Leukemia Research xxx (2015) xxx–xxx

Contents lists available at ScienceDirect

Leukemia Research

journal homepage: www.elsevier.com/locate/leukres

Next-generation-sequencing of recurrent childhood high hyperdiploid acute lymphoblastic leukemia reveals mutations typically associated with high risk patients

Cai Chen[a,b], Christoph Bartenhagen[c], Michael Gombert[a], Vera Okpanyi[a], Vera Binder[a], Silja Röttgers[d], Jutta Bradtke[d], Andrea Teigler-Schlegel[d], Jochen Harbott[d], Sebastian Ginzel[a,e], Ralf Thiele[e], Peter Husemann[a], Pina F.I. Krell[a], Arndt Borkhardt[a,1], Martin Dugas[c,1], Jianda Hu[b,**,1], Ute Fischer[a,*,1]

[a] Department of Pediatric Oncology, Hematology and Clinical Immunology, University Children's Hospital, Medical Faculty, Heinrich Heine University, Düsseldorf, Germany
[b] Fujian Institute of Hematology, Fujian Medical University Union Hospital, Fuzhou, China
[c] Institute of Medical Informatics, University of Münster, Münster, Germany
[d] Oncogenetic Laboratory, Department of Pediatric Hematology and Oncology, Justus Liebig University, Giessen, Germany
[e] Department of Computer Science, Bonn-Rhine-Sieg University of Applied Sciences, Sankt Augustin, Germany

ARTICLE INFO

Article history:
Received 14 April 2015
Received in revised form 8 June 2015
Accepted 10 June 2015
Available online xxx

Keywords:
Acute lymphoblastic leukemia
High hyperdiploidy
Relapse
RAS
CREBBP
Ikaros
TP53

ABSTRACT

20% of children suffering from high hyperdiploid acute lymphoblastic leukemia develop recurrent disease. The molecular mechanisms are largely unknown. Here, we analyzed the genetic landscape of five patients at relapse, who developed recurrent disease without prior high-risk indication using whole-exome- and whole-genome-sequencing. Oncogenic mutations of RAS pathway genes (*NRAS*, *KRAS*, *FLT3*, $n = 4$) and deactivating mutations of major epigenetic regulators (*CREBBP*, *EP300*, each $n = 2$ and *ARID4B*, *EZH2*, *MACROD2*, *MLL2*, each $n = 1$) were prominent in these cases and virtually absent in non-recurrent cases ($n = 6$) or other pediatric acute lymphoblastic leukemia cases ($n = 18$).

In relapse nucleotide variations were detected in cell fate determining transcription factors (*GLIS1*, *AKNA*). Structural genomic alterations affected genes regulating B-cell development (*IKZF1*, *PBX1*, *RUNX1*). Eleven novel translocations involved the genes *ART4*, *C12orf60*, *MACROD2*, *TBL1XR1*, *LRRN4*, *KIAA1467*, and *ELMO1/MIR1200*. Typically, patients harbored only single structural variations, except for one patient who displayed massive rearrangements in the context of a germline tumor suppressor *TP53* mutation and a Li-Fraumeni syndrome-like family history. Another patient harbored a germline mutation in the DNA repair factor *ATM*. In summary, the relapse patients of our cohort were characterized by somatic mutations affecting the RAS pathway, epigenetic and developmental programs and germline mutations in DNA repair pathways.

Abbreviations: ALL, acute lymphoblastic leukemia; AML, acute myeloid leukemia; BWA, Burrows Wheeler Alignment tool; CGH, comparative genomic hybridization; CNV, copy number variations; DGV, database of genomic variants; EGA, European Genome-phenome Archive; FISH, fluorescence in situ hybridization; GASV, geometric analysis of structural variants software; indels, insertions or deletions; lincRNA, large intergenic non-coding RNA; LTR, long terminal repeat; SNP, single nucleotide polymorphism; SNV, somatic nucleotide variation; RAG, recombination-activating gene; RSS, recombination signal sequences.

