Identification of hub genes and molecular mechanisms in infant acute lymphoblastic leukemia with *MLL* gene rearrangement

Hao Zhang [Equal first author, 1], **Juan Cheng** [Equal first author, 1], **Zijian Li** [1], **Yaming Xi** [Corresp. 1]

[1] Department of hematology, The First Hospital of Lanzhou University, Lanzhou, Gansu, China

Corresponding Author: Yaming Xi
Email address: ldyyxyk@gmail.com

Infant acute lymphoblastic leukemia (ALL) with the mixed lineage leukemia (*MLL*) gene rearrangement (*MLL*-R) is considered a distinct leukemia from childhood or non-*MLL*-R infant ALL. To elucidate the molecular mechanisms of *MLL*-R infant ALL. Microarray expression data were downloaded from the Gene Expression Omnibus (GEO) database, and differentially expressed genes (DEGs) between *MLL*-R and non-*MLL*-R infant ALL were identified. Gene ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analyses were carried out. Then, we constructed a protein-protein interaction (PPI) network and identified the hub genes. Finally, drug-gene interactions were mined. 139 cases of *MLL*-R infant ALL including 77 (55.4%) fusions with *AF4*, 38 (27.3%) with *ENL*, 14 (10.1%) with *AF9*, and 10 (7.2%) other gene fusions were characterized. 236 up-regulated and 84 down-regulated DEGs were identified. The up-regulated DEGs were enriched in 62 GO biological process terms, 21 in GO cellular components terms, and 22 in GO molecular functions terms, and down-regulated DEGs were enriched in 34, 22, and 17 terms, respectively. The up-regulated DEGs were enriched in 7 KEGG pathways, mainly involving transcriptional regulation and signaling pathways, and down-regulated DEGs were involved in 3 main KEGG pathways. The PPI network included 297 nodes and 410 edges, with *MYC*, *ALB*, *CD44*, *PTPRC* and *TNF* identified as hub genes. 23 drug-gene interactions including 4 up-regulated hub genes and 24 drugs were identified by DGIdb. In conclusion, *MYC, ALB, CD44, PTPRC* and *TNF* may be potential bio-markers for the diagnosis and therapy of *MLL*-R infant ALL.

Identification of hub genes and molecular mechanisms in infant acute lymphoblastic leukemia with *MLL* gene rearrangement

Hao Zhang*, Juan Cheng*, Zijian Li, YaMing Xi

Department of hematology, The First Hospital of Lanzhou University, Lanzhou, Gansu, China

*Hao Zhang and Juan Cheng contributed equally to this work

Corresponding Author:

YaMing Xi

No.1, Donggangxi Rd, Chengguan District, Lanzhou, Gansu, 730000, China

Email address: ldyyxyk@gmail.com

Abstract

Infant acute lymphoblastic leukemia (ALL) with the mixed lineage leukemia (MLL) gene rearrangement (MLL-R) is considered a distinct leukemia from childhood or non-MLL-R infant ALL. To elucidate the molecular mechanisms of MLL-R infant ALL. Microarray expression data were downloaded from the Gene Expression Omnibus (GEO) database, and differentially expressed genes (DEGs) between MLL-R and non-MLL-R infant ALL were identified. Gene ontology (GO) and Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway enrichment analyses were carried out. Then, we constructed a protein-protein interaction (PPI) network and identified the hub genes. Finally, drug-gene interactions were mined. 139 cases of MLL-R infant ALL including 77 (55.4%) fusions with AF4, 38 (27.3%) with ENL, 14 (10.1%) with AF9, and 10 (7.2%) other gene fusions were characterized. 236 up-regulated and 84 down-regulated DEGs were identified. The up-regulated DEGs were enriched in 62 GO biological process terms, 21 in GO cellular components terms, and 22 in GO molecular functions terms, and down-regulated DEGs were enriched in 34, 22, and 17 terms, respectively. The up-regulated DEGs were enriched in 7 KEGG pathways, mainly involving transcriptional regulation and signaling

pathways, and down-regulated DEGs were involved in 3 main KEGG pathways. The PPI network included 297 nodes and 410 edges, with MYC, ALB, CD44, PTPRC and TNF identified as hub genes. 23 drug-gene interactions including 4 up-regulated hub genes and 24 drugs were identified by DGIdb. In conclusion, MYC, ALB, CD44, PTPRC and TNF may be potential biomarkers for the diagnosis and therapy of MLL-R infant ALL.

Keywords acute lymphoblastic leukemia, infant, mixed-lineage leukemia, gene expression profiles, differentially expressed genes

Introduction

Infant acute lymphoblastic leukemia (ALL) refers to ALL arising in infants prior to 12 months of age. Infant ALL is less common but more aggressive than pediatric ALL, generally with a poorer outcome (Brown, Pieters, & Biondi, 2019; Pieters et al., 2007). Despite advances in the treatment of pediatric ALL, >50% of patients with infant ALL relapse within five years of diagnosis, and the four-year event-free survival is <50% (Hilden et al., 2006; Nagayama et al., 2006). Approximately 80% of infant ALL cases are characterized genetically by rearrangements in the mixed lineage leukemia gene (*MLL*, also known as *KMT2A*, located on chromosome 11q23) - (Krivtsov & Armstrong, 2007). These rearrangements occur in nearly 100% of infants with congenital leukemia and approximately 5% of pediatric ALL patients, with predicted inferior outcomes (van der Linden et al., 2009). Previous studies have shown that infant ALL with MLL rearrangements (*MLL*-R) were clinically distinct, and were characterized by high white blood cell counts, hepatosplenomegaly, and central nervous system and skin involvement (Hilden et al., 2006). The 4-5-year event-free survival for *MLL*-R infant ALL patients was 29.1%-43.2%, compared with 56.9%-95.5% for patients without MLL rearrangements (non-*MLL*-R) (Guest & Stam, 2017). *MLL*-R infant ALL thus requires novel treatment strategies compared with non-*MLL*-R infant ALL and pediatric ALL, to improve outcomes in these patients.

More than 90 different *MLL* partner genes have been identified to date. Frequent *MLL* rearrangements in infant ALL include fusions with *AF4* (49%), *ENL* (22%), *AF9* (16%), and *AF10* (6%) (Meyer et al., 2018). All types of rearrangements in *MLL* were independently associated with an unfavorable prognosis (Pieters et al., 2007). Several studies found that *MLL* rearrangements occurred in utero, resulting in rapid progression to leukemia (Brown et al., 2019; Ford et al., 1993; Gale et al., 1997). Notably, this phenomenon showed a high concordance rate between identical twins (Greaves, Maia, Wiemels, & Ford, 2003; Guest & Stam, 2017). These

findings revealed that *MLL* rearrangements may initiate leukemogenesis for *MLL*-R infant ALL. However, *MLL-AF4* expression alone was not sufficient to induce leukemia in human embryonic stem cell-derived hematopoietic cells, and additional genetic candidates were required (Stam, 2013). These results suggested that the mechanisms responsible for *MLL*-R infant ALL are distinct from those acting during leukemogenesis in childhood and non-*MLL*-R infant ALL.

We investigated the molecular mechanisms of *MLL*-R infant ALL by determining differentially expressed genes (DEGs) between *MLL*-R and non-*MLL*-R infant ALL, using available microarray datasets, followed by bio-functional enrichment of identified DEGs. These results will provide new insights into the molecular mechanisms behind ALL in infants with *MLL* rearrangements, and help identify new diagnostic bio-markers and candidate therapeutic targets.

Materials & Methods

Microarray data collection

Microarray expression data in the Gene Expression Omnibus (GEO) database (http://www.ncbi.nlm.nih.gov/geo) (Barrett et al., 2013), ArrayExpress database (https://www.ebi.ac.uk/arrayexpress/) and The Cancer Genome Atlas (TCGA) database (https://cancergenome.nih.gov/), were searched using the keywords “acute lymphoblastic leukemia”, and data containing expression profiles of *MLL*-R compared with non-*MLL*-R infant ALL cases were selected manually. Raw CEL files were downloaded for further analysis.

Identification of DEGs

Gene expression profile data were preprocessed using the Affy package in R (version 1.58.0) (Gautier, Cope, Bolstad, & Irizarry, 2004), including background adjustment, normalization, and summarization. According to annotation files, the mean value was computed for several probes matched to a specific gene, and used as the expression value of that gene. DEGs between *MLL*-R and non-*MLL*-R infant ALL cases were identified using the Limma package in R (version 3.36.5) (Ritchie et al., 2015). |log2 fold change (FC)| >1 (log2FC >1 defined as upregulated genes, log2FC <−1 defined as downregulated genes) and a P value < 0.01 were considered as threshold points.

