Characterization of PAX5 Intragenic Tandem Multiplication in Pediatric B-Lymphoblastic Leukemia by Optical Genome Mapping

Tracking no: ADV-2021-006328R1

Jeffrey Jean (LAC+USC Medical Center, United States) Alexandra Kovach (Children’s Hospital Los Angeles, United States) Andrew Doan (Children’s Hospital Los Angeles, United States) Matthew Oberley (Caris Life Sciences, United States) Jianling Ji (Children’s Hospital Los Angeles, University of Southern California, United States) Ryan Schmidt (Children’s Hospital Los Angeles, United States) Jaclyn Biegel (Children’s Hospital Los Angeles, United States) Deepa Bhojwani (Children’s Hospital Los Angeles, United States) Gordana Raca (Children’s Hospital Los Angeles, United States)

Abstract:

Conflict of interest: No COI declared

COI notes:

Preprint server: No;

Author contributions and disclosures: GR, AEK and DB initiated the cohort study, JJean and GR analyzed OGM data, all authors analyzed other cohort data, GR, JJean, AEK and DB wrote the manuscript; AEK, AD, MJO, JJ, RJS, JAB and DB took care of patients or performed clinical laboratory testing; all authors reviewed and approved the final manuscript.

Non-author contributions and disclosures: No;

Agreement to Share Publication-Related Data and Data Sharing Statement: Emails to corresponding author

Clinical trial registration information (if any):
Characterization of PAX5 Intragenic Tandem Multiplication in Pediatric B-Lymphoblastic Leukemia by Optical Genome Mapping

Jeffrey Jean1, Alexandra E. Kovach2, Andrew Doan3, Matthew Oberley4, Jianling Ji5, Ryan J. Schmidt2, Jaclyn A.Biegel2, Deepa Bhojwani3 and Gordana Raca2

1. Keck School of Medicine, University of Southern California, Los Angeles, CA, USA.
2. Department of Pathology and Laboratory Medicine, Children's Hospital Los Angeles, Keck School of Medicine, University of Southern California, Los Angeles, CA, USA.
3. Division of Hematology and Oncology, Cancer and Blood Disease Institute, Children's Hospital Los Angeles, Department of Pediatrics, Keck School of Medicine, University of Southern California, Los Angeles, CA, USA.
4. Caris Life Sciences, Phoenix, AZ, USA

Author contributions and disclosures: GR, AEK and DB initiated the cohort study, JJean and GR analyzed OGM data; all authors analyzed other cohort data; GR, JJean, AEK and DB wrote the manuscript; AEK, AD, MJO, JJi, RJS, JAB and DB took care of patients or performed clinical laboratory testing; all authors reviewed and approved the final manuscript.

COI: Authors have no relevant conflicts

Data Sharing Statement: For data sharing, please contact the corresponding author at graca@chla.usc.edu.
The PAX5 (Paired box domain gene 5) gene encodes a transcription factor with a key role in regulating B-cell differentiation.\(^1\) PAX5 abnormalities are observed in approximately 30% of B-lymphoblastic leukemia (B-ALL) cases, occurring as secondary changes in association with different subtype-defining genetic drivers,\(^2\) but recently also described as primary oncogenic alterations in two novel genetic subtypes of the disease: PAX5 P80R and PAX5 Alt.\(^3,4\)

A rare but recurrent PAX5 abnormality in B-ALL is the presence of multiple additional copies of several exons within the 5’ end of the gene, referred to as “PAX5 intragenic tandem multiplication (PAX5-ITM)” or “PAX5 intragenic amplification” in the literature.\(^5,6\)

While PAX5-ITM has been investigated by a variety of methods, many of its characteristics remain unknown. These include the number of extra copies of the affected region, their position and orientation within the gene, and the effects of the extra copies on the structure and function of the PAX5 gene and protein.

Optical genomic mapping (OGM) is a powerful new technology that allows genome-wide, high-resolution, high-throughput detection of balanced and unbalanced structural variants (SVs).\(^7,8\) It is based on fluorescent labeling of high-molecular weight genomic DNA at a short nucleotide motif, followed by linearization and imaging of labeled DNA and comparison of the genome-wide labeling patterns with reference genomes to identify SVs. Direct visualization of long DNA molecules allows for successful utilization of OGM to study repetitive regions.\(^9,10\) These characteristics led us to hypothesize that OGM may be an optimal technique for characterization of PAX5-ITM in B-ALL.

