



Research Paper

Genomic Profiling of Adult and Pediatric B-cell Acute Lymphoblastic Leukemia



Yuan-Fang Liu ^{a,1}, Bai-Yan Wang ^{a,m,1}, Wei-Na Zhang ^{a,1}, Jin-Yan Huang ^{a,*,1}, Ben-Shang Li ^{c,1}, Ming Zhang ^a, Lu Jiang ^a, Jian-Feng Li ^a, Ming-Jie Wang ^a, Yu-Jun Dai ^a, Zi-Guan Zhang ^b, Qiang Wang ^a, Jie Kong ^a, Bing Chen ^a, Yong-Mei Zhu ^a, Xiang-Qin Weng ^a, Zhi-Xiang Shen ^a, Jun-Min Li ^a, Jin Wang ^a, Xiao-Jing Yan ⁱ, Yan Li ⁱ, Ying-Min Liang ^k, Li Liu ^k, Xie-Qun Chen ^j, Wang-Gang Zhang ^m, Jin-Song Yan ^l, Jian-Da Hu ⁿ, Shu-Hong Shen ^c, Jing Chen ^c, Long-Jun Gu ^c, Deqing Pei ^h, Yongjin Li ^f, Gang Wu ^f, Xin Zhou ^f, Rui-Bao Ren ^a, Cheng Cheng ^h, Jun J. Yang ^g, Kan-Kan Wang ^a, Sheng-Yue Wang ^d, Jinghui Zhang ^f, Jian-Qing Mi ^{a,*}, Ching-Hon Pui ^{e,*}, Jing-Yan Tang ^{c,*}, Zhu Chen ^{a,b,*}, Sai-Juan Chen ^{a,b,*}

^a State Key Laboratory of Medical Genomics, Shanghai Institute of Hematology, Rui Jin Hospital Affiliated to Shanghai Jiao Tong University School of Medicine, 197 Rui Jin Road II, Shanghai 200025, China

^b Key Laboratory of Ministry of Education for Systems Biomedicine, Shanghai Center for Systems Biomedicine, Shanghai Jiao Tong University, 800 Dongchuan Road, Shanghai 200240, China

^c Key Laboratory of Pediatric Hematology & Oncology, Ministry of Health, Department of Hematology and Oncology, Shanghai Children's Medical Center, Shanghai Jiao Tong University School of Medicine, Shanghai 200127, China

^d Chinese National Human Genome Center at Shanghai, Shanghai 201203, China

^e Department of Oncology, St. Jude Children's Research Hospital, Memphis, TN, USA

^f Computational Biology and Bioinformatics, St. Jude Children's Research Hospital, Memphis, TN, USA

^g Pharmaceutical Sciences, St. Jude Children's Research Hospital, Memphis, TN, USA

^h Biostatistics, St. Jude Children's Research Hospital, Memphis, TN, USA

ⁱ Department of Hematology, First Hospital of China Medical University, Shenyang 110001, China

^j Department of Hematology, Xi Jing Hospital affiliated to the Fourth Military Medical University, Xi'an, Shaan Xi 710032, China

^k Department of Hematology, Tang Du Hospital affiliated to the Fourth Military Medical University, Xi'an, Shaan Xi 710038, China

^l Dalian Key Laboratory of Hematology, Department of Hematology, Second Hospital of Dalian Medical University, Dalian, Liaoning 116027, China

^m Department of Hematology, The Second Affiliated Hospital of Xi'an Jiaotong University School of Medicine, Xi'an, Shaan Xi 710004, China

ⁿ Fujian Institute of Hematology, Fujian Medical University Union Hospital, Fuzhou 350000, China

ARTICLE INFO

Article history:

Received 21 February 2016

Received in revised form 11 April 2016

Accepted 29 April 2016

Available online 13 May 2016

Keywords:

Adult B-ALL

Pediatric B-ALL

Next-generation sequencing

MEF2D fusions

ZNF384 fusions

ABSTRACT

Genomic landscapes of 92 adult and 111 pediatric patients with B-cell acute lymphoblastic leukemia (B-ALL) were investigated using next-generation sequencing and copy number alteration analysis. Recurrent gene mutations and fusions were tested in an additional 87 adult and 93 pediatric patients. Among the 29 newly identified in-frame gene fusions, those involving *MEF2D* and *ZNF384* were clinically relevant and were demonstrated to perturb B-cell differentiation, with *EP300-ZNF384* inducing leukemia in mice. Eight gene expression subgroups associated with characteristic genetic abnormalities were identified, including leukemia with *MEF2D* and *ZNF384* fusions in two distinct clusters. In subgroup G4 which was characterized by *ERG* deletion, *DUX4-IGH* fusion was detected in most cases. This comprehensive dataset allowed us to compare the features of molecular pathogenesis between adult and pediatric B-ALL and to identify signatures possibly related to the inferior outcome of adults to that of children. We found that, besides the known discrepancies in frequencies of prognostic markers, adult patients had more cooperative mutations and greater enrichment for alterations of epigenetic modifiers and genes linked to B-cell development, suggesting difference in the target cells of transformation between adult and pediatric patients and may explain in part the disparity in their responses to treatment.

© 2016 The Authors. Published by Elsevier B.V. This is an open access article under the CC BY-NC-ND license (<http://creativecommons.org/licenses/by-nc-nd/4.0/>).

1. Introduction

Acute lymphoblastic leukemia (ALL) results from the clonal proliferation of lymphoid stem or progenitor cells, with more than 80% being originated from B-cell progenitors (B-ALL) (Pui, 2010). Recurrent

* Corresponding authors.

E-mail address: sjchen@stn.sh.cn (S.-J. Chen).

¹ Y.-F.L., B.-Y.W., W.-N.Z., J.-Y.H. and B.-S.L. contributed equally to this work.

cytogenetic and molecular abnormalities have been identified to play key roles in ALL pathogenesis, frequently by targeting vital molecular components of hematopoietic differentiation, cell cycling, tumor suppression and stem cell self-renewal (Roberts et al., 2012). The identification of these abnormalities not only reveals underlying molecular pathology, but also provides important therapeutic targets, as exemplified by the improved outcome achieved with ABL tyrosine kinase inhibitors in patients with Philadelphia chromosome-positive or Philadelphia chromosome-like ALL patients with ABL class fusion transcripts (Schultz et al., 2014; Chalandon et al., 2015; Roberts et al., 2014). Recent technological advances have enabled detailed characterization of the genomic landscape of childhood ALL, including DNA sequence abnormalities such as single nucleotide variations (SNV), small insertions or deletions (indels) and copy number variations (CNV) (Mullighan et al., 2007); gene expression anomalies (Roberts et al., 2014; Den Boer et al., 2009); gene fusions due to cryptic chromosomal rearrangements (Roberts et al., 2014; Gocho et al., 2015); and aberrant epigenetic modifications including abnormal DNA methylation, histone modifications and mutations of epigenetic modifier genes (Gabriel et al., 2015). The precise contribution of these genetic or epigenetic abnormalities to leukemogenesis, the development of drug resistance and leukemic clone evolution remains to be defined (Ma et al., 2015).

Remarkable progress has been made in the treatment of ALL. Currently, the five-year overall survival rates exceed 85% in pediatric patients in developed countries (Pui et al., 2015), but remain less than 45% in adults (Jabbour et al., 2015; Bassan and Hoelzer, 2011). The poor outcome in adults has been attributed in part to a high frequency of unfavorable genetic subtypes of ALL, pre-existing co-morbidities, and poor tolerance of intensive treatment.

In our previous report on large B-ALL cohorts in China, the outcome of both adult and pediatric patients seemed less favorable than that of the best centers in Western countries (Mi et al., 2012). The higher frequency of unfavorable prognostic factors such as *BCR-ABL1* and the lower frequency of favorable factors such as *ETV6-RUNX1* and hyperdiploidy in Chinese pediatric ALL could contribute to this difference (Chen et al., 2012). However, within the same genetic subtypes of ALL, the leukemic cells of adult patients are more resistant to treatment than those of pediatric patients. For example, cures can be achieved with intensive chemotherapy and an ABL tyrosine kinase inhibitor in up to 70% of children with Philadelphia chromosome-positive ALL with *BCR-ABL1* fusion (Chalandon et al., 2015), but in less than 50% of adults even with the addition of transplantation (Chalandon et al.,

2015; Jabbour et al., 2015). In a recent study of *MLL*-rearranged B-ALL, older children had more somatic mutations and had a higher frequency of mutated epigenetic regulators than did infants (Andersson et al., 2015), suggesting important differences in the development and prognosis of leukemia between infants and older children. Thus, we undertook a detailed analysis of B-ALL leukemic cell genomics in adults and children to identify genetic abnormalities in a systematic way and to discover alterations that might explain the inferior prognosis of adult B-ALL and to identify potential therapeutic targets for this high-risk cancer.

2. Methods

2.1. Patients and Samples

All of the B-ALL patients enrolled in this study were diagnosed and/or treated in the Shanghai Institute of Hematology (SIH)-based hospital network or Multicenter Hematology-Oncology Protocols Evaluation System (M-HOPES) in China. A total of 383 patients including 179 adults (>18 years) and 204 children (≤18 years) with newly diagnosed B-cell ALL and tissue samples available for genomic analyses were enrolled in this study.