Gene functional enrichment analysis

Gene ontology (GO) functional annotation analyses including biological processes (BP), cellular components (CC), and molecular function (MF) terms. A Kyoto Encyclopedia of Genes and Genomes (KEGG) pathway analysis was performed using the Database for Annotation, Visualization and Integrated Discovery (DAVID) v6.8 (Huang da, Sherman, & Lempicki, 2009), with a default cut-off criterion of count=2 and P value=0.1.

Protein - protein interaction network construction and analysis

A protein - protein interaction (PPI) network of DEGs was constructed using the STRING (version 10.5, http://www.string-db.org/) database (von Mering et al., 2003), with minimum required interaction score >0.4 (median confidence). The PPI network was visualized using Cytoscape (version 3.7.1, http://www.cytoscape.org/) (Shannon et al., 2003). Bio-functional modules in the PPI network were explored using a plug-in MCODE (version 1.4.2, http://apps.cytoscape.org/apps/MCODE) in Cytoscape with Node Score Cutoff of 0.2 and K-Core of 2. Hub genes were screened using the plug-in CytoHubba (version 2.1.6, http://apps.cytoscape.org/apps/cytohubba) in Cytoscape with methods including maximal clique centrality, degree, and betweenness.

Drug-gene interactions analyses

Drug-gene interactions were mined from the Drug-Gene Interaction database (DGIdb, v3.0.2, http://www.dgidb.org/) (Cotto et al., 2018) with preset filter antineoplastic and default advanced filters of source databases, gene categories and interaction types, the interactions were visualized using Cytoscape.

Results

Microarray datasets and patient characteristics

Based on searches in the GEO, ArrayExpress and TCGA databases, two microarray datasets, GSE68720 and GSE19475, that met the criteria, were selected for analysis. Both datasets were generated using the GPL570 Affymetrix Human Genome U133 Plus 2.0 Array platform. There were 80 *MLL*-R and 17 non-*MLL*-R infant ALL samples in the GSE68720 dataset, and 59 and 14, respectively, in the GSE19475 dataset. The age range of the 139 *MLL*-R infant ALL cases

(65 male, 73 female and 1 unknown) was 0 - 363 days, and the age range of the 31 non-*MLL*-R infant ALL cases (22 male, 9 female) was 0 - 365 days. *MLL*-R infant ALL displayed 77 (55.4%) *AF4*, 38 (27.3%) *ENL*, 14 (10.1%) *AF9*, and 10 (7.2%) other gene fusions.

Identification of DEGs

Based on the cut-off criteria, a total of 320 DEGs, including 236 up-regulated and 84 down-regulated genes, were identified between *MLL*-R and non-MLL-R infant ALL samples. The top five up-regulated DEGs were *LAMP5*, *PROM1*, *CCNA1*, *MEIS1*, and *KCNK12*, and the top five down-regulated DEGs were *MME*, *CMTM8*, *ELK3*, *NRN1*, and *PLEKHG4B* (Figure 1).

GO functional annotation and KEGG pathway analyses

GO functional annotation analysis showed that up-regulated DEGs were enriched in 62 BP, 21 CC, and 22 MF terms, and down-regulated DEGs were enriched in 34 BP, 22 CC, and 17 MF terms. The top five significant terms are shown in Figure 2, Figure 3 and Figure 4. The up-regulated DEGs were also enriched in seven KEGG pathways, mainly involving transcriptional regulation and signaling pathways, and down-regulated DEGs were enriched in three KEGG pathways (Table 1).

PPI network construction and analysis

The PPI network of DEGs included 297 nodes and 410 edges (Figure 5). Five bio-functional modules were screened out, based on the cut-off criteria, and the maximal MCODE score of module, which consisted of 15 nodes and 75 edges, was 10.714. This module contained 11 up-regulated (e.g. *MYC*, *CD44*, and *PTPRC*, etc.) and four down-regulated (e.g. *TNF*, *MME*, and *RAG1*, etc.) DEGs (Figure 6). The top 10 hub genes identified by Cytoscape, using maximal clique centrality, degree and betweenness, are listed in Table 2. *MYC*, *ALB*, *CD44*, *PTPRC* and TNF were identified as hub genes based on all three methods.

Drug-gene interactions analyses

23 drug-gene interactions including 4 up-regulated hub genes (*MYC*, *ALB*, *CD44*, *PTPRC*) and 24 drugs were identified by DGIdb, as shown in Figure 7.

Discussion

MLL is expressed during embryonic development and hematopoiesis by the histone-lysine N-methyltransferase 2A (*KMT2A*), which acts as a transcriptional coactivator that binds with other proteins in complex and methylates histone H3 lysine 4(Mohan, Lin, Guest, & Shilatifard, 2010). Rearrangements of *MLL* results in the fusion of its N-terminus with the C-terminus of a partner gene, leading to transcription dysregulation (Armstrong, Golub, & Korsmeyer, 2003). Previous research identified more than 90 fusion genes of *MLL*, including the frequent *MLL* rearrangement in infant ALL t(4;11)(q21;q23), which results in the *MLL-AF4* fusion (Meyer et al., 2018; Meyer et al., 2013). In the present study, three partner genes accounted for 92.8% of 139 cases: *MLL-AF4* (55.4%), *MLL-ENL* (27.3%), and *MLL-AF9* (10.1%). To confirm that the molecular mechanisms responsible for MLL-R infant ALL differs from those of childhood or non-*MLL*-R, we analyzed differences in gene expression profiles between *MLL*-R and non-*MLL*-R infant ALL samples, based on microarray datasets obtained from the GEO, ArrayExpress, and TCGA databases. Two microarray datasets were selected, and a total of 320 DEGs with fold-change >2 were screened out, including 236 up-regulated genes (e.g. *LAMP5*, *PROM1*, *CCNA1*, *MEIS1*, and *KCNK12*) and 84 down-regulated genes (e.g. *PLEKHG4B*, *NRN1*, *ELK3*, *CMTM8*, and *MME*). These results provide preliminary evidence for the distinct mechanisms of *MLL*-R infant ALL. We further investigated the functions of these DEGs and their possible roles in *MLL*-R infant ALL using DAVID functional annotations. In GO-BP, the most up-regulated DEGs were enriched in processes involved in the negative regulation of apoptosis, monocyte differentiation, and homophilic cell adhesion, suggesting that inhibition of apoptosis and monocyte differentiation may be involved in leukemogenesis. Similarly, homophilic cell adhesion, mediated by the up-regulation of protocadherin (*PCDH*) ɣ subfamily of genes (e.g. *PCDHGA*, *PCDHGB*, *PCDHGC*) may contribute to cell migration and invasion. Previous studies have confirmed that *PCDH* genes were involved in tumorigenesis and metastasis of gastric cancer, follicular lymphoma and non-small cell lung cancer (Mukai et al., 2017; Zhang et al., 2016; Zhou et al., 2017), although other studies have found that *PCDH* genes were implicated in tumor suppression (Chen et al., 2017; Weng et al., 2018). However, the role of *PCDH* genes in *MLL*-R ALL is currently unclear. *MEF2C*, *RAP2A*, *PDE4A*, *PPP3CA*, *PRNP*, and *MYC* were enriched in cellular responses to drugs in GO-BP terms, while down-regulated DEGs were enriched in extracellular matrix organization, protein kinase C signaling, and neuron projection extension. In GO-CC, up-regulated DEGs were mostly enriched in cell membranes or components of membrane-related terms.

We further uncovered molecular mechanistic insights of *MLL*-R ALL by KEGG pathway analysis, using the DAVID tool. The most significant pathway affected was transcriptional misregulation in cancer, which was enriched with up-regulated DEGs, including *MEF2C*,

PROM1, *CEBPA*, *LMO2*, *FLT3*, *RUNX1*, *HMGA2*, *RUNX2*, *MEIS1*, *MYC*, and *WT1*. Some of these genes have been defined as MLL targets genes. MEIS1, which is known as a key regulator in transcriptional regulation, cellular differentiation, and cell-cycle control, exhibited significant increases in both *MLL* fusion protein binding and mRNA expression on *MLL-ENL* activation in *MLL-ENL* leukemia cases and in an inducible cellular model (Wang et al., 2011). Further research also indicated that MEIS1 was essential for the development of *MLL* leukemia, by promoting cell differentiation resistance, and it was also confirmed to be involved in chemotherapy resistance (Rosales-Avina et al., 2011). As a downstream gene, *HMGA2* was positively regulated by *MLL* fusion proteins in infant *MLL-AF4* ALL leukemic cells (Wu et al., 2015). *PROM1* is a commonly used stem cell and cancer stem cell marker, and *MLL-AF4* was shown to promote *PROM1* transcription, which is required for *MLL-AF4*-driven leukemia cell growth (Mak, Nixon, & Moffat, 2012). *FLT3* is a class III receptor tyrosine kinase that plays an important role in hematopoietic stem cell development; high levels of *FLT3* are a common cooperating event in *MLL-AF4* ALL (Bueno et al., 2013). The presence of activating *FLT3* mutations in *MLL* is in keeping with a multistep pathway of leukemogenesis, suggesting that *FLT3* mutations may act as a second hit to lead to leukemogenesis in *MLL*-R infants (Kang et al., 2012). *LMO2* plays an important role in hematopoiesis and leukemogenesis, Bégay-Müller' research indicated that AF6, a recurrent fusion partner of *MLL*, bound to *LMO2* and may be involved in mixed lineage leukemia (Begay-Muller, Ansieau, & Leutz, 2002). *RUNX1* is known to be a putative target gene of *MLL* fusions, and was highly expressed in *MLL-AF4* ALL cases, when compared with normal bone marrow cells (Guenther et al., 2008; Krivtsov et al., 2008). The *BET* inhibitor I-BET151, arrested the growth of *MLL-AF4* infant leukemic cells in vitro through gene deregulation, including *RUNX1* (Bardini et al., 2018).