In the present study, we used OGM to perform the first direct visualization of PAX5-ITM, determine the specific number of extra copies of the multiplicated region, and confirm their intragenic position and orientation.

We utilized a retrospective clinical cohort of 412 consecutive pediatric B-ALL cases tested at the Children’s Hospital Los Angeles (CHLA) Center for Personalized Medicine (CPM) at the time of diagnosis or relapse. Genetic evaluation of leukemia samples included karyotype analysis, fluorescence in situ hybridization (FISH) studies, Chromosomal Microarray (CMA) analysis and the OncoKids® NGS panel, which were conducted as described previously.\(^11\) This testing had detected the following subtypes of B-ALL: hyperdiploid (n=94, 22.8%), ETV6::RUNXI (n=61, 14.8%), Philadelphia (Ph)-positive (n=21, 5%), Ph-like, CRLF2-positive (n=63, 15%), Ph-like, ABL-type (n=12, 2.9%), KMT2A-rearranged (n=17, 4%), iAMP21 (n=15, 3.6%), hypodiploid (n=12, 3%), TCF3::PBX1 (n=11, 2.6%), dic(9;20) (n=15, 3.6%), ERG deleted/DUX4-rearranged (n=5, 1.2%), ZNF384-rearranged (n=5, 1.2%), and IGH-rearranged, MEF2D-rearranged, PAX5 P80R and PAX5Alt (≤1% each) (Supplemental Table 1). In 65 cases (15.8%), the subtype-defining genetic driver remained unknown. OGM processing of 42 selected cases (including those positive for PAX5-ITM) was performed either at Bionano Genomics Service Laboratory (Bionano Genomics, San Diego, CA) (19 cases) or at CHLA CPM (23 cases), and the data were analyzed using the Bionano Solve software. Technical details of OGM testing are described in Supplemental Materials.
PAX5-ITM was detected by CMA in 6 of 412 B-ALL cases for a prevalence of 1.4%, consistent with previous reports. Basic clinical characteristics, immunophenotype and other genetic findings for the 6 PAX5-ITM positive cases are provided in Supplemental Materials (Supplemental Table 2). CMA findings for the PAX5-ITM abnormality in the 6 positive cases are shown in Table 1 and Figure 1. CMA results were consistent with two extra copies (triplication) of a region that was approximately 25 kb in length in 5 cases and approximately 50 kb in one case (#5). Based on the breakpoints determined by CMA, the affected region included PAX5 exons 2-5 in 5 cases and exons 1-5 in one case (#5).

Material for OGM was available for 3 of 6 PAX5-ITM cases. OGM results are displayed in Figure 1. OGM showed the presence of 4 additional copies of the 5' PAX5 region in Cases 1 and 6, and 5 additional copies in Case 4. These insertions were not detected by OGM in any of the remaining 39 B-ALL cases in our cohort, or in more than 300 normal control samples tested at the Bionano Genomics Laboratory. The multiplied copies were inserted “in situ” and in direct orientation, and their approximate size was concordant between OGM and CMA. Notably, in all three cases CMA analysis showed only two extra copies of the affected region, demonstrating that CMA analysis, as well as previously published Multiplex Ligation-dependent Probe Amplification (MLPA)-based testing, may significantly underestimate the number of extra copies of the affected region in PAX5-ITM.

PAX5-ITM was first comprehensively studied by Schwab et al., who identified this abnormality by MLPA in approximately 1% of B-ALL cases in their cohort. The MLPA data suggested the presence of more than one extra copy of the affected region, but the exact copy number was not determined. The PAX5-ITM positive cases lacked other primary, subtype-defining molecular changes, leading to the hypothesis that PAX5-ITM may be associated with a novel genetic subtype of B-ALL.

