

Assessing the clinical utility of a breakpoint analysis method in identification of structural variants in targeted sequencing genetic tests

Affiliation: Blueprint Genetics, Espoo, Finland: Lotta Koskinen, Margarita Andreevskaya, Antti Väisänen, Mikko Muona, Tuuli Pietilä, Janica Djupsjöbacka, Ville Kytölä, Kati Kämpjärvi, Samuel Myllykangas, Pertteli Salmenperä, Juha Koskenvuo, Miko Valori.

P15.097.B

Introduction

Structural variants (SVs) are a diverse group of alterations that includes deletions, duplications, insertions (novel or retrotransposon element [RTE]), inversions, translocations, and other complex rearrangements. SVs are defined as impacting at least 50 bp of genomic sequence. They can affect gene copy number, structure, and regulation¹. Even though SVs are rare compared with single-nucleotide variants, an estimated 10% to 25% of them are pathogenic variants, making their efficient detection clinically relevant². However, certain categories of SVs are typically not detected by commonly used exome variant callers but may be detected by breakpoint analysis³. We assessed the frequency and types of clinically relevant SVs identified using a breakpoint analysis method for targeted NGS gene panels.

Methods

Samples referred to panel testing were target-sequenced using whole exome- and clinical exome-based assays and a short-read Illumina NovaSeq 6000 platform at a Blueprint Genetics laboratory. SV calling was done using a read-breakpoint analysis method developed in-house for clinical use. The method combined read-pair orientation-based (RP) and split-read/clipped-read (SR)-based methods. The method can detect deletions, tandem duplications, and inversions > 50 bp, as well as RTE insertions such as Alu, SVA, and L1 with resolution of 1 to a few base pairs. Rare events with a control population frequency of < 0.001 were evaluated further; only reported variants classified as pathogenic (P), likely pathogenic (LP), or of uncertain significance (VUS) that were not detected by other variant calling methods were included in this analysis. Reported SVs were confirmed using quantitative PCR, digital PCR, long-range PCR, Sanger sequencing or additional bioinformatic analyses.

Variant interpretation was performed in accordance with ACMG/AMP guidelines. A diagnostic finding was defined as the identification of a pathogenic (P) or likely pathogenic (LP) variant(s) consistent with the patient's reported phenotype and disease inheritance.

Results

In the analysis of order-specific target regions of more than 80,000 samples, we detected 162 clinically relevant SVs (97 unique) using the breakpoint analysis method. These variants were not detected with other variant calling methods. They constituted 0.2% of all panel orders processed and 0.3% of the orders with reported findings.

Of the clinically relevant SVs detected with this method, 79% contributed to molecular diagnosis and 21% were considered a carriership related to an autosomal recessive disorder. Of these SVs, 85% were classified as pathogenic or likely pathogenic; 15% were classified as variants of uncertain significance.

RTE insertions were the most common type of SV, and Alu insertions were the most common type of RTE identified (74% of all unique RTEs). The proportions of unique SVA, L1, and unknown RTE types are the same for each: 8.5% (3 of each). The most commonly identified RTE insertions were 2 Alu insertions in retinal dystrophy-related genes: MAK c.1297_1298insAlu and RP1 c.2321_2322insAlu.

The second most commonly identified SV type was deletions (39 in total, unique 32) with the length of 77 to 27,506 bp and a median length of 1,355 bp (for unique findings). These deletions included almost exclusively partial exons. Genes with 3 or more unique SVs were USH2A (7), RP1 (4), CHM (3), and NF1 (3).

Complex SVs are not directly annotated by the breakpoint script but can be identified by inspection of other called events. Fourteen complex events were identified, all of them were unique. Eight complex events were of a similar deletion-insertion-inversion type, where the inserted sequence was derived either from the deleted sequence or up-/downstream of the deletion and it was inverted.

Read breakpoint analysis increases diagnostic yield even in targeted sequencing and may provide diagnoses which are missed by standard bioinformatic approaches.

