TO THE EDITOR:

Characterization of *PAX5* intragenic tandem multiplication in pediatric B-lymphoblastic leukemia by optical genome mapping

Jeffrey Jean,[1] Alexandra E. Kovach,[2] Andrew Doan,[3,4] Matthew Oberley,[5] Jianling Ji,[2] Ryan J. Schmidt,[2] Jaclyn A. Biegel,[2] Deepa Bhojwani,[3,4] and Gordana Raca[2]

[1]Department of Pathology, Keck School of Medicine, University of Southern California, Los Angeles, CA; [2]Department of Pathology and Laboratory Medicine, and [3]Division of Hematology and Oncology, Cancer and Blood Disease Institute, Children's Hospital Los Angeles, Los Angeles, CA; [4]Department of Pediatrics, Keck School of Medicine, University of Southern California, Los Angeles, CA; and [5]Caris Life Sciences, Phoenix, AZ

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 ~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 2 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]

Although *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 enables genome-wide, high-resolution, high-throughput detection of balanced and unbalanced structural variants.[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 structural variants. 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 multiplied region, and confirm their intragenic position and orientation.

We used a retrospective clinical cohort of 412 consecutive pediatric B-ALL cases tested at the Children's Hospital Los Angeles Center for Personalized Medicine at the time of diagnosis or relapse. Genetic evaluation of leukemia samples included karyotype analysis, fluorescence in situ hybridization studies, chromosomal microarray (CMA) analysis, and the OncoKids® next-generation–sequencing panel, which were conducted as described previously.[11] This testing had detected the following subtypes of B-ALL: hyperdiploid (n = 94; 22.8%); *ETV6::RUNX1* (n = 61; 14.8%); Philadelphia (Ph)-positive (n = 21; 5%); Ph-like, *CRLF2*[+] (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 *PAX5*Alt (≤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 at either Bionano Genomics Service

Laboratory (Bionano Genomics, San Diego, CA) (19 cases) or Children's Hospital Los Angeles CPM (23 cases), and the data were analyzed with 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.[5] Basic clinical characteristics, immunophenotype, and other genetic findings for the 6 PAX5-ITM[+] cases are provided in 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 2 extra copies (triplication) of a region that was ~25 kb in length in 5 cases and ~50 kb in the remaining case (no. 5). Based on the breakpoints determined by CMA, the affected region included PAX5 exons 2 to 5 in 5 cases and exons 1 to 5 in 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 >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 3 cases, CMA analysis showed only 2 extra copies of the affected region, demonstrating that CMA analysis, as well as previously published Multiplex Ligation-dependent Probe Amplification (MLPA)–based testing,[5] 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 ~1% of B-ALL cases in their cohort. The MLPA data suggested the presence of >1 extra copy of the affected region, but the exact copy number was not determined. The PAX5-ITM[+] 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.[5]

In a more recent study, Gu et al used integrated genomic analysis of 1988 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, whereas no classification was provided for the remaining 2.[3] 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.[5]

Gu et al performed detailed characterization of 1 PAX5-ITM[+] case using genome, exome, and transcriptome sequencing.[3] The authors showed that multiplication involved exons 2 to 5, which code for the paired-box/DNA binding domain of the PAX5-encoded protein. Transcriptome sequencing confirmed expression of the multiplied exons, whereas reverse transcription-polymerase chain reaction and Sanger sequencing showed that the extra copies resulted in an in-frame junction between exons 5 and 2. Gu et al hypothesized that the abnormality resulted in the presence of extra copies of the paired-box/DNA binding domain within the mutant protein. However, accurate determination of copy number for small gains and losses from short-read sequencing data remains challenging,[12] 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.

Herein, we describe the first direct visualization and accurate determination of the structure of the PAX5-ITM abnormality. Although the number of the studied cases is small, to our knowledge, this is one of the largest cohorts of B-ALL cases with 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 been inferred only indirectly by previous studies. In the 3 characterized cases, OGM showed >1 extra copy of the multiplied region. Although CMA analysis indicated the total copy number to be 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 $\log_2$ ratio increase for high copy gains) and by the presence of normal (nonleukemic) 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 PAX5-ITM consisted of >1 extra copy of the multiplied region.[5] Consistent presence of multiple additional copies of the critical region is intriguing, and may be related to the mechanisms through which the abnormality originates or may reflect the need for multiple copies for the PAX5-ITM abnormality to be oncogenic. Accurate determination of the number of extra copies in a large series of PAX5-ITM[+]cases, as well as functional studies of the abnormal protein, will be needed to address these questions. Considering that it is likely to result in the presence of extra copies of the DNA-binding region in the PAX5-encoded transcription factor, PAX5-ITM may change its binding affinity or specificity, thus altering its transcription program and disrupting B-cell differentiation.

