RAG1 co-expression signature identifies ETV6-RUNX1-like B-cell precursor acute lymphoblastic leukemia in children

Dongfeng Chen1,2 | Alessandro Camponeschi2 | Jessica Nordlund3 | Yanara Marinevic-Zuniga3 | Jonas Abrahamsson4 | Gudmar Lönnerholm5 | Linda Fogelstrand6,7 | Inga-Lill Mårtensson2

1Institute of Life Sciences, Jiangsu University, Zhenjiang, China
2Department of Rheumatology and Inflammation Research, Institute of Medicine, Sahlgrenska Academy, University of Gothenburg, Gothenburg, Sweden
3Department of Medical Sciences, Molecular Medicine and Science for Life Laboratory, Uppsala University, Uppsala, Sweden
4Department of Pediatrics, Institute of Clinical Sciences, Sahlgrenska University Hospital, Gothenburg, Sweden
5Department of Women and Children’s Health, Uppsala University, Uppsala, Sweden
6Department of Clinical Chemistry, Sahlgrenska University Hospital, Gothenburg, Sweden
7Department of Clinical Chemistry and Transfusion Medicine, University of Gothenburg, Gothenburg, Sweden

Abstract

B-cell precursor acute lymphoblastic leukemia (BCP-ALL) can be classified into subtypes according to the genetic aberrations they display. For instance, the translocation t(12;21)(p13;q22), representing the ETV6-RUNX1 fusion gene (ER), is present in a quarter of BCP-ALL cases. However, around 10% of the cases lack classifying chromosomal abnormalities (B-other). In pediatric ER BCP-ALL, rearrangement mediated by RAG (recombination-activating genes) has been proposed as the predominant driver of oncogenic rearrangement. Herein we analyzed almost 1600 pediatric BCP-ALL samples to determine which subtypes express RAG. We demonstrate that RAG1 mRNA levels are especially high in the ETV6-RUNX1 (ER) subtype and in a subset of B-other samples. We also define 31 genes that are co-expressed with RAG1 (RAG1-signature) in the ER subtype, a signature that also identifies this subset of B-other samples. Moreover, this subset shares leukemia and pro-B gene expression signatures as well as high levels of the ETV6 target genes (BIRC7, WBP1L, CLIC5, ANGPTL2) with the ER subtype, indicating that these B-other cases are the recently identified ER-like subtype. We validated our results in a cohort where ER-like has been defined, which confirmed expression of the RAG1-signature in this recently described subtype. Taken together, our results demonstrate that the RAG1-signature identifies the ER-like subtype. As there are no definitive genetic markers to identify...
1  INTRODUCTION

Acquired chromosomal aberrations have been linked to the overall survival of patients with B-cell precursor acute lymphoblastic leukemia (BCP-ALL), which is the most common cancer in children.\(^1\) The many different subtypes of BCP-ALL have been classified according to the genetic aberrations they display, allowing correlations between disease type and prognosis to be made.\(^2\) For instance, the translocation t(12;21) (p13;q22), representing the $ETV6$-$RUNX1$ fusion gene (ER), is present in a quarter of BCP-ALL cases. Unclassified cases form a heterogeneous group referred to as B-other, which research efforts have reduced over the last 10 years from 25% to 5% of all cases.\(^3\) The characterization of disease genotype in such cases remains a priority, as it provides the means for diagnosis, prognosis, risk assessment, and targeted treatment.

During the early stages of B cell development, B-cell precursors (BCP) undergo immunoglobulin ($Ig$) gene rearrangement that is necessary for the production of a membrane-bound B-cell antigen receptor (BCR) on more mature B cells. Mouse knockout models,\(^4,5\) have shown the recombination-activating genes ($RAG$) to be essential for the rearrangement process, in which $Ig$ V(D)J gene segments are rearranged to provide instructions for a unique BCR. RAG activity may also play a role in leukemogenesis, as it has been proposed for the ER subtype, where it appears that RAG introduces mutations and aberrant rearrangements in non-Ig loci.\(^6\) However, RAG is likely not responsible for the ER translocation, but rather for introducing additional genetic modifications that drive leukemogenesis. Herein, we examined 1582 BCP-ALL cases, both microarray (DS1-6-M) and RNAseq (DS7-8-R) data sets (Table 1, Table S1), to determine on a large scale which subtypes express $RAG$ and whether we could find any co-expressed genes that would allow us to identify new subtypes.

