Human Genomics

BRIEF REPORT Open Access

Resolving non-coding splice-altering variants using an integrative genomic and transcriptomic workflow: application to *FOXP1*

Pauline Planté-Bordeneuve[1,2,3], Anne-Sophie Jourdain[1,2], Caroline Thuillier[2], Aurélien Caux[2], Marine Tessarech[1,2,3], Jean-Pascal Meneboo[5], Emilie Ait-Yahya[4], Sara Costantini[6], Perrine Brunelle[1,2,3] and Thomas Smol[1,2,3*]

Abstract

Genome sequencing (GS) often reveals non-coding variants of uncertain significance, especially those predicted to affect pre-mRNA splicing. We present an integrative workflow combining splice prediction, a minigene assay and long-read RNA sequencing to assess their functional impact. Applied to a *de novo* heterozygous 17-bp intronic deletion in *FOXP1* from a child with a neurodevelopmental disorder, this approach showed impaired exon recognition and multiple aberrant transcript isoforms, supporting a loss-of-function mechanism consistent with *FOXP1* haploinsufficiency. Our results illustrate how experimental and isoform-resolved transcriptomic analyses can refine the interpretation of non-coding splice-altering variants detected by GS in clinical genomics.

Keywords Non-coding variants, RNA splicing, Genome diagnostics, Minigene assay, Long-read RNA sequencing, FOXP1

Introduction

The use of genome sequencing in diagnostic practice has increased the detection of non-coding genetic variants. A large fraction of these variants remain classified as variants of uncertain significance (VUS), especially when located in intronic regions predicted to affect pre-mRNA splicing. Although splicing prediction algorithms provide valuable prioritisation, they remain limited in their ability to reliably predict cryptic splice-site activation, partial intron retention, or the simultaneous production of multiple aberrant transcript isoforms [1–3].

In routine clinical workflows, short-read RNA sequencing and targeted RT-PCR approaches are widely used to investigate splicing abnormalities and often provide valuable insights. However, some splicing configurations,

*Correspondence:
Thomas Smol
thomas.smol@chu-lille.fr
[1]Univ. Lille, ULR7364 RADEME, FHU-G4 Génomique, Lille F-59000, France
[2]CHU Lille, Institut de Génétique Médicale, F-59000 Lille, France
[3]Laboratoire de biologie médicale multisites Seqoia-PFMG2025, F-75000 Paris, France
[4]Cellule de Bioinformatique, CHU Lille, Plateau Commun de Séquençage, CHU Lille, Lille F-59000, France
[5]Univ. Lille, CNRS, Inserm, CHU Lille, Institut Pasteur de Lille, US 41-UAR 2014-PLBS, Lille, France
[6]CHU Lille, Clinique de Génétique, Lille F-59000, France

particularly those involving multiple alternative isoforms or partial intron retention, may remain difficult to resolve without complementary transcript-level analyses. As a result, functional assays play an increasingly important role in variant interpretation by bridging the gap between genomic data and molecular consequence [4]. Minigene splicing assays offer a controlled experimental system to test whether a specific variant is sufficient to perturb exon recognition [5]. In parallel, long-read RNA sequencing technologies allow direct observation of full-length transcripts, providing isoform-level information that is inaccessible to short-read methods [6].

FOXP1 encodes a forkhead box transcription factor that is critical for neurodevelopmental processes, including neural circuit formation, synaptic plasticity, and the regulation of gene expression in the developing brain [7]. Monoallelic loss-of-function variants in *FOXP1* cause an autosomal dominant neurodevelopmental disorder characterised by intellectual.

disability, speech and language impairment, and behavioural abnormalities (MIM #613670) [7]. While many pathogenic coding and canonical splice-site variants have been described, the interpretation of non-coding intronic variants in *FOXP1* remains challenging and often requires functional confirmation.