* Corresponding author at: Department of Pediatric Oncology, Hematology and Clinical Immunology, Center for Child and Adolescent Health, Heinrich Heine University, Düsseldorf, Moorenstr. 5, 40225 Düsseldorf, Germany. Tel.: +49 211 811 6568; fax: +49 211 811 6707.
** Corresponding author at: Fujian Institute of Hematology, Fujian Provincial Key Laboratory of Hematology, Fujian Medical University Union Hospital, 29 Xinquan Road, Fuzhou 350000, Fujian, China. Tel.: +86 591 83357896; fax: +86 591 83324116.
E-mail addresses: jdhu@medmail.com.cn (J. Hu), ute.fischer@med.uni-duesseldorf.de (U. Fischer).
[1] These authors contributed equally to this work.

http://dx.doi.org/10.1016/j.leukres.2015.06.005
0145-2126/

1. Introduction

High hyperdiploidy (51–67 chromosomes) is the most frequent numerical cytogenetic alteration in pediatric B-cell precursor acute lymphoblastic leukemia (ALL), occurring in 25–30% of patients. It is characterized by nonrandom gains of chromosomes X, 4, 6, 10, 14, 17, 18, or 21 [1]. In general, children suffering from high hyperdiploid ALL have a favorable prognosis, nevertheless 15–20% relapse [1,2].

High hyperdiploid preleukemic cell clones arise *in utero* [3–5], but additional cooperating oncogenic lesions are required for the development of overt leukemia. Employing cytogenetic, fluorescence in situ hybridization (FISH)-, comparative genomic hybridization (CGH)- or single nucleotide polymorphism (SNP)-analyses, mutations in the receptor tyrosine kinase-RAS-signaling pathway, microdeletions in the genes *IKZF1*, *CDKN2A*, *PAX5*, *ETV6*, *RB1*, *TCF3* or unbalanced translocations as well as gains of chromosome 1q, deletions of 6q and isochromosomes of 7q and 17q, have been observed in some cases [1,2]. Rare cases (approximately 1–3%) harbor typical leukemia-associated translocations, including t(1;19)(q23;p13)/*TCF3-PBX1*, t(9;22)(q34;q11)/*BCR-ABL1*, t(12;21)(p13;q21)/*ETV6-RUNX1* or gene fusions involving 11q23/*MLL* [2,6,7]. Mutations of the histone acetyltransferase *CREBBP* have been suggested to be associated with recurrent ALL and high hyperdiploid ALL in particular [8,9]. Currently, no other molecular markers are known that are associated with poor outcome in high hyperdiploid ALL [10,11].

We chose a small cohort of patients with recurrent disease who were initially stratified into a low risk group. They were 2–5 years of age at diagnosis, had no known adverse translocations, a low number of blasts at diagnosis, a good response to induction therapy and achieved remission after less than 30 days of therapy. We performed low coverage whole-genome sequencing to analyze structural variations and high coverage whole-exome sequencing to identify single nucleotide variations in parallel for all five patients at relapse and remission and (depending on availability) also in the corresponding diagnosis samples (patient 4 and 5, see Supplemental Table 1 for an overview) to gain insight into the mechanisms underlying the unpredicted recurrence of the disease in low risk cases. We validated major findings in validation cohorts comprising non-recurrent high hyperdiploid ALL, as well as other recurrent and non-recurrent pediatric ALL cases.

2. Materials and methods

2.1. Patients and controls

DNA and RNA were isolated from bone marrow of five pediatric patients with recurrent high hyperdiploid ALL (Table 1) and validation cohorts (non-recurrent high hyperdiploid ALL ($n = 6$), recurrent *ETV6-RUNX1*-positive ALL ($n = 7$), non-recurrent Down-syndrome associated ($n = 6$) and *TCF3-PBX1*-rearranged ($n = 5$) ALL) at diagnosis, remission and/or relapse (blast content >70%). The local ethics committee approved the research – written informed consent was given by all parents. High hyperdiploidy was confirmed by cytogenetic karyotype analysis.

2.2. Library preparation, whole-genome and whole-exome sequencing

Whole-genome sequencing was performed at a low coverage that is sufficient to detect copy number variations and structural alterations such as translocations. Higher coverage is needed for the identification of single nucleotide variations and indels in coding genes. These were analyzed in depth after enrichment of the exonic gene regions (about 2% of the genome) and sequencing of those with high coverage (84% of nucleotides were at least 30 fold covered). Four of the patients (patients 1–3, 5) were included as a validation cohort in a previous publication [12]. There, loss of heterozygosity was analyzed in a near haploid cell line masking as high hyperdiploid and compared to the patients' high hyperdiploid cells. Here, we analyzed the generated whole-genome and whole-exome sequencing data with respect to structural and nucleotide variations and indels, and included another patient (patient 4) to complete a set of five patients. Briefly, exome capture was performed using the SeqCap EZ Exome Library 2.0 kit (Roche/Nimblegen, Madison, WI) and 100 bp single-read sequencing was performed on a HiSeq2000 (Illumina). For whole-genome sequencing libraries of 350 bp fragments were generated from sheared genomic DNA by PCR amplification after ligation of sequencing adapters (Illumina, San Diego, CA). Paired-end sequencing was performed using a Genome Analyzer IIx or a HiSeq2000 (Illumina). Sequencing statistics are provided in Supplemental Table 1.