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	Extra copies by OGM (n)
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	NA
4	chr9:36998965-37024122	25.1	2-5	2	NA
5	chr9:36984128-37034669	50.5	1-5	2	NA
6	chr9:36999180-37020988	21.8	2-5	2	4

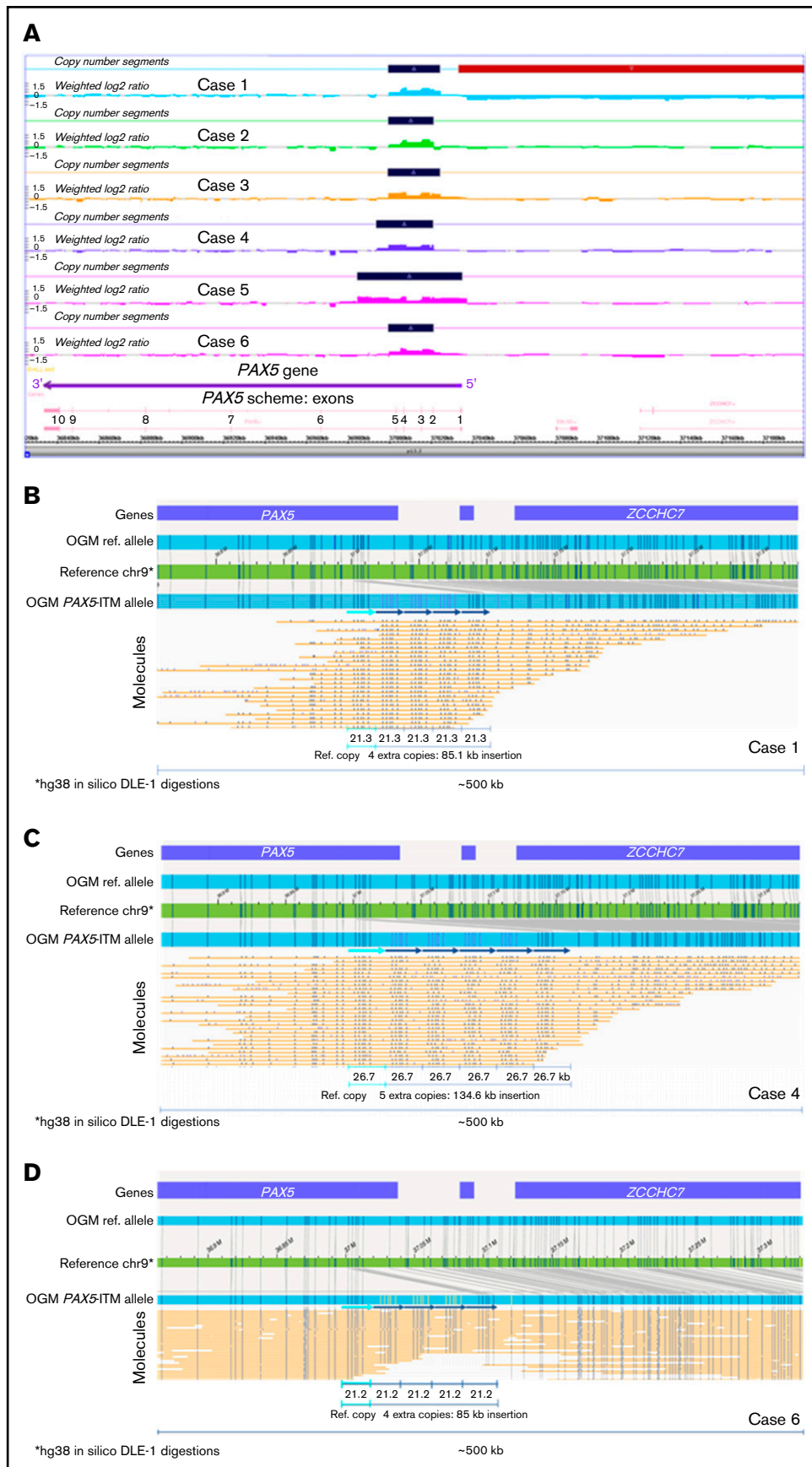NA, not applicable; material was not available to perform OGM.

Figure 1.

In summary, we have shown the feasibility and advantages of using OGM to determine the exact number of copies and accurate structure of *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.

ORCID profiles: M.O., 0000-0001-6419-2513; J.J., 0000-0003-4941-3538; R.J.S., 0000-0001-8789-9917; G.R., 0000-0002-3521-9585.

Correspondence: Gordana Raca, Children's Hospital Los Angeles, 4650 Sunset Blvd #173, Los Angeles, CA 90027; e-mail: graca@chla.usc.edu.

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, Lilljebjörn 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. Öfverholm 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. *Proc Natl Acad Sci USA.* 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.

Figure 1 (continued) *PAX5*-**ITM characterization by CMA and OGM.** (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⁺ cases (dark blue bars on the Copy Number Segments plots). OGM characterization of *PAX5*-ITM: case 1, showing 4 extra copies of the affected region in direct orientation, each ∼21 kb in size (B); case 4, showing 5 extra copies of the multiplied region in direct orientation, each ∼26 kb in size (C); and case 6, showing 4 extra copies of the affected region in direct orientation, each ∼21 kb in size (D). Arrows indicate copies of the multiplied region in the Bionano long-molecule assembly for the mutant allele, as compared with 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).