Here, we present an integrative genomic and transcriptomic workflow designed to resolve the functional impact of non-coding splice-altering variants identified by genome sequencing. Using a *de novo* intronic *FOXP1* variant as a representative example, we illustrate how combining predictive, experimental, and long-read transcriptomic approaches enables coherent mechanistic interpretation and robust clinical classification.

Materials and methods

Ethics statement

Our study was performed following the Declaration of Helsinki protocol. Clinical data have been collected from patients' medical files, conserved in Lille University Hospital. Written informed consents were obtained from the parents.

Patient

The patient was a male child born to non-consanguineous parents. He is the second of three siblings, which are all healthy. The pregnancy was uneventful. He achieved independent walking at 23 months of age. At the last examination, at 9 years old, the patient shows normal height and weight (0 SD), but a reduced and non-progressive head circumference (−2 SD at 5 years old, − 1 SD at 9 years old). The patient presented with marked language delay and behavioral disturbances with autistic features. There are no dysmorphic facial features, renal or genital abnormalities and no seizures. Brain MRI and echocardiography yielded normal results. Previous genetic testing for neurodevelopmental disorders (chromosomal microarray, Fragile X analysis, and exome sequencing) revealed no abnormalities.

DNA short read sequencing

Genome sequencing (GS) was performed on DNA from whole blood samples of the proband and parents using an Illumina NovaSeq 6000 system (Illumina, San Diego, US-CA). The library was prepared without amplification (NEBNext Ultra II End-repair/A-tailing-module, New England Biolabs, Ipswich, US-MA). Reads were aligned to the GRCh38.92 Human reference using BWA-MEM (v0.7.15). Single nucleotide variations (SNV) and small insertions/deletions were detected using GATK (v4.1.7.0) and annotated with SNPeff (v4.3t) and SnpSift (v4.3t). *FOXP1* gene and genetic variation were annotated according to the MANE transcript NM_001349338.3. Specific PCR primers were designed to target the region of the candidate variation *FOXP1*. Primers are available on request.

Minigene assays and splicing effect predictions

The splicing minigene assay was adapted from the method using the pcDNA3.1-based pCAS2 reporter plasmid [5]. This vector contains constitutive donor and acceptor sites flanking a multiple cloning site that allows insertion of genomic fragments of interest. Genomic segments encompassing *FOXP1* exon 18 and its flanking intronic regions (745 bp total) were amplified by PCR from both wild-type and patient genomic DNA using AmpliTaq Gold DNA Polymerase (Thermo Fisher Scientific, Waltham, MA, USA). Primers available on request.

PCR products corresponding to the wild-type (NM_001349338.3) and mutant (c.1531-31_1531-15del) alleles were cloned into the pCAS2 vector using the In-Fusion Snap Assembly Kit (Takara, Tokyo, Japan), generating pCAS2-FOXP1 wild-type and mutant hybrid constructs. HeLa cells were transfected with each construct using the Effectene Transfection Reagent (Qiagen, Hilden, Germany) and harvested 24 h after transfection. Total RNA was extracted with the NucleoSpin RNA kit (Macherey-Nagel, Düren, Germany) and reverse-transcribed using the SuperScript VILO cDNA Synthesis Kit (Thermo Fisher).

Splicing products were analysed by fragment analysis and Sanger sequencing to identify differences between wild-type and mutant transcripts. Primer sequences and PCR conditions are available upon request. In silico predictions of splicing effects were performed using SpliceAI, SPiP and MaxEntScan [1–3].

Nanopore long read cDNA sequencing

RNA was extracted from peripheral blood using the NucleoSpin RNA Blood (Macherey-Nagel). For the ONT sequencing experiment, the Direct cDNA sequencing protocol (SQK-LSK114 kit, Oxford Nanopore Technologies, Oxford, UK) was followed according to the manufacturer's instructions. Total RNA was reverse transcribed using SuperScript VILO cDNA Synthesis followed by strand switching to produce full-length cDNA. The prepared cDNA libraries were then loaded onto PromethION flow cells, for the patient and for the control samples. Base calling was performed using the Guppy basecaller (version 6.4.6) in High Accuracy (HAC) mode. The resulting reads were aligned to the human reference genome GRCh38 using Minimap2 with splice-aware alignment settings.

