A pitfall of whole exome sequencing: variants in the 5 ' UTR Splice Site of BTK causing XLA

Takada, S.; Pico-Knijnenburg, I.; Pac, M.; Warris, A.; Burg, M. van der

Journal of Clinical Immunology (2022) 42:709–712
https://doi.org/10.1007/s10875-021-01198-x

LETTER TO EDITOR

A Pitfall of Whole Exome Sequencing: Variants in the 5′UTR Splice Site of *BTK* Causing XLA

Sanami Takada[1] · Ingrid Pico-Knijnenburg[1] · Malgorzata Pac[2] · Adilia Warris[3] · Mirjam van der Burg[1]

Received: 30 September 2021 / Accepted: 12 December 2021 / Published online: 29 January 2022

To the Editor,

X-linked agammaglobulinemia (XLA, OMIM300755) is an X-linked primary immunodeficiency characterized by absent or a very low number of B cells in peripheral blood and a- or hypo-gammaglobulinemia. XLA is caused by pathogenic variants in the Bruton's tyrosine kinase (*BTK*) gene, which is crucial for B cell differentiation. XLA is known as the most frequent cause of agammaglobulinemia, as *BTK* pathogenic variants were identified in over 85% of boys with early-onset infections, agammaglobulinemia and less than 2% B cells in periphery [1]. The diagnosis is confirmed by detection of a *BTK* variant by genetic analysis; *BTK*-targeted sanger sequencing and whole exome sequencing (WES) are widely used.

Here, we present a case of XLA with a *BTK* variant in the 5′ untranslated region (5′UTR), which was not detected by WES. We further performed the molecular analysis using 2 XLA cases with a *BTK* splicing site variant in the 5′UTR. We highlight the importance of sequencing non-coding regions in which pathogenic variants can be identified.

A 2-year-old boy (Patient-1) presented with arthritis of his knees and hip. He had a history of a number of respiratory tract infections without notable family history. Laboratory test showed agammaglobulinemia (IgG, 0.1 g/L; IgA, undetectable; IgM, undetectable) and absence of CD19 + B cells in peripheral blood while the number of CD8 + T cells (1.27×10^9/L) and NK cells (0.14×10^9/L) were in the normal range and CD4 + T cells was slightly elevated (3.32×10^9/L) (Supplemental Table). After he received immunoglobulin supplementation and antibiotics, the arthritis disappeared gradually in 3 months. Since this, the patient is in a good clinical condition with immunoglobulin supplementation.