A more recent study by Gu et al. used integrated genomic analysis of 1,988 childhood and adult B-ALL cases to delineate 23 genetic subtypes, some of them novel. PAX5-ITM was detected in 10 cases in this cohort (0.5%), of which 8 were classified based on their expression profiles as PAX5Alt, while no classification was provided for the remaining two. This finding suggested that PAX5-ITM could be one of the abnormalities defining PAX5Alt B-ALL, supporting its association with the unique molecular subtype of B-ALL proposed by Schwab et al.

Gu et al. performed detailed characterization of one PAX5-ITM-positive case using genome, exome and transcriptome sequencing. The authors showed that multiplication involved exons 2-5, which code for the Paired-box/DNA binding domain of the PAX5-encoded protein. Transcriptome sequencing confirmed expression of the multiplied exons, while RT-PCR and Sanger sequencing showed that the extra copies resulted in an in-frame junction between exons 5 and 2. The authors hypothesized that the abnormality resulted in the presence of extra copies of the Paired-box/DNA binding domain within the mutant protein. However, accurate copy number determination for small gains and losses from short read sequencing data remains challenging, and even this comprehensive genomic analysis did not detect the exact number of copies of the multiplied region; furthermore, the structure of the mutant PAX5 allele was indirectly inferred.
Here we describe the first direct visualization and accurate determination of the structure of the \textit{PAX5-ITM} abnormality. Although the number of the studied cases is small, this is one of the largest cohorts of B-ALL cases with \textit{PAX5-ITM} to date at a single institution. Using OGM, we showed intragenic position and tandem orientation of the repeated segments; both of these features had only been inferred indirectly by previous studies. In the 3 characterized cases, OGM showed more than one extra copy of the multiplied region. While CMA analysis indicated the total copy number to be approximately 4 in all the cases (2 extra copies), OGM showed 4 extra copies in 2 cases and 5 extra copies in 1 case, for a total of 5 and 6 copies, respectively. The underestimation of the copy number by CMA may be related to intrinsic limitations of CMA technology (loss of linearity for the log2 ratio increase for high-copy gains) and by the presence of normal (non-leukemic) cells in the tested bone marrow aspirate samples. The presence of normal cells did not interfere with OGM analysis, which allowed separate visualization of mutant and wild-type DNA molecules. Interestingly, previous studies also reported that \textit{PAX5-ITM} consisted of more than one extra copy of the multiplied region. \textsuperscript{5} Consistent presence of multiple additional copies of the critical region is intriguing, and may be either related to the mechanisms through which the abnormality originates or alternatively may reflect the need for multiple copies in order for the \textit{PAX5-ITM} abnormality to be oncogenic. Accurate determination of the number of extra copies in a large series of \textit{PAX5-ITM} positive cases, as well as functional studies of the abnormal protein, will be required to address these questions. Considering that it likely results in the presence of extra copies of the DNA-binding region in the \textit{PAX5}-encoded transcription factor, \textit{PAX5-ITM} may change its binding affinity or specificity, thus altering its transcription program and disrupting B-cell differentiation.

In summary, we have shown the feasibility and advantages of using OGM to determine the exact number of copies and accurate structure of \textit{PAX5-ITM} through direct visualization of the repeat region, suggesting OGM as the optimal approach for characterization of this and similar abnormalities in further studies.
References:

1. Cobaleda C, Schebesta A, Delogu A, Busslinger M. Pax5: the guardian of B cell identity and function. Nat Immunol. 2007; 8(5): 463-470.

2. Coyaud E, Struski S, Prade N, et al. Wide diversity of PAX5 alterations in B-ALL: a Groupe Francophone de Cytogenetique Hematologique study. Blood. 2010;115(15): 3089-3097.

3. Gu Z, Churchman ML, Roberts KG, et al. PAX5-driven subtypes of B-progenitor acute lymphoblastic leukemia. Nat Genet. 2019; 51(2): 296-307.

4. Li JF, Dai YT, Lilljebjorn H, et al. Transcriptional landscape of B cell precursor acute lymphoblastic leukemia based on an international study of 1,223 cases. Proc Natl Acad Sci USA. 2018;115(50): E11711-E11720.

5. Schwab C, Nebral K, Chilton L, et al. Intragenic amplification of PAX5: a novel subgroup in B-cell precursor acute lymphoblastic leukemia? Blood Adv. 2017; 1(19): 1473-1477.