2  MATERIALS AND METHODS

2.1  Gene expression microarray data

Gene expression data from BCP-ALL and healthy B-cells were gathered from published studies (Table 1, Table S1). All gene expression microarray data were log2 transformed and normalized using the Robust Multichip Average (RMA) method.

2.2  Gene expression RNA-sequencing data

Strand-specific RNA sequencing libraries were constructed from rRNA-depleted RNA using the ScriptSeq V2 Kit

---

TABLE 1  Data sets used in this study

| GEO accession | Dataset# | Country | Platform | Patient # | References |
|---------------|----------|---------|----------|-----------|------------|
| GSE45460      | DS0-M    | South Korea | GPL6244  | 8         | [7]        |
| GSE26281      | DS1-M    | USA     | GPL96    | 127       | [8]        |
| GSE28497      | DS2-M    | USA     | GPL96    | 239       | [9]        |
| GSE47051      | DS3-M    | Sweden  | GPL570   | 75        | [10]       |
| GSE12995      | DS4-M    | USA     | GPL96    | 175       | [11]       |
| GSE33315      | DS5-M    | USA     | GPL96    | 483       | [12]       |
| GSE26366      | DS6-M    | USA     | GPL96    | 172       | [13]       |
| RNA-seq1      | DS7-R    | Sweden  | Hiseq2000/2500 | 116       | [14]       |
| RNA-seq2      | DS8-R    | Sweden  | NextSeq 500 | 195       | [15]       |
| GSE34861      | DS9-M    | USA     | GPL15088 | 194       | [16]       |
| B95           | DS10-M   | USA     | GPL8300  | 95        | [17]       |
(A) RAG1

Relative Gene Expression

RAG2

Relative Gene Expression

DS1-M

(B) Leukemia signatures

RAG1

DS1-M

(C) RAG1-signature

DS1-M

(D) Leukemia signatures

DS1-M
RESULTS AND DISCUSSION

We have previously shown that the components of the pre-BCR complex, assembled from Ig heavy chain and surrogate light chain, are differentially expressed in the ETV6-RUNX1 and TCF3-PBX1 BCP-ALL subtypes. To determine the expression pattern of RAG1 and RAG2 that regulate Ig gene rearrangements, we first analyzed the expression of the RAG1 and RAG2 genes in the microarray dataset DS1-M that was used in the aforementioned study, which includes 127 BCP-ALL samples (Table 1, Table S1). We found both genes expressed in the ER subtype in this (Figure 1A) and the additional five microarray datasets (DS2-6-M) analyzed (Figure S1A). In particular, the expression of RAG1 was consistently higher in ER compared to all other genetic subtypes except B-other. Next, we performed a genome-wide screen for genes co-expressed with RAG1 and RAG2 in DS1-M using Pearson co-efficiency correlation analysis. Although none consistently appeared with RAG2, we identified a set of 31 genes that were the highest-ranked co-expressed genes with RAG1, henceforth referred to as the RAG1-signature (Figure 1B, Tables S2 and S3). Using the RAG1-signature as an identifier in the DS1-M data set distinguished the ER from all other BCP-ALL subtypes, except for four B-other and one CRLF2 samples that also expressed the RAG1-signature (Figure 1C). By contrast, although pro- and pre-B cells from healthy donors (DS0-M, Table 1) express RAG1 and RAG2, they do not express the RAG1-signature (Figure S2), demonstrating that this signature is specific for certain BCP-ALL.

Based on the above results, we hypothesized that the B-other samples with a RAG1-signature could represent the recently defined new subtype termed ER-like, which usually carry ETV6 fusions and IKZF1 aberrations. However, they vary and hence lack the definitive ER fusion gene, predicting a similar gene expression pattern to the ER subtype. Therefore, to pinpoint the relationships between genetic subtypes, we performed an unsupervised PCA analysis based on all genes expressed in DS1-M. This showed that the four B-other samples with the RAG1-signature clustered together with the ER samples with a unique leukemia signature (Figure 1D, labeled ER-like to distinguish them from the remaining B-other samples). We could confirm these results by analyzing the other five microarray data sets (DS2-M to DS6-M) with a total of 1145 samples, where the RAG1-signature distinguished the ER from the other subtypes, and in each data set a few B-other samples clustered with the ER samples (Figure S3). Moreover, we found that these samples expressed a leukemia-signature and clustered with the ER samples also in these microarray data sets (Figure S4).