The 92 adults and 111 children with sufficient bone marrow and matched remission samples or saliva samples for one or more of the following analyses formed the discovery cohort: whole-exome sequencing (201 patients), whole-genome sequencing (9), whole-transcriptome sequencing (172) and copy number alteration analysis (202) (Fig. 1). The features for these patients are reported in Table 1 (Table S1 for individual patients) and are in concordance with the data we previously reported (Mi et al., 2012; Chen et al., 2012).

Targeted deep sequencing for gene mutations and reverse transcriptase (RT)-PCR for recurrent fusion genes were performed in an additional 87 adults and 93 children, designated the recurrent cohort. Adult patients in discovery cohort were mostly enrolled in an SIH protocol [Chinese Clinical Trial Registry, number ChiCTR-RNC-14004969 (for sample collection) and ChiCTR-ONRC-14004968 (for treatment)] which were basically a modified VDLCF regimen. Pediatric patients in discovery cohort were mostly enrolled in the Shanghai Children's Medical Center ALL-2005 protocol (Chinese Clinical Trial Registry, number ChiCTR-ONC-14005003). Other patients in the recurrent cohort were treated as described in detail previously (Mi et al., 2012). Fifteen patients 17 to 18 years old were treated with the adult clinical trial. The

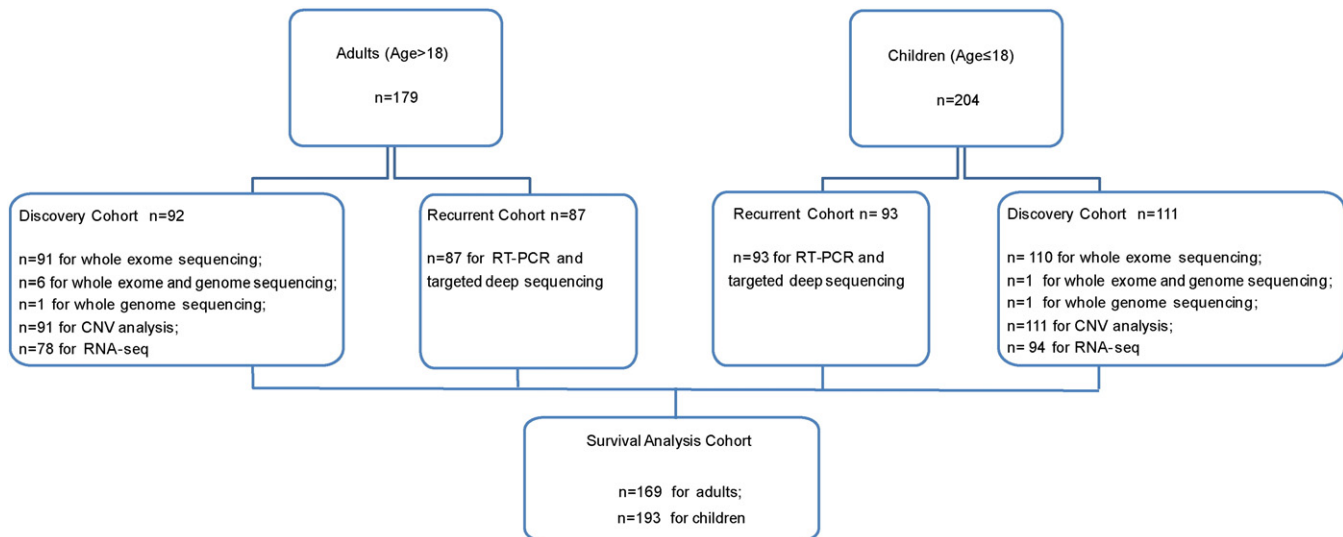


Fig. 1. Flow chart of B-ALL patients in this study. A total of 383 patients including 179 adults (>18 years) and 204 children (≤18 years) with newly diagnosed B-ALL were enrolled in this study. The 92 adults and 111 children with sufficient samples for next-generation sequencing formed the discovery cohort. An additional 87 adults and 93 children, designated the recurrent cohort, were screened for recurrent gene mutations and fusion genes.

Table 1
Clinical characteristics and genetic types of patient cohorts.

	Discovery cohort		Recurrent cohort	
	Adult	Pediatric	Adult	Pediatric
Number	92	111	87	93
Age at diagnosis (year)				
Median	31.4	5.1	36.7	6.8
Range	18.1–68.9	0.4–18.0	18.1–63.4	1.2–17.8
Gender, no. (%)				
Male	48(52.2%)	63(56.8%)	43(49.4%)	63(67.7%)
Female	44(47.8%)	48(43.2%)	44(50.6%)	30(32.3%)
WBC count at diagnosis ($\times 10^9/L$)				
Median	22.5	10.8	18.3	15.2
Range	0.4–438.6	0.9–508.8	1.1–420.0	1.1–767.7
Specific genetic abnormalities, no. (%)				
<i>BCR-ABL1</i>	24(26.1%)	10(9.0%)	25(28.7%)	14(15.1%)
<i>BCR-ABL1</i> -like ^a	8(10.3%)	5(5.3%)	–	–
<i>MLL</i> rearranged	7(7.6%)	4(3.6%)	5(5.7%)	2(2.2%)
<i>TCF3-PBX1</i>	9(9.8%)	8(7.2%)	0(0.0%)	4(4.3%)
<i>ETV6-RUNX1</i>	0(0.0%)	22(19.8%)	1(1.1%)	13(14.0%)
<i>ZNF384</i> fusions ^b	9(10.0%)	6(5.7%)	4(4.6%)	2(2.2%)
<i>MEF2D</i> fusions ^b	3(3.3%)	4(3.8%)	9(10.3%)	2(2.2%)
hyperdiploidy >50	0(0.0%)	3(2.7%)	1(1.1%)	11(11.8%)
hypodiploidy	6(6.5%)	1(0.9%)	6(6.9%)	3(3.2%)
others	26(28.3%)	48(43.2%)	36(41.4%)	42(45.2%)

^a The *BCR-ABL1*-like signature was identified with gene expression data, which was available in 78 adults and 94 children subject to RNA-seq.

^b The data of *MEF2D* and *ZNF384* fusions were available in 90 adults and 106 children in the discovery cohort, and 87 adults and 93 children in the recurrent cohort.

study was approved by the institution review board of each participating center. All patients, parents or guardians provided informed consent for sample collection and research in accord with the Declaration of Helsinki.

Bone marrow (BM) aspiration was conducted at diagnosis and mononuclear cells were enriched by density gradient centrifugation with Ficoll solution. Genomic DNA and total RNA were extracted using AllPrep DNA/RNA/Protein Mini Kit (Qiagen) or TRIzol reagent (Invitrogen). The germline control DNA was obtained from matching peripheral blood during complete remission after standard chemotherapy and prepared by QuickGene DNA whole blood Kit L (FUJIFILM, Life Science), or from saliva samples collected using Oragene kits according to the manufacturer's instructions (DNA Genotek Inc.). Morphological, immunophenotypic and cytogenetic analyses were performed at the time of diagnosis. The transcripts of *BCR-ABL1*, *ETV6-RUNX1*, *TCF3-PBX1*, *MLL-MLL1* and *MLL-AFF1* fusion genes, as well as the transcriptional expression of *CRLF2* and *IKZF1* were detected as previously described (Chen et al., 2012).

2.2. Genomic Analysis and Functional Experiments

The details of next-generation sequencing, gene expression and single-nucleotide-polymorphism microarray profiling and functional experiments are provided in Supplementary methods.

2.3. Statistical Analysis

Comparisons of categorical variables were determined by Pearson's Chi-square test or Fisher's exact test. Survival was measured from the date of diagnosis of B-ALL to the date of death from any cause or to the date of last contact. The database frozen on December 3, 2015 was used for analysis. The Kaplan-Meier method was used to calculate estimates of survival probabilities, which were compared by the log-rank test. The Welch's *t*-test was used for the luciferase reporter assay, chromatin immunoprecipitation (ChIP) assay, flow cytometry assay and comparisons of mutation numbers in adults and children. Two-sided *P* values less than 0.05 were considered statistically significant. Analyses were performed with the use of SPSS 22.0 software and R3.2.2.

2.4. Acknowledgements and Role of the Funding Source

We are indebted to colleagues from SIH and Department of Hematology, Rui Jin Hospital. We are grateful to all the patients who participated in the study. We also thank Drs. Wei-Wu He and Ke-Hu Yuan (OriGene Technologies) for kindly providing the plasmids of wild-type full-length *ZNF384*, *MEF2D*, *HNRNPUL1*, *BCL9*, *EP300* and *HDAC9* cDNAs constructed into the CMV6 vector. This work was supported by Chinese National Key Basic Research Project 973 grant (2013CB966800), National High Tech Program for Biotechnology grant 863 (2012AA02A505), Ministry of Health grant (201202003), Mega-projects of Scientific Research for the 12th Five-Year Plan Grant (2013ZX09303302), National Natural Science Foundation of China grants (81123005, 81200373, 81570122), the Samuel Waxman Cancer Research Foundation Co-Principal Investigator Program, The Program for Professor of Special Appointment (Eastern Scholar) at Shanghai Institutions of Higher Learning (QD2015005) and the Program of Shanghai Subject Chief Scientist (16XD1402000). These funding sources played key supportive role for data and sample collection, experiments of genomic profiling and molecular analysis of patient samples, *in vitro* and *in vivo* studies of gene functions, bioinformatics and data analysis, as well as data interpretation. This work was also supported by Grants No. CA21765, U01GM92666, and P50 GM115279 from the National Institutes of Health, Cancer Center support grant P30 CA021765 from the US National Cancer Institute, and the American Lebanese Syrian Associated Charities, in data analyses and interpretation.

