg.cancer.gov/programs/target), and 256 *de novo* speci-

mens were included in analysis after removal of induction failure or recurrent samples from childhood AML cases. Clinical data and *DOCK1* relative expression of samples in Beat AML programme were downloaded from the Vizome (http://www.vizome.org/aml/) [17]. Of the 288 *de novo* samples in the Beat AML cohort, 65 specimens without *DOCK1* expression level, 22 cases with missing OS data, and 2 duplications were excluded, leaving 199 cases for analysis.

DOCK1 methylation and expression level of different cell subpopulations

SOFT file of GSE63409 containing genome-wide methylation profiles of LSC and non-LSC fractions from bone marrow of AML patients performed by Illumina HumanMethylation450 BeadChip was downloaded from GEO. *DOCK1* DNA methylation levels of 20 LSC and 24 non-LSC fractions (blast cells) were compared using the limma package in R. The preprocessed DNA methylation data of TCGA and TARGET downloaded from the same databases were defined as gene expression data. We included 170 and 113 cases with genomic methylation data and corresponding expression profiles in TCGA and TARGET AML projects for analysis.

GSE30029 non-normalized TXT file that contains non-normalized summary values and detection *P*-values for regular probes from Illumina BeadChip Arrays (HT12 v3) was downloaded from GEO [18]. Raw intensities were corrected using normexp background, quantile normalized and log2-transformed by limma package [19]. We filtered out probes that were not detected in 3 samples or less (detection *P*-values < 0.05). The paired CD34+ and CD34- cell subfraction specimens from 44 AML cases were included for analysis in GSE30029 dataset. The processed expression data of GSE24759 and Beat-AML dataset were downloaded from GEO and Vizome databases respectively.

Statistical analysis

All statistical analyses were performed in R language (3.3.3 version). Pearson chi square or Fisher's exact test were applied to compare categorical data, and student's t test or Wilcoxon test for two continuous variables. The Pearson or Spearman rank method of correlation was used. Univariate survival analysis was performed by the Kaplan-Meier with log-rank tests for comparing two survival curves. Gene expression level divided into high and low expression using a median threshold. Hazard ratios (HR) for OS and disease-free survival (DFS) were estimated by Cox proportional hazards model. All survival analyses were performed by the survival 2.42-3 package.

Results

Differentially expressed genes between LSC and non-LSC populations

Differentially expressed (DE) genes between LSC and non-LSC fractions were screened in GSE30377 and GSE76008, in which AML blast cells were sorted by surface marker expression of CD34 and CD38, and then LSC activity of the subpopulations was assessed by mice xenotransplant assay [13, 14]. There were 132 probes in GSE30377, and 1004 probes in GSE76008 with over 1.5-fold difference ($P < 0.001$, **Figure 1A**, **1B**). We defined reference profile of LSC as average expression level of the 37 differential genes in LSC fractions, and found that engraftment ability of individual cell subpopulations showed strong correlations with the similarity between gene expression and the corresponding reference profile of LSC in each dataset (GSE30377: r=0.58, $P=4.8\times10^{-6}$; GSE76008: r=0.33, $P=3.8\times10^{-7}$, **Figure 1C**, **1D**). Ultimately, a total of 37 overlapping DE genes in the two microarray datasets were listed in **Table 1** (**Figure 2**). In the cross validations, there were also strong correlations between the ability to engraft and its expression level similarity to the reference profile of LSC generated by the other dataset (GSE30377: r=0.47, $P=0.00038$; GSE76008: r=0.42, $P=4.0\times10^{-11}$, Figure S1). The above findings suggest that the 37 common DE genes are associated with the LSC expression features.

Extraction of LSC gene related to AML disease prognosis

To further identify the gene expression level associated with clinical outcomes in these DE genes between LSCs and non-LSCs, we investigated the association between expression level of these 37 genes and prognosis of AML patients in TCGA dataset. Among them, the

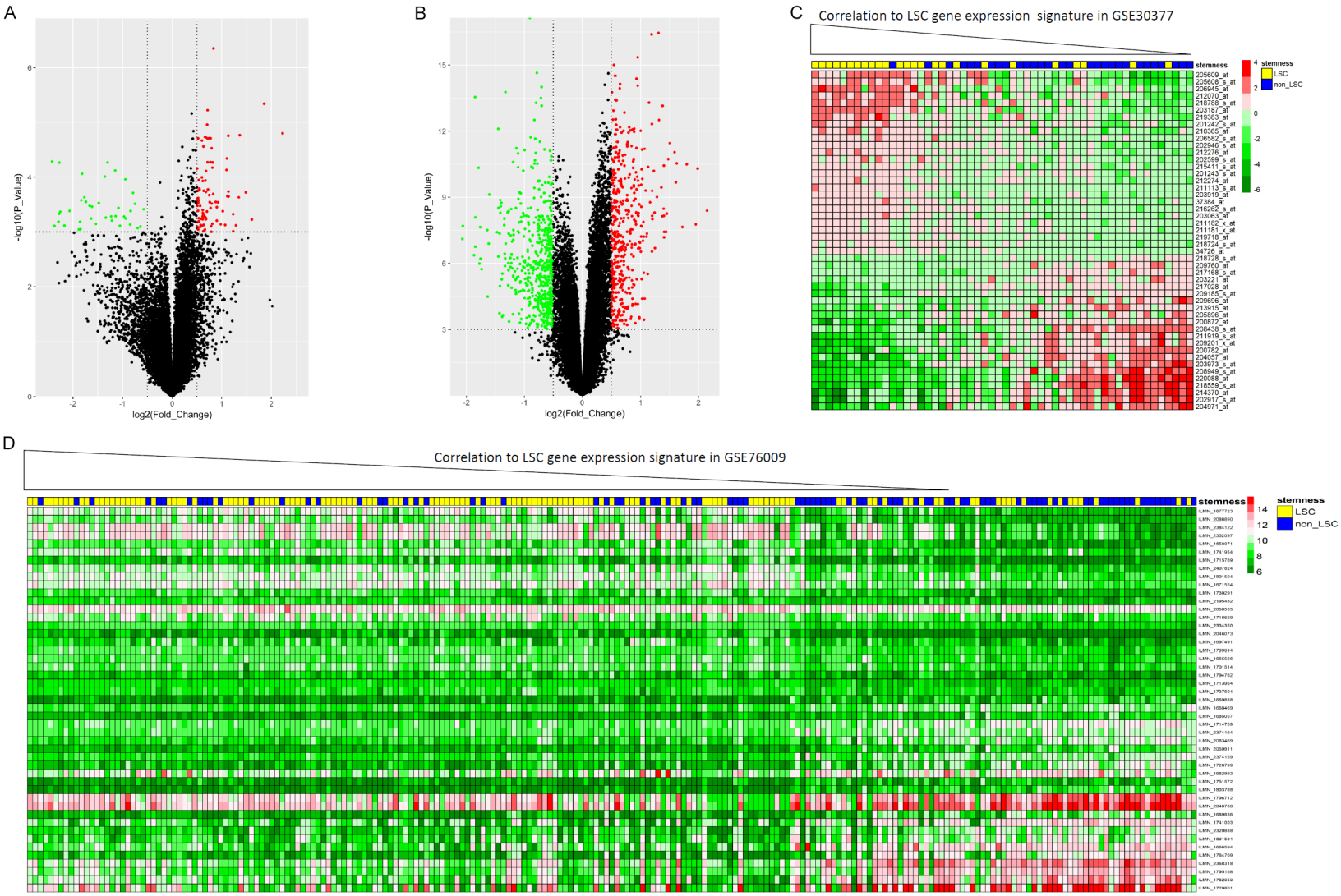

Figure 1. Screening of genes involved with leukemia stem cells (LSCs). (A, B) Volcano plots of differentially expressed (DE) genes between LSC and non_LSC fractions from bone marrow of AML patients. There are 93 up and 39 down regulated probes in GSE30377 (A, 25 vs 29 samples), 453 up and 551 down regulated probes in GSE76008 (B, 138 vs 89 samples). (C, D) Heat map of the common DE genes (rows) in the GSE30377 (C) and GSE76008 (D) datasets. Cases (columns) are ranked by correlation coefficients between expression profile of DE genes and the corresponding reference profile of LSC. Yellow and blue bars denote LSC and non_LSC fractions in the top row, respectively.