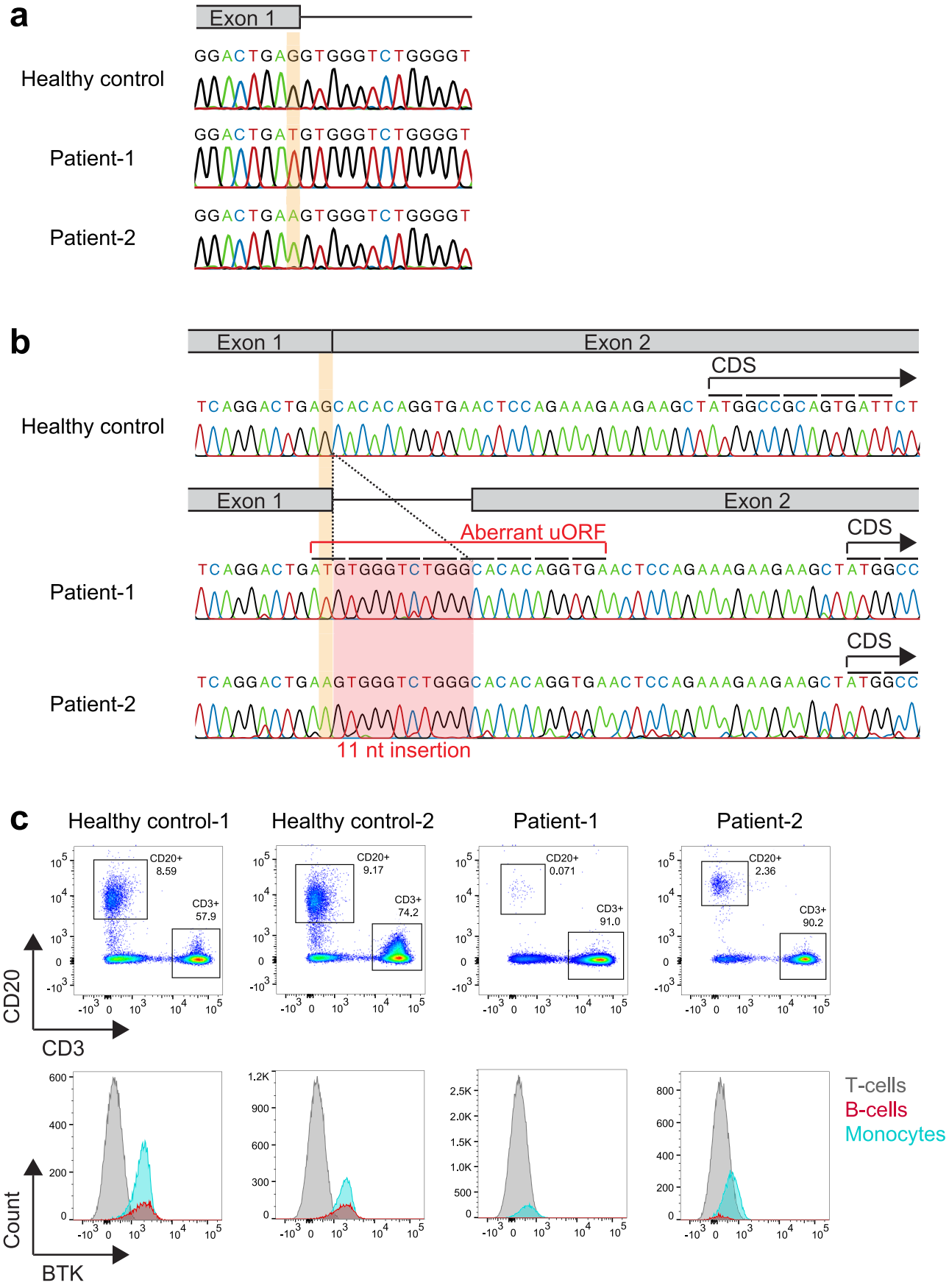
Because of his agammaglobulinemia and absent B cells, XLA was clinically suspected. For genetic analysis, WES was performed, but the *BTK* gene was intact as well as other autosomal recessive agammaglobulinemia-related genes (*IGHM*, *CD79A*, *CD79B*, *IGLL1*, *BLNK*, *PIK3R1*, *SLC39A7*, *TCF3*). Copy number analysis showed no evidence of large deletions or duplications in *BTK*. However, bone marrow analysis supported XLA since there was a block in precursor B cell differentiation before the pre-B-II-cell stage (CD19 + / cytoplasmic IgM +), which is characteristic for XLA patients [2] (Supplemental Fig. 1a). Subsequently, whole genome sequencing (WGS) showed a hemizygous *BTK* variant c.-31G > T in exon 1 in the 5′UTR, which was validated by Sanger sequencing (Fig. 1a). This *BTK* variant was located at the end of exon 1 in the splicing donor site. Although the variant was absent on the gnomAD database (https://gnomad.broadinstitute.org/), one patient (Patient-2, immunological data available in Supplemental Table) in our original XLA database ($n = 125$) had the c.-31G > A variant that was at the exact same position but with a different nucleotide substitution (Fig. 1a). To assess the effect on splicing, we studied *BTK* transcripts of these 2 patients. Sanger sequencing of the *BTK* transcripts amplified by reverse transcription PCR (RT-PCR) encompassing exon 1 and 4 demonstrated that both patients had an 11-nucleotide insertion from intron 1 (Fig. 1b, Supplemental Fig. 1b). This insertion in Patient-1 resulted in a new upstream open-reading frame (ORF) as upstream start codon was inserted, while Patient-2 did not have the upstream start codon (Fig. 1b). Both patients showed absent expression of BTK protein in monocytes by flow cytometric analysis (Fig. 1c).