Results

Identification of a non-coding FOXP1 variant by genome sequencing

Trio genome sequencing identified a heterozygous 17-bp intronic deletion in *FOXP1* c.1531-31_1531-15del, located upstream of the canonical acceptor site of exon 18. According to ACMG criteria, this variant is *de novo* (PS2) and absent from population databases (PM2). However, the criterion 'multiple lines of computational evidence support a deleterious effect on the gene or gene product' cannot be applied, as the prediction scores do not clearly support a pathogenic impact (PP3). Therefore, this variant, due to its non-coding localisation, should be classified as a VUS (Fig. 1A). The variant has been deposited in the ClinVar database under the accession number SUB15731157.

In silico prediction of splicing disruption

SpliceAI predicted a strong loss of the canonical acceptor site of exon 18 (Δ score = 0.76), indicating a high likelihood of impaired exon recognition. In addition, a moderate probability of donor site loss was predicted at chr3:70972555 (NM_001349338.3:c.1653; Δ score = 0.36), located 136 bp downstream of the variant, suggesting a broader destabilization of the local splicing environment [1]. SpliceAI also identified weak gains of two upstream cryptic acceptor sites at chr3:70972712 (NM_001349338.3:c.1531−36; Δ score = 0.17) and chr3:70972725 (NM_001349338.3:c.1531−49; Δ score = 0.09), supporting the potential activation of non-canonical splice sites [1].

Consistent with these findings, SPiP predicted a severe alteration of the polypyrimidine tract, with a risk score of 96.71% (95% CI: 92.49–98.92). MaxEntScan analysis showed that the wild-type 3'acceptor site scored − 13.56 in the mutant context, consistent with a loss of canonical splice-site strength, whereas a cryptic acceptor site created by the deletion scored 3.77, representing a 127.8% increase in predicted splicing efficiency [2, 3].

Taken together, these predictions indicate that the c.1531-31_1531-15del variant is likely to disrupt normal exon 18 inclusion, either through exon skipping or the use of upstream cryptic splice acceptor sites, leading to the generation of aberrant transcripts (Fig. 1B, C).

Minigene splicing assay demonstrates impaired exon recognition

To assess the functional impact of the variant in vitro, we constructed wild-type and mutant splicing reporter minigenes encompassing the *FOXP1* exon 18 region. Compared with the wild-type construct, the mutant minigene produced additional splice products. Two aberrant amplicons were observed, 19 bp and 32 bp longer than the canonical splice product (355 bp), consistent with the inclusion of intronic sequences driven by the activation of cryptic splice sites predicted in silico (Fig. 1C, D).

A relative increase in the abundance of the backbone minigene product (237 bp) was observed in the mutant condition. Although this fragment may reflect vector-derived splicing, its increased representation in the presence of the variant is consistent with reduced splicing fidelity at the canonical exon 18 acceptor site. Overall, these results indicate that the intronic deletion leads to inefficient exon recognition and the production of aberrant splice products with altered coding potential.

Long-read RNA sequencing reveals complex splicing outcomes in vivo

Long-read cDNA sequencing of peripheral blood RNA identified several aberrant *FOXP1* transcript isoforms in the patient that were not observed in the control sample. On a total of 69 long reads covering the transcript, 51% ($n = 35$) corresponds to the canonical transcript, we identified 20% ($n = 14$) of complete skipping of exon 18, 6% ($n = 4$) of combined skipping of exons 18 and 19, 21% ($n = 15$) of partial intron 17 retention, spanning from the 3′ boundary at the deletion site to the canonical 5′ splice junction of exon 18, and 2% ($n = 1$) of activation of a cryptic splice acceptor site located 32 bp upstream of the canonical exon 18 site. Approximately half of the sequenced transcripts corresponded to the canonical isoform, consistent with a haploinsufficiency mechanism rather than complete splice abrogation (Fig. 1E).