We examined interrelationships among the DEGs by constructing a PPI network using the STRING database and Cytoscape, to identify *MYC*, *ALB*, *CD44*, *PTPRC* and *TNF* as hub genes. *MYC* has been shown to be a direct target of *MLL-AF9* and is differentially expressed between neonatal and adult cells expressing *MLL-AF9* (Zuber et al., 2011). Notably, expression of the neonatal, but not the adult *MLL-AF9* signature, was also enriched in a core *MYC* network, suggesting that neonatal cells are inherently more prone to *MLL-AF9*-mediated immortalization than adult cells (Horton et al., 2013). Furthermore, MYC was essential for *MLL-ENL* to promote differentiation arrest of myelomonocytic precursor cells, and constitutive MYC expression cooperated with *MLL-ENL* to transform cells, with irreversible maturation arrest (Schreiner et al., 2001). *CD44* is a type I transmembrane glycoprotein and a leukocyte marker expressed on hematopoietic cells, various epithelial cell types, fibroblasts, and endothelial cells. *MLL*-R ALL had a unique genetic profile clearly distinguishable from those of other types of leukemia, with very high *CD44* levels (Tsutsumi et al., 2003). Furthermore, *CD44* expression represented the early steps of lymphoid development; high levels of *CD44* were associated with maturation arrest at an early lymphoid progenitor stage of development, and levels decreased with

maturation (Armstrong et al., 2002). *PTPRC* gene encodes for the protein tyrosine phosphatase *CD45,* which acts as a haematopoietic *JAK* phosphatase required for lymphocyte activation and development (Irie-Sasaki et al., 2001; Trowbridge & Thomas, 1994). High levels of *PTPRC* (*CD45*) expression were associated with a poor prognosis of B-cell-precursor ALL and T-cell ALL, which may be caused by *MLL-AF4* positivity and poor prednisone response (Cario et al., 2014). However, ALB up-regulation and *TNF* down-regulation in *MLL*-R ALL remain unclear.

In addition, 24 antineoplastic drugs were forecasted based on up-regulated hub genes. Some of these drugs have been used in the treatment of acute leukemia. The rest of them could be potential treatment options for infant *MLL*-R ALL, which will be the focus of our future research.

Conclusions

We performed differential gene expression analysis to uncover possible mechanisms underlying MLL-R infant ALL. The results indicated that GO-BP, including negative regulation of apoptotic processes, monocyte differentiation, and homophilic cell adhesion, and the KEGG pathway transcriptional misregulation in cancer, with hub genes including *MYC*, *ALB*, *CD44*, *PTPRC* and *TNF*, contributed to leukemogenesis, migration, and invasion in *MLL*-R infant ALL. And several drugs were predicted based on the hub genes. The roles of some of these genes were unambiguous based on previous studies, but further validation studies are needed to clarify the underlying genetic mechanisms of *MLL*-R infant ALL, and to provide evidence to support the development of novel therapeutic strategies. Similarly, further studies are also needed to elucidate the molecular mechanisms of other genes, such as *ALB* and *TNF*, in MLL-R infant ALL.

References

Armstrong SA, Golub TR, Korsmeyer SJ. 2003. MLL-rearranged leukemias: insights from gene expression profiling. Semin Hematol 40:268-273.

Armstrong SA, Staunton JE, Silverman LB, Pieters R, den Boer ML, Minden MD, Sallan SE, Lander ES, Golub TR, Korsmeyer SJ. 2002. MLL translocations specify a distinct gene expression profile that distinguishes a unique leukemia. Nat Genet 30:41-47. 10.1038/ng765.

Bardini M, Trentin L, Rizzo F, Vieri M, Savino AM, Garrido Castro P, Fazio G, Van Roon EHJ, Kerstjens M, Smithers N, Prinjha RK, Te Kronnie G, Basso G, Stam RW, Pieters R, Biondi A, Cazzaniga G. 2018. Antileukemic Efficacy of BET Inhibitor in a Preclinical Mouse Model of MLL-AF4(+) Infant ALL. Mol Cancer Ther 17:1705-1716. 10.1158/1535-7163.MCT-17-1123.

Barrett T, Wilhite SE, Ledoux P, Evangelista C, Kim IF, Tomashevsky M, Marshall KA, Phillippy KH, Sherman PM, Holko M, Yefanov A, Lee H, Zhang N, Robertson CL, Serova N, Davis S, Soboleva A. 2013. NCBI GEO: archive for functional genomics data sets--update. Nucleic Acids Res 41:D991-995. 10.1093/nar/gks1193.

Begay-Muller V, Ansieau S, Leutz A. 2002. The LIM domain protein Lmo2 binds to AF6, a translocation partner of the MLL oncogene. FEBS Lett 521:36-38.

Brown P, Pieters R, Biondi A. 2019. How I treat infant leukemia. Blood 133:205-214. 10.1182/blood-2018-04-785980.

Bueno C, Ayllon V, Montes R, Navarro-Montero O, Ramos-Mejia V, Real PJ, Romero-Moya D, Arauzo-Bravo MJ, Menendez P. 2013. FLT3 activation cooperates with MLL-AF4 fusion protein to abrogate the hematopoietic specification of human ESCs. Blood 121:3867-3878, S3861-3863. 10.1182/blood-2012-11-470146.

Cario G, Rhein P, Mitlohner R, Zimmermann M, Bandapalli OR, Romey R, Moericke A, Ludwig WD, Ratei R, Muckenthaler MU, Kulozik AE, Schrappe M, Stanulla M, Karawajew L. 2014. High CD45 surface expression determines relapse risk in children with precursor B-cell and T-cell acute lymphoblastic leukemia treated according to the ALL-BFM 2000 protocol. Haematologica 99:103-110. 10.3324/haematol.2013.090225.

Chen HF, Ma RR, He JY, Zhang H, Liu XL, Guo XY, Gao P. 2017. Protocadherin 7 inhibits cell migration and invasion through E-cadherin in gastric cancer. Tumour Biol 39:1010428317697551. 10.1177/1010428317697551.

Cotto KC, Wagner AH, Feng YY, Kiwala S, Coffman AC, Spies G, Wollam A, Spies NC, Griffith OL, Griffith M. 2018. DGIdb 3.0: a redesign and expansion of the drug-gene interaction database. Nucleic Acids Res 46:D1068-D1073. 10.1093/nar/gkx1143.

Ford AM, Ridge SA, Cabrera ME, Mahmoud H, Steel CM, Chan LC, Greaves M. 1993. In utero rearrangements in the trithorax-related oncogene in infant leukaemias. Nature 363:358-360. 10.1038/363358a0.

Gale KB, Ford AM, Repp R, Borkhardt A, Keller C, Eden OB, Greaves MF. 1997. Backtracking leukemia to birth: identification of clonotypic gene fusion sequences in neonatal blood spots. Proc Natl Acad Sci U S A 94:13950-13954.

Gautier L, Cope L, Bolstad BM, Irizarry RA. 2004. affy--analysis of Affymetrix GeneChip data at the probe level. Bioinformatics 20:307-315. 10.1093/bioinformatics/btg405.

Greaves MF, Maia AT, Wiemels JL, Ford AM. 2003. Leukemia in twins: lessons in natural history. Blood 102:2321-2333. 10.1182/blood-2002-12-3817.

Guenther MG, Lawton LN, Rozovskaia T, Frampton GM, Levine SS, Volkert TL, Croce CM, Nakamura T, Canaani E, Young RA. 2008. Aberrant chromatin at genes encoding stem cell regulators in human mixed-lineage leukemia. Genes Dev 22:3403-3408. 10.1101/gad.1741408.