6. Ofverholm I, Tran AN, Heyman M, et al. Impact of IKZF1 deletions and PAX5 amplifications in pediatric B-cell precursor ALL treated according to NOPHO protocols. Leukemia. 2013;27(9): 1936-1939.

7. Lam ET, Hastie A, Lin C, et al. Genome mapping on nanochannel arrays for structural variation analysis and sequence assembly. Nat Biotechnol. 2012; 30(8): 771-776.

8. Barseghyan H, Tang W, Wang RT, et al. Next-generation mapping: a novel approach for detection of pathogenic structural variants with a potential utility in clinical diagnosis. Genome Med. 2017;9(1): 90.

9. Zhang Q, Xu X, Ding L, et al. Clinical application of single-molecule optical mapping to a multigeneration FSHD1 pedigree. Mol Genet Genomic Med. 2019; 7(3): e565.

10. Wight DJ, Aimola G, Aswad A, et al. Unbiased optical mapping of telomere-integrated endogenous human herpesvirus 6. PNAS. 2020;117(49): 31410-31416.

11. Hiemenz MC, Ostrow DG, Busse TM, et al. OncoKids: A Comprehensive Next-Generation Sequencing Panel for Pediatric Malignancies. J Mol Diagn. 2018;20(6): 765-776.

12. Whitford W, Lehnert, K, Snell RG, Jacobsen JC. Evaluation of the performance of copy number variant prediction tools for the detection of deletions from whole genome sequencing data. J Biomed Inform. 2019; 94: 103174.
### Table 1. PAX5-ITM characteristics determined by CMA and OGM

| Case | Genomic Coordinates by CMA (GRCh37/hg19) | Approximate Size (kb) | Exons (NM_016734.3) | Number of extra copies by CMA | Number of extra copies by OGM |
|------|------------------------------------------|-----------------------|----------------------|-------------------------------|-------------------------------|
| 1    | chr9:36999179-37020645                   | 21.4                  | 2-5                  | 2                             | 4                             |
| 2    | chr9:36993414-37020645                   | 27.2                  | 2-5                  | 2                             | 5                             |
| 3    | chr9:36999179-37024122                   | 24.9                  | 2-5                  | 2                             | N/A                           |
| 4    | chr9:36998965-37024122                   | 25.1                  | 2-5                  | 2                             | N/A                           |
| 5    | chr9:36984128-37034669                   | 50.5                  | 1-5                  | 2                             | N/A                           |
| 6    | chr9:36999180-37020988                   | 21.8                  | 2-5                  | 2                             | 4                             |

CMA: Chromosome Microarray; OGM: Optical Genome Mapping; N/A: Not applicable; material was not available to perform OGM.

### Figure Legend

**Figure 1:** PAX5-ITM characterization by CMA (A) and OGM (B-D).

- **(A)** A screenshot from the Chromosome Analysis Suite (ChAS) CMA analysis software showing copy number gains in the 5’ portion of the PAX5 gene in the 6 PAX5-ITM positive cases (dark blue bars on the ‘Copy number segments’ plots).

- **OGM characterization of PAX5-ITM:** (B) Case 1, showing 4 extra copies of the affected region in direct orientation, each approximately 21 kb in size. (C) Case 4, showing 5 extra copies of the multiplied region in direct orientation, each approximately 26 kb in size. (D) Case 6, showing 4 extra copies of the affected region in direct orientation, each approximately 21 kb in size. Arrows indicate copies of the multiplied region in the Bionano long-molecules assembly for the mutant allele, as compared to the chromosome 9 map (green) and the assembly from reference samples (top, light blue bar). Additional copies of the affected region are also visible in individual long molecules (yellow lines).
Figure 1

A. Copy Number Segments

B. Genes
OGM Ref. allele
Reference Chr9*
OGM PAX5-ITM allele

C. Genes
OGM Ref. allele
Reference Chr9*
OGM PAX5-ITM allele

D. Genes
OGM Ref. allele
Reference Chr9*
OGM PAX5-ITM allele

*hrB in silico E1 digestion
~500 kb