3. Results

3.1. Overview of Sequence Mutations, Copy Number Alterations and Gene Fusions

We addressed the frequencies of single nucleotide transitions and transversions of the coding regions and entire genomic regions in the whole-exome sequencing and whole-genome sequencing datasets, respectively. It was noted that in both sequencing analyses, the most common mutation types were transitions including C/G to T/A and A/T to G/C (Fig. 2a). In total, 2336 non-silent mutations within 1779 genes were confirmed in the primary leukemia samples of 203 cases (Table S2). A

tendency of the mutation load increasing with the age was observed ($P < 0.001$, Fig. 2b). These mutations consisted predominantly of missense ones ($n = 1954$; 83.6%), followed by nonsense mutations ($n = 132$; 5.7%), frameshift due to small indels ($n = 128$; 5.5%), in-frame indels ($n = 66$; 2.8%) and splice-site variants ($n = 56$; 2.4%) (Fig. 2c).

Across 203 sequenced samples, the number of non-silent coding mutations (Table S2) in each individual varied significantly (range, 0–88), with a median of 11 mutations (range, 3–88) per adult case and 9 mutations (range, 0–45) per pediatric case ($P = 0.004$, Fig. 3a). In addition to the well-established genes involved in B-ALL pathogenesis such as *PAX5* and *RAS* family members, to identify possible “driver” mutations related to leukemogenesis, we mainly focused on the recurrent ones. We also used PolyPhen2, SIFT and PROVEAN tools to predict the possible effect of amino acid changes on the normal functions of the proteins (Adzhubei et al., 2010; Choi and Chan, 2015). Indeed, 76.4% (1481/1938) of the missense mutations available for either of these tools was predicted to be deleterious (Table S2). Three hundred and twenty-five genes were recurrently mutated, 105 (32.3%) of which were not described before in B-ALL (Table S3). The most frequently mutated genes are listed in Fig. 3b, after exclusion of very large size genes with lower selection pressure (Greenman et al., 2007). Notably, mutations affecting *ZFH4*, a member of the DNA-damage-repair pathway that may interact with *CHD4* to modulate *TP53* (Chudnovsky et al., 2014), were not previously reported in leukemia.

Single nucleotide polymorphism array profiling identified genes with somatic copy number gains or losses that were recurrent across 91 adult patients and 111 pediatric patients analyzed (Table S4). Copy losses of *CDKN2A/2B* (9p21), *PAX5* (9p13) and *ETV6* (12p13) were prevalent in both adults and children, while copy gains of *RUNX1* (21q22.3) were more enriched in children and deletions of *IKZF1* (7p12.2) were more common in adults (Fig. 4).

Besides commonly observed fusions (*BCR-ABL1*, *E2A-PBX1*, *ETV6-RUNX1*, *MLL*-rearranged), RNA-seq of 78 adults and 94 children identified 41 fusions which were not reported previously, including 29 in-frame fusions and 12 out-of-frame fusions (Fig. 3c, Table S5 and Table S6). Two classes of in-frame fusions – one with *MEF2D* as the N-terminus partner and the other with *ZNF384* as the C-terminus partner – were noteworthy for their frequent occurrence in our cohort and the disruption of partner genes involved in transcriptional regulation. Another class of fusion transcript with *DUX4* and *IGH* rearrangements was also identified.

3.2. Characterization of *MEF2D* and *ZNF384* Related Gene Fusions

From the combined data of the discovery and recurrent cohorts, *MEF2D* fusions were found in 12 of 177 (6.8%) adult and 7 of 199 (3.5%) pediatric patients, with *ZNF384* fusions identified in 13 of 177

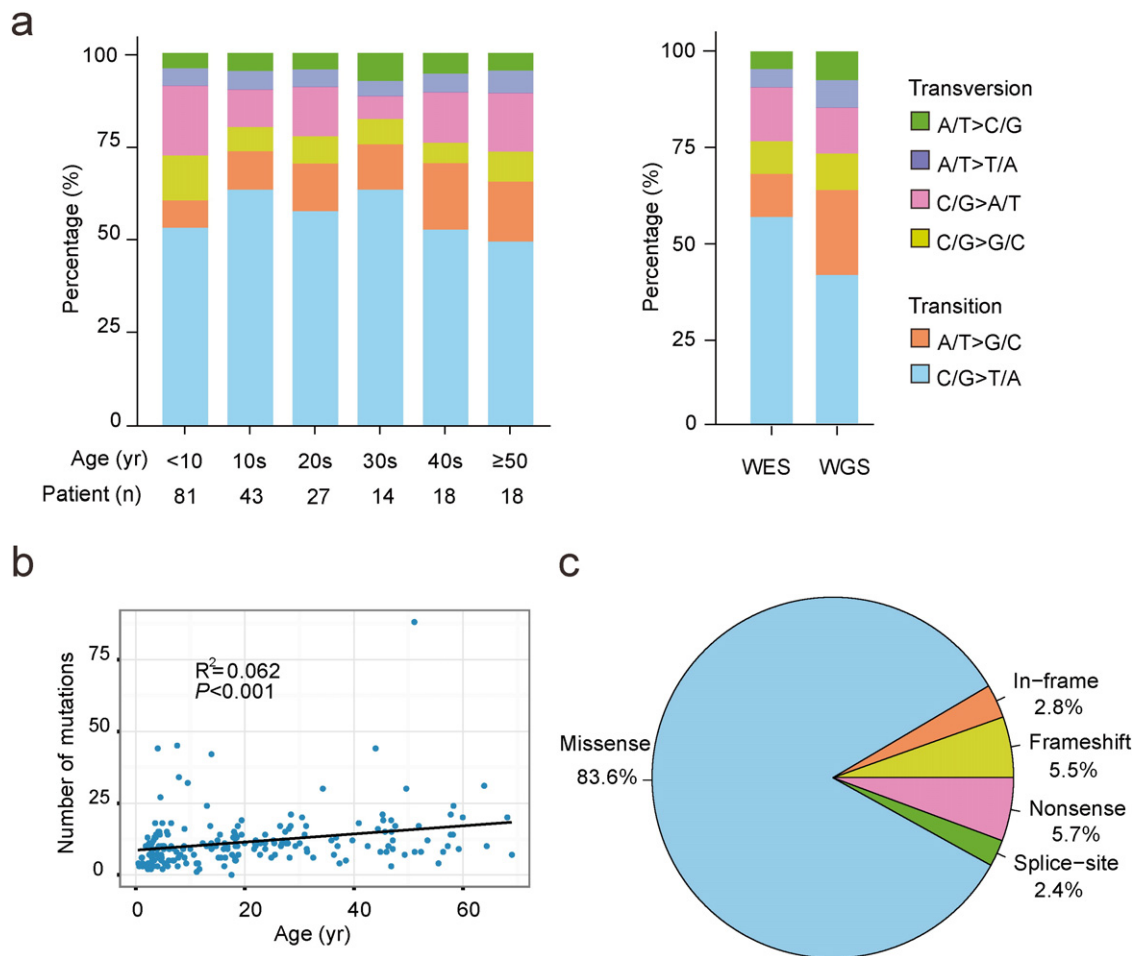


Fig. 2. Patterns of somatic non-silent mutations identified by whole-exome sequencing (WES) and whole-genome sequencing (WGS) in B-ALL. (a) The percentages of distinct transitions and transversions of all non-silent SNVs in WES and WGS. (b) Correlation of mutation burdens and the age of B-ALL patients. A linear regression model was applied to calculate R^2 and significant level, and a fitting curve was drawn to indicate the trend. (c) Proportions of non-silent mutation types according to their effects on protein coding.