✉ Mirjam van der Burg
m.van_der_burg@lumc.nl

[1] Department of Pediatrics, Laboratory for Pediatric Immunology, Willem-Alexander Children's Hospital, Leiden University Medical Center, Albinusdreef 2, 2333 ZA Leiden, the Netherlands

[2] Department of Immunology, The Children's Memorial Health Institute, Warsaw, Poland

[3] MRC Center for Medical Mycology, University of Exeter and Great Ormond Street Hospital, London, UK


a
Exon 1
Healthy control
GGACTGAGGTGGGTCTGGGGT
Patient-1
GGACTGATGTGGGTCTGGGGT
Patient-2
GGACTGAAGTGGGTCTGGGGT
b
Exon 1
Exon 2
CDS
Healthy control
TCAGGACTGAGCACACAGGTGAACTCCAGAAAGAAGAAGCTATGGCCGCAGTGATTCT
Exon 1
Exon 2
Aberrant uORF
CDS
Patient-1
TCAGGACTGATGTGGGTCTGGGCACACAGGTGAACTCCAGAAAGAAGAAGCTATGGCC
CDS
Patient-2
TCAGGACTGAAGTGGGTCTGGGCACACAGGTGAACTCCAGAAAGAAGAAGCTATGGCC
11 nt insertion
c
Healthy control-1
Healthy control-2
Patient-1
Patient-2
CD20+ 8.59
CD3+ 57.9
CD20+ 9.17
CD3+ 74.2
CD20+ 0.071
CD3+ 91.0
CD20+ 2.36
CD3+ 90.2
CD20
CD3
Count
BTK
T-cells
B-cells
Monocytes

◂**Fig. 1** **a** Sanger sequencing analysis of *BTK* gene. Patient-1 showed hemizygous c.-31G > T variant and Patient-2 showed hemizygous c.-31G > A variant, which were nucleotide substitution at the last nucleotide in exon 1. **b** Sanger sequencing analysis of *BTK* transcripts amplified by reverse transcription PCR (RT-PCR) encompassing exon 1 and 4. The variants were highlighted in orange shadow. Patient-1 and Patient-2 demonstrated 11 nucleotide (nt) insertion (highlighted in red shadow) between exon 1 and exon 2. In Patient-1, together with the 11-nucleotide insertion and the nucleotide substitution, aberrant upstream open-reading frame (uORF) was created in the upstream region of coding sequences (CDS). **c** Flowcytometric analysis of cytoplasmic BTK. Upper plots show CD3 + (T cells) and CD20 + (B cells) gated for lymphocytes. Patient-1 and Patient-2 had low composition of B cells. Lower histograms show cytoplasmic BTK expression in T cells (gray), B cells (red) and monocytes (light blue, gated by CD14 +). BTK in T cells was used as negative control. BTK was not expressed in monocytes in Patient-1 and Patient-2, while 2 healthy controls showed BTK expression in B cells and monocytes

WES is a frequently used method in combination with filters for specific disease categories to identify variants in rare genetic diseases, such as primary immunodeficiency diseases (PID). WES is exhaustive and has recently become a rapid, low cost and high accessibility approach comparable to traditional Sanger sequencing. However, the pitfall of WES is that non-coding regions are not covered, which might contain pathogenic variants. For Patient-1, WES in combination with a filter for PID genes caused a diagnostic delay, which could have been prevented by flow cytometry or Western blot analysis for BTK protein expression at an earlier stage. A diagnostic delay is not uncommon in XLA patients as El-Sayed *et al*. described that 34% of centers in 32 countries reported the delay of over 24 months between the first complaint and diagnosis of XLA [1]. Although our patient did receive appropriate treatment, a genetic diagnosis is important for prognostic reasons as well as providing the family an answer to where the disease comes from.