2.3. Bioinformatic data analysis

Data were analyzed as previously reported and detailed in the online supplementary material and methods section [12]. Briefly, paired-end sequencing data was aligned against the reference genome hg19 (GRCh 37) using the Burrows Wheeler Alignment tool (BWA). Unique reads served as input for the software GASV (Geometric Analysis of Structural Variants) used for identification of structural variants. Somatic variations not detected in the database of genomic variants (DGV) were reported. Copy number variations were analyzed using the tool FREEC. Exome sequencing data was processed employing BWA, Picardtools, and Samtools. Variation calls were obtained employing the datasets GATK, HapMap, Omni Array and dbSNP134 provided by The Broad Institute (Cambridge, MA). Calls were annotated using the variant effect predictor, NGS-SNP, and the Ensembl database. The algorithms SIFT and PolyPhen-2 were used to analyze the impact of a mutation on protein structure or function as predicted by Ensembl (v70). SNPs were imported into an SQL database. Somatic nucleotide variations (SNV) and indels identified using the tools VarScan2 and MuTect (with an allele frequency $\leq$1% in the data sets of the HapMap and the 1000 Genomes project) were retained as potentially tumorigenesis-associated. Sequencing data were deposited at the European Genome-phenome Archive (EGA, http://www.ebi.ac.uk/ega/), hosted by the EBI, under the accession number EGAS00001000670.

2.4. Validation of variations

Variations were validated by PCR/Sanger sequencing as described [12] (Supplemental Tables 2 and 3). Breakpoint sequences were deposited at the EMBL-EBI Databank (http://www.ebi.ac.uk/cgi-bin/embl/webin/; Supplemental Table 4).

2.5. Fluorescence in situ hybridization (FISH)

FISH was performed as described [12] and detailed in the online supplementary material and methods.

2.6. Quantitative real-time PCR (qRT-PCR)

cDNA was transcribed using the Quantitect Reverse-Transcription Kit (Qiagen, Hilden, Germany). *IKZF1* expression was assessed by quantitative PCR (7900 HT Fast Real-time PCR system, Applied Biosystems, Darmstadt, Germany) employing SYBR Green. *β-Actin* served as a control. Primer sequences are listed in Supplemental Table 5.

3. Results

3.1. Identification of a cytogenetically undetected unbalanced translocation t(14;21) associated with a gain of RUNX1 in patient 1

Copy number profiling based on next-generation-sequencing revealed a karyotype of 52 chromosomes with additional copies of chromosomes 6, 7, 21, X and Y in the relapse sample of patient 1 (Fig. 1A, Table 1). Trisomies of chromosomes 10 and 18 indicated in the initial cytogenetic analysis were not observed. Partial gains of chromosome 14p and 21q indicated a cytogenetically undetected translocation t(14;21). The breakpoints were validated by PCR and Sanger sequencing (Fig. 1B). On chromosome 21 the gene locus of a large intergenic non-coding RNA (lincRNA, Supplemental Table 6) and on chromosome 14 a long terminal repeat (LTR) were affected by the translocation. The fusion of the LTR to intron 1/2 of the lincRNA might lead to expression of an open reading frame encoded by exon 2 and 3 of the lincRNA. The protein would have homology to tetratricopeptide repeats containing proteins (e.g. 70% homology to the histone demethylase UTY isoform 1).

The resulting derivative chromosome +der(14)t(14;21)(q21.2;q21.3) carries an additional copy of the *RUNX1* oncogene at 21q22.12. FISH analysis confirmed the presence of four *RUNX1* copies per cell (Fig. 1C). In the cytogenetic analysis the derivative chromosome was misinterpreted as a chromosome 18 (Fig. 1D). All patients in this study harbored four *RUNX1* copies per cell.

Please cite this article in press as: C. Chen, et al., Next-generation-sequencing of recurrent childhood high hyperdiploid acute lymphoblastic leukemia reveals mutations typically associated with high risk patients, Leuk Res (2015), http://dx.doi.org/10.1016/j.leukres.2015.06.005