To validate our observations based on microarray datasets, we analyzed RNA-seq data from a cohort of 116 BCP-ALL samples (DS7-R). Here we found not only RAG1 but also RAG2 expressed at higher levels in the ER compared to the other subtypes (Figure S1B). Moreover, as in the microarray data sets, a few samples belonging to the B-other group also expressed higher levels of RAG1 and RAG2. Further, using the RAG1-signature as an identifier in this RNAseq data set distinguished the ER from the other subtypes (Figure 2A and Table S3). We found also four B-other and one hyperdiploid (HH) samples that clustered with the ER samples (Figure 2A). This would be consistent with our previous analysis of the DNA methylation pattern of the patient samples in DS7-R in which three of the four samples identified here showed a pattern similar to that of the ER samples. Thereafter, we performed unsupervised PCA analyses based on gene expression, which confirmed that the same B-other samples...
(A) RAG1-signature

(B) Leukemia signatures

(C) Pro-B signature

(D) ETV6-RUNX1-like

(E) ETV6-targets

(F) Remaining ALL

- **3.1%** Remaining ALL
- **89.8%** B-other
- **3.5%** ETV6-RUNX1-like

- **3.1%** Remaining ALL
- **89.8%** B-other
- **3.5%** ETV6-RUNX1-like

- **116**
- **85.3%**
- **127**

- **1583**
- **77.8%**
- **2.7%**
ETV6 that involve have been found to harbor deletions and in-frame fusions. Here, we confirmed the pro-B signature, whereas the t(1;19) TCF3/PBX1 ALL subtype, which resembles pre-B cells. 18 Here, we confirmed the pro-B signature in DS2-M, 19 allowing for the definition of the ER- like subtype. Validating this in adult BCP-ALL of the ER subtype are infrequent. We therefore asked whether we could find any ER- like samples based on the RAG1-signature. However, among a total of 285 samples (DS9-10-M) with only one ER sample (Table 1, Table S1), which we could clearly distinguish, we were unable to define any ER- like samples (Figure S8). Thus, this indicates that not only ER but also ER- like subtypes are infrequent in adult BCP-ALL.

In this study, we show that a subset of BCP-ALL patient samples with unclassified chromosomal abnormalities (B- other) can be defined by the expression of RAG1 in conjunction with an additional 31 genes, the RAG1-signature. This signature as well as leukemia and pro-B gene expression signatures and the high levels of ETV6 target genes were all shared with the ER subtype, and suggested that these B- other cases belong to the ER- like subtype. Validating this in samples previously defined as ER- like, we could confirm this notion. Taken together, our results demonstrate that the RAG1-signature identifies ER-like BCP-ALL in children. As there are no consistent translocations or other definitive genetic markers, using the RAG1-signature could represent a means to screen for the ER-like subtype.

ACKNOWLEDGMENTS

We are grateful to Drs Henrik Liljebjörn and Thoas Fioretos for providing BCP-ALL RNAseq data (DS8-R). We also thank Dr Fiona McConnell for critically reading and editing the manuscript. Sequencing was performed by the SNP&SEQ Technology Platform in Uppsala. The facility is part of the National Genomics Infrastructure (NGI) Sweden at Science for Life Laboratory and supported by the Swedish Research Council and the Knut and Alice Wallenberg Foundation.

CONFLICT OF INTEREST

The authors declare no conflict of interest.
ETHICAL CONSIDERATIONS
Not applicable, meta-analyses, all ethics linked to respective data set, see Table 1.

DATA AVAILABILITY STATEMENT
Gene expression data from BCP-ALL and healthy B-cells were gathered from published studies (Table 1).