Guest EM, Stam RW. 2017. Updates in the biology and therapy for infant acute lymphoblastic leukemia. Curr Opin Pediatr 29:20-26. 10.1097/MOP.0000000000000437.

Hilden JM, Dinndorf PA, Meerbaum SO, Sather H, Villaluna D, Heerema NA, McGlennen R, Smith FO, Woods WG, Salzer WL, Johnstone HS, Dreyer Z, Reaman GH, Children's Oncology G. 2006. Analysis of prognostic factors of acute lymphoblastic leukemia in infants: report on CCG 1953 from the Children's Oncology Group. Blood 108:441-451. 10.1182/blood-2005-07-3011.

Horton SJ, Jaques J, Woolthuis C, van Dijk J, Mesuraca M, Huls G, Morrone G, Vellenga E, Schuringa JJ. 2013. MLL-AF9-mediated immortalization of human hematopoietic cells along different lineages changes during ontogeny. Leukemia 27:1116-1126. 10.1038/leu.2012.343.

Huang da W, Sherman BT, Lempicki RA. 2009. Systematic and integrative analysis of large gene lists using DAVID bioinformatics resources. Nat Protoc 4:44-57. 10.1038/nprot.2008.211.

Irie-Sasaki J, Sasaki T, Matsumoto W, Opavsky A, Cheng M, Welstead G, Griffiths E, Krawczyk C, Richardson CD, Aitken K, Iscove N, Koretzky G, Johnson P, Liu P, Rothstein DM, Penninger JM. 2001. CD45 is a JAK phosphatase and negatively regulates cytokine receptor signalling. Nature 409:349-354. 10.1038/35053086.

Kang H, Wilson CS, Harvey RC, Chen IM, Murphy MH, Atlas SR, Bedrick EJ, Devidas M, Carroll AJ, Robinson BW, Stam RW, Valsecchi MG, Pieters R, Heerema NA, Hilden JM, Felix CA, Reaman GH, Camitta B, Winick N, Carroll WL, Dreyer ZE, Hunger SP, Willman CL. 2012. Gene expression profiles predictive of outcome and age in infant acute lymphoblastic leukemia: a Children's Oncology Group study. Blood 119:1872-1881. 10.1182/blood-2011-10-382861.

Krivtsov AV, Armstrong SA. 2007. MLL translocations, histone modifications and leukaemia stem-cell development. Nat Rev Cancer 7:823-833. 10.1038/nrc2253.

Krivtsov AV, Feng Z, Lemieux ME, Faber J, Vempati S, Sinha AU, Xia X, Jesneck J, Bracken AP, Silverman LB, Kutok JL, Kung AL, Armstrong SA. 2008. H3K79 methylation profiles define murine and human MLL-AF4 leukemias. Cancer Cell 14:355-368. 10.1016/j.ccr.2008.10.001.

Mak AB, Nixon AM, Moffat J. 2012. The mixed lineage leukemia (MLL) fusion-associated gene AF4 promotes CD133 transcription. Cancer Res 72:1929-1934. 10.1158/0008-5472.CAN-11-3589.

Meyer C, Burmeister T, Groger D, Tsaur G, Fechina L, Renneville A, Sutton R, Venn NC, Emerenciano M, Pombo-de-Oliveira MS, Barbieri Blunck C, Almeida Lopes B, Zuna J, Trka J, Ballerini P, Lapillonne H, De Braekeleer M, Cazzaniga G, Corral Abascal L, van der Velden VHJ, Delabesse E, Park TS, Oh SH, Silva MLM, Lund-Aho T, Juvonen V, Moore AS, Heidenreich O, Vormoor J, Zerkalenkova E, Olshanskaya Y, Bueno C, Menendez P, Teigler-Schlegel A, Zur Stadt U, Lentes J, Gohring G, Kustanovich A, Aleinikova O, Schafer BW, Kubetzko S, Madsen HO, Gruhn B, Duarte X, Gameiro P, Lippert E, Bidet A, Cayuela JM, Clappier E, Alonso CN, Zwaan CM, van den Heuvel-Eibrink MM, Izraeli S, Trakhtenbrot L, Archer P, Hancock J, Moricke A, Alten J, Schrappe M, Stanulla M, Strehl S, Attarbaschi A, Dworzak M, Haas OA, Panzer-Grumayer R, Sedek L, Szczepanski T, Caye A, Suarez L, Cave H, Marschalek R. 2018. The MLL recombinome of acute leukemias in 2017. Leukemia 32:273-284. 10.1038/leu.2017.213.

Meyer C, Hofmann J, Burmeister T, Groger D, Park TS, Emerenciano M, Pombo de Oliveira M, Renneville A, Villarese P, Macintyre E, Cave H, Clappier E, Mass-Malo K, Zuna J, Trka J, De Braekeleer E, De Braekeleer M, Oh SH, Tsaur G, Fechina L, van der Velden VH, van Dongen JJ, Delabesse E, Binato R, Silva ML, Kustanovich A, Aleinikova O, Harris MH, Lund-Aho T, Juvonen V, Heidenreich O, Vormoor J, Choi WW, Jarosova M, Kolenova A, Bueno C, Menendez P, Wehner S, Eckert C, Talmant P, Tondeur S, Lippert E, Launay E, Henry C, Ballerini P, Lapillone H, Callanan MB, Cayuela JM, Herbaux C, Cazzaniga G, Kakadiya PM, Bohlander S, Ahlmann M, Choi JR, Gameiro P, Lee DS, Krauter J, Cornillet-Lefebvre P, Te Kronnie G, Schafer BW, Kubetzko S, Alonso CN, zur Stadt U, Sutton R, Venn NC, Izraeli S, Trakhtenbrot L, Madsen HO, Archer P, Hancock J, Cerveira N, Teixeira MR, Lo Nigro L, Moricke A, Stanulla M, Schrappe M, Sedek L, Szczepanski T, Zwaan CM, Coenen EA, van den Heuvel-Eibrink MM, Strehl S, Dworzak M, Panzer-Grumayer R, Dingermann T, Klingebiel T, Marschalek R. 2013. The MLL recombinome of acute leukemias in 2013. Leukemia 27:2165-2176. 10.1038/leu.2013.135.

Mohan M, Lin C, Guest E, Shilatifard A. 2010. Licensed to elongate: a molecular mechanism for MLL-based leukaemogenesis. Nat Rev Cancer 10:721-728. 10.1038/nrc2915.

Mukai S, Oue N, Oshima T, Imai T, Sekino Y, Honma R, Sakamoto N, Sentani K, Kuniyasu H, Egi H, Tanabe K, Yoshida K, Ohdan H, Yasui W. 2017. Overexpression of PCDHB9 promotes peritoneal metastasis and correlates with poor prognosis in patients with gastric cancer. J Pathol 243:100-110. 10.1002/path.4931.

Nagayama J, Tomizawa D, Koh K, Nagatoshi Y, Hotta N, Kishimoto T, Takahashi Y, Kuno T, Sugita K, Sato T, Kato K, Ogawa A, Nakahata T, Mizutani S, Horibe K, Ishii E, Japan Infant

Leukemia Study G. 2006. Infants with acute lymphoblastic leukemia and a germline MLL gene are highly curable with use of chemotherapy alone: results from the Japan Infant Leukemia Study Group. Blood 107:4663-4665. 10.1182/blood-2005-11-4728.

Pieters R, Schrappe M, De Lorenzo P, Hann I, De Rossi G, Felice M, Hovi L, LeBlanc T, Szczepanski T, Ferster A, Janka G, Rubnitz J, Silverman L, Stary J, Campbell M, Li CK, Mann G, Suppiah R, Biondi A, Vora A, Valsecchi MG. 2007. A treatment protocol for infants younger than 1 year with acute lymphoblastic leukaemia (Interfant-99): an observational study and a multicentre randomised trial. Lancet 370:240-250. 10.1016/S0140-6736(07)61126-X.

Ritchie ME, Phipson B, Wu D, Hu Y, Law CW, Shi W, Smyth GK. 2015. limma powers differential expression analyses for RNA-sequencing and microarray studies. Nucleic Acids Res 43:e47. 10.1093/nar/gkv007.

Rosales-Avina JA, Torres-Flores J, Aguilar-Lemarroy A, Gurrola-Diaz C, Hernandez-Flores G, Ortiz-Lazareno PC, Lerma-Diaz JM, de Celis R, Gonzalez-Ramella O, Barrera-Chaires E, Bravo-Cuellar A, Jave-Suarez LF. 2011. MEIS1, PREP1, and PBX4 are differentially expressed in acute lymphoblastic leukemia: association of MEIS1 expression with higher proliferation and chemotherapy resistance. J Exp Clin Cancer Res 30:112. 10.1186/1756-9966-30-112.