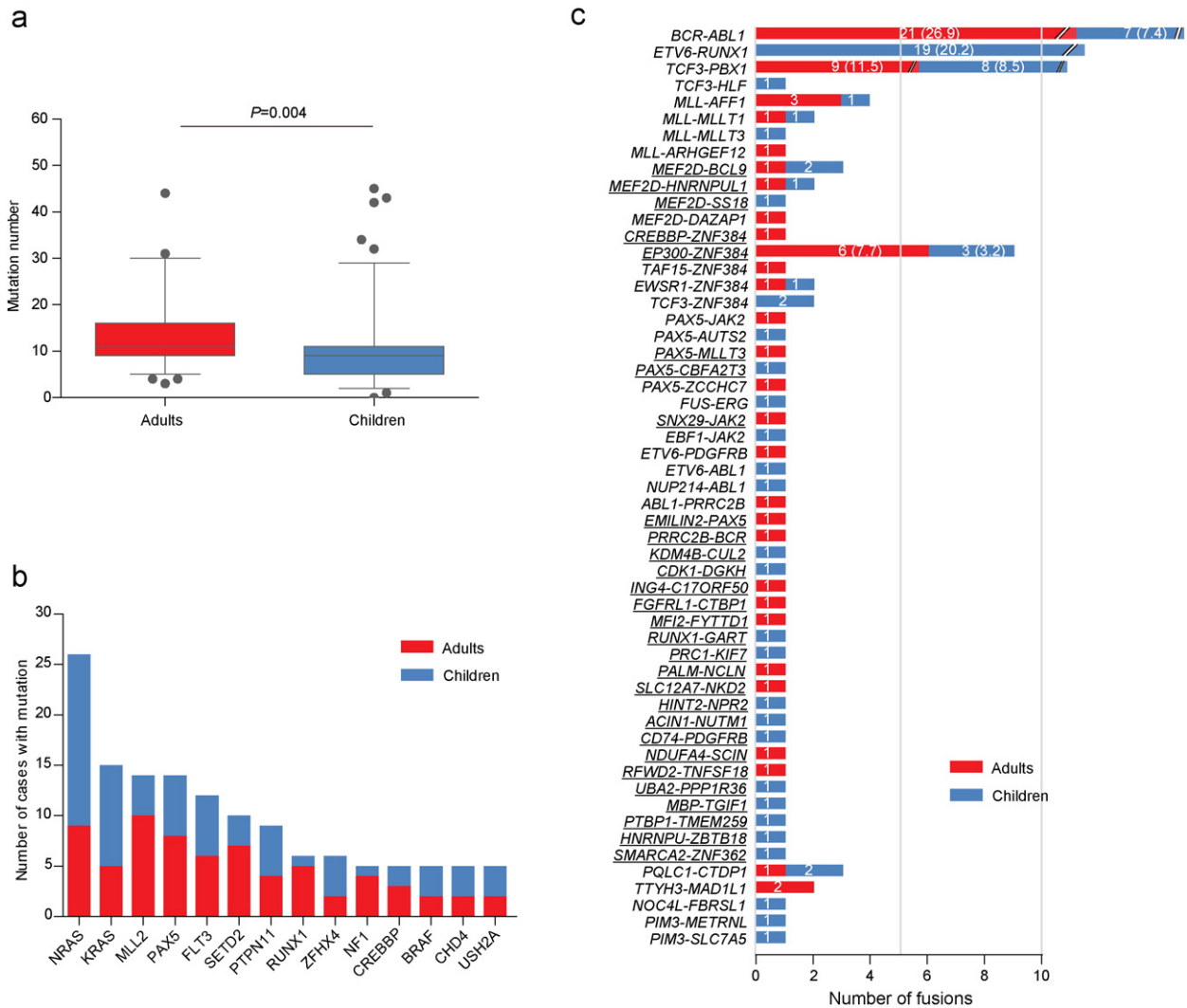


Fig. 3. Comparison of non-silent mutations identified by whole-exome and whole-genome sequencing, and gene fusions identified by RNA-seq between adult and childhood samples. (a) Box plot of the numbers of non-silent mutations detected by whole-exome and whole-genome sequencing. (b) The distribution of the most frequently mutated genes. (c) All of the in-frame fusions identified by RNA-seq. The fusion events underlined represent novel fusion genes. The numbers in the bars are the exact numbers of cases with each fusion. For more commonly identified fusion genes, frequencies are indicated in the parenthesis after the numbers.

(7.3%) adult and 8 of 199 (4.0%) pediatric patients. Pediatric patients with *MEF2D* fusions tended to be older (median 12.1 years old, $P = 0.08$). The leukemic cells with *MEF2D* fusions mostly had a pre-B immunophenotype (Table S7). Among patients with available data, the 7 children with *MEF2D* fusions had a significantly worse five-year survival than did the remaining 186 without this feature (33.3% vs. 71.2%, $P = 0.01$); a similar tendency was observed in adult B-ALL (15.6% [$n = 12$] vs. 31.3% [$n = 157$], $P = 0.08$) (Table S8).

Leukemic cells from patients with *ZNF384* fusions were more likely to be CD10-negative than were those from other patients in both the adult (15.8% vs. 5.3%, $P = 0.04$) and pediatric (18.8% vs. 2.9%, $P = 0.02$) cohorts; and to co-express myeloid-associated antigen CD13 and/or CD33 (13.1% vs. 0, $P = 0.001$ in adults; 12.1% vs. 1.7%, $P = 0.02$ in children) (Table S7). There were no significant survival differences between patients with or without *ZNF384* fusions in either the adult (38.6% [$n = 12$] vs. 29.6% [$n = 157$]) or pediatric cohort (75.0% [$n = 8$] vs. 69.4% [$n = 185$]) (Table S8). Our observation on the clinical impact of the above two fusions (Fig. 5) was in concordance with a very recent report (Yasuda et al., 2016).

We identified 3 new fusion partners of *MEF2D*, including *BCL9* (8 cases), *HNRNPUL1* (8) and *SS18* (2), and a single case with the known fusion partner *DAZAP1* (Prima et al., 2005), some of which could express different fusion transcripts with distinct exon usage (Table S7). We also

identified five partners for *ZNF384* fusions: *EP300* (14 cases), *EWSR1* (2), *TCF3* (2) and *TAF15* (2) and *CREBBP* (1) (Table S7). *CREBBP*, a histone acetyltransferase, represented a fusion partner of *ZNF384* newly identified in this study (Fig. 6a), while its homolog, *EP300*, was described recently (Gocho et al., 2015). The protein domain compositions of *MEF2D* and *ZNF384* gene fusions are depicted in Fig. 6a and Fig. S1, while the nuclear localizations of *MEF2D-HNRNPUL1*, *MEF2D-BCL9* and *EP300-ZNF384* proteins were confirmed (Fig. S2).

When co-cultured with SCF, interleukin 7 and Flt-3 ligand (Flt3L) to evaluate the effects of *MEF2D* fusions on hematopoietic development (Hirose et al., 2002), *MEF2D-HNRNPUL1* and *MEF2D-BCL9* significantly inhibited the differentiation of mouse Lin⁻ c-Kit^{low} cells into CD19⁺ B-cells in vitro (Fig. S3a). In parallel, in a retrovirus-mediated BM transplantation model, ectopic expression of *MEF2D* fusions resulted in a striking increase in the proportion of early immature B-cells (B220⁺, CD43⁺) and decrease in B220⁺, CD19⁺ subset in BM (Opferman et al., 2003) (Fig. 6b) with a concomitant decrease of B220⁺ cells in peripheral blood (Fig. 6c). These results suggested an impairment of B-cell development. Compared to wild-type *MEF2D*, fusion proteins showed markedly higher transactivating activities for *HDAC9*, a known transcriptional target of *MEF2D* (Haberland et al., 2007) (Fig. 6d). Both chimeric proteins directly interacted with the *HDAC9* promoter region as evidenced by ChIP, and this interaction was much stronger for fusion proteins

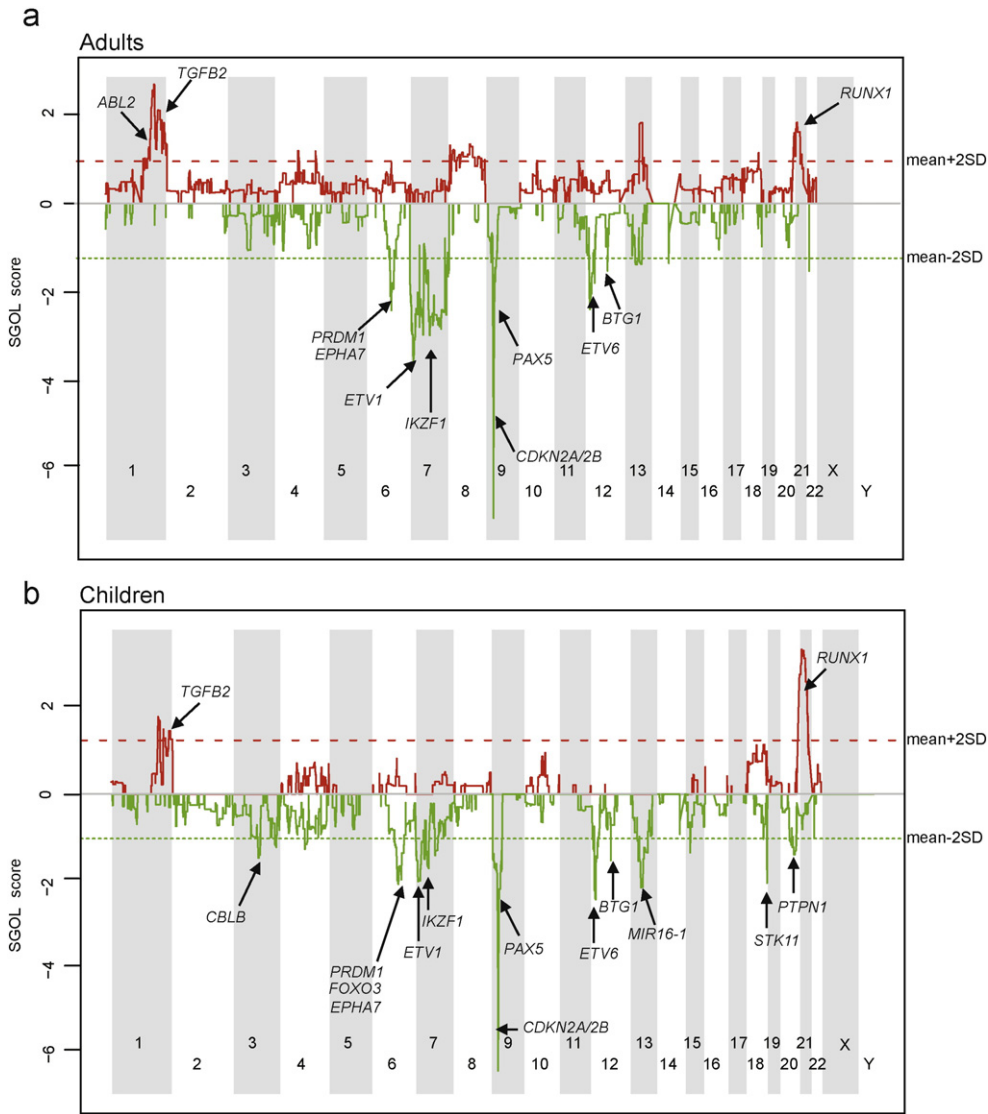


Fig. 4. Spectrum of acquired CNVs between adult (a) and childhood (b) samples. Copy number gain and loss were indicated red or green separately.