Discussion

Interpreting non-coding splice-altering variants remains a major challenge in genome-based diagnostics. Although genome sequencing allows their detection, splicing prediction tools alone are frequently insufficient to resolve their functional consequences, particularly

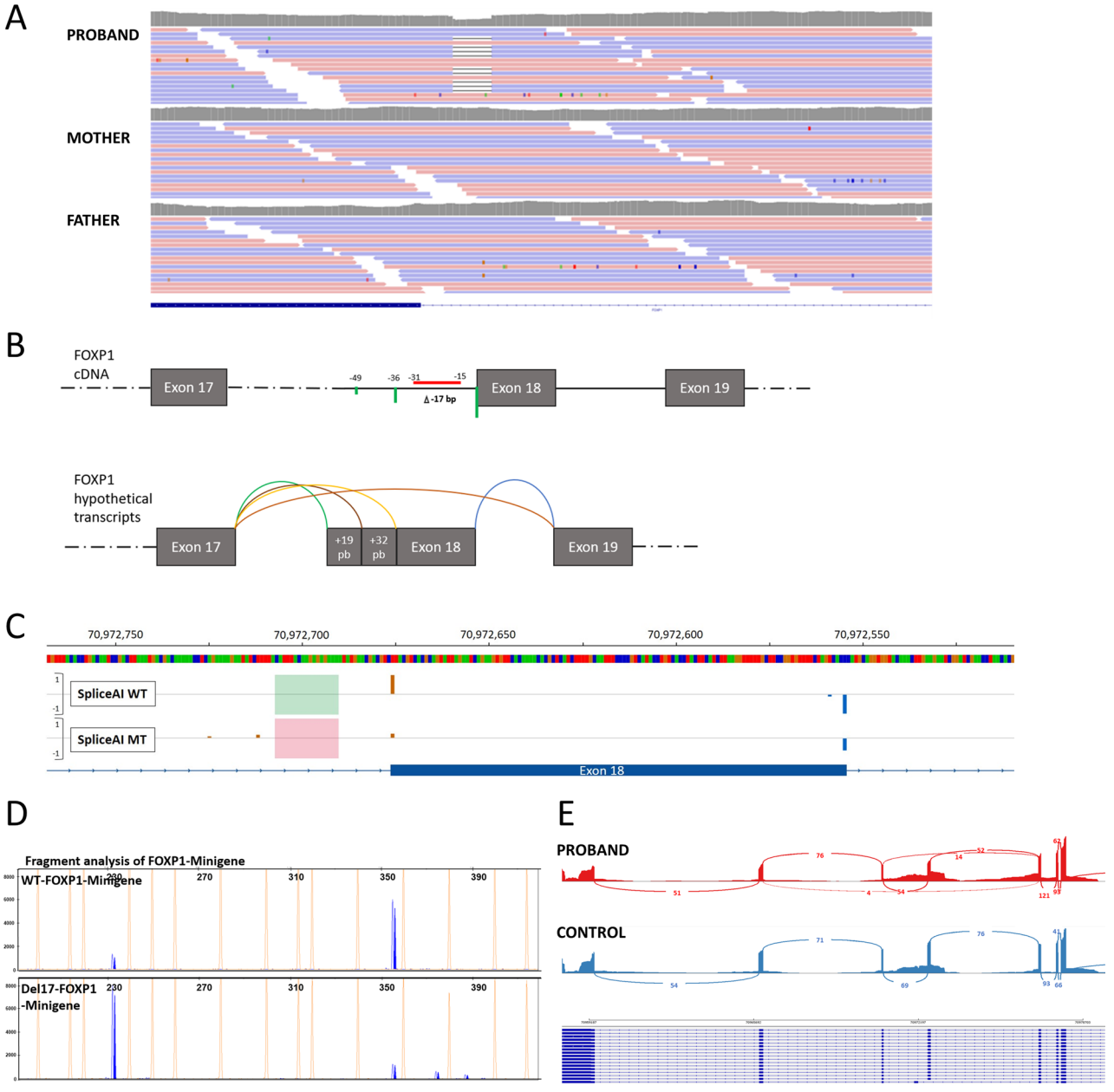

Fig. 1 **A** Integrative Genome Viewer (IGV) visualization of the trio genome sequences of the variant. **B** Schematic representation of the *FOXP1* wild type (WT) transcript and hypothetical alternative transcripts from in silico approaches. **C** SpliceAI visualization of splicing alterations in *FOXP1*. The 17 bp intronic deletion is depicted in pink, causing splicing alterations depicted in orange for acceptor splice site and in blue for donor splice site. The transcript Wild-type is noted WT and the mutant one is noted MT. **D** Results of fragment analyses of the *FOXP1* splicing reporter minigene. The wild-type (WT) minigene is shown at the top with 2 fragments: one fragment of 237pb corresponding to the backbone minigene and one of 355pb corresponding of the *FOXP1* exon 18 minigene; the mutant minigene is shown at the bottom with the creation of alternative transcripts that are 19 bp and 32 bp longer compared to the 355 bp WT *FOXP1* exon 18 construction. A higher proportion of the empty minigene transcript suggests exon skipping in the mutant minigene. The duplicated peaks are presumed to result from PCR-related artifacts (such as stutter or partial extension) rather than true genetic duplications. **E** Sashimi plot of total aligned reads of the patient (red) and control (blue) for *FOXP1*. Fourteen reads show exon 18 skipping in the patient, which is not observed in the control or any of the transcripts referenced in Ensembl

when complex or competing splicing events are involved. This limitation is a major contributor to the persistence of VUS in clinical practice.

The workflow we describe was designed to combine prediction, functional testing, and transcript-level confirmation. In this case, the variant was first highlighted by prediction tools suggesting a loss of the canonical acceptor site of *FOXP1* exon 18 and activation of upstream cryptic sites [1, 2]. These hypotheses were useful but not sufficient. The algorithms could not anticipate