Schreiner S, Birke M, Garcia-Cuellar MP, Zilles O, Greil J, Slany RK. 2001. MLL-ENL causes a reversible and myc-dependent block of myelomonocytic cell differentiation. Cancer Res 61:6480-6486.

Shannon P, Markiel A, Ozier O, Baliga NS, Wang JT, Ramage D, Amin N, Schwikowski B, Ideker T. 2003. Cytoscape: a software environment for integrated models of biomolecular interaction networks. Genome Res 13:2498-2504. 10.1101/gr.1239303.

Stam RW. 2013. The ongoing conundrum of MLL-AF4 driven leukemogenesis. Blood 121:3780-3781. 10.1182/blood-2013-03-491738.

Trowbridge IS, Thomas ML. 1994. CD45: an emerging role as a protein tyrosine phosphatase required for lymphocyte activation and development. Annu Rev Immunol 12:85-116. 10.1146/annurev.iy.12.040194.000505.

Tsutsumi S, Taketani T, Nishimura K, Ge X, Taki T, Sugita K, Ishii E, Hanada R, Ohki M, Aburatani H, Hayashi Y. 2003. Two distinct gene expression signatures in pediatric acute lymphoblastic leukemia with MLL rearrangements. Cancer Res 63:4882-4887.

van der Linden MH, Valsecchi MG, De Lorenzo P, Moricke A, Janka G, Leblanc TM, Felice M, Biondi A, Campbell M, Hann I, Rubnitz JE, Stary J, Szczepanski T, Vora A, Ferster A, Hovi L, Silverman LB, Pieters R. 2009. Outcome of congenital acute lymphoblastic leukemia treated on the Interfant-99 protocol. Blood 114:3764-3768. 10.1182/blood-2009-02-204214.

von Mering C, Huynen M, Jaeggi D, Schmidt S, Bork P, Snel B. 2003. STRING: a database of predicted functional associations between proteins. Nucleic Acids Res 31:258-261.

Wang QF, Wu G, Mi S, He F, Wu J, Dong J, Luo RT, Mattison R, Kaberlein JJ, Prabhakar S, Ji H, Thirman MJ. 2011. MLL fusion proteins preferentially regulate a subset of wild-type MLL target genes in the leukemic genome. Blood 117:6895-6905. 10.1182/blood-2010-12-324699.

Weng J, Xiao J, Mi Y, Fang X, Sun Y, Li S, Qin Z, Li X, Liu T, Zhao S, Zhou L, Wen Y. 2018. PCDHGA9 acts as a tumor suppressor to induce tumor cell apoptosis and autophagy and inhibit the EMT process in human gastric cancer. Cell Death Dis 9:27. 10.1038/s41419-017-0189-y.

Wu Z, Eguchi-Ishimae M, Yagi C, Iwabuki H, Gao W, Tauchi H, Inukai T, Sugita K, Ishii E, Eguchi M. 2015. HMGA2 as a potential molecular target in KMT2A-AFF1-positive infant acute lymphoblastic leukaemia. Br J Haematol 171:818-829. 10.1111/bjh.13763.

Zhang X, Takata K, Cui W, Miyata-Takata T, Sato Y, Noujima-Harada M, Yoshino T. 2016. Protocadherin gamma A3 is expressed in follicular lymphoma irrespective of BCL2 status and is associated with tumor cell growth. Mol Med Rep 14:4622-4628. 10.3892/mmr.2016.5808.

Zhou X, Updegraff BL, Guo Y, Peyton M, Girard L, Larsen JE, Xie XJ, Zhou Y, Hwang TH, Xie Y, Rodriguez-Canales J, Villalobos P, Behrens C, Wistuba, II, Minna JD, O'Donnell KA. 2017. PROTOCADHERIN 7 Acts through SET and PP2A to Potentiate MAPK Signaling by EGFR and KRAS during Lung Tumorigenesis. Cancer Res 77:187-197. 10.1158/0008-5472.CAN-16-1267-T.

Zuber J, Rappaport AR, Luo W, Wang E, Chen C, Vaseva AV, Shi J, Weissmueller S, Fellmann C, Taylor MJ, Weissenboeck M, Graeber TG, Kogan SC, Vakoc CR, Lowe SW. 2011. An integrated approach to dissecting oncogene addiction implicates a Myb-coordinated self-renewal program as essential for leukemia maintenance. Genes Dev 25:1628-1640. 10.1101/gad.17269211.

Figure 1(on next page)

Heat map of the top 20 up-regulated and down-regulated DEGs

Each row represents a single gene, each column represents a sample. The gradual color change from green to red represents the gene expression values change from low to high. DEG: differentially expressed genes; *MLL*-R: Mixed Lineage Leukemia rearrangement.

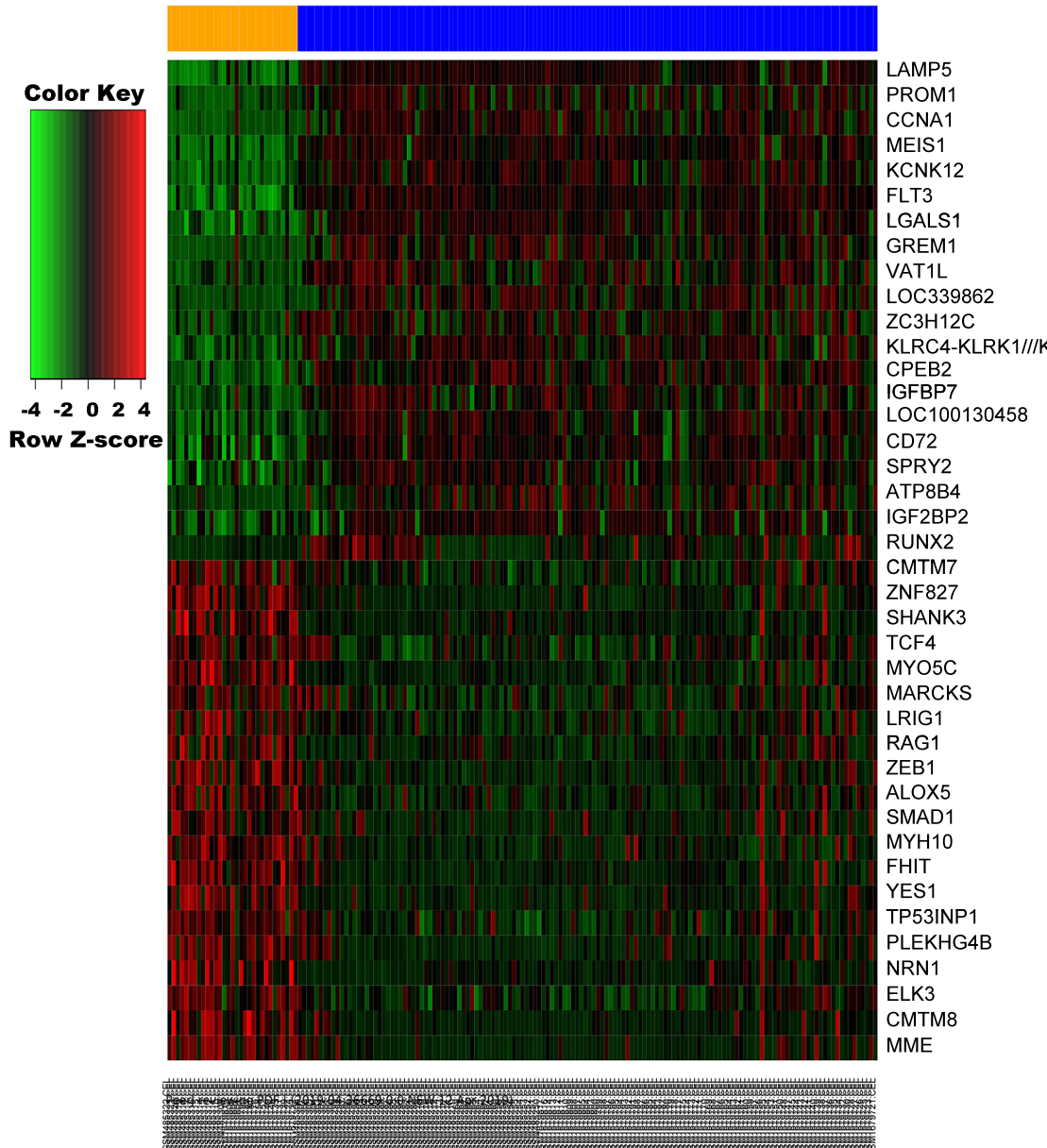
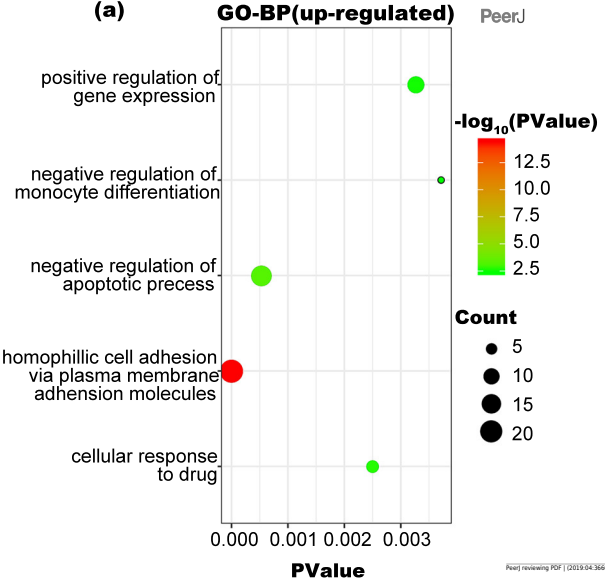