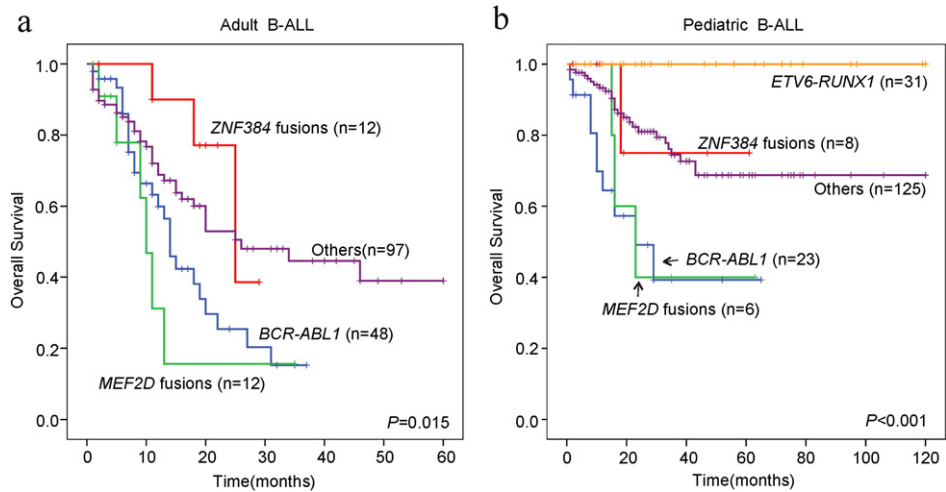


Fig. 5. Overall survival of adult and pediatric B-ALL with the fusions involving *MEF2D* and *ZNF384* genes. (a) Kaplan-Meier survival curves of adult B-ALL patients. (b) Kaplan-Meier survival curves of pediatric B-ALL patients.

compared to wild-type MEF2D (Fig. 6e). *RAG1*, an executor of V(D)J recombination, was regulated by HDAC9, which was verified in JM-1, a pre-B leukemia cell line (Fig. 6f). In fact, B-ALL blasts harboring *MEF2D* fusions showed extremely high levels of *HDAC9* transcripts and down-regulation of *RAG1* (Fig. S3b). Similarly, EP300-ZNF384 completely blocked the differentiation of mouse Lin⁻ c-Kit^{Low} cells into CD19⁺ B-cells *in vitro* (Fig. S4a), and mice transplanted with EP300-ZNF384-expressing BM cells were largely void of B-cells in peripheral blood (Fig. 6g). Notably, EP300-ZNF384 mice developed acute leukemia with a median survival of 100 days (Fig. 6h), with characteristic leukocytosis, anemia, splenomegaly (Fig. S4b, Fig. S4c) and an accumulation of blast cells in BM (Fig. 6i). The leukemic blasts were monoblastic in appearance with a significant inhibition of naphthol AS-D acetate esterase activities by sodium fluoride (Fig. 6j). Flow cytometry analysis further confirmed a striking expansion of c-Kit⁺ blasts with a severe reduction of lymphoid compartments (Fig. 6k). According to a recent redefinition of the hematopoietic hierarchy, hematopoiesis does not follow a rigid model of myeloid-lymphoid segregation and the myelomonocytic lineage is tied to lymphoid fate (Notta et al., 2016). Secondary transplantation of the EP300-ZNF384 mouse leukemic blasts led to even more pronounced disease progression (Fig. 6h).

3.3. Gene Expression Profiles and Associations with Genetic Abnormalities

Unsupervised clustering of gene expression derived from RNA-seq of 78 adult and 94 pediatric B-ALL patients (Table S9) identified 8 distinct subgroups G1–G8 (Fig. 7) showing strong associations with oncogenic gene fusion, karyotype abnormality or intra-genic deletion. Leukemia with *MEF2D* or *ZNF384* gene fusions were clustered in G1 and G5, respectively, providing strong evidence that these gene fusions are oncogenic drivers of two distinct leukemia subtypes. The associated expression signatures of these two subgroups were described below.

The G1 subgroup with *MEF2D* fusions was the closest to G2 subgroup with *TCF3-PBX1*, as both showed upregulation of genes encoding pre-B-cell receptor signaling (pre-BCR) molecules (*IGHM*, *IGLL1* and *ZAP70*) (Geng et al., 2015) and transcription factors promoting B-cell differentiation (*IRF4*, *PAX5*, *EBF1*, *BCL6* and *IKZF3*), together with the downregulation of CD34 and STAT5 pathway related genes (*ITGA6*, *CCND2* and *SOCS2*). However, G1 was separated from G2 by upregulated expression in several outlier genes, such as *HDAC9* (Haberland et al., 2008), one of the downstream target genes of *MEF2D*, and *PTPRZ1* that shape the B-cell repertoire (Cohen et al., 2012). The G5 subgroup driven by *ZNF384* fusions exhibited high expression of transcription factors such as *GATA3* (Heavey et al., 2003), *CEBPA* and *CEBPB* (Xie et al., 2004), all of which play key roles in the reprogramming of B-cells into myeloid cells (Fig. S5a). This is consistent with our observation of CD13/CD33 expression on primary B-ALL blasts with *ZNF384* fusions (Table S7). Strikingly, G5 had few mutations or deletions of genes involved in B-cell development (e.g. *IKZF1*, *PAX5*, *RUNX1*, *ETV6*) or the cell cycle (e.g. *CDKN2A/2B*). The JAK-STAT signaling pathway was upregulated in leukemic cells with *ZNF384* fusions, in contrast to its downregulation in leukemic cells with *MEF2D* fusions, which may rely on pre-BCR signaling for survival (Fig. S5b).

In other subgroups including G3, G4, G6, G7 and G8, the dominant genetic subtypes were *ETV6-RUNX1*, *ERG* deletion, hyperdiploidy, *MLL*-rearranged, *BCR-ABL1* or *BCR-ABL1*-like, respectively. The characteristic expression profiles were summarized in Supplementary Results. Of note, all cases of G4 subgroup (dominated by *ERG* deletion) were found to exhibit over-expression of *DUX4* gene mostly due to the *DUX4-IGH* fusion as described by a very recent work (Yasuda et al., 2016), suggesting a possible cooperative relationship between the two events.

3.4. Comparison of Mutation Frequencies Between Adult and Pediatric Cases

To explore the possible genetic differences between adults and children, we first compared the total number of non-silent mutations in the

discovery cohort and found that it was significantly higher in adults than in children (Fig. 3a), a finding that was also consistent across gene expression cluster subgroups, especially G4, G7 and G8 (Fig. S6). The combined analysis of the findings of SNV or indels from the discovery and the recurrent cohorts (Table S10) revealed an interesting trend of higher frequencies of *IKZF1* ($P = 0.03$), *MLL2* ($P = 0.03$), *JAK3* ($P = 0.047$) mutations in adults, and *PTPN11* ($P = 0.03$) mutations in children (Table S11). Hyperdiploidy was found in 6 of our 14 pediatric patients with *PTPN11* mutations but in only 36 of the other 204 without such mutations ($P = 0.03$). Of our 4 adult patients with *PTPN11* mutations, 3 had pseudodiploidy and one hypodiploidy. Moreover, epigenetic modifiers involved in histone or DNA methylation and demethylation (*MLL2*, *SETD2*, *ASXL1*, *EZH2*, *TET2* and *KDM5C*) were mutated more often in adults than in children (22.9% vs. 12.7%, $P = 0.009$, Table S11).

3.5. Comparisons of Clinical Relevance and Cooperating Mutations in Different Gene Expression Subgroups

We then compared the clinical relevance of the gene expression subgroups in the combined group. Patients in the recurrent cohort were assigned to each of the eight gene expression subgroups according to the dominant genetic subtypes. Overall, adult group had a worse five-year survival estimate than childhood one (Table S12, Table S13). With regard to gene expression cluster subgroups, G3 (*ETV6-RUNX1*) and G6 (hyperdiploid) were seen mostly in children and with quite favorable prognosis. In each of the remaining subgroups, adult cases tended to have inferior outcome to pediatric counterparts. We then examined, in subgroups G1, G2, G4, G5, G7 and G8, the proportions of adult versus pediatric patients with gene alterations combining the data of SNV or indels and CNVs in each of the six significantly mutated pathways – B-cell development, cell cycle, epigenetic modifiers, RAS signaling, JAK signaling and other signaling. The genes involved in the pathway analysis were listed in Table S14. We found that pediatric patients ($n = 10$) in G5 (*ZNF384* fusion) lacked alterations of B-cell development genes as compared to adult patients (42.9%, $n = 14$, $P = 0.02$). Adult patients in G8 (*BCR-ABL1* or *BCR-ABL1*-like) had significantly higher rates of somatic alterations in B-cell development genes (78.6%, $n = 56$) than did pediatric patients (53.6%, $n = 28$, $P = 0.02$), but lower frequencies of alterations affecting signaling genes other than those in the RAS or JAK pathway (5.4% vs. 21.4%, $P = 0.05$). For patients in G7 (*MLL*-rearranged), adults had more mutations in epigenetic modifiers compared to pediatric cases (42.9% of 28 vs. 11.8% of 17, $P = 0.046$). In this regard, adults had more frequent alterations in epigenetic modifiers in all subgroups except G1 (Fig. 8a, Fig. S7). When all patients were combined, genes involved in B-cell development and epigenetics were mutated more frequently in adults than in children ($P = 0.01$ and 0.03, respectively, Fig. 8b).

