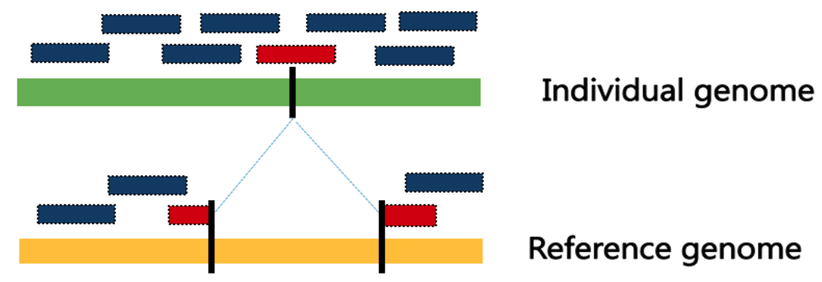
Identifying pathogenic long indels in patients with intellectual disability

Sander Pajusalu[1, 2], Rolph Pfundt[3], Lisenka E.L.M. Vissers[3], Michael P. Kwint[3], Tiia Reimand[1,2,4], Katrin Õunap[1,4], Joris A. Veltman[3,5], Jayne Y. Hehir-Kwa[3]

1. Department of Genetics, United Laboratories, Tartu University Hospital, Tartu, Estonia, 2. Department of Biomedicine, Institute of Biomedicine and Translational Medicine, University of Tartu, Tartu, Estonia, 3. Department of Human Genetics, Donders Institute, Radboudumc, Nijmegen, The Netherlands, 4. Department of Pediatrics, Institute of Clinical Medicine, University of Tartu, Tartu, Estonia, 5. Department of Clinical Genetics, Maastricht University Medical Centre, Maastricht, The Netherlands

Aim: We analysed the exome sequencing (WES) data from 98 patients with intellectual disability (ID) to identify pathogenic long indels located in exons of 650 ID genes[1] using Pindel[2] and Platypus[3]. All patients had previously screened negative for WES based pathogenic single nucleotide, small indel and large copy number variants.



Background: Sensitive methods exist for the identification of single nucleotide variants and indels less than 20bp in size from NGS data. The discovery of long indels 20 - 200bp in size is challenging even in non-repetitive regions of the genome such as exons. As a result the role of small deletions (long indels) is currently unknown and under reported in many studies.

Results: First we calculated a detection sensitivity of 74% based on 26 common exonic indels from a public dataset[4] (**Fig 1**). Then analysis of rare variants within the patient cohort identified two clinically relevant indels (diagnostic yield 2%) – a 42bp homozygous deletion of exon-intron border in *PGAP3* gene (**Fig 2**), and a 115bp heterozygous complex indel disrupting the *MECP2* gene (**Fig 3**).

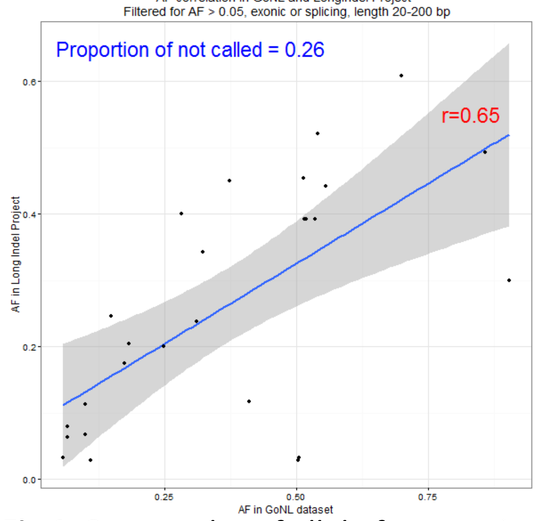


Fig 1. Scatter plot of allele frequencies (AF) found in this study and GoNL dataset.