non-MLL-R
MLL-R
Color Key
-4 -2 0 2 4
Row Z-score
LAMP5
PROM1
CCNA1
MEIS1
KCNK12
FLT3
LGALS1
GREM1
VAT1L
LOC339862
ZC3H12C
KLRC4-KLRK1///K
CPEB2
IGFBP7
LOC100130458
CD72
SPRY2
ATP8B4
IGF2BP2
RUNX2
CMTM7
ZNF827
SHANK3
TCF4
MYO5C
MARCKS
LRIG1
RAG1
ZEB1
ALOX5
SMAD1
MYH10
FHIT
YES1
TP53INP1
PLEKHG4B
NRN1
ELK3
CMTM8
MME

Figure 2(on next page)

GO-BP function annotation the DEGs

The gradual color change from green to red represents the -log10(PValue) change from low to high, the size of point represents the the count of genes. (a) The top 5 significantly enriched GO-BP terms for up-regulated DEGs. (b) The top 5 significantly enriched GO-BP terms for down-regulated DEGs. DEG: differentially expressed genes; GO: gene ontology; BP: biological process.



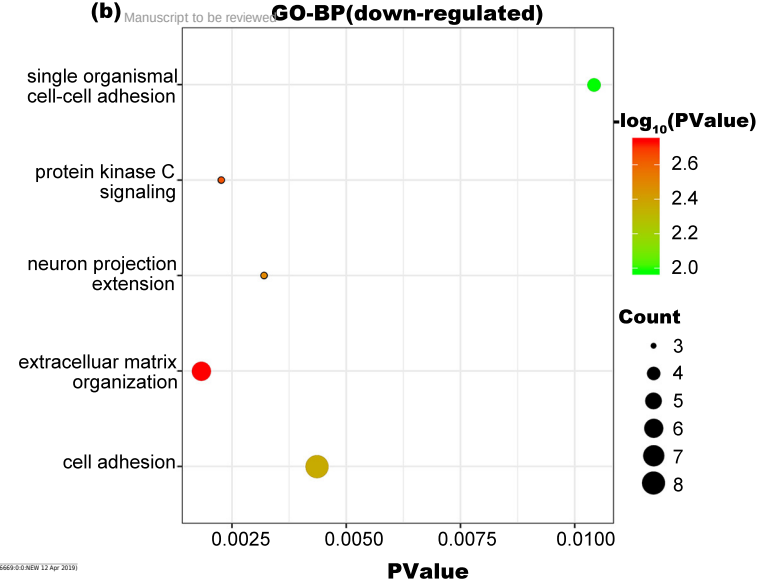


Figure 3(on next page)

GO-CC function annotation the DEGs

The gradual color change from green to red represents the -log10(PValue) change from low to high, the size of point represents the the count of genes. (a) The top 5 significantly enriched GO-CC terms for up-regulated DEGs. (b) The top 5 significantly enriched GO-CC terms for down-regulated DEGs. DEG: differentially expressed genes; GO: gene ontology; CC: cellular component.

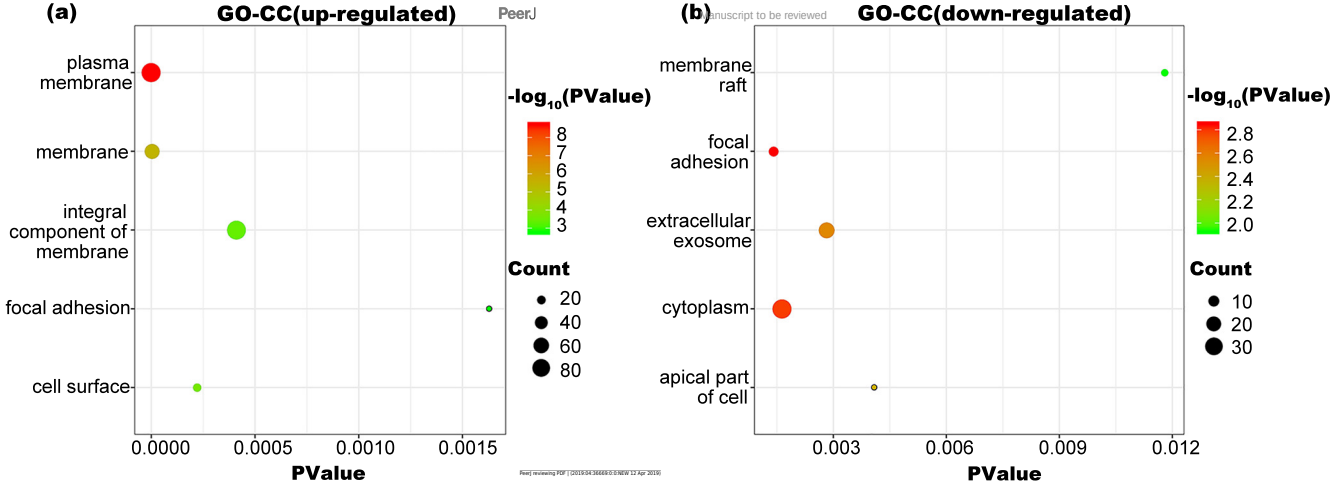
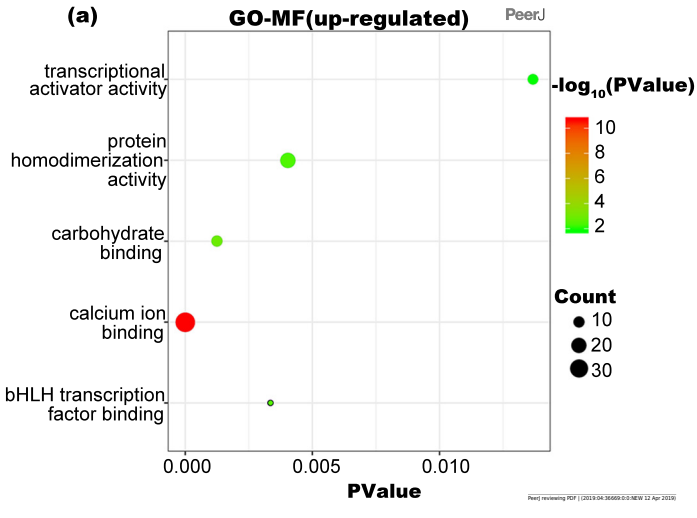


Figure 4(on next page)

GO-MF function annotation the DEGs

The gradual color change from green to red represents the -log10(PValue) change from low to high, the size of point represents the the count of genes. (a) The top 5 significantly enriched GO-MF terms for up-regulated DEGs. (b) The top 5 significantly enriched GO-MF terms for down-regulated DEGs. DEG: differentially expressed genes; GO: gene ontology; MF: molecular function.



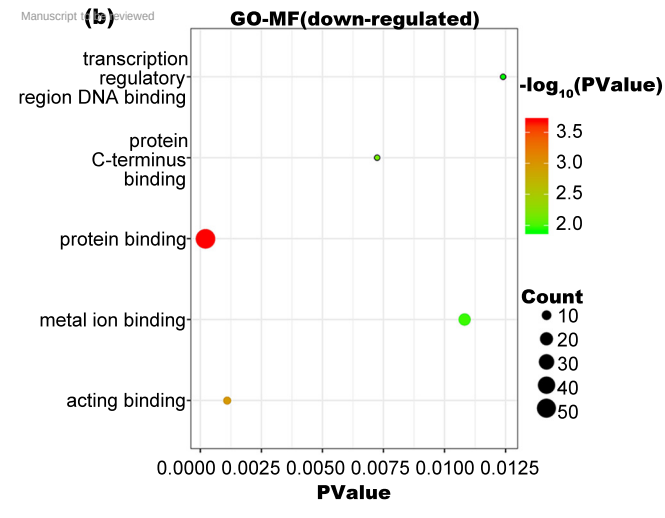


Figure 5(on next page)

PPI network of DEGs

The size of edges change from small to large represents the combined score of nodes change from low to high, the size of nodes represents the count. Red nodes represent up-regulated genes, yellow nodes represent down-regulated genes. DEG: differentially expressed genes.