the full spectrum of transcripts observed in experimental data, a limitation already noted in comparative evaluations of splicing predictors [8].

The minigene assay provided a first functional validation. It demonstrated that the 17-bp deletion alone was enough to alter exon recognition and produce aberrant splice products. This approach, although performed in a heterologous context, remains one of the most direct approaches to demonstrate causality between a single variant and its splicing consequence [4]. It circumvents the requirement for endogenous gene expression in the tissue under investigation. However minigenes do not fully reproduce the native genomic environment, because incorporating the full *FOXP1* context, including long-range intronic sequences (spanning ~ 6.2 kb, including intron 17, exon 18, intron 18, and exon 19), chromatin context, and tissue-specific regulatory elements, are not possible. Some splicing events observed in this system may not solely reflect the effect of the variant itself. Future studies using midigenes, which better preserve the native genomic context, would be valuable [9]. Besides, certain products, such as increased backbone-derived transcripts, likely arise from the artificial architecture of the construct, justified further transcriptomic analysis.

Long-read RNA sequencing proved essential for in vivo validation and for resolving the complexity of splicing outcomes through direct analysis of full-length *FOXP1* transcripts. Several abnormal *FOXP1* isoforms were detected, including exon skipping, partial intron retention, and cryptic splice-site usage, co-existing with normal transcripts. This isoform-level resolution cannot be achieved with short-read RNA-seq or RT-PCR, confirming the added value of long-read sequencing in resolving ambiguous splicing events [6]. These splicing outcomes were consistent with a loss-of-function mechanism, supporting *FOXP1* haploinsufficiency as the pathogenic mechanism [7].