4. Discussion

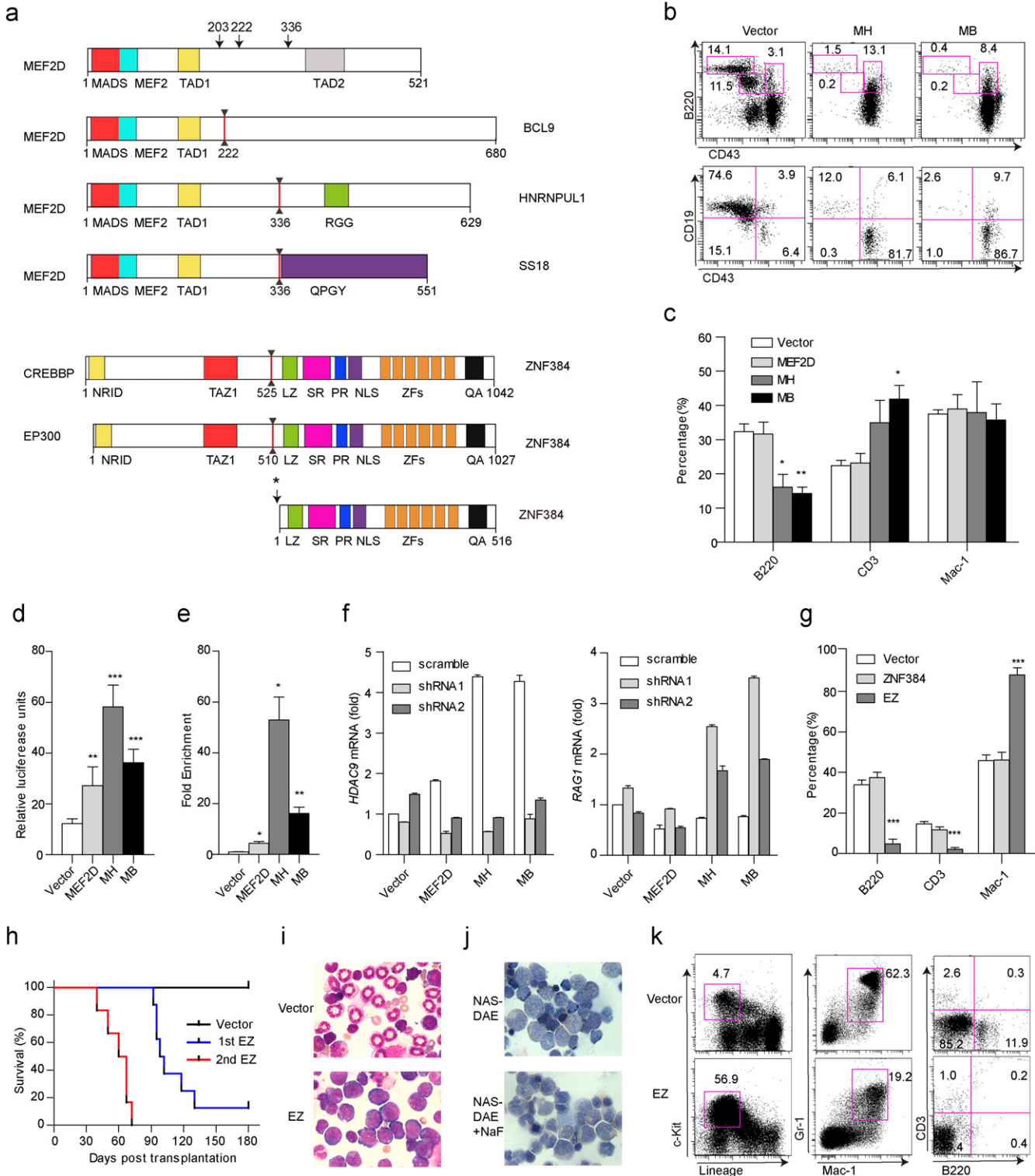
In this comprehensive study on the genetic landscape of B-ALL, we found 105 genes with recurrent sequence mutations and identified 41 gene fusions including 29 in-frame ones which had not been reported previously. Particularly, RNA-seq analysis allowed us to identify 8 gene expression subgroups which demonstrated strong correlation with major gene fusions or genomic copy number abnormalities. These subgroups also showed certain relationship to the immunophenotypes.

Of note, *MEF2D* and *ZNF384* fusions were highly recurrent affecting 6.7% and 7.3% of adults, 3.4% and 3.9% of pediatric patients, respectively. Leukemia cases with these two fusions comprised distinct gene expression subgroups (G1 and G5), suggesting that the fusions are oncogenic drivers. Indeed, functional studies confirmed that *MEF2D-BCL9*, *MEF2D-HNRNPUL1* and *EP300-ZNF384* fusion genes profoundly disrupted B-cell development *in vivo* and *in vitro*. Interestingly, *MEF2D* fusions up-regulated *HDAC9*, a class II histone deacetylase, which in turn could cooperate with the fusion transcription factors in the

repression of genes essential for B-lineage differentiation (*RAG1*). Moreover, *EP300-ZNF384* fusion alone rapidly gave rise to overt acute leukemia. In our series, *EP300-ZNF384* fusion conferred an intermediate prognosis and the gene set enrichment analysis showed significant up-regulation of JAK-STAT pathway, suggestive of a potential benefit from treatment with inhibitors of this pathway. By contrast, *MEF2D* fusions appeared to be associated with a poor prognosis in both adults and children. It may be worthwhile to test histone deacetylase inhibitors in patients with this genotype. Moreover, we discovered a close association

between *ERG* deletion and *DUX4* overexpression in G4 subgroup, and future therapies may benefit from a deeper understanding of this cooperative mechanism of the two genetic defects.

While analyzing recurrent sequence abnormalities in adult and pediatric B-ALL groups, we found that adult patients had more gene mutations, especially *IKZF1*, *MLL2*, and *JAK3*, but fewer alterations of *PTPN11* than did pediatric patients. A number of sequence mutations could cooperate with gene fusions or aberrant expression patterns in disease mechanisms and could exert effect on distinct clinical outcomes



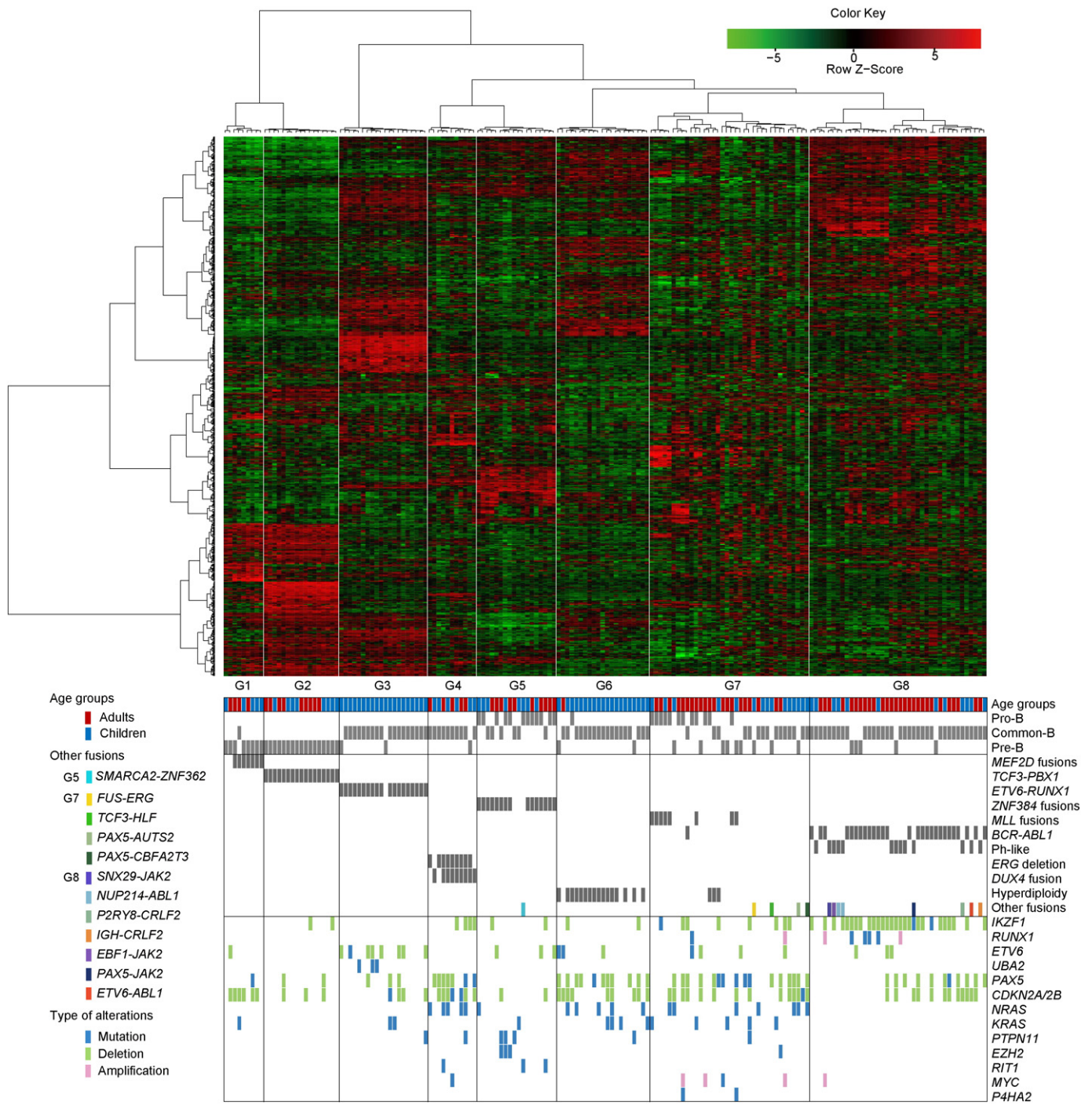


Fig. 7. Unsupervised hierarchical clustering identified specific subgroups of patients with shared gene expression patterns. Columns indicate ALL patients and rows are genes. The bottom panels show immunophenotype and genotype for each sample as well as significantly altered genes (Fisher's exact $P < 0.05$) within each of the eight unique gene expression subgroups. The immunophenotypes were determined according to the recommendation of European Group for the Immunological Characterization of Leukemias (EGIL).