Table 1. The list of 37 differentially expressed LSC genes

Symbol	Entrez ID	GSE76008			GSE30377			TCGA AML	
		Probe ID	logFC	*P* Value	Probe ID	logFC	*P* Value	HR	*P* Value
ANGPT1	284	ILMN_1677723	1.51	3.5×10^{-12}	205609_at	1.60	0.00059	1.45	0.04827
ANGPT1	284	ILMN_2086890	1.09	5.0×10^{-10}	205608_s_at	1.49	0.00019		
GPR56	9289	ILMN_2384122	1.07	4.0×10^{-5}	212070_at	1.28	0.00076	1.45	0.04711
GPR56	9289	ILMN_2352097	1.03	4.0×10^{-5}	206582_s_at	0.80	2.96×10^{-5}		
ATP1B1	481	ILMN_1658071	0.96	5.2×10^{-9}	201242_s_at	1.10	4.54×10^{-5}	0.67	0.03737
SMYD3	64754	ILMN_1741954	0.95	1.8×10^{-10}	218788_s_at	1.23	0.00097	0.94	0.72797
DOCK1	1793	ILMN_1715789	0.94	1.2×10^{-10}	203187_at	1.23	0.00047	2.48	1.54×10^{-6}
ATP1B1	481	ILMN_2407824	0.90	1.4×10^{-7}	201243_s_at	0.71	1.10×10^{-5}		
RUNX1	861	ILMN_1801504	0.89	7.7×10^{-13}	210365_at	0.98	0.00073	0.86	0.42225
RUNX1	861				211182_x_at	0.57	0.00081		
LPIN1	23175	ILMN_1671554	0.73	4.0×10^{-7}	212276_at	0.76	0.00029	1.79	0.00197
LPIN1	23175				212274_at	0.67	0.00055		
ATP1B1	481	ILMN_1730291	0.70	5.0×10^{-7}					
CACNB3	784	ILMN_2195482	0.66	1.6×10^{-10}	34726_at	0.52	0.00024	1.22	0.29670
PPM1F	9647	ILMN_2059535	0.65	1.7×10^{-9}	37384_at	0.62	0.00040	1.61	0.01175
PPM1F	9647				203063_at	0.58	0.00057		
NRIP1	8204	ILMN_1718629	0.65	4.6×10^{-5}	202599_s_at	0.75	0.00021	1.37	0.09019
BTBD3	22903	ILMN_2334350	0.64	7.2×10^{-15}	202946_s_at	0.79	0.00079	0.86	0.44048
LCT	3938	ILMN_2046073	0.61	8.7×10^{-7}	206945_at	1.35	0.00023	1.48	0.03509
FLJ14213	79899	ILMN_1697491	0.59	0.00026	219383_at	1.11	0.00022	NA	NA
TGIF2	60436	ILMN_1709044	0.57	1.5×10^{-11}	216262_s_at	0.60	2.15×10^{-5}	0.75	0.12154
TGIF2	60436				218724_s_at	0.53	0.00026		
TCEA2	6919	ILMN_1665526	0.54	8.4×10^{-9}	203919_at	0.64	0.00053	1.23	0.26682
TRAF3IP2	10758	ILMN_1701514	0.53	4.3×10^{-8}	215411_s_at	0.72	5.99×10^{-6}	0.79	0.20048
ABCG1	9619	ILMN_1794782	0.52	4.7×10^{-6}	211113_s_at	0.66	1.85×10^{-5}	1.40	0.07410
BTBD3	22903	ILMN_1713964	0.51	1.7×10^{-14}					
FLJ10986	55277	ILMN_1737604	0.50	1.8×10^{-9}	211181_x_at	0.56	0.00034	NA	NA
FLJ10986	55277	ILMN_1737604	0.50	1.8×10^{-9}	219718_at	0.55	0.00016		
CSTA	1475	ILMN_1669888	-0.51	0.00015	204971_at	-2.42	5.18×10^{-5}	0.95	0.76276
KIAA0922	23240	ILMN_1668469	-0.54	1.5×10^{-9}	209760_at	-0.64	0.00080	1.00	0.99373
SLC22A4	6583	ILMN_1685057	-0.55	9.4×10^{-8}	205896_at	-1.30	5.38×10^{-5}	0.73	0.08834
CNIH4	29097	ILMN_1714759	-0.57	8.4×10^{-7}	218728_s_at	-0.57	0.00039	0.95	0.76857
HERPUD1	9709	ILMN_2374164	-0.57	1.2×10^{-10}	217168_s_at	-0.69	0.00084	0.98	0.92180
IRS2	8660	ILMN_2083469	-0.64	4.1×10^{-5}	209185_s_at	-0.97	0.00052	0.84	0.34355
SLC22A4	6583	ILMN_2050911	-0.66	5.1×10^{-8}					
HERPUD1	9709	ILMN_2374159	-0.71	1.4×10^{-8}					
FBP1	2203	ILMN_1728799	-0.74	0.00023	209696_at	-1.05	0.00077	0.92	0.65840
NKG7	4818	ILMN_1682993	-0.76	8.9×10^{-5}	213915_at	-1.22	0.00062	1.07	0.72274
TLE1	7088	ILMN_1751572	-0.80	9.7×10^{-12}	203221_at	-0.75	0.00031	0.89	0.53744
LGALS3	3958	ILMN_1803788	-0.89	3.0×10^{-9}	208949_s_at	-1.88	0.00091	1.10	0.62431
S100A10	6281	ILMN_1796712	-0.89	0.00053	200872_at	-1.48	0.00030	1.24	0.24474
S100A10	6281	ILMN_2046730	-0.92	0.00063					
C5AR1	728	ILMN_1689836	-0.94	2.9×10^{-7}	220088_at	-2.07	0.00041	1.11	0.58857
ANXA5	308	ILMN_1741003	-0.96	0.00030	200782_at	-1.77	0.00023	1.25	0.23461
CXCR4	7852	ILMN_2320888	-0.99	9.0×10^{-7}	217028_at	-0.94	0.00011	1.22	0.28692
CXCR4	7852	ILMN_1801584	-1.04	2.1×10^{-6}	211919_s_at	-1.62	0.00034		

CXCR4	7852				209201_x_at	-1.66	0.00033		
IRF8	3394	ILMN_1666594	-1.14	5.1×10^{-6}	204057_at	-1.83	0.00026	1.21	0.30717
MAFB	9935	ILMN_1764709	-1.15	1.6×10^{-5}	218559_s_at	-2.17	0.00077	1.23	0.27097
FGR	2268	ILMN_2368318	-1.19	4.0×10^{-6}	208438_s_at	-1.54	0.00072	1.39	0.07916
FGR	2268	ILMN_1795158	-1.22	4.0×10^{-7}					
CEBPD	1052	ILMN_1782050	-1.22	1.3×10^{-5}	203973_s_at	-1.86	0.00065	0.81	0.27274
S100A8	6279	ILMN_1729801	-1.63	3.2×10^{-5}	214370_at	-2.26	0.00043	1.30	0.16699
S100A8	6279				202917_s_at	-2.28	0.00047		

results of cox model analysis showed that the only gene associated with a shorter OS in AML patients was an increased *DOCK1* expression level (HR=2.48, $P < 0.001$, **Table 1**). Using gene set enrichment analysis (GSEA), we examined the associations between *DOCK1* expression pattern in the TCGA AML expression profile and the LSC gene signatures sets that were generated by studies of GSE30377 and GSE76008 [13, 14]. These LSC signatures were enriched in the genes ranked by Pearson's correlation coefficients with *DOCK1* expression level (enrichment score, ES=0.60, 0.59, P=0.032, 0.004, respectively, **Figure 3A**, **3B**). Therefore, the high expression of *DOCK1* was considered as a biomarker of LSC and a prognosis factor for AML.

Next, the influence of *DOCK1* expression on the clinical features and disease prognosis in the TCGA AML cohort was analyzed. As shown in **Table 2**, AML patients with a high *DOCK1* expression level were older ($P=2.1\times10^{-4}$) and less frequent in the good but over-represented in the poor cytogenetic risk stratification ($P=3.7\times10^{-4}$). High *DOCK1* expression was significantly associated with inferior OS and DFS in AML cases (**Figure 3C**, median OS 10.0 vs 46.5 months, $P=1.5\times10^{-6}$; **Figure 3D**, median DFS 12.0 vs 24.2 months, P=0.003). In addition, the results of cox proportional hazards model suggested that *DOCK1* expression indicated AML prognosis significance independent of other outcomes predictors including age at diagnosis, initial WBC count and cytogenetic risk stratifications (**Table 3**, OS: HR=1.06, P=0.03; DFS: HR=1.09, P=0.02).

Confirmation of the association between DOCK1 expression and AML prognosis

Next, the association of *DOCK1* expression level with disease prognosis in other five independent AML cohorts were evaluated, whose gene expression profile and survival data are publicly available. Three cohorts of non-M3 AML patients with normal karyotype were from GEO accession GSE12417 and GSE71014 [14, 15]. The results exhibited that high *DOCK1* expression was significantly associated with short OS of AML cases in GSE12417-GPL96 (**Figure 4A**, median OS 7.9 vs 24.7 months, $P=9.8\times10^{-5}$). Similarly, cases with high *DOCK1* expression also had marginally poor OS in GSE12417-GPL570 cohort (**Figure 4B**, median OS 13.3 vs 37.3 months, P=0.053). In the GSE71014 cohort, the median OS was not reached in patients with low *DOCK1* expression, but was 20.0 months in patients with high *DOCK1* expression (**Figure 4C**, P=0.0075). Similar results were observed in the TARGET AML cohort [16]: the childhood AML cases with a high *DOCK1* expression level had an inferior OS and event-free survival (EFS) than those with low *DOCK1* expression (**Figure 4D**, median OS 45.4 months vs Not Reach, P=0.0020; **Figure 4E**, median EFS 12.8 vs 25.4 months, $P=4.6\times10^{-5}$). The Beat AML programme that mapped functional genomic landscape of AML was newly published [17]. As expected, a significantly short OS of *de novo* AML cases with high *DOCK1* expression was observed in the programme (**Figure 4F**, median OS 14.6 vs 33.5 months, P=0.0019).

