APPLICATION NOTE

Enhanced insights into hearing loss genetics using Illumina Complete Long Reads

Advanced DRAGEN™ analysis pipeline increases insights for unresolved disease

illumina®

Introduction

For most children with congenital hearing loss, the causes are genetic and highly heterogeneous with over 125 associated genes.[1] Whole-exome sequencing (WES) has been successfully used to identify the genomic underpinnings in ~30% of hearing loss cases.[2] Whole-genome sequencing (WGS) can raise this percent by helping detect more genetic variants associated with hearing loss. However, some genomic regions with known associations to hearing loss, such as the *STRC* locus, present mapping challenges for standard, short-read WGS applications alone.[1]

To help resolve difficult-to-map regions, researchers can use Illumina Complete Long Read Prep, Human or Illumina Complete Long Read Prep with Enrichment, Human Comprehensive Panel as a reflex solution to complement standard short-read WGS for enhanced insights.[3–5] Illumina Complete Long Reads can generate highly accurate long reads with N50 of 6–7 kb for low DNA inputs and various sample types (Figure 1).[6]

Eliot Shearer, MD, PhD and the Translational Hearing Genomics Lab at Boston Children's Hospital and Harvard Medical School study the genetics of hearing loss and work to expand genomic testing for families with suspected genetic pediatric hearing loss. Illumina scientists assisted Dr Shearer and Boston Children's Hospital, using Illumina Complete Long Reads to sequence samples from 10 children with sensorineural hearing loss unexplained by previous exome sequencing and short-read WGS tests.

This application note demonstrates how the DRAGEN Illumina Complete Long Reads analysis pipelines for WGS and enrichment can help identify variants related to hearing loss. These pipelines use a comprehensive suite of analytical tools including aligners, variant callers, and specialized targeted callers to leverage both long and short reads for variant detection.

Methods

Boston Children's Hospital clinicians provided genomic DNA isolated from blood or saliva samples from 10 children with sensorineural hearing loss who were negative when analyzed using short-read WGS. Eight of the 10 samples, including BCH-01, BCH-02, and BCH-03, were blood-derived samples. DNA was extracted from 1–2 ml blood, collected in 3.0-ml BD Vacutainer plastic tubes with K2 EDTA 5.4 mg (Becton, Dickinson, Catalog no. 366473), using the QIAamp DNA Blood Midi kit (Qiagen, Catalog no. 51185). Two of the 10 samples were derived from saliva. DNA was collected from saliva using the Oragene-Discover Collection Kit (DNA Genotek, Catalog no. OGR-500) and extracted using the prepIT-L2P kit (DNA Genotek, Catalog no. PT-L2P-5).

Libraries were prepared from 50 ng DNA per sample using Illumina Complete Long Read Prep, Human (Illumina, Catalog no. 20089108) or Illumina Complete Long Read Prep with Enrichment, Human Comprehensive Panel (Illumina, Catalog no. 20113834) plus 300 ng DNA per sample using Illumina DNA PCR-Free Prep (Illumina, Catalog no. 20041794).