To understand the mechanism why these two variants caused XLA while leaving the *BTK* open reading frame (ORF) intact, we performed additional analyses. In our 2 patients, the splicing site variant caused an aberrant 11-nucleotide insertion into the 5′UTR in *BTK* transcripts. Although the exact function of 5′UTR is still uncovered, it is known that the 5′UTR containing an internal ribosome entry site (IRES), riboswitch and upstream ORF, is crucial for post-transcriptional regulation via the secondary structure and stability of mRNA, and for translational efficiency as entry sites for the ribosome. In line with this, variants in 5′UTR are potentially associated with aberrant stability of mRNA and/or regulation of protein translation [3]. We already demonstrated that BTK expression was absent, but to check for mRNA stability we performed a semi-quantified RT-PCR on a part of the coding sequence (CDS) of *BTK* in peripheral blood mononuclear cells. This revealed that both patients had decreased levels of *BTK* transcripts. Patient-1 had slightly decreased levels and Patient-2 had strongly decreased levels (Supplemental Fig. 1c and d). Since RT-PCR was performed using RNA extracted from total peripheral blood mononuclear cells, *BTK* transcripts were amplified only from patients' monocytes while *BTK* transcripts were derived from B cells and monocytes in the healthy control. Therefore, the expression level of *BTK* transcripts could not be fully compared between the healthy control and the patients. However, the *BTK* transcript analysis suggested that both pathogenic mechanisms might be involved in both patients: a translational defect and/or mRNA instability.

In Patient-1, the T to G single nucleotide substitution along with the aberrant insertion resulted in a new upstream start codon, creating a new short upstream ORF. In general, 49% of human gene transcripts has at least 1 upstream ORF that regulate translation of the downstream CDS [4]. Upstream ORF is known to regulate the downstream CDS via protein translation regulation and/or mRNA decay. In line with this, new upstream ORFs created by variants are reported to be associated with human diseases [5]. Thus, the new upstream ORF shown in Patient-1 was potentially pathogenic to suppress BTK protein expression. On the other hand, Patient-2 who had a T to A substitution leading to the same aberrant insertion, but without the new upstream ORF, which also resulted in suppressed BTK protein expression. Based our data of Patient-1 and Patient-2, the elongation of the 5′UTR by an 11-nucleotide insertion seems to be critical to suppress BTK protein expression, causing XLA. We hypothesized that the aberrant insertion influenced the 5′UTR secondary structure of mRNA, which is crucial for RNA-binding proteins such as ribosomes. In the gnomAD database, several variants in splice site in the 5′UTR and in intron 1 were reported in *BTK*, supporting that the 5′UTR and intron 1 have an important role in regulation of BTK expression and transcription, respectively.

The two variants in 5′UTR in the *BTK* gene reported in this Letter emphasize the importance of being aware of sequencing non-coding regions. Further analysis of *BTK* variants in untranslated regions may uncover the complex mechanisms of BTK regulation during B cell development.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s10875-021-01198-x.

Declarations

Ethics Approval This study was part of diagnostics under the guideline of the local Ethic Committees.

Consent to Participate Informed consent was obtained from the patients' guardians. The healthy control samples were provided from healthy volunteers after receiving their informed consent.

Consent for Publication Consent for publication in the journal article was obtained from the patients' guardians.

Conflicts of Interest The authors declare that they have no conflict of interest.

References

1. El-Sayed ZA, Abramova I, Aldave JC, Al-Herz W, Bezrodnik L, Boukari R, et al. X-linked agammaglobulinemia (XLA): phenotype, diagnosis, and therapeutic challenges around the world. World Allergy Organ J. 2019;12(3):100018.
2. Noordzij JG, de Bruin-Versteeg S, Comans-Bitter WM, Hartwig NG, Hendriks RW, de Groot R, et al. Composition of precursor B-cell compartment in bone marrow from patients with X-linked agammaglobulinemia compared with healthy children. Pediatr Res. 2002;51(2):159–68.
3. Mignone F, Gissi C, Liuni S, Pesole G. Untranslated regions of mRNAs. Genome biology. 2002;3(3):Reviews0004.
4. Calvo SE, Pagliarini DJ, Mootha VK. Upstream open reading frames cause widespread reduction of protein expression and are polymorphic among humans. Proc Natl Acad Sci. 2009;106:7507–12.
5. Fisher EMC, Barbosa C, Peixeiro I, Romão L. Gene expression regulation by upstream open reading frames and human disease. PLoS Genet. 2013;9(8):e1003529.