DOCK1 DNA methylation pattern in AML cell subpopulations

The GSE63409 dataset that provided genome-wide methylation profiles of LSC and non-LSC subpopulations was interrogated here [13]. The methylation levels of 8 CpG sites near *DOCK1* transcription start site (TSS) in LSC populations were significantly lower than those in non-LSC (**Figures 5A** and S2). It was noteworthy that these differentially methylated sites are all located in a CpG island. Thus, this region could be served as a differentially meth-

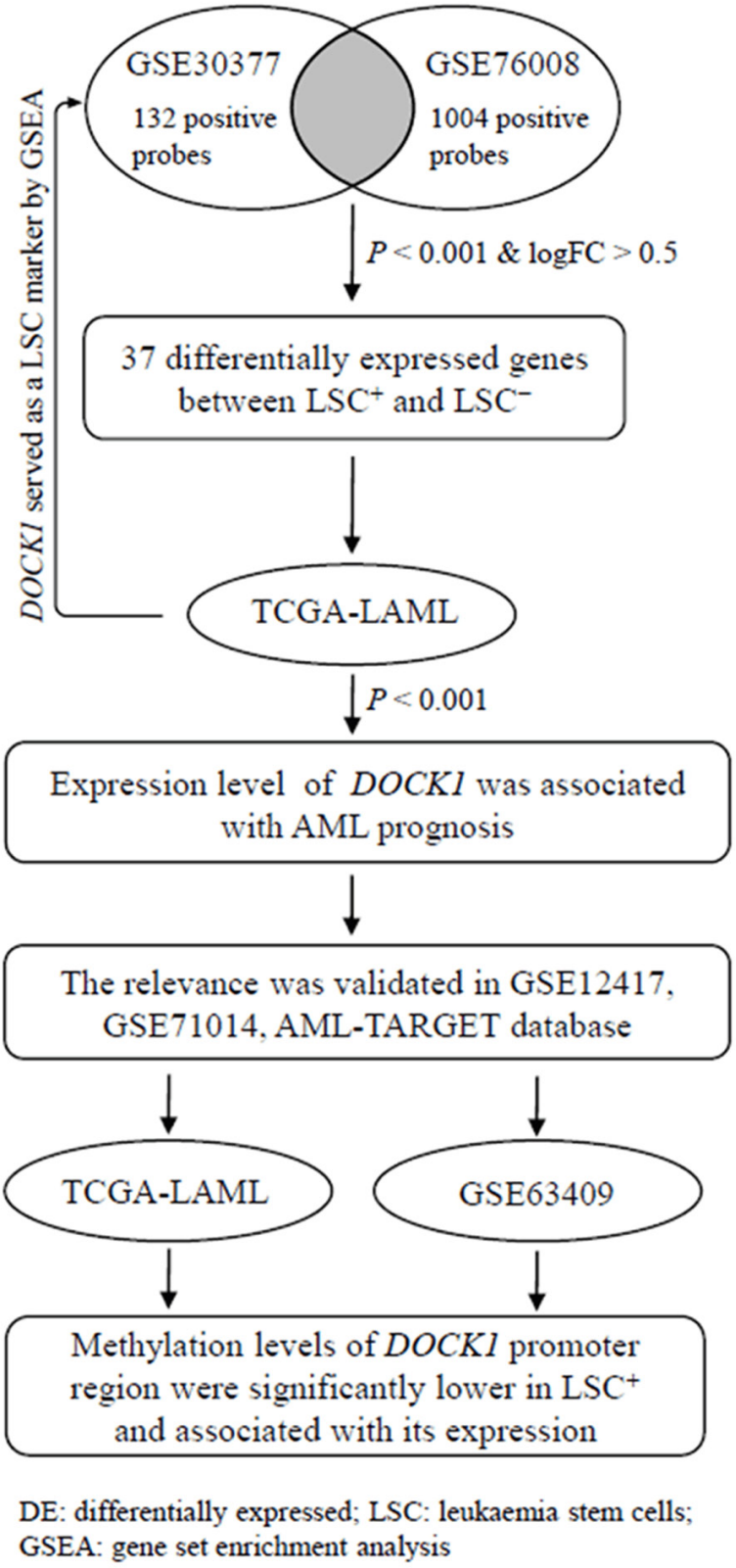

Figure 2. Flow chart of the study.

ylated region (DMR) between LSC and non-LSC subpopulations.

In order to explore whether methylation levels of the region were involved in *DOCK1* transcriptional regulation, we investigated the relationship between *DOCK1* promoter region methylation status and the gene expression in AML samples. The results indicated that DNA methylation levels of each CpG site in the region were all negatively correlated with the gene expression in TCGA and TARGET AML cohorts (Figures S3 and S4). Of course, the strong negative correlation between *DOCK1* expression and average methylation levels of these differentially methylated sites were observed in the above-mentioned two cohorts (**Figure 5B**, **5C**).

DOCK1 expression in different hematopoietic cell subpopulations

Given that LSCs have been found in CD34+ subpopulations [8, 20], the expression level of *DOCK1* in different fractions purified by cell surface markers was investigated. AML CD34+ cells had a significantly high *DOCK1* expression level relative to CD34- cells in the GSE30029 dataset (P=1.0×10^{-7}; **Figure 6A**), in which there were expression profiles of 44 paired CD34- and CD34+ fractions from AML patient's mononuclear cells [18]. Besides, we compared *DOCK1* expression level in normal human hematopoietic cell subpopulations. Significant difference in *DOCK1* expression between hematopoietic stem cells and granulocytes and monocytes from umbilical cord blood was detected in GSE24759 (P= 1.0×10^{-7}, **Figure 6B**) [21]. Similarly, *DOCK1* gene was also expressed at high level in CD34+

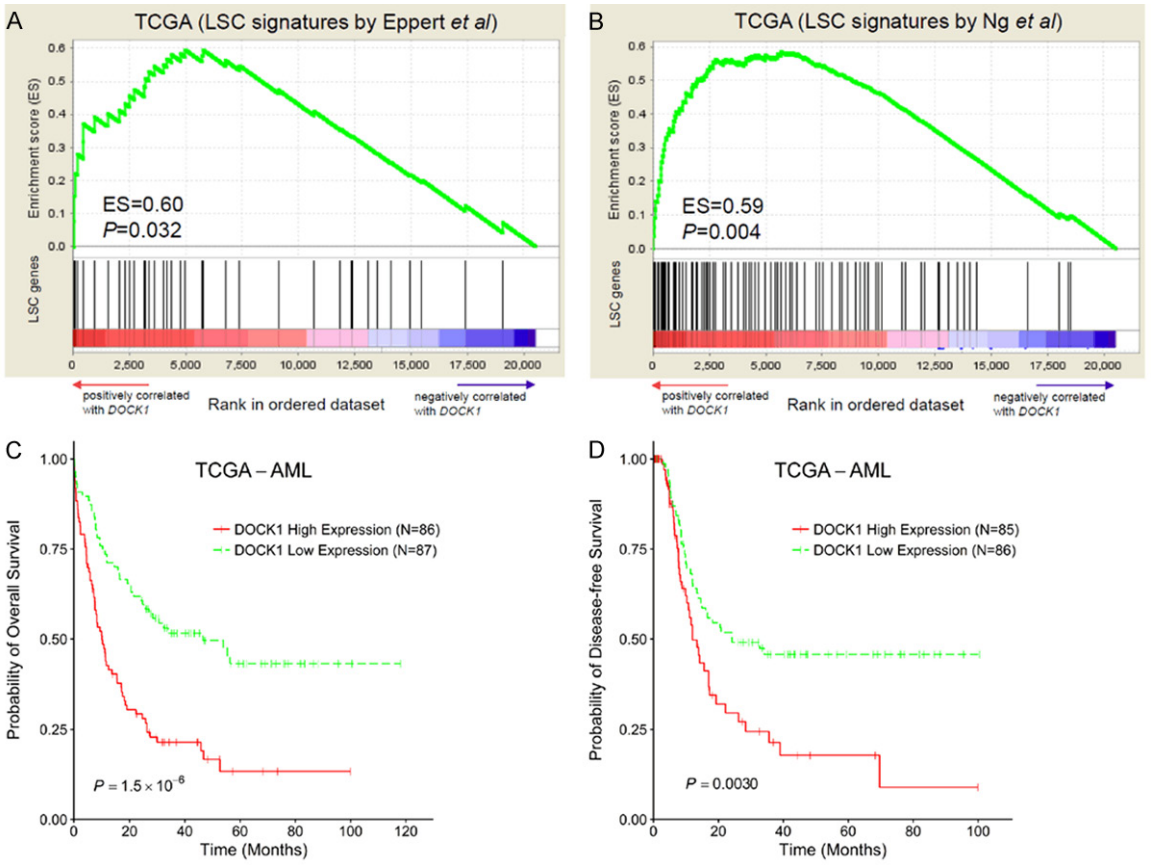

Figure 3. Extraction of AML prognostic genes in TCGA. (A, B) GSEA plot of enrichment of LSC gene signatures in AML expression profile from TCGA. Red-to-indigo bars denote the lists of genome-wide genes ranked by Pearson's correlation coefficients with *DOCK1* expression level. (C, D) The high *DOCK1* expression level is associated with short overall survival (C) and disease-free survival (D) in TCGA AML cohort.

cell fractions when compared with mononuclear cells from bone marrow of healthy donators in Beat AML programme (P=5.8×10^{-9}, **Figure 6C**) [17]. These results provide further support that *DOCK1* is implicated in stemness that is shared between LSCs and normal HSPCs.

Discussion

DOCK1 is a Rac-specific guanine nucleotide exchange factor [22]. In mammalian, the most prominent member among the Rho GTPases family is Rac, which is implicated in activation of protein kinases, cytoskeletal reorganization and control of cell growth [23]. As molecular switches, Rac GTPases regulate cellular processes through cycling between active GTP-bound and inactive GDP-bound states. DOCK1 proteins can activate Rac through catalysis of GDP-GTP exchange [22].

