Long-Read Sequencing of a *STRC* deletion

Maria Wong[1,2], Maisie Dantuma[3], Henry Keen[3], Hela Azaiez[1], Richard JH Smith[1,2,3]

[1]Molecular Otolaryngology and Renal Research Laboratories, Department of Otolaryngology, University of Iowa, Iowa City, IA 52242, USA; [2]Interdisciplinary Graduate Program in Informatics, University of Iowa, Iowa City, IA 52242, USA; [3]Iowa Institute of Human Genetics, University of Iowa, Iowa City, IA 52242, USA

Introduction

- *STRC* mutations account for ~13% of all positive diagnoses, making genetic variation in this gene the most common cause of autosomal recessive mild-to-moderate hearing loss in the non-Finnish European population.
- Due to the presence of a large tandem duplication, the most common pathogenic change is a copy number variant (CNV).
- The most prevalent CNV is a contiguous gene deletion that includes *STRC* and the neighboring gene, *CATSPER2*.
- Biallelic *STRC-CATSPER2* deletions result in Deafness Infertility Syndrome (DIS) in males and DFNB16 in females.
- Short-read sequencing is inadequate to resolve the complexities of this region.
- We explore using long-read sequencing (LRS) technology to characterize this genomic region.

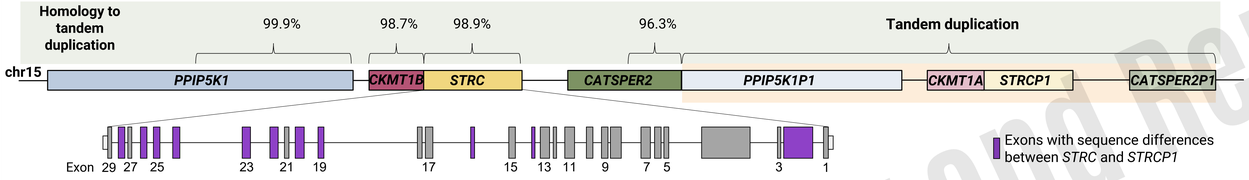


Figure 1. Schematic of *STRC* region. The tandem duplication involves 4 genes. Of the 29 exons in *STRC*, 18 (62%) are ***identical*** to the pseudogene *STRCP1*.

Subject and Methods

Subject
A child of consanguineous parentage (parents are second cousins once removed) with homozygous contiguous gene deletions of *STRC* and *CATSPER2* identified through gene panel testing (OtoSCOPE®)

Genetic Testing
We completed sequencing using:

1- Oxford Nanopore Technologies (ONT) with adaptive sampling to evaluate the size and breakpoints of this deletion. Read alignment and variant calling were performed using ONT bioinformatics pipeline, with manual alignment where necessary.

2- Whole genome sequencing (WGS) was done with Illumina short-read sequencing and analyzed with Illumina DRAGEN software.

Bioinformatic Analysis Pipeline for ONT Data

Long Read Data
Super Accurate Basecalling

wf-cas9
Generate statistics/plots of coverage of targets
Alignment to Reference genome = minimap2

samtools → IGV for visualization (See Figure 5)

wf-human-variation
Variant Calling:
Structural Variant = Sniffles2
CNV = QDNASeq & Spectre
SNP = Clair3
Variant Phasing: WhatsHap