Fig. 6. Schema of the wild-type and fusion proteins involving MEF2D and ZNF384 and results of functional studies. (a) Structural and functional domains of wild-type proteins and the most frequently identified fusion proteins. Arrows indicate breakpoints of the wild-type proteins. * labeled beside ZNF384 indicated two fusion points upstream of the coding region of ZNF384 (5 bp or 65 bp). (b) Representative flow cytometry results of the B-cell population of GFP⁺ bone marrow (BM) cells in vector control (Vector), MEF2D-HNRPUL1 (MH) and MEF2D-BCL9 (MB) mice. The upper panel showed different B-cell subsets in total GFP⁺ cells (B220 vs. CD43), and bottom panel showed subsets in B220⁺ B-cell fraction (CD19 vs. CD43). (c) The percentages of B220, CD3 and Mac-1 in GFP⁺ peripheral blood (PB) cells in Vector, MEF2D, MH and MB mice. * and ** denote differences between MH and MB with Vector. (d) Responsiveness of the HDAC9 promoter to wild-type MEF2D and MEF2D fusions. 293T cells were cotransfected with a pGL4.15-Luc reporter containing the promoter region of HDAC9 and wild-type MEF2D or MEF2D fusions. Compared to wild-type MEF2D, MEF2D fusions displayed stronger transcriptional activity (MH vs. MEF2D, $P < 0.001$; MB vs. MEF2D, $P = 0.03$). (e) ChIP assays revealed MEF2D fusions had enhanced binding activity of HDAC9 promoter in 293T cells (MH vs. MEF2D, $P = 0.01$; MB vs. MEF2D, $P = 0.02$). ChIP DNA, immunoprecipitated with an anti-Myc tag antibody or goat IgG, was quantified with primers flanking HDAC9 promoter. (f) Cotransfection of wild-type MEF2D or MEF2D fusions plasmids and HDAC9 shRNA in JM1. Transfection of MEF2D fusions led to the upregulation of HDAC9 and the downregulation of RAG1, by contrast, reduced expression of HDAC9 mRNA level with transient transfection of HDAC9 shRNA caused remarkable rebound of RAG1 expression. (g) Flow cytometry analysis of different lineage markers of GFP⁺ PB in Vector, ZNF384, EP300-ZNF384 (EZ) four weeks after transplantation. (h) Kaplan-Meier survival curves of EZ mice (1st transplantation, $n = 8$; 2nd transplantation, $n = 6$). (i) Wright's staining of BM cytopsin samples from control and EZ mice. (j) Naphthol AS-D acetate esterase (NAS-DAE) staining (up) and inhibition of NAS-DAE staining by sodium fluoride (NaF) (bottom) of the BM of EZ mice. (k) Flow cytometry analysis of the BM of EZ mice versus vector control. * $P < 0.05$; ** $P < 0.01$; *** $P < 0.001$.

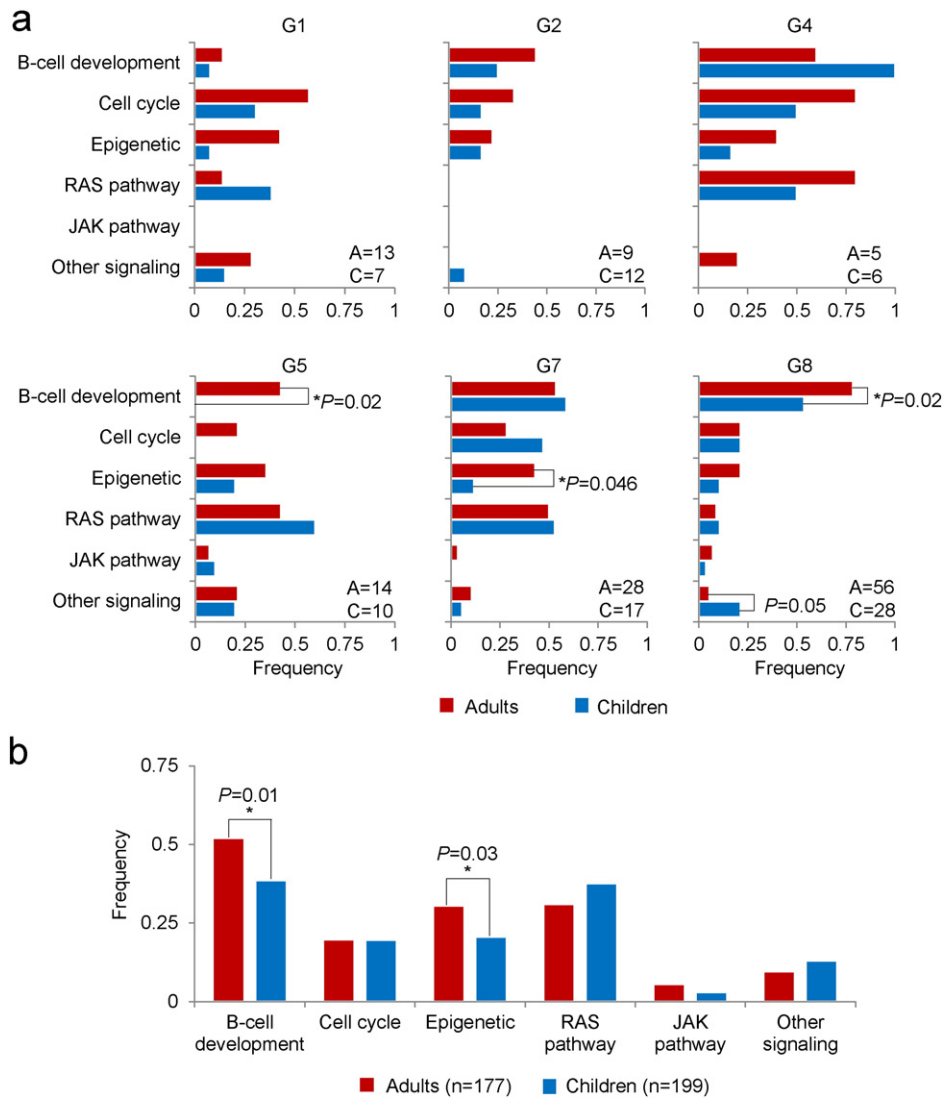


Fig. 8. Comparisons between adult and pediatric B-ALL patients with regard to gene pathways. (a) Comparisons within each cluster subgroup excluding G3 and G6. A: number of adults; P: number of children. (b) Comparison of all the patients.

between different age groups. For example, the strong association of *IKZF1* alteration with Philadelphia chromosome-positive and Philadelphia chromosome-like B-ALL and *PTPN11* mutation with hyperdiploid B-ALL suggests that the enrichment for *IKZF1* in adults, and *PTPN11* in children, is associated with biases in leukemia subtype distribution between the two age groups. On the other hand, our findings of more mutations within each gene expression cluster subgroup, and the enrichment of genetic alterations affecting B-cell development and epigenetic modifier genes in adult B-ALL as compared to pediatric patients in several subgroups, suggest increased complexity of pathway involvements and different target cells for malignant transformation with increasing age, which may partly, albeit not fully, explain the dismal prognosis in adults with B-ALL. Recent studies of clonal evolution of pediatric (Ma et al., 2015; Mullighan et al., 2011) and adult (Xiao et al., 2016) ALL found enrichment of mutations in epigenetic regulators from diagnosis to relapse. It is thus possible that profound chromatin changes owing to more frequent epigenetic modifier abnormalities in adult B-ALL cells confer greater drug-resistance.

We would stress that even though the total number of patients in our study is relatively large, the number in each subgroup varies, compromising our ability to generate statistically significant results in some analyses. Second, patients are arbitrarily divided into adult and

pediatric age groups using 18 years as the threshold, even though young children, adolescents, young adults and older adults differ in their responses to treatment. Further comparisons with larger sample sizes would allow better discrimination among a broader spectrum of age groups. Third, patients 17 to 18 years mostly receive adult treatment in China. New trials based on pediatric-type protocols would be expected to further improve outcome in these patients in the near future. In spite of the above limitations, our genomic profiling of adult and pediatric B-ALL provides useful clues to precise molecular classifications of this heterogeneous disease group and to identify new molecular vulnerabilities that could be exploited in the design of more effective targeted therapies.