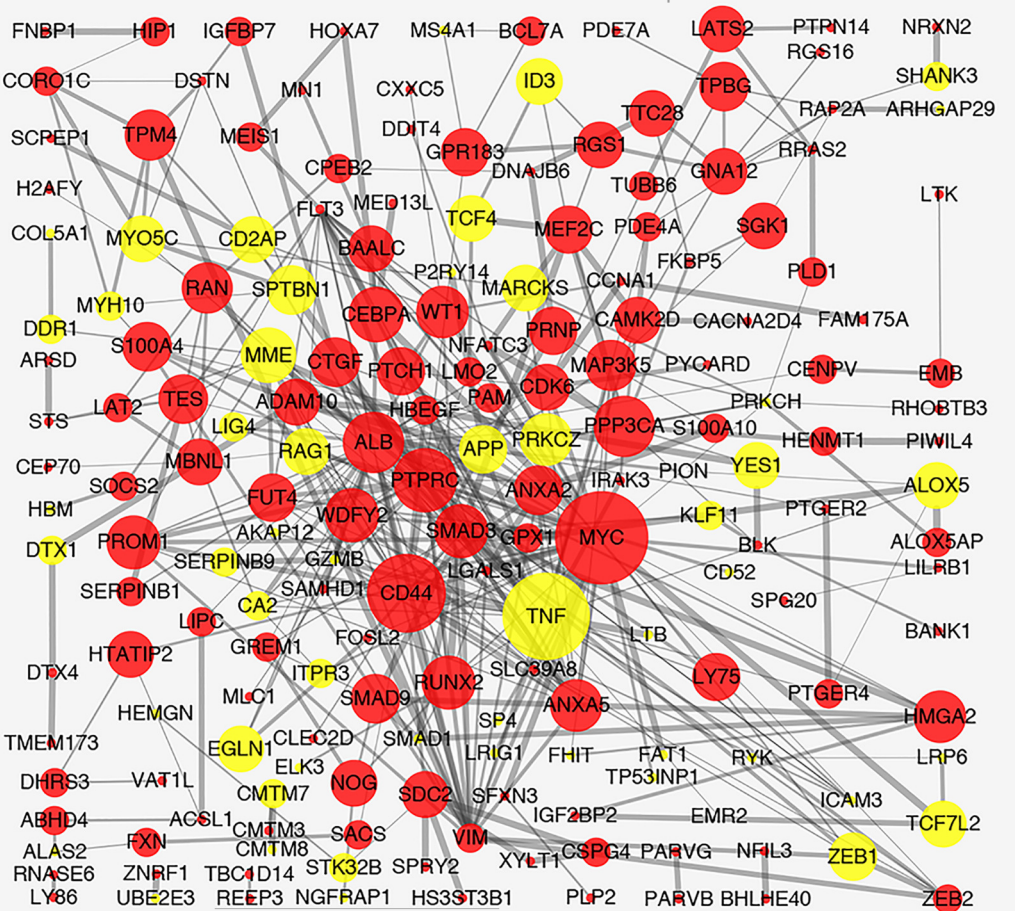

FNBP1
HIP1
IGFBP7
HOXA7
MS4A1
BCL7A
PDE7A
LATS2
PTPN14
NRXN2
RGS16
CORO1C
DSTN
MN1
CXXC5
ID3
TPBG
SHANK3
SCRIB1
TPM4
MEIS1
DDIT4
GPR183
RGS1
TTC28
RAP2A
ARHGAP29
RRAS2
H2AFY
CPEB2
DNAJB6
TUBB6
GNA12
LTK
COL5A1
MYO5C
CD2AP
FLT3
MED13L
TCF4
MEF2C
PDE4A
SGK1
BAALC
FKBP5
PLD1
P2RY14
MARCKS
CCNA1
RAN
SPTBN1
MYH10
CEBPA
WT1
PRNP
CAMK2D
CACNA2D4
FAM175A
DDR1
S100A4
MME
CTGF
NFATC3
ARSD
PTCH1
LMO2
MAP3K5
PYCARD
CENPV
EMB
LAT2
TES
ADAM10
HBEGF
PAM
CDK6
PRKCH
RHOBTB3
STS
LIG4
PPP3CA
S100A10
HENMT1
PIWIL4
CEP70
MBNL1
RAG1
ALB
APP
PRKCZ
YES1
SOCS2
PTPRC
ANXA2
IRAK3
PION
ALOX5
HBM
FUT4
WDFY2
KLF11
PTGER2
DTX1
PROM1
AKAP12
SMAD3
GPX1
MYC
BLK
ALOX5AP
SERPINB9
GZMB
LGALS1
CD52
LILRB1
SERPINB1
SAMHD1
CD44
SPG20
CA2
TNF
LIPC
LTB
BANK1
FOSL2
HTATIP2
GREM1
DTX4
ITPR3
SLC39A8
LY75
RUNX2
MLC1
SMAD9
ANXA5
PTGER4
HEMGN
SP4
HMGA2
TMEM173
CLEC2D
SMAD1
EGLN1
LRIG1
FHIT
FAT1
RYK
LRP6
ELK3
TP53INP1
DHRS3
VAT1L
NOG
SDC2
SFXN3
CMTM7
ICAM3
ABHD4
ACSL1
IGF2BP2
EMR2
TCF7L2
CMTM3
SACS
VIM
CMTM8
CSPG4
PARVG
NFIL3
ZEB1
ALAS2
FXN
RNASE6
ZNRF1
TBC1D14
STK32B
SPRY2
XYLT1
LY86
UBE2E3
REEP3
NGFRAP1
HS3ST3B1
PLP2
PARVB
BHLHE40
ZEB2

Figure 6(on next page)

Bio-functional module of DEGs

The size of edges change from small to large represents the combined score of nodes change from low to high, the size of nodes represents the count. Red nodes represent up-regulated genes, yellow nodes represent down-regulated genes. DEG: differentially expressed genes.