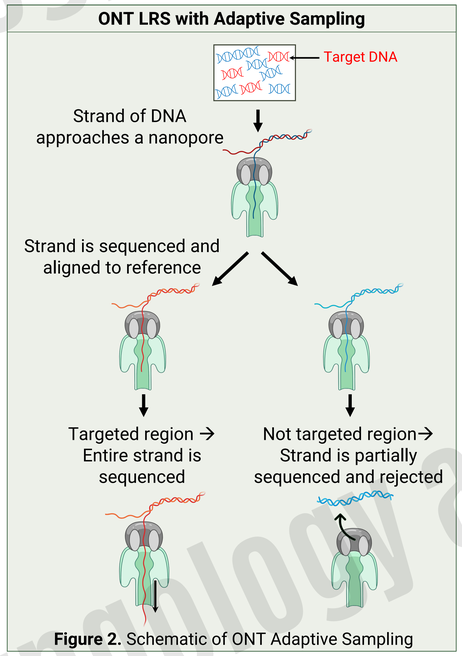
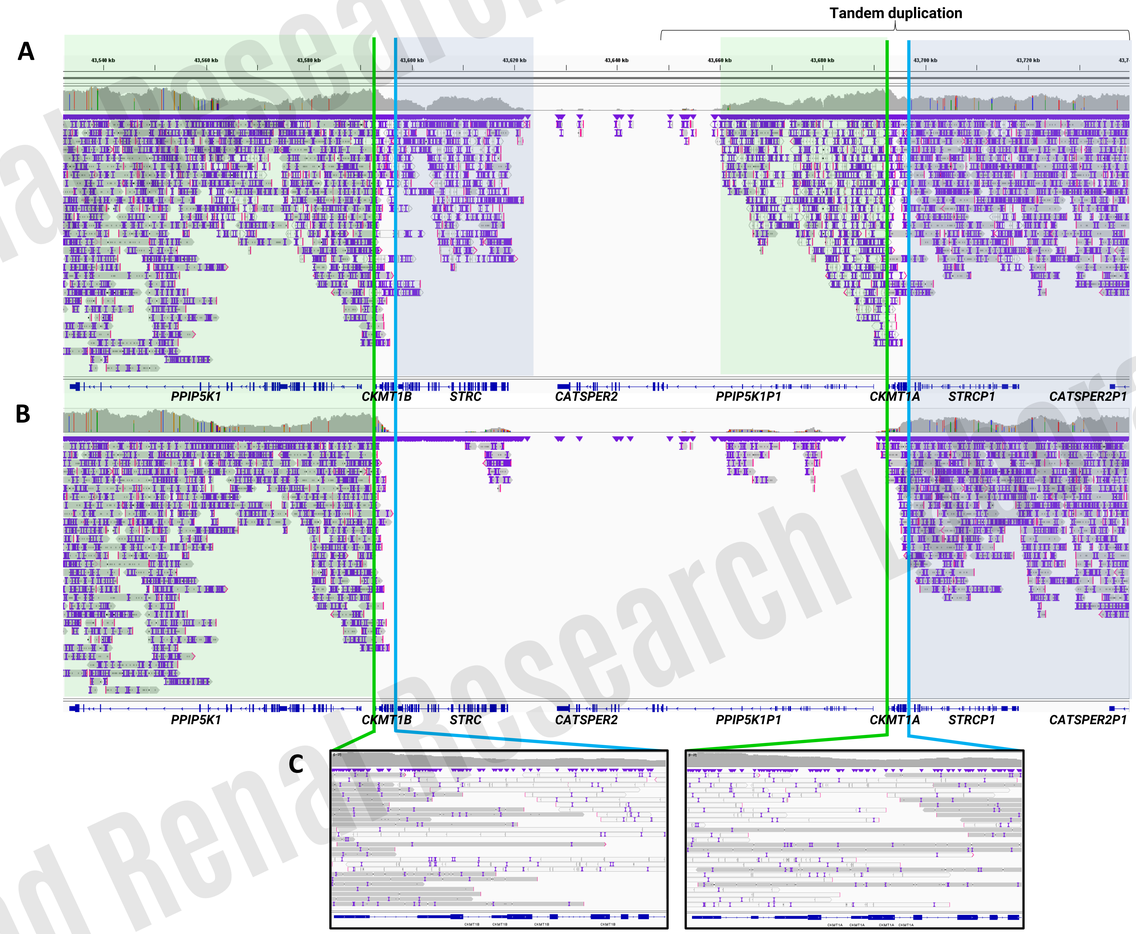


Figure 2. Schematic of ONT Adaptive Sampling

Long-Read Sequencing Results



Identical region between gene (*CKMT1B*) and duplicated copy (*CKMT1A*).

Figure 5. Breakpoint mapping with ONT long-read sequencing. **A.** Long-read sequencing reads were aligned using minimap2 and visualized in IGV. Despite an average read length of ~6.5 kb, alignment of highly homologous regions was challenging, and many reads had a mapping quality score of 0, indicating low confidence in their placement. **B.** Filtering out poorly mapped reads provided a clearer picture of the deletion in this region. Structural variant callers failed to detect the large deletion and its breakpoints, however manual inspection of sequence differences between the gene and its duplicated copy (highlighted in green and blue) revealed that green-marked reads align to the gene, while blue-marked reads map to the duplicated copy. This finding suggests that the deletion breakpoints are located between these sequence differences. **C.** A comparison of the two regions shows that the sequences of the gene and its duplicated copy are 100% identical.