Contributions

S.-J.C. and Z.C. were the principal investigators of the study. S.-J.C., Z.C., J.-Y.T., C.-H.P., J.-Q.M and J.-Y.H. coordinated and oversaw the study. Y.-F.L., B.-Y.W. and W.-N.Z. collected samples, performed data analyses and most of the experiments; J.-Y.H. led and performed the bioinformatics and data analyses; J.-F.L., M.-J.W., Z.-G.Z., and J.K. participated in bioinformatics analyses; K.-K.W. provided helps on bioinformatics analyses; M.Z., Q.W., Y.-M.Z., and L.J. participated in sample

processing, PCR and targeted deep sequencing. B.C. performed cytogenetic analyses; X.-Q.W. helped in flow cytometry analyses; B.-S.L., J.-Y.T., J.C., L.-J.G., J.-M.L., J.W., J.-Q.M., Z.-X.S., B.C., X.-J.Y., Y.L., X.-Q.C., Y.-M.L., L.L., W.-G.Z., J.-S.Y. and J.-D.H. participated in sample collection and/or treatment to patients. S.-Y.W. and L.J. helped to perform next generation sequencing. M.Z. and Y.-J.D. helped in functional experiments. C.-H.P., J.Z., J.-J.Y., Y.L., G.W., X.Z., D.P. and C.C. of St. Jude Children's Research Hospital helped in data analyses and interpretation. Z.C., S.-J.C., C.-H.P., J.-Q.M., J.-Y.H., Y.-F.L., B.-Y.W. and W.-N.Z. wrote the manuscript.

Declaration of Interests

The authors declare no conflicts of interests.

Appendix A. Supplementary Data

Supplementary data to this article can be found online at <http://dx.doi.org/10.1016/j.ebiom.2016.04.038>.

References

- Adzhubei, I.A., Schmidt, S., Peshkin, L., et al., 2010. A method and server for predicting damaging missense mutations. *Nat. Methods* 7, 248–249.
- Andersson, A.K., Ma, J., Wang, J., et al., 2015. The landscape of somatic mutations in infant MLL-rearranged acute lymphoblastic leukemias. *Nat. Genet.* 47, 330–337.
- Bassan, R., Hoelzer, D., 2011. Modern therapy of acute lymphoblastic leukemia. *J. Clin. Oncol.* 29, 532–543.
- Chalandon, Y., Thomas, X., Hayette, S., et al., 2015. Randomized study of reduced-intensity chemotherapy combined with imatinib in adults with Ph-positive acute lymphoblastic leukemia. *Blood* 125, 3711–3719.
- Chen, B., Wang, Y.-Y., Shen, Y., et al., 2012. Newly diagnosed acute lymphoblastic leukemia in China (I): abnormal genetic patterns in 1346 childhood and adult cases and their comparison with the reports from Western countries. *Leukemia* 26, 1608–1616.
- Choi, Y., Chan, A.P., 2015. PROVEAN web server: a tool to predict the functional effect of amino acid substitutions and indels. *Bioinformatics* 31, 2745–2747.
- Chudnovsky, Y., Kim, D., Zheng, S., et al., 2014. ZFX4 interacts with the NuRD core member CHD4 and regulates the glioblastoma tumor-initiating cell state. *Cell Rep.* 6, 313–324.
- Cohen, S., Shoshana, O.-y., Zelman-Toister, E., et al., 2012. The cytokine midline and its receptor RPTP ζ regulate B cell survival in a pathway induced by CD74. *J. Immunol.* 188, 259–269.
- Den Boer, M.L., van Slegtenhorst, M., De Menezes, R.X., et al., 2009. A subtype of childhood acute lymphoblastic leukaemia with poor treatment outcome: a genome-wide classification study. *Lancet Oncol.* 10, 125–134.
- Gabriel, A.S., Lafta, F.M., Schwalbe, E.C., et al., 2015. Epigenetic landscape correlates with genetic subtype but does not predict outcome in childhood acute lymphoblastic leukemia. *Epigenetics* 10, 717–726.
- Geng, H., Hurtz, C., Lenz, K.B., et al., 2015. Self-enforcing feedback activation between BCL6 and pre-B cell receptor signaling defines a distinct subtype of acute lymphoblastic leukemia. *Cancer Cell* 27, 409–425.
- Gocho, Y., Kiyokawa, N., Ichikawa, H., et al., 2015. A novel recurrent EP300-ZNF384 gene fusion in B-cell precursor acute lymphoblastic leukemia. *Leukemia* 29, 2445–2448.
- Greenman, C., Stephens, P., Smith, R., et al., 2007. Patterns of somatic mutation in human cancer genomes. *Nature* 446, 153–158.
- Haberland, M., Arnold, M.A., McAnally, J., Phan, D., Kim, Y., Olson, E.N., 2007. Regulation of HDAC9 gene expression by MEF2 establishes a negative-feedback loop in the transcriptional circuitry of muscle differentiation. *Mol. Cell Biol.* 27, 518–525.
- Haberland, M., Montgomery, R.L., Olson, E.N., 2008. The many roles of histone deacetylases in development and physiology: implications for disease and therapy. *Nat. Rev. Genet.* 10, 32–42.
- Heavey, B., Charalambous, C., Cobaleda, C., Busslinger, M., 2003. Myeloid lineage switch of Pax5 mutant but not wild-type B cell progenitors by C/EBP α and GATA factors. *EMBO J.* 22, 3887–3897.
- Hirose, J., Kouro, T., Igarashi, H., Yokota, T., Sakaguchi, N., Kincade, P.W., 2002. A developing picture of lymphopoiesis in bone marrow. *Immunol. Rev.* 189, 28–40.
- Jabbar, E., O'Brien, S., Konopleva, M., Kantarjian, H., 2015. New insights into the pathophysiology and therapy of adult acute lymphoblastic leukemia. *Cancer* 121, 2517–2528.
- Ma, X., Edmonson, M., Yergeau, D., et al., 2015. Rise and fall of subclones from diagnosis to relapse in pediatric B-acute lymphoblastic leukaemia. *Nat. Commun.* 6, 6604.
- Mi, J.Q., Wang, X., Yao, Y., et al., 2012. Newly diagnosed acute lymphoblastic leukemia in China (II): prognosis related to genetic abnormalities in a series of 1091 cases. *Leukemia* 26, 1507–1516.
- Mullighan, C.G., Goorha, S., Radtke, L., et al., 2007. Genome-wide analysis of genetic alterations in acute lymphoblastic leukaemia. *Nature* 446, 758–764.
- Mullighan, C.G., Zhang, J., Kasper, L.H., et al., 2011. CREBBP mutations in relapsed acute lymphoblastic leukaemia. *Nature* 471, 235–239.
- Notta, F., Zandi, S., Takayama, N., et al., 2016. Distinct routes of lineage development reshape the human blood hierarchy across ontogeny. *Science* 351, aab2116.
- Opferman, J.T., Letai, A., Beard, C., Sorcinelli, M.D., Ong, C.C., Korsmeyer, S.J., 2003. Development and maintenance of B and T lymphocytes requires antiapoptotic MCL-1. *Nature* 426, 671–676.
- Prima, V., Gore, L., Cairns, A., et al., 2005. Cloning and functional characterization of MEF2D/DAZAP1 and DAZAP1/MEF2D fusion proteins created by a variant t(1;19)(q23;p13.3) in acute lymphoblastic leukemia. *Leukemia* 19, 806–813.
- Pui, C.-H., 2010. PART XI: NEOPLASTIC LYMPHOID DISEASES, 93. Acute lymphoblastic leukemia. In: Kenneth Kaushansky, M.L., Beutler, E., Kipps, T., Prchal, J., Seligsohn, U. (Eds.), *Williams Hematology*. 2460, eighth ed. McGraw-Hill Education, USA (2010-07-09).
- Pui, C.-H., Yang, J.J., Hunger, S.P., et al., 2015. Childhood acute lymphoblastic leukemia: progress through collaboration. *J. Clin. Oncol.* 33, 2938–2948.
- Roberts, K.G., Morin, R.D., Zhang, J., et al., 2012. Genetic alterations activating kinase and cytokine receptor signaling in high-risk acute lymphoblastic leukemia. *Cancer Cell* 22, 153–166.
- Roberts, K.G., Li, Y., Payne-Turner, D., et al., 2014. Targetable kinase-activating lesions in Ph-like acute lymphoblastic leukemia. *N. Engl. J. Med.* 371, 1005–1015.
- Schultz, K.R., Carroll, A., Heerema, N.A., et al., 2014. Long-term follow-up of imatinib in pediatric Philadelphia chromosome-positive acute lymphoblastic leukemia: Children's Oncology Group study AALL0031. *Leukemia* 28, 1467–1471.
- Xiao, H., Wang, L.M., Luo, Y., et al., 2016. Mutations in epigenetic regulators are involved in acute lymphoblastic leukemia relapse following allogeneic hematopoietic stem cell transplantation. *Oncotarget* 7, 2696–2708.
- Xie, H., Ye, M., Feng, R., Graf, T., 2004. Stepwise reprogramming of B cells into macrophages. *Cell* 117, 663–676.
- Yasuda, T., Tsuzuki, S., Kawazu, M., et al., 2016. Recurrent DUX4 fusions in B cell acute lymphoblastic leukemia of adolescents and young adults. *Nat. Genet.* 48, 569–574.