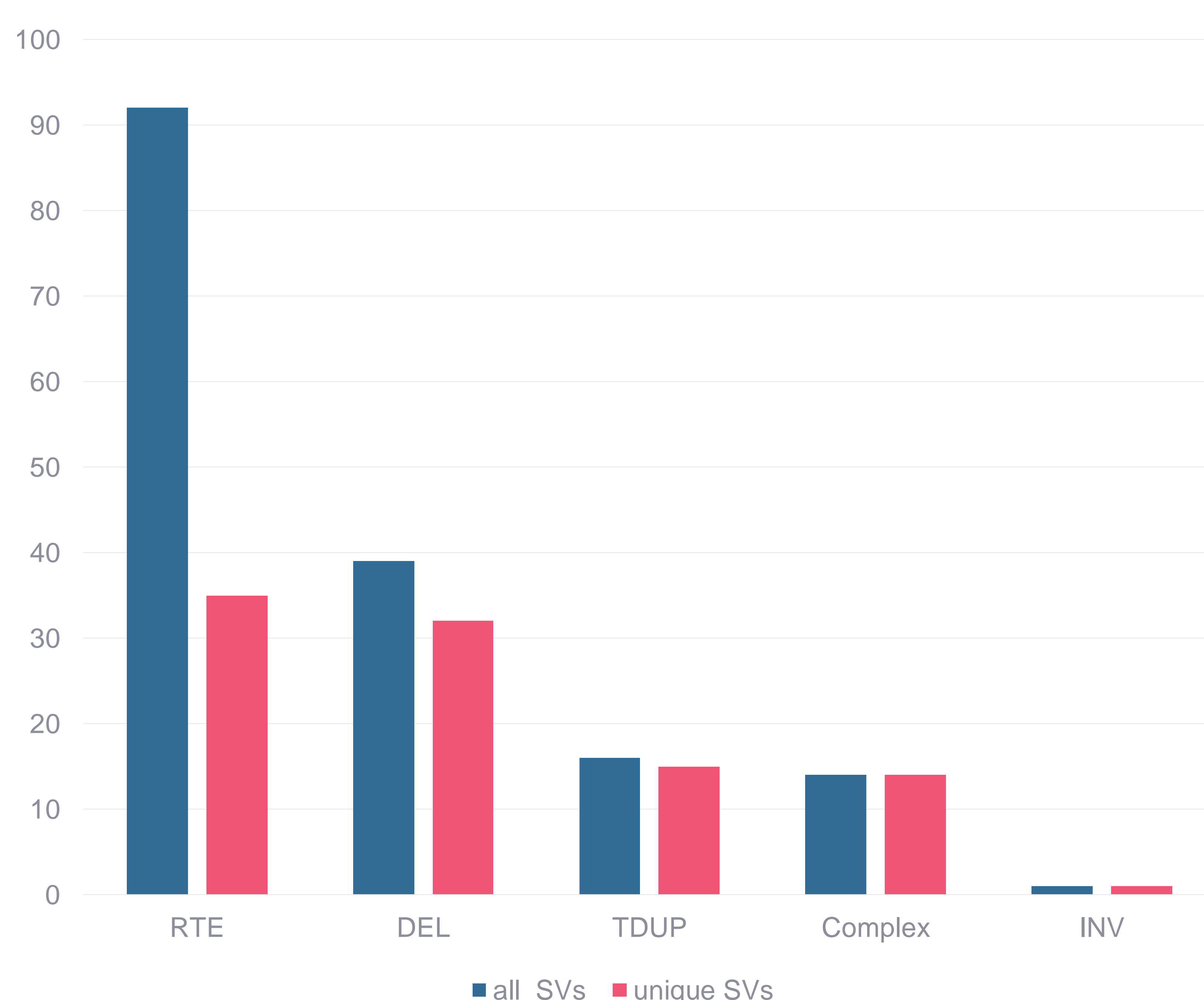


Figure. The number of different structural variants (SVs) detected by the breakpoint analysis, but not detected by other variant calling methods.

RTE: retrotransposon element; DEL: deletion; TDUP: tandem duplication; Complex: complex events; INV: inversion

Conclusions

- Breakpoint SV detection based on RP and SR approaches enables the detection of certain categories of clinically relevant SVs that are not recognized by other commonly used exome variant callers; this method does this despite having limitations for exome sequencing compared to genome sequencing.
- The most commonly identified SV types were RTE insertions and partial-exon deletions.
- Breakpoint SV detection increases diagnostic yield and could provide answers to patients with previous uninformative genetic testing results.
- Besides detecting additional SVs, breakpoint analysis allows defining exact breakpoints for the called CNVs, which can be used to design more specific confirmation tests or even decrease the number of confirmation tests by providing additional orthogonal evidence to CNV:s detected with the coverage-based method.
- In rare cases, the breakpoint analysis allowed us to split a CNV into 2 smaller adjacent events.

References:

1. Guan P, Sung WK. Structural variation detection using next-generation sequencing data: A comparative technical review. *Methods*. 2016;102:36-49. doi:10.1016/j.jmeth.2016.01.020
2. van den Akker J, Hon L, Ondov A, et al. Intronic Breakpoint Signatures Enhance Detection and Characterization of Clinically Relevant Germline Structural Variants. *J Mol Diagn*. 2021;23(5):612-629. doi:10.1016/j.jmoldx.2021.01.015
3. Tattini L, D'Aurizio R, Magi A. Detection of Genomic Structural Variants from Next-Generation Sequencing Data. *Front Bioeng Biotechnol*. 2015;3:92. doi:10.3389/fbioe.2015.00092