Our results provided direct evidences that the high expression of *DOCK1*, as a stem cell marker, could become an adverse prognostic factor in AML. First, increased expression level of *DOCK1* was found in LSCs, and the GSEA analysis indicated that *DOCK1* was involved with gene signature of LSCs. There were also significant expression differences between AML CD34+ and CD34- cell fractions, normal CD34+ and mononuclear cells, as well as hematopoietic stem cells and granulocytes/monocytes. In addition, the high expression of *DOCK1* was associated with a poor prognosis in TCGA AML cohort, which could be validated in other five independent cohorts globally. Finally, methylation levels of the *DOCK1* promoter region in LSCs were significantly lower than non-LSCs, which were all strongly negatively correlated with the gene expression.

Rac GTPases, whose activation is regulated by DOCK1 [22], play central roles in the hematopoietic system and hemopathies [24]. Rac GTPases are required for homing and migration of HSCs through integrating β1-integrins and

Table 2. Clinical characteristics of the TCGA AML cohort according to *DOCK1* expression level

Characteristic	TCGA AML cohort (n=173)	*DOCK1* high expression subset (n=86)	*DOCK1* low expression subset (n=87)	*P* value
Age at diagnosis, years, median (range)	58 (18-88)	62 (21-88)	51 (18-81)	2.1×10^{-4}†
Female sex, n (%)	81 (46.8)	40 (46.0)	41 (47.7)	0.94‡
Parameters at Diagnosis, median (range)				
WBC count, $\times10^9$/L	17.0 (0.4-297.4)	18.8 (0.6-297.0)	14.9 (0.4-224.0)	0.80§
PB Blast, %	39 (0-98)	36 (0-98)	40 (0-97)	0.56§
BM Blast, %	72 (30-100)	71 (30-100)	74 (32-100)	0.23§
Cytogenetic risk stratifications, n (%)	n=170*	n=85	n=85	
Good	32 (18.8)	8 (9.4)	24 (28.2)	3.7×10^{-4}‡
Intermediate	101 (59.4)	50 (58.8)	51 (60.0)	
Poor	37 (21.8)	27 (31.8)	10 (11.8)	
Karyotype	n=170*	n=85	n=85	
Normal Karyotype	75 (44.1)	35 (41.2)	40 (47.1)	0.54‡
Abnormal Karyotype	95 (55.9)	50 (58.8)	45 (52.9)	
FLT3 status, n (%)	n=166*	n=81	n=85	
Negative	116 (69.9)	55 (67.9)	61 (71.8)	0.71
Positive	50 (30.1)	26 (32.1)	24 (28.2)	

*Number of patients was based on the available clinical information. †*P* value calculated by Student's t-test. ‡*P* value calculated by Pearson's chi-squared test. §*P* value calculated by Wilcoxon rank-sum test.

Table 3. Multivariate survival analysis of *DOCK1* expression levels in TCGA AML cohort

Variables	TCGA AML OS (n=173)		TCGA AML DFS (n=171*)	
	HR (95% CI)	*P* Value	HR (95% CI)	*P* Value
DOCK1 expression level	1.06 (1.00-1.12)	0.03	1.09 (1.02-1.16)	0.02
Age, years	1.04 (1.02-1.05)	4.7×10^{-6}	1.00 (0.99-1.02)	0.70
WBC count, 10^9/L	1.01 (1.00-1.01)	0.002	1.01 (1.00-1.01)	0.006
Good cytogenetic risk	0.52 (0.27-1.00)	0.05	0.41 (0.21-0.82)	0.01
Poor cytogenetic risk	1.70 (1.09-2.64)	0.02	0.90 (0.48-1.67)	0.73

*Number of patients was based on the available clinical information.

tyrosine kinase c-Kit signals [25], both of which are also critical regulators of stem self-renewal [26]. Furthermore, Rac GTPases promote cell survival and proliferation in mutated c-Kit induced AML [27]. Rac signal transduction is required for maintenance and expansion of LSCs through mediating interaction with stromal cells [28]. Moreover, DOCK1 was required to bind engulfment and cell motility 1 (ELMO1) for Rac activation [29]. The high expression of *ELMO1* was found in AML CD34+ cells when compared to CD34- cells and was associated with poor prognosis of AML patients as well. The progenitors replacing capacity and long-term growth ability was impaired in AML CD34+ cells after knocking down ELMO1 [30]. More importantly, the Rac-specific inhibitor significantly impacts leukemia cell proliferation, survival and cell-cycle, and Rac GTPases are expected to become a therapeutic target [31]. Therefore, it is speculated that high expression of DOCK1 might confer or enhance stemness of leukemia cells via activation of Rac signaling pathway.

Stem cell properties often lead to high recurrence incidence and poor prognosis in cancer [9]. Lee and colleagues have shown that high DOCK1 expression was associated with poor prognosis in AML patients, and suggested that the DOCK1 expression level was implicated in stem cell functions [32]. Zhang et al. have indicated that high expression of DOCK1 adversely influenced the prognosis of AML patients

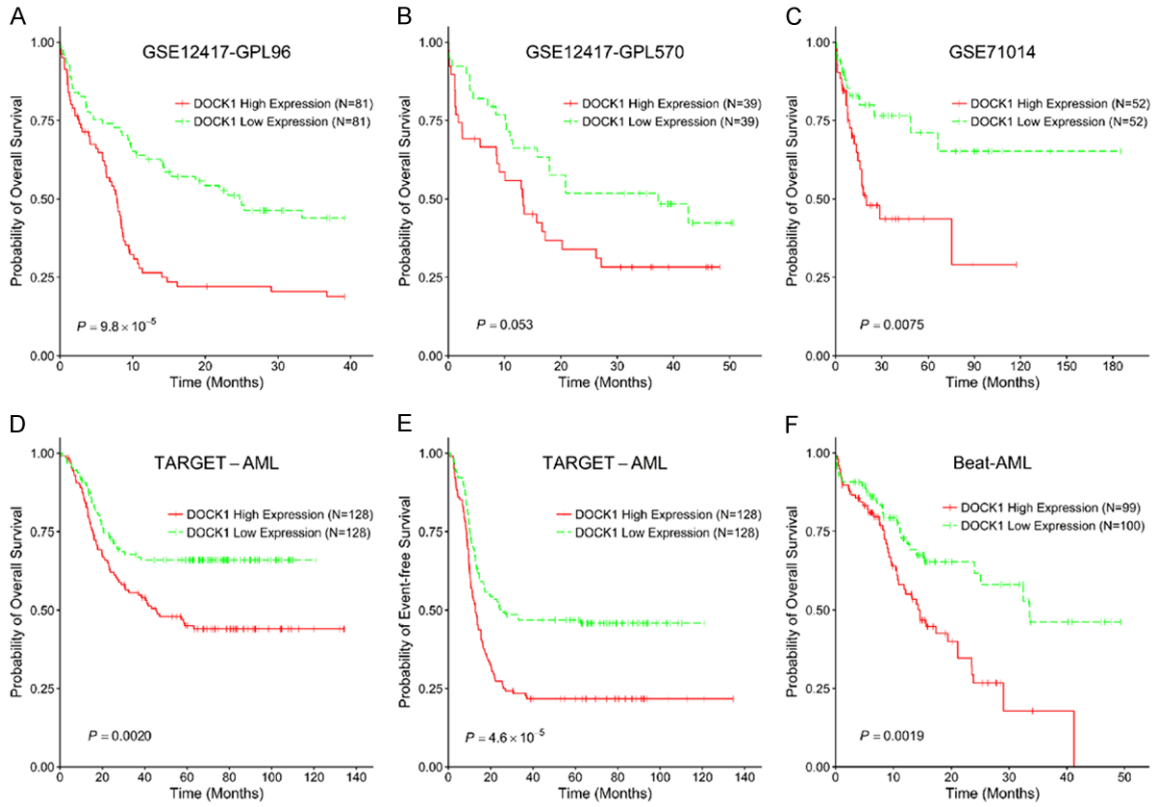

Figure 4. The association of high *DOCK1* expression level with a worse disease prognosis in multiple independent AML cohorts. (A-F) Kaplan-Meier estimates of overall survival (A-D, F) or event-free survival (E) based on the median of *DOCK1* expression level detected by microarray (A-C) or RNA-seq (B-F).

undergoing allogeneic hematopoietic stem cell transplantation [33]. Consistent with them, this paper presented that increased levels of DOCK1 was indeed enriched with gene signature of LSC in AML and confirmed that patients with high DOCK1 expression had a shorter OS, DFS and EFS in multiple AML cohorts. In fact, several lines of evidence have shown that DOCK1 is involved with malignant phenotypes in other cancers. Elevated DOCK1 expression was related to an inferior survival time in patients with HER2+ or basal breast cancer through enhancement of HER2-induced Rac activation and cell migration [34]. Circular RNA DOCK1 could promote bladder carcinoma progression via modulating circDOCK1/hsa-miR-132-3p/Sox5 signaling pathway [35]. In addition, the association between high *DOCK1* expression and poor prognosis of ovarian cancer was observed [36]. It was also reported that upregulation of DOCK1 contributed to cancer cell migration and invasion in human colorectal cancer [37]. Moreover, DOCK1 could promote invasive migration and metastasis of glioma cells by regulation of Rac1 activation [38, 39]. It is worth noting that cancer stem cells all have been identified in these tumors, suggesting that the impact of DOCK1 on the cancers might be involved with stem cell properties. Based on the present findings, it is reasonable to develop a series of innovative therapies aiming at eliminating the high-DOCK1 content for AML patients. For example, a recent study has reported that DOCK2 enhanced the survival of leukemia cells through interaction with FLT3 protein [40]. Thus, the DOCK1 selective inhibitor that has been newly developed to suppress growth and metastasis of cancer cells might be a promising candidate drug for AML treatment [41].