STRC Mutation Prevalence by Type

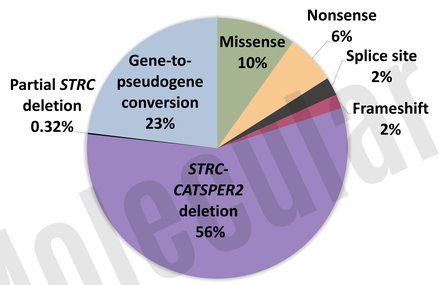


Figure 3. The majority of *STRC* mutations are CNVs. Of the 305 positive diagnoses of *STRC*-related hearing loss in our cohort, we identified 347 copies of *STRC-CATSPER2* deletions (116 homozygous).

Patient CNV Analysis

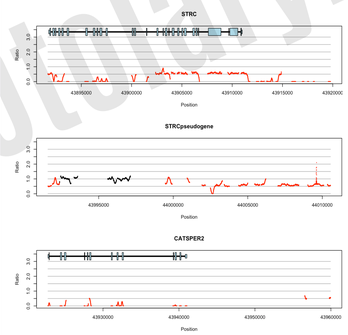


Figure 4. CNV analysis image from OtoSCOPE® targeted panel testing of proband's homozygous *STRC-CATSPER2* deletion.

Illumina WGS vs ONT LRS

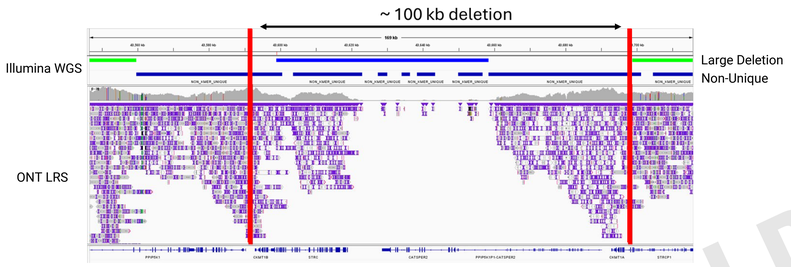


Figure 6. Detection of STRC deletion using Illumina depth-based DRAGEN CNV caller. The DRAGEN CNV caller identified a ~59 kb deletion (denoted by the top blue bar) in *STRC* using short-read whole-genome sequencing (WGS) data. However, the pipeline excludes genomic regions that are not uniquely mappable, leading to ambiguities in CNV analysis (shown in dark blue bars). When comparing the CNV results from WGS with LRS data, the reported deletion from WGS does not extend into the highly homologous regions as effectively as LRS data (highlighted in red), suggesting that short-read sequencing underestimates the full extent of the deletion.

Conclusions

- LRS showed that the CNV deletion in this patient extended beyond *STRC-CATSPER2*, affecting adjacent genes and tandem duplication regions.
- While ONT Technology allowed for the sequencing of larger continuous regions, alignment challenges persisted due to mis-mapping between the genes and their duplicated counterparts.
- Structural variant callers could not identify precise deletion breakpoints; however, manual inspection suggested the breakpoints lie within completely identical regions between *CKMT1B* and its tandem duplicate, *CKMT1A*.
- **This study highlights that while LRS enhances resolution in the *STRC* region, genomic complexities continue to hinder precise breakpoint mapping.**

References

Loose M, Malla S, Stout M. Real-time selective sequencing using nanopore technology. *Nat Methods*. 2016;13(9):751-754. doi:10.1038/nmeth.3930 · Shearer AE, Kolbe DL, Azaiez H, et al. Copy number variants are a common cause of non-syndromic hearing loss. *Genome Med*. 2014;6(5):37. Published 2014 May 22. doi:10.1186/gm554 · Sloan-Heggen CM, Bierer AO, Shearer AE, et al. Comprehensive genetic testing in the clinical evaluation of 1119 patients with hearing loss. *Hum Genet*. 2016;135(4):441-450. doi:10.1007/s00439-016-1648-8 · Figure 2 created with BioRender.com

This study was supported in part by NIDCDs R01s DC003544, DC002842 and DC012049 to RJHS. Please send questions to Maria Wong: maria-wong@uiowa.edu || https://morl.lab.uiowa.edu/