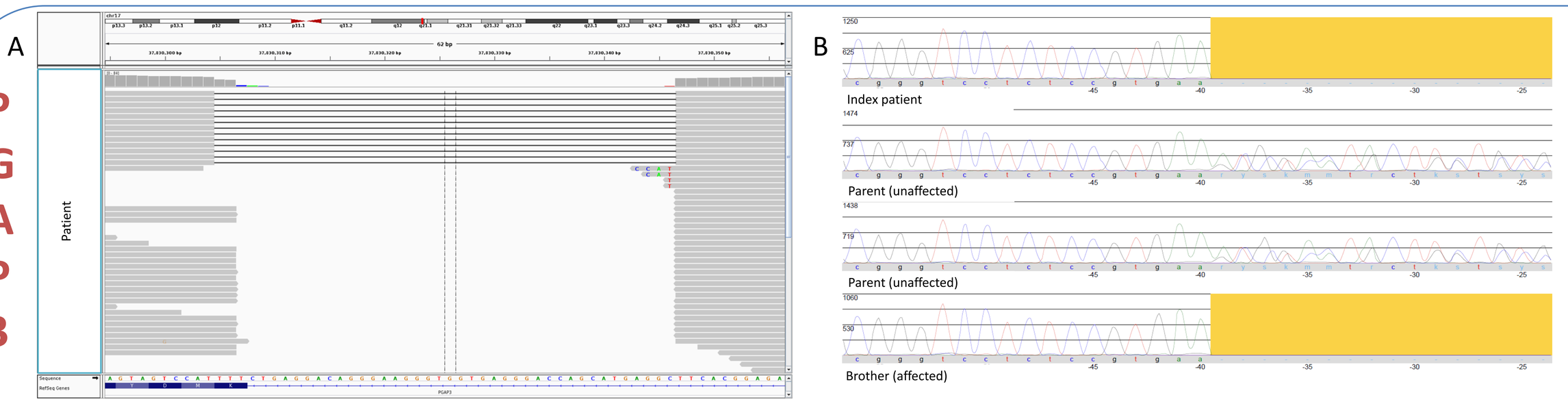


Fig 2. 42bp homozygous exon-intron deletion (c.496-39_498del) in the *PGAP3* gene. A) Identified in WES data by Pindel, Platypus and Haplotype Caller, but not Unified Genotyper; B) was validated via Sanger. The patient has severe ID, microcephaly, epilepsy, and palatoschisis compatible with *PGAP3*-related hyperphosphatasia with ID syndrome (OMIM #615716). His parents are first cousins and have another child with a similar phenotype, and the same mutation.

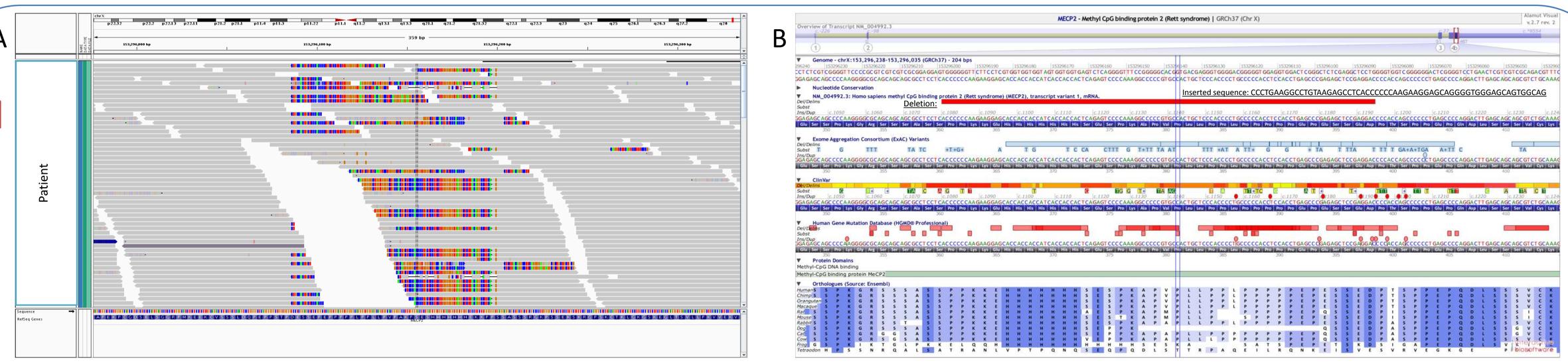


Fig 3. 115bp indel in *MECP2* gene: c.1080_1194delins60 p.(Pro362Glufs*4). A) The indel was identified in WES data by Pindel, but not Platypus, Haplotype Caller or Unified Genotyper; B) in a mutation hotspot region. The patient is a 3 year-old girl with severe ID, epilepsy, and hypotonia which can be explained by the disruption of *MECP2* gene and associated Rett syndrome (OMIM #312750). The mutation was confirmed by Sanger sequencing.

Conclusion. WES data is fragmented making long indel identification difficult. However, using specific software tools for long indel detection increases diagnostic yield in patients with ID.

References:

1. http://www.genomediagnosticsnijmegen.nl/index.php/en/services/exome-sequencing-diagnostics
2. Ye, K., Schulz, M. H., Long, Q., Apweiler, R. & Ning, Z. Pindel: a pattern growth approach to detect break points of large deletions and medium sized insertions from paired-end short reads. Bioinformatics 25, 2865–71 (2009).
3. Rimmer A, Phan H, Mathieson I, Iqbal Z, Twigg SR, Wilkie AO, et al. Integrating mapping-, assembly- and haplotype-based approaches for calling variants in clinical sequencing applications. Nat Genet.2014;46(8):912–8.
4. Hehir-Kwa, J., Marschall, T., Kloosterman, W., et al. A high-quality reference panel reveals the complexity and distribution of structural genome changes in a human population. Biorxiv 2016

Contact information:
Sander Pajusalu, MD
Sander.Pajusalu@kliinikum.ee

Handout:

Funding: SP was supported by national scholarship program Kristjan Jaak (Archimedes Foundation & Ministry of Education and Research of Estonia). JHK was supported by NWO 016.166.015. JV was supported by ERC DENOVO 281964.