Although the potential mechanisms underlying this distinct LSC-related high DOCK1 expression pattern in AML are still unclear, it is generally agreed that further investigation of its molecular and epigenetic characterization can

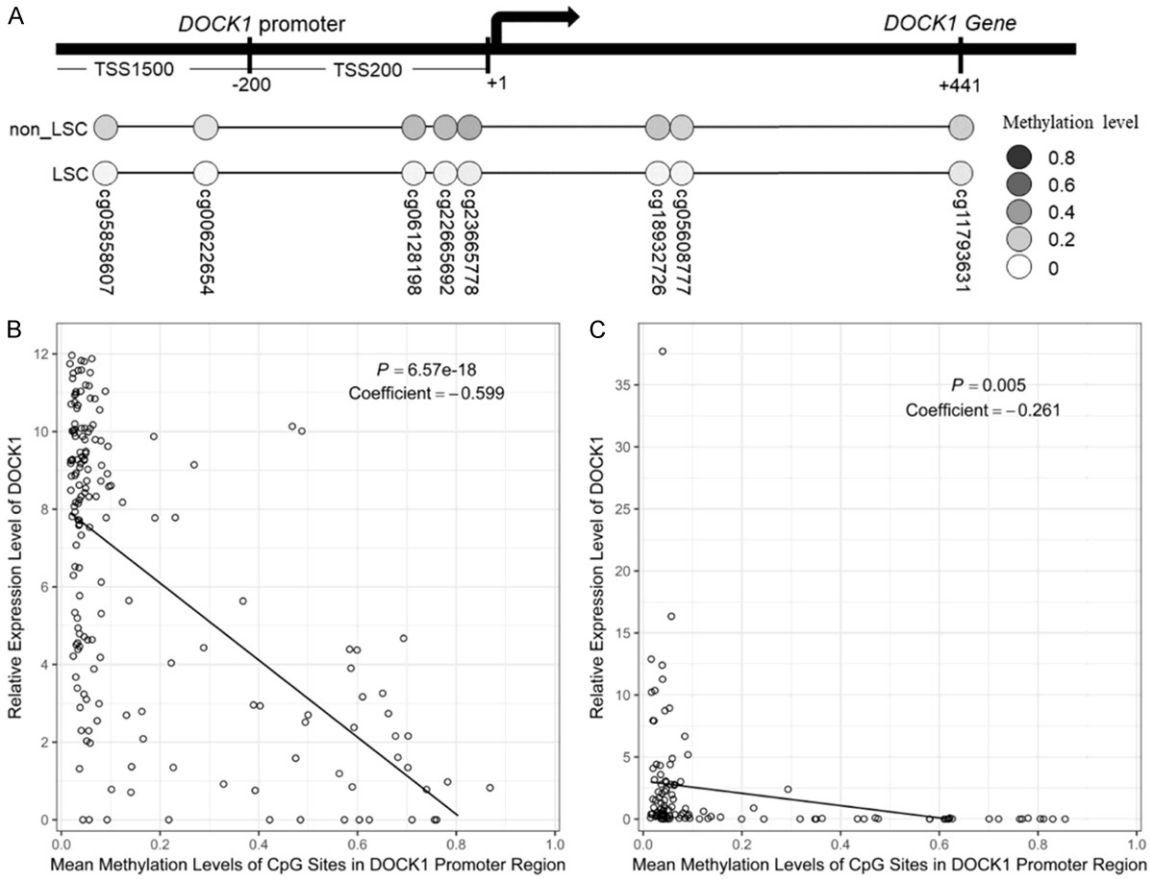

Figure 5. Analysis of *DOCK1* methylation pattern in AML cell fractions. (A) Schematic map of differentially methylated sites within the CpG islands in the *DOCK1* promoter between LSC (n=20) and non-LSC (n=24) subpopulations. (B, C) DNA methylation levels of *DOCK1* promoter are negatively correlated with the gene expression in TCGA (B, n=170) and TARGET (C, n=113) AML projects.

help in understanding the pathology and drug development of AML. Current progress in epigenetic research for AML has shown that epigenetic modification plays important roles in normal hematopoiesis, and aberrant DNA methylation patterns contribute to etiology and development of AML [42]. It was reported the DNA methylation status of specific genes is associated with clinical outcomes of AML [43]. Lee and colleagues have shown that higher DOCK1 expression was closely associated with *DNMT3A* mutations (encoding DNA methyltransferase 3A), suggesting that methylation status may affect the activated expression of *DOCK1* gene. In this article, we found that methylation level is the main mechanism of regulation of *DOCK1* gene expression in *de novo* AML patients, and the promoter region was barely methylated in LSCs relative to non-LSCs. Nowadays, epigenetic treatments that target aberrant DNA methylation, such as using azacitidine and decitabine, have been approved by the FDA. Consequently, not merely the abnormal methylation of *DOCK1* gene could be given extra understanding to the characteristic of LSCs, but such finding would also be used to develop novel treatment strategies for AML. Finally, there is no doubt that these findings are worthy of attention as well as further experimental examination in the following years.

Conclusions

In summary, the high expression of *DOCK1*, which was found in LSCs, could predict inferior prognosis of AML. Besides, we also found that methylation was involved in *DOCK1* transcriptional regulation. DOCK1 protein activates Rac GTPases, an essential regulator of HSC and LSC, which might be the function mechanism of DOCK1 on stem cells. Taken together, DOCK1

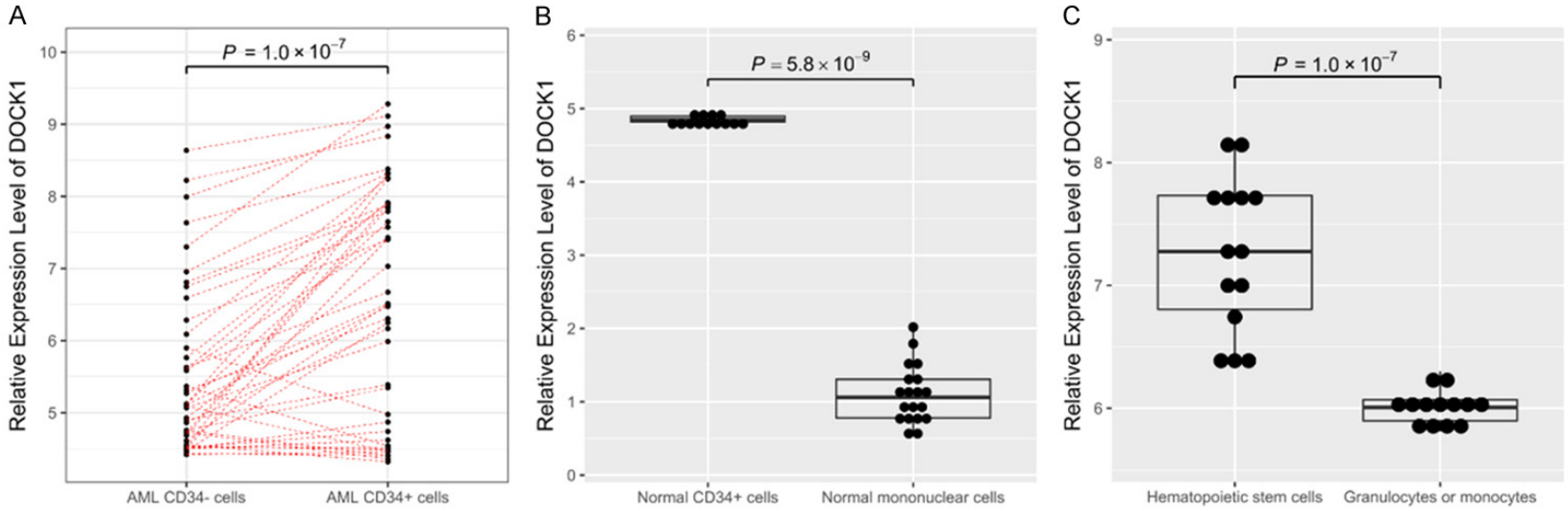

Figure 6. Comparison of *DOCK1* expression level between hematopoietic cell subpopulations. A. *DOCK1* expression of CD34+ subfractions is high than paired CD34- from AML mononuclear cells (n=44) in GSE30029 dataset. B. High *DOCK1* gene expression in hematopoietic stem cells (n=14) relative to granulocytes and monocytes (n=13) from human umbilical cord blood is found in GSE24759 dataset. C. There is a significant difference in *DOCK1* levels between normal CD34+ cells (n=13) and bone marrow mononuclear cells (n=19) from healthy donators in Beat AML programme.

plays an important role in AML and would be a potential therapeutic target for AML.

Acknowledgements

The authors thank all of the researchers and study participants for their contributions.

Disclosure of conflict of interest

None.

Address correspondence to: Daiqiang Li, Department of Pathology, Xiangya Second Hospital, Central South University, No. 139 Middle Renmin Road, Changsha 410011, Hunan, P. R. China. E-mail: lidaiqiang@csu.edu.cn

References

[1] Dohner H, Weisdorf DJ and Bloomfield CD. Acute myeloid leukemia. N Engl J Med 2015; 373: 1136-1152.

[2] Sant M, Minicozzi P, Mounier M, Anderson LA, Brenner H, Holleczek B, Marcos-Gragera R, Maynadié M, Monnereau A, Osca-Gelis G, Visser O and De Angelis R. Survival for haematological malignancies in Europe between 1997 and 2008 by region and age: results of EUROCARE-5, a population-based study. Lancet Oncol 2014; 15: 931-42.

[3] 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, Dohner K, Schlenk RF, Dohner H and Campbell PJ. Genomic classification and prognosis in acute myeloid leukemia. N Engl J Med 2016; 374: 2209-2221.

[4] Schwind S, Marcucci G, Maharry K, Radmacher MD, Mrozek K, Holland KB, Margeson D, Becker H, Whitman SP, Wu YZ, Metzeler KH, Powell BL, Kolitz JE, Carter TH, Moore JO, Baer MR, Carroll AJ, Caligiuri MA, Larson RA and Bloomfield CD. BAALC and ERG expression levels are associated with outcome and distinct gene and microRNA expression profiles in older patients with de novo cytogenetically normal acute myeloid leukemia: a cancer and leukemia group B study. Blood 2010; 116: 5660-5669.

[5] Zhang DY, Yan H, Cao S, Zhang W, Li XL, Zeng H and Chen XP. Wilms Tumor 1 rs16754 predicts favorable clinical outcomes for acute myeloid leukemia patients in South Chinese population. Leuk Res 2015; 39: 568-574.

[6] Niederwieser C, Kohlschmidt J, Volinia S, Whitman SP, Metzeler KH, Eisfeld AK, Maharry K, Yan P, Frankhouser D, Becker H, Schwind S, Carroll AJ, Nicolet D, Mendler JH, Curfman JP, Wu YZ, Baer MR, Powell BL, Kolitz JE, Moore JO, Carter TH, Bundschuh R, Larson RA, Stone RM, Mrozek K, Marcucci G and Bloomfield CD. Prognostic and biologic significance of DNMT3B expression in older patients with cytogenetically normal primary acute myeloid leukemia. Leukemia 2015; 29: 567-575.

[7] Wellbrock J, Latuske E, Kohler J, Wagner K, Stamm H, Vettorazzi E, Vohwinkel G, Klokow M, Uibeleisen R, Ehm P, Riecken K, Loges S, Thol F, Schubert C, Amling M, Jucker M, Bokemeyer C, Heuser M, Krauter J and Fiedler W. Expression of hedgehog pathway mediator gli represents a negative prognostic marker in human acute myeloid leukemia and its inhibition exerts antileukemic effects. Clin Cancer Res 2015; 21: 2388-98.

[8] Bonnet D and Dick JE. Human acute myeloid leukemia is organized as a hierarchy that originates from a primitive hematopoietic cell. Nat Med 1997; 3: 730-737.

[9] Dean M, Fojo T and Bates S. Tumour stem cells and drug resistance. Nat Rev Cancer 2005; 5: 275-284.

[10] Yang C, Jin K, Tong Y and Cho WC. Therapeutic potential of cancer stem cells. Med Oncol 2015; 32: 619.

[11] Eppert K, Takenaka K, Lechman ER, Waldron L, Nilsson B, van Galen P, Metzeler KH, Poeppl A, Ling V, Beyene J, Canty AJ, Danska JS, Bohlander SK, Buske C, Minden MD, Golub TR, Jurisica I, Ebert BL and Dick JE. Stem cell gene expression programs influence clinical outcome in human leukemia. Nat Med 2011; 17: 1086-1093.

[12] Ng SW, Mitchell A, Kennedy JA, Chen WC, McLeod J, Ibrahimova N, Arruda A, Popescu A, Gupta V, Schimmer AD, Schuh AC, Yee KW, Bullinger L, Herold T, Gorlich D, Buchner T, Hiddemann W, Berdel WE, Wormann B, Cheok M, Preudhomme C, Dombret H, Metzeler K, Buske C, Lowenberg B, Valk PJ, Zandstra PW, Minden MD, Dick JE and Wang JC. A 17-gene stemness score for rapid determination of risk in acute leukaemia. Nature 2016; 540: 433-437.

[13] Jung N, Dai B, Gentles AJ, Majeti R and Feinberg AP. An LSC epigenetic signature is largely mutation independent and implicates the HOXA cluster in AML pathogenesis. Nat Commun 2015; 6: 8489.

[14] Metzeler KH, Hummel M, Bloomfield CD, Spiekermann K, Braess J, Sauerland MC, Heinecke A, Radmacher M, Marcucci G, Whitman SP, Maharry K, Paschka P, Larson RA, Berdel WE, Buchner T, Wormann B, Mansmann U, Hid-

demann W and Bohlander SK; Christian Buske for Cancer and Leukemia Group B and the German AML Cooperative Group. An 86-probeset gene-expression signature predicts survival in cytogenetically normal acute myeloid leukemia. Blood 2008; 112: 4193-4201.

[15] Chuang MK, Chiu YC, Chou WC, Hou HA, Tseng MH, Kuo YY, Chen Y, Chuang EY and Tien HF. An mRNA expression signature for prognostication in de novo acute myeloid leukemia patients with normal karyotype. Oncotarget 2015; 6: 39098-39110.

[16] 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.

[17] Tyner JW, Tognon CE, Bottomly D, Wilmot B, Kurtz SE, Savage SL, Long N, Schultz AR, Traer E, Abel M, Agarwal A, Blucher A, Borate U, Bryant J, Burke R, Carlos A, Carpenter R, Carroll J, Chang BH, Coblentz C, d'Almeida A, Cook R, Danilov A, Dao KT, Degnin M, Devine D, Dibb J, Edwards DKt, Eide CA, English I, Glover J, Henson R, Ho H, Jemal A, Johnson K, Johnson R, Junio B, Kaempf A, Leonard J, Lin C, Liu SQ, Lo P, Loriaux MM, Luty S, Macey T, MacManiman J, Martinez J, Mori M, Nelson D, Nichols C, Peters J, Ramsdill J, Rofelty A, Schuff R, Searles R, Segerdell E, Smith RL, Spurgeon SE, Sweeney T, Thapa A, Visser C, Wagner J, Watanabe-Smith K, Werth K, Wolf J, White L, Yates A, Zhang H, Cogle CR, Collins RH, Connolly DC, Deininger MW, Drusbosky L, Hourigan CS, Jordan CT, Kropf P, Lin TL, Martinez ME, Medeiros BC, Pallapati RR, Pollyea DA, Swords RT, Watts JM, Weir SJ, Wiest DL, Winters RM, McWeeney SK and Druker BJ. Functional genomic landscape of acute myeloid leukaemia. Nature 2018; 562: 526-531.

[18] de Jonge HJ, Woolthuis CM, Vos AZ, Mulder A, van den Berg E, Kluin PM, van der Weide K, de Bont ES, Huls G, Vellenga E and Schuringa JJ. Gene expression profiling in the leukemic stem cell-enriched CD34(+) fraction identifies target genes that predict prognosis in normal karyotype AML. Leukemia 2011; 25: 1825-1833.

[19] Shi W, Oshlack A and Smyth GK. Optimizing the noise versus bias trade-off for Illumina whole genome expression beadchips. Nucleic Acids Res 2010; 38: e204.

[20] Lapidot T, Sirard C, Vormoor J, Murdoch B, Hoang T, Caceres-Cortes J, Minden M, Paterson B, Caligiuri MA and Dick JE. A cell initiating human acute myeloid leukaemia after transplantation into SCID mice. Nature 1994; 367: 645-648.

[21] Novershtern N, Subramanian A, Lawton LN, Mak RH, Haining WN, McConkey ME, Habib N, Yosef N, Chang CY, Shay T, Frampton GM, Drake AC, Leskov I, Nilsson B, Preffer F, Dombkowski D, Evans JW, Liefeld T, Smutko JS, Chen J, Friedman N, Young RA, Golub TR, Regev A and Ebert BL. Densely interconnected transcriptional circuits control cell states in human hematopoiesis. Cell 2011; 144: 296-309.

[22] Kiyokawa E, Hashimoto Y, Kobayashi S, Sugimura H, Kurata T and Matsuda M. Activation of Rac1 by a Crk SH3-binding protein, DOCK180. Genes Dev 1998; 12: 3331-3336.

[23] Hall A. Rho family GTPases. Biochem Soc Trans 2012; 40: 1378-1382.

[24] Durand-Onayli V, Haslauer T, Harzschel A and Hartmann TN. Rac GTPases in hematological malignancies. Int J Mol Sci 2018; 19: 4041.

[25] Gu Y, Filippi MD, Cancelas JA, Siefring JE, Williams EP, Jasti AC, Harris CE, Lee AW, Prabhakar R, Atkinson SJ, Kwiatkowski DJ and Williams DA. Hematopoietic cell regulation by Rac1 and Rac2 guanosine triphosphatases. Science 2003; 302: 445-449.

[26] Olabi S, Ucar A, Brennan K and Streuli CH. Integrin-Rac signalling for mammary epithelial stem cell self-renewal. Breast Cancer Res 2018; 20: 128.

[27] Martin H, Mali RS, Ma P, Chatterjee A, Ramdas B, Sims E, Munugalavadla V, Ghosh J, Mattingly RR, Visconte V, Tiu RV, Vlaar CP, Dharmawardhane S and Kapur R. Pak and Rac GTPases promote oncogenic KIT-induced neoplasms. J Clin Invest 2013; 123: 4449-4463.

[28] Rozenveld-Geugien M, Baas IO, van Gosliga D, Vellenga E and Schuringa JJ. Expansion of normal and leukemic human hematopoietic stem/progenitor cells requires rac-mediated interaction with stromal cells. Exp Hematol 2007; 35: 782-792.

[29] Brugnera E, Haney L, Grimsley C, Lu M, Walk SF, Tosello-Trampont AC, Macara IG, Madhani H, Fink GR and Ravichandran KS. Unconventional Rac-GEF activity is mediated through the Dock180-ELMO complex. Nat Cell Biol 2002; 4: 574-582.

[30] Capala ME, Vellenga E and Schuringa JJ. ELMO1 is upregulated in AML CD34+ stem/progenitor cells, mediates chemotaxis and predicts poor prognosis in normal karyotype AML. PLoS One 2014; 9: e111568.

[31] Muller LU, Schore RJ, Zheng Y, Thomas EK, Kim MO, Cancelas JA, Gu Y and Williams DA. Rac guanosine triphosphatases represent a

potential target in AML. Leukemia 2008; 22: 1803-1806.
[32] Lee SH, Chiu YC, Li YH, Lin CC, Hou HA, Chou WC and Tien HF. High expression of dedicator of cytokinesis 1 (DOCK1) confers poor prognosis in acute myeloid leukemia. Oncotarget 2017; 8: 72250-72259.
[33] Zhang G, Zhang J, Yang X, Zhang X, Yang S, Wang J, Hu K, Shi J, Ke X and Fu L. High expression of dedicator of cytokinesis 1 adversely influences the prognosis of acute myeloid leukemia patients undergoing allogeneic hematopoietic stem cell transplantation. Cancer Manag Res 2019; 11: 3053-3060.
[34] Laurin M, Huber J, Pelletier A, Houalla T, Park M, Fukui Y, Haibe-Kains B, Muller WJ and Cote JF. Rac-specific guanine nucleotide exchange factor DOCK1 is a critical regulator of HER2-mediated breast cancer metastasis. Proc Natl Acad Sci U S A 2013; 110: 7434-7439.
[35] Liu P, Li X, Guo X, Chen J, Li C, Chen M, Liu L, Zhang X and Zu X. Circular RNA DOCK1 promotes bladder carcinoma progression via modulating circDOCK1/hsa-miR-132-3p/Sox5 signalling pathway. Cell Prolif 2019; 52: e12614.
[36] Zhao F, Siu MK, Jiang L, Tam KF, Ngan HY, Le XF, Wong OG, Wong ES, Chan HY and Cheung AN. Overexpression of dedicator of cytokinesis I (Dock180) in ovarian cancer correlated with aggressive phenotype and poor patient survival. Histopathology 2011; 59: 1163-1172.
[37] Jing X, Wu H, Ji X, Wu H, Shi M and Zhao R. Cortactin promotes cell migration and invasion through upregulation of the dedicator of cytokinesis 1 expression in human colorectal cancer. Oncol Rep 2016; 36: 1946-1952.
[38] Jarzynka MJ, Hu B, Hui KM, Bar-Joseph I, Gu W, Hirose T, Haney LB, Ravichandran KS, Nishikawa R and Cheng SY. ELMO1 and Dock180, a bipartite Rac1 guanine nucleotide exchange factor, promote human glioma cell invasion. Cancer Res 2007; 67: 7203-7211.
[39] Feng H, Hu B, Jarzynka MJ, Li Y, Keezer S, Johns TG, Tang CK, Hamilton RL, Vuori K, Nishikawa R, Sarkaria JN, Fenton T, Cheng T, Furnari FB, Cavenee WK and Cheng SY. Phosphorylation of dedicator of cytokinesis 1 (Dock180) at tyrosine residue Y722 by Src family kinases mediates EGFRvIII-driven glioblastoma tumorigenesis. Proc Natl Acad Sci U S A 2012; 109: 3018-3023.
[40] Wu M, Hamaker M, Li L, Small D and Duffield AS. DOCK2 interacts with FLT3 and modulates the survival of FLT3-expressing leukemia cells. Leukemia 2017; 31: 688-696.
[41] Tajiri H, Uruno T, Shirai T, Takaya D, Matsunaga S, Setoyama D, Watanabe M, Kukimoto-Niino M, Oisaki K, Ushijima M, Sanematsu F, Honma T, Terada T, Oki E, Shirasawa S, Maehara Y, Kang D, Cote JF, Yokoyama S, Kanai M and Fukui Y. Targeting ras-driven cancer cell survival and invasion through selective inhibition of DOCK1. Cell Rep 2017; 19: 969-980.
[42] Guillamot M, Cimmino L and Aifantis I. The Impact of DNA methylation in hematopoietic malignancies. Trends Cancer 2016; 2: 70-83.
[43] Bullinger L, Ehrich M, Dohner K, Schlenk RF, Dohner H, Nelson MR and van den Boom D. Quantitative DNA methylation predicts survival in adult acute myeloid leukemia. Blood 2010; 115: 636-642.

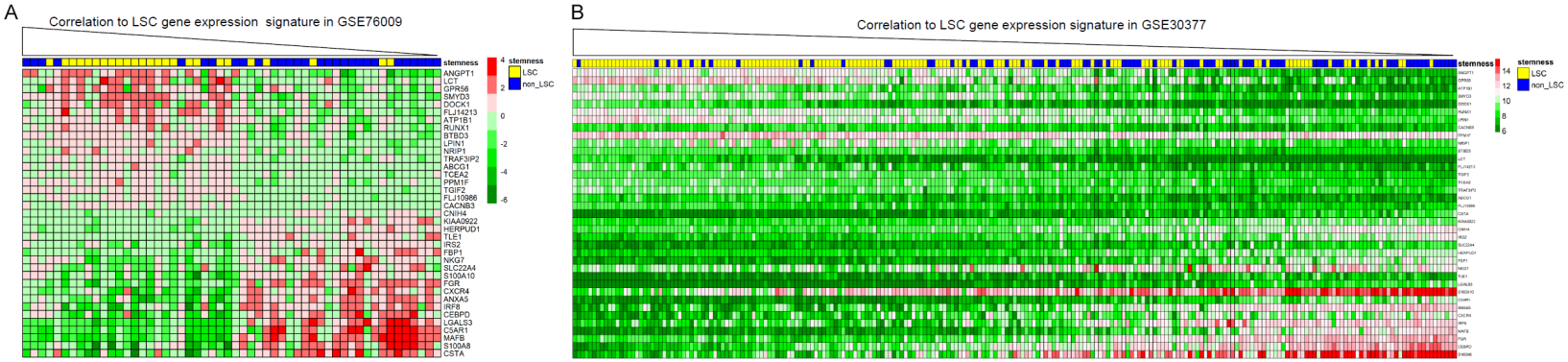

Figure S1. Heat map of DE LSC genes expression (rows) in the GSE30377 (A) and GSE76008 (B). Cases (columns) are ranked by correlation coefficients between expression profile of DE genes and reference profile of LSC generated by GSE76008 and GSE30377 datasets, respectively.