Although *FOXP1* is not maximally expressed in peripheral blood, transcript levels were sufficient to allow informative analysis, supporting the feasibility of blood-based RNA studies in a diagnostic context (4.6 Transcripts Per Kilobase Million (TPM) on GTEX but detectable on our internal RNAseq controls). A skin biopsy was considered, but the detection of transcripts in the blood was weighed against the discomfort caused to the patient. This observation aligns with previous studies showing that peripheral blood RNA can provide meaningful splicing data for many neurodevelopmental genes [10]. Integration of *de novo* occurrence (PS2), functional assays (PS3), and transcript-level evidence, added with absent from population databases (PM2), provided strong support for pathogenicity according to current ACMG/AMP and RNA-based interpretation frameworks.

Taken together, this study shows that no single method is sufficient to resolve the functional impact of complex non-coding splice-altering variants. Instead, integration of genome sequencing, functional assays, and isoform-resolved transcriptomic analysis enables robust mechanistic interpretation and confident clinical classification. As genome sequencing continues to expand the landscape of detected non-coding variation, such integrative workflows will be increasingly important for reducing diagnostic uncertainty and improving the clinical utility of genomic data.

Acknowledgements
The authors acknowledge the patient and his family. We thank Alban Lermine and Pierre Blanc from the SeqOIA genomic platform for generating the genome data.

Author contributions
Conceptualization: PP, TS; Formal Analysis: PP, AJ, CT, AC, MT, PB, SC, TS; Funding acquisition: TS; Investigation: PP, AJ, CT, TS; Methodology: PP, AJ, CT, TS; Project administration: TS; Software: PP, EAY, JPM; Supervision: TS; Validation: PP, AJ, CT, AC, MT, PB, EAY, SC, TS; Visualization: PP, AJ, CT; Writing – original draft: PP, TS; Writing – review & editing: PP, TS.

Data availability
The data that support the findings of this study are available from the corresponding author upon reasonable request. Processed isoform annotations and summary splicing count tables can be shared to facilitate reproducibility; raw long-read sequencing data cannot be publicly deposited due to consent and privacy restrictions.

Declarations

Competing interests
The authors declare no competing interests.

Received: 23 January 2026 / Accepted: 6 March 2026
Published online: 13 March 2026

References

1. Jaganathan K, et al. Predicting splicing from primary sequence with deep learning. Cell. 2019;176:535-548.e24.
2. Leman R, et al. SPiP: splicing prediction pipeline, a machine learning tool for massive detection of exonic and intronic variant effects on mRNA splicing. Hum Mutat. 2022;43:2308–23.
3. Yeo G, Burge CB. Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals. J Comput Biol. 2004;11:377–94.
4. Anna A, Monika G. Splicing mutations in human genetic disorders: examples, detection, and confirmation. J Appl Genet. 2018;59:253–68.
5. Gaildrat P. Use of splicing reporter minigene assay to evaluate the effect on splicing of unclassified genetic variants. In: Webb M, editor. Cancer susceptibility, vol. 653. Totowa: Humana Press; 2000.
6. Damaraju N, Miller AL, Miller DE. Long-read DNA and RNA sequencing to streamline clinical genetic testing and reduce barriers to comprehensive genetic testing. J Appl Lab Med. 2024;9:138–50.
7. Sollis E, et al. Identification and functional characterization of de novo FOXP1 variants provides novel insights into the etiology of neurodevelopmental disorder. Hum Mol Genet. 2016;25:546–57.
8. Jang W, Park J, Chae H, Kim M. Comparison of in silico tools for splice-altering variant prediction using established spliceogenic variants: an end-user's point of view. Int J Genomics. 2022;2022:5265686.

9. Sangermano R, et al. ABCA4 midigenes reveal the full splice spectrum of all reported noncanonical splice site variants in Stargardt disease. Genome Res. 2018;28:100–10.
10. Dekker J, et al. Web-accessible application for identifying pathogenic transcripts with RNA-seq: increased sensitivity in diagnosis of neurodevelopmental disorders. Am J Hum Genet. 2023;110:251–72.

Publisher's Note