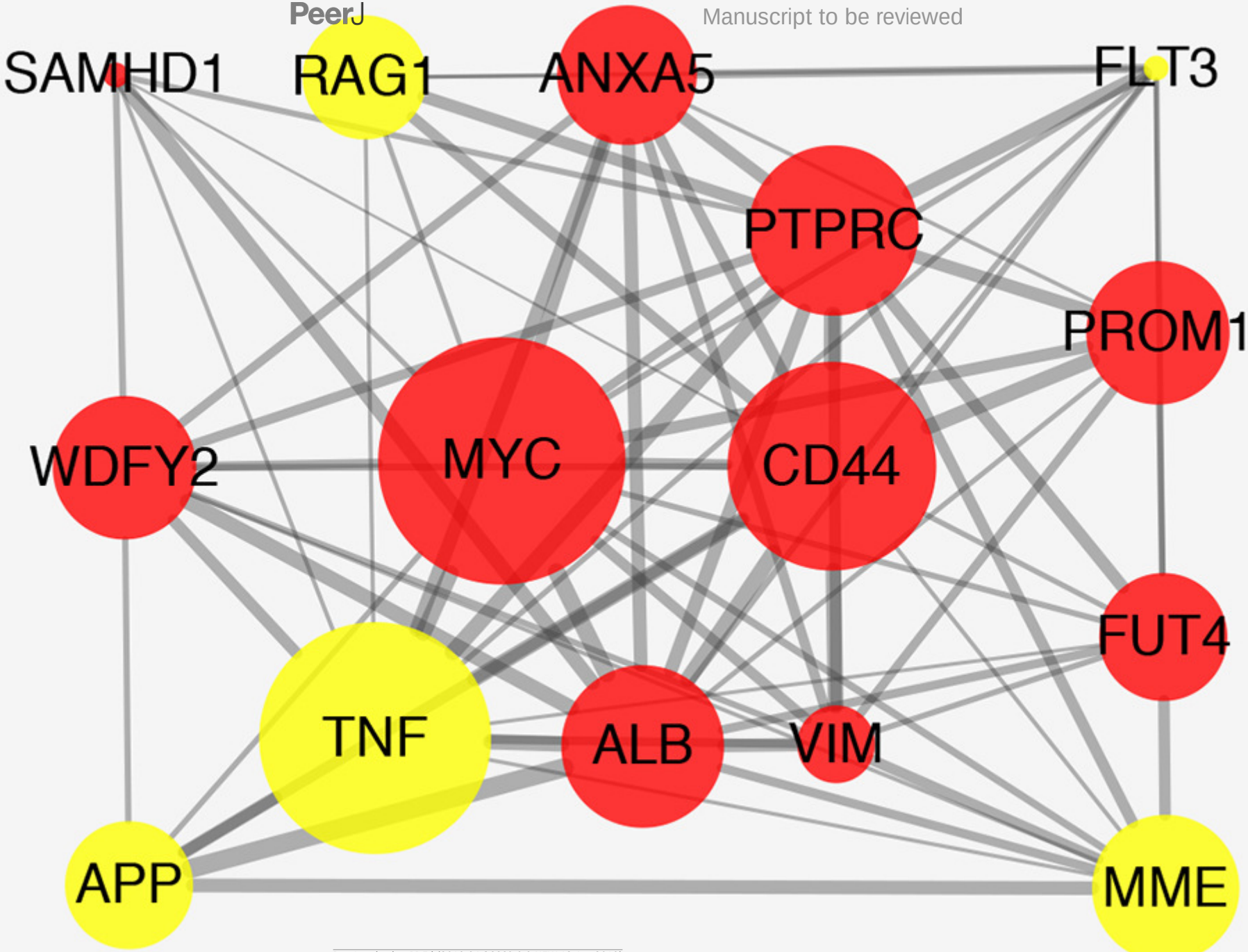

SAMHD1
RAG1
ANXA5
FLT3
PTPRC
PROM1
WDFY2
MYC
CD44
FUT4
TNF
ALB
VIM
APP
MME

Figure 7(on next page)

Drug-gene interactions

Red nodes represent up-regulated hub genes and green nodes represent the drugs.

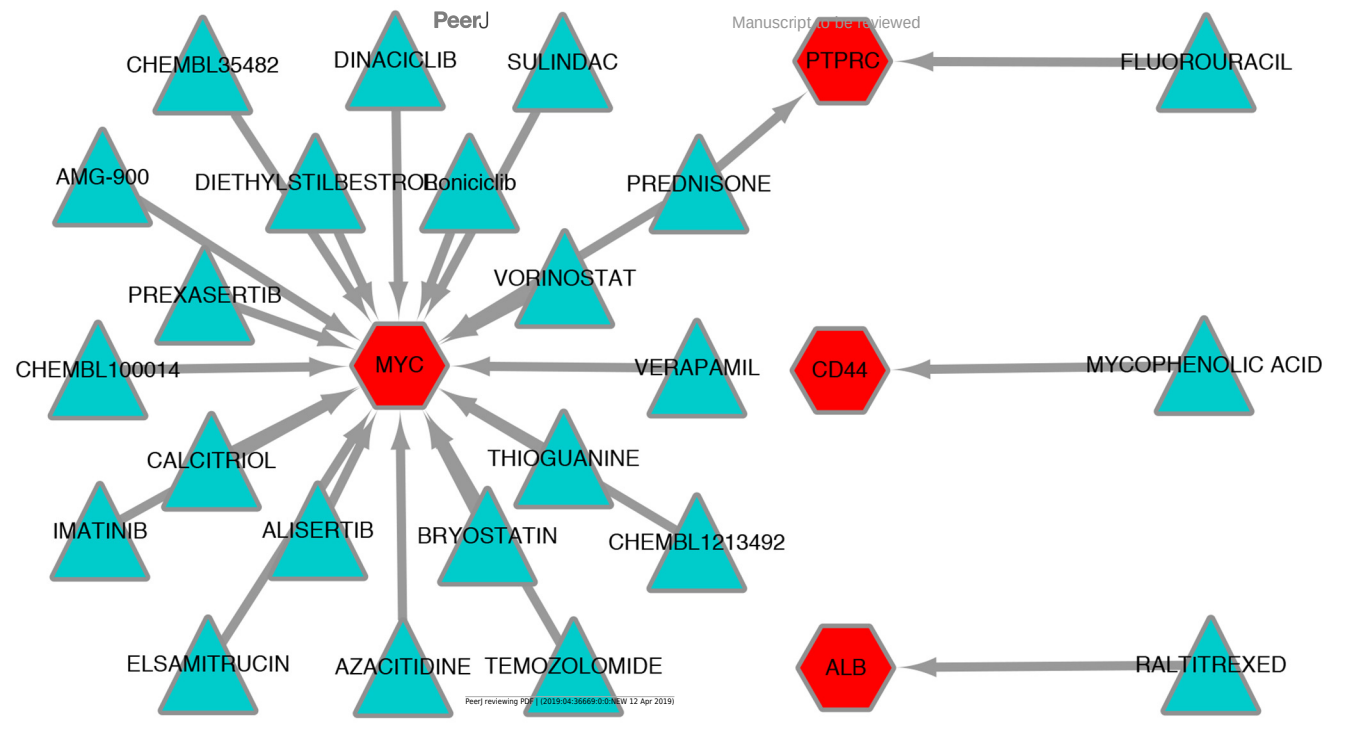

CHEMBL35482
DINACICLIB
SULINDAC
PTPRC
FLUOROURACIL
AMG-900
DIETHYLSTILBESTROL
Boniciclib
PREDNISONE
PREXASERTIB
VORINOSTAT
CHEMBL100014
MYC
VERAPAMIL
CD44
MYCOPHENOLIC ACID
CALCITRIOL
THIOGUANINE
IMATINIB
ALISERTIB
BRYOSTATIN
CHEMBL1213492
ELSAMITRUCIN
AZACITIDINE
TEMOZOLOMIDE
ALB
RALTITREXED

Table 1(on next page)

KEGG pathways of DEGs

KEGG: Kyoto encyclopedia of genes and genomes; DEG: differentially expressed genes

1 Table 1. KEGG pathways of DEGs

Term	Genes	Type of DEGs	PValue
Transcriptional misregulation in cancer	*MEF2C, PROM1, CEBPA, LMO2, FLT3, RUNX1, HMGA2, RUNX2, MEIS1, MYC, WT1*	up-regulated	0.000
Acute myeloid leukemia	*CEBPA, FLT3, RUNX1, MYC*	up-regulated	0.042
Oxytocin signaling pathway	*MEF2C, CAMK2D, GUCY1A3, PPP3CA, NFATC3, CACNA2D4*	up-regulated	0.056
Proteoglycans in cancer	*CD44, RRAS2, CAMK2D, HBEGF, PTCH1, MYC, SDC2*	up-regulated	0.058
Renin secretion	*PTGER2, PTGER4, GUCY1A3, PPP3CA*	up-regulated	0.058
MAPK signaling pathway	*MEF2C, MAP3K5, RRAS2, GNA12, PPP3CA, NFATC3, MYC, CACNA2D4*	up-regulated	0.059
Pathways in cancer	*CEBPA, PTGER2, PTGER4, FLT3, GNA12, SMAD3, PTCH1, CDK6, RUNX1, MYC*	up-regulated	0.092
MicroRNAs in cancer	*SPRY2, CD44, VIM, ZEB2, CDK6, HMGA2, MYC, DDIT4*	up-regulated	0.099
Alzheimer's disease	*APP, TNF, MME, ITPR3*	down-regulated	0.053
TGF-beta signaling pathway	*TNF, ID3, SMAD1*	down-regulated	0.068
Hematopoietic cell lineage	*TNF, MS4A1, MME*	down-regulated	0.072

2 KEGG: Kyoto encyclopedia of genes and genomes; DEG: differentially expressed genes

3

4

5

6

Table 2(on next page)

Top 10 hub genes that evaluated by the methods including MCC (Maximal Clique Centrality), Degree and Betweenness

Table 2 Top 10 hub genes that evaluated by the methods including MCC (Maximal Clique Centrality), Degree and Betweenness

Gene	MCC Score	Gene	Degree Score	Gene	Betweenness Score
MYC	22218	*MYC*	43	*ALB*	8542.027612
CD44	22075	*ALB*	42	*MYC*	8398.777703
ALB	21713	*TNF*	33	*TNF*	4702.068242
PTPRC	21310	*CD44*	26	*CD44*	4164.511092
TNF	19364	*PTPRC*	22	*PPP3CA*	2262.216619
VIM	16980	*VIM*	20	*SDC2*	2006
WDFY2	11570	*SMAD3*	18	*FLT3*	1790.958256
MME	10849	*FLT3*	17	*MAP3K5*	1762.347815
ANXA5	8334	*ANXA5*	15	*SMAD3*	1738.446661
FUT4	7200	*MME*	14	*PTPRC*	1709.765874