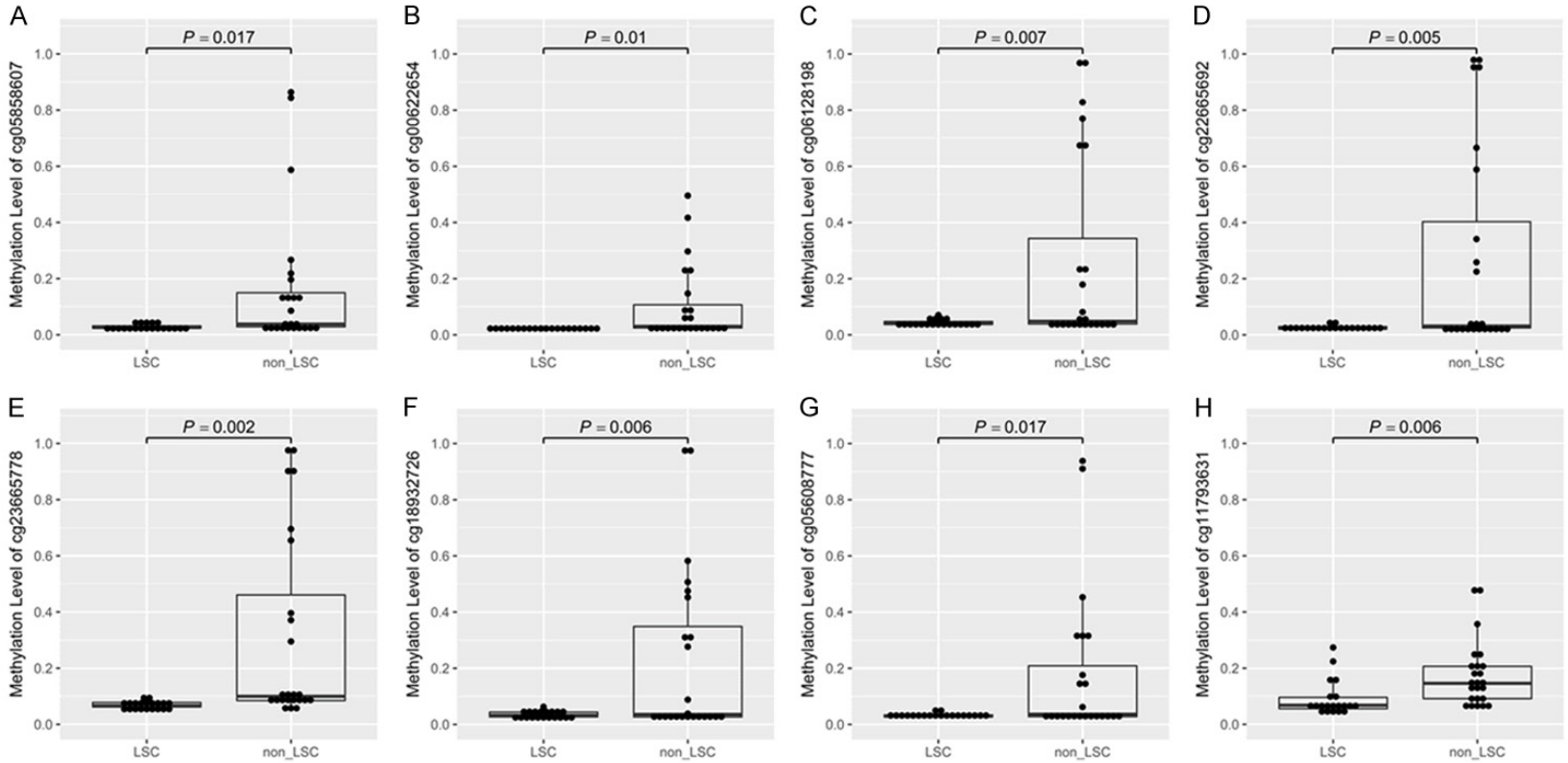

Figure S2. The differential methylation levels of each CpG site in the *DOCK1* promoter between LSC+ (n=20) and LSC- (n=24) fractions.

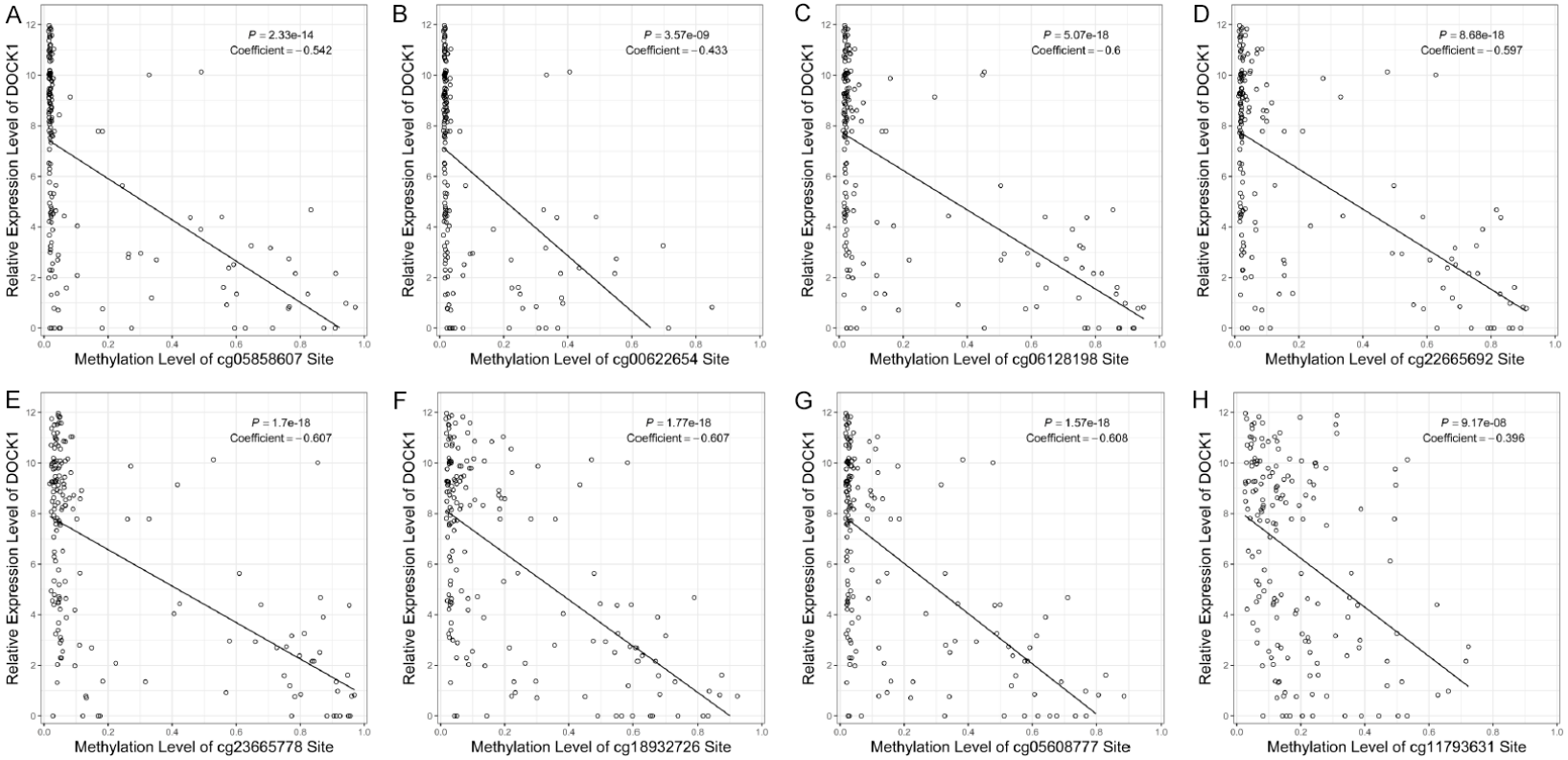

Figure S3. DNA methylation levels of each CpG site in *DOCK1* promoter are negatively correlated with the gene expression level in TCGA AML dataset (n=170).

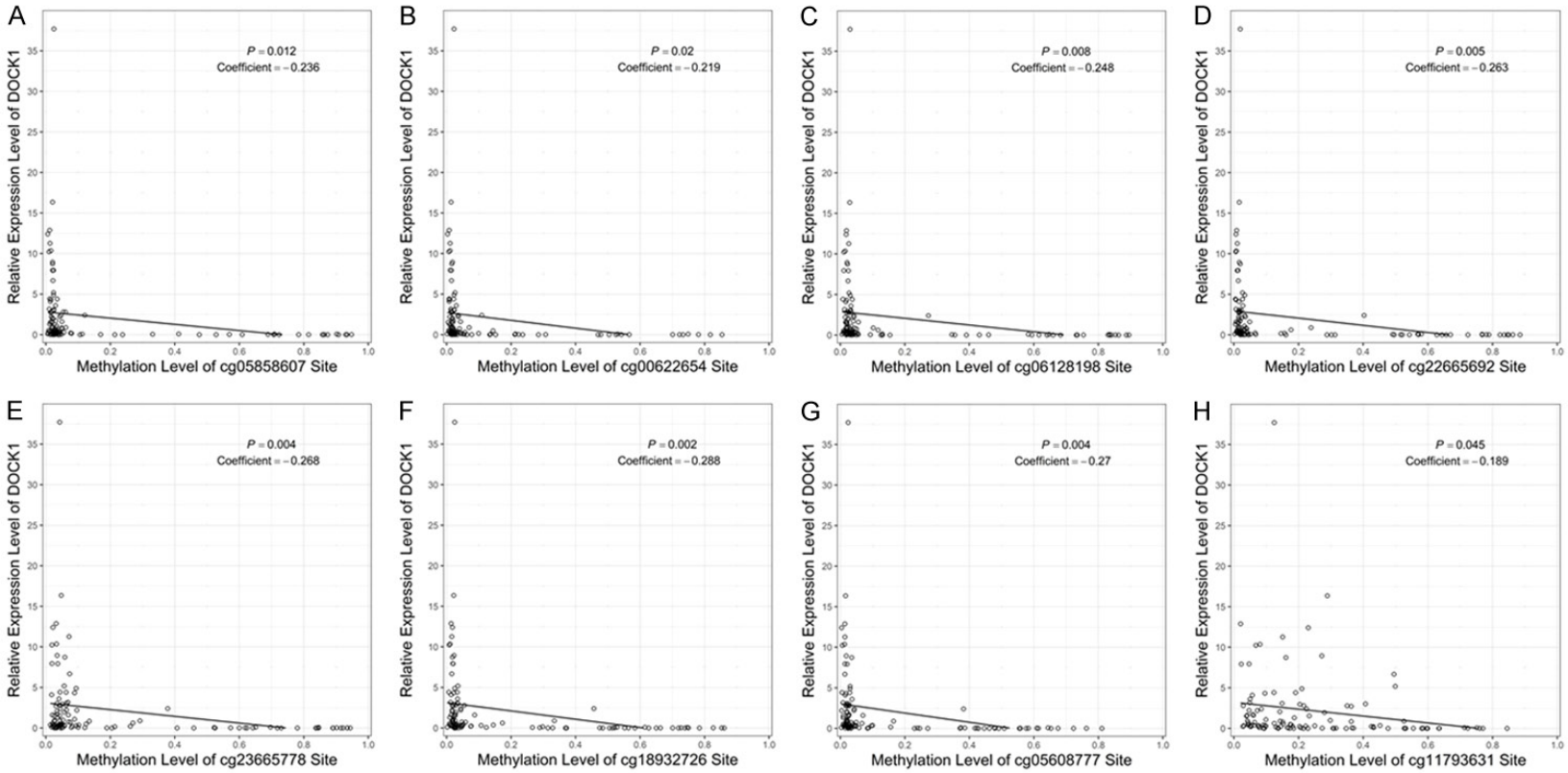

Figure S4. Methylation levels of each CpG site in *DOCK1* promoter are negatively correlated with the gene expression level in TARGET AML projects (n=113).