ORCID
Inga-Lill Mårtensson https://orcid.org/0000-0003-3415-0560

REFERENCES
1. Mullighan CG. Molecular genetics of B-precursor acute lymphoblastic leukemia. *J Clin Invest.* 2012;122:3407-3415.
2. Roberts KG, Mullighan CG. Genomics in acute lymphoblastic leukemia: insights and treatment implications. *Nat Rev Clin Oncol.* 2015;12:344-357.
3. Lilljebjörn H, Fioretos T. New oncogenic subtypes in pediatric B-cell precursor acute lymphoblastic leukemia. *Blood.* 2017;130:1395-1401.
4. Mombaerts P, Iacomini J, Johnson RS, Herrup K, Tonegawa S, Papaioannou VE. RAG-1-deficient mice have no mature B and T lymphocytes. *Cell.* 1992;68:869-877.
5. Shinkai Y, Rathbun G, Lam KP, et al. RAG-2-deficient mice lack mature lymphocytes owing to inability to initiate V(D)J rearrangement. *Cell.* 1992;68:855-867.
6. Papaemmanuil E, Rapado I, Li Y, et al. RAG-mediated recombination is the predominant driver of oncogenic rearrangement in ETV6-RUNX1 acute lymphoblastic leukemia. *Nat Genet.* 2014;46:116-125.
7. Lee S-T, Xiao Y, Muench MO, et al. A global DNA methylation and gene expression analysis of early human B-cell development reveals a demethylation signature and transcription factor network. *Nucleic Acids Res.* 2012;40:11339-11351.
8. Figueroa ME, Chen SC, Andersson AK, et al. Integrated genetic and epigenetic analysis of childhood acute lymphoblastic leukemia. *J Clin Invest.* 2013;123:3099-3111.
9. Coustan-Smith E, Song G, Clark C, et al. New markers for minimal residual disease detection in acute lymphoblastic leukemia. *Blood.* 2011;117:6267-6276.
10. Nordlund J, Bäcklin CL, Wahlberg P, et al. Genome-wide signatures of differential DNA methylation in pediatric acute lymphoblastic leukemia. *Genome Biol.* 2013;14:105.
11. Mullighan CG, Su X, Zhang J, et al. Deletion of IKZF1 and prognosis in acute lymphoblastic leukemia. *N Engl J Med.* 2009;360:470-480.
12. Zhang J, Ding LI, Holmfeldt L, et al. The genetic basis of early T-cell precursor acute lymphoblastic leukemia. *Nature.* 2012;481:157-163.
13. Kannan S, Fang W, Song G, et al. Notch/HES1-mediated PARP1 activation: a cell type-specific mechanism for tumor suppression. *Blood.* 2011;117:2891-2900.
14. Marinčević-Zuniga Y, Dahlberg J, Nilsson S, et al. Transcriptome sequencing in pediatric acute lymphoblastic leukemia identifies fusion genes associated with distinct DNA methylation profiles. *J Hematol Oncol.* 2017;10:148.
15. Lilljebjörn H, Henningsson R, Hyrenius-Wittsten A, et al. Identification of ETV6-RUNX1-like and DUX4-rearranged subtypes in pediatric B-cell precursor acute lymphoblastic leukemia. *Nat Commun.* 2016;7:11790.
16. Geng H, Brennan S, Milne TA, et al. Integrative epigenomic analysis identifies biomarkers and therapeutic targets in adult B-acute lymphoblastic leukemia. *Cancer Discov.* 2012;2:1004-1023.
17. Chiaretti S, Li X, Gentleman R, et al. Gene expression profiles of B-lineage adult acute lymphocytic leukemia reveal genetic patterns that identify lineage derivation and distinct mechanisms of transformation. *Clin Cancer Res.* 2005;11:7209-7219.
18. Chen D, Zheng J, Gerasimčik N, et al. The expression pattern of the pre-B cell receptor components correlates with cellular stage and clinical outcome in acute lymphoblastic leukemia. *PLoS One.* 2016;11:e0162638.
19. Zaliova M, Kotrova M, Bresolin S, et al. ETV6/RUNX1-like acute lymphoblastic leukemia: a novel B-cell precursor leukemia subtype associated with the CD27/CD44 immunophenotype. *Genes Chromosomes Cancer.* 2017;56:608-616.
20. Nordlund J, Bäcklin CL, Zachariadis V, et al. DNA methylation-based subtype prediction for pediatric acute lymphoblastic leukemia. *Clin Epigenetics.* 2015;7:11.
21. Chen D, Gerasimčik N, Camponeschi A, et al. CD27 expression and its association with clinical outcome in children and adults with pro-B acute lymphoblastic leukemia. *Blood Cancer J.* 2017;7:e575.
22. Neveu B, Spinella JF, Richer C, et al. CLIC5: a novel ETV6 target gene in childhood acute lymphoblastic leukemia. *Haematologica.* 2016;101:1534-1543.

SUPPORTING INFORMATION
Additional supporting information may be found online in the Supporting Information section.

How to cite this article: Chen D, Camponeschi A, Nordlund J, et al. RAG1 co-expression signature identifies ETV6-RUNX1-like B-cell precursor acute lymphoblastic leukemia in children. *Cancer Med.* 2021;10:3997–4003. https://doi.org/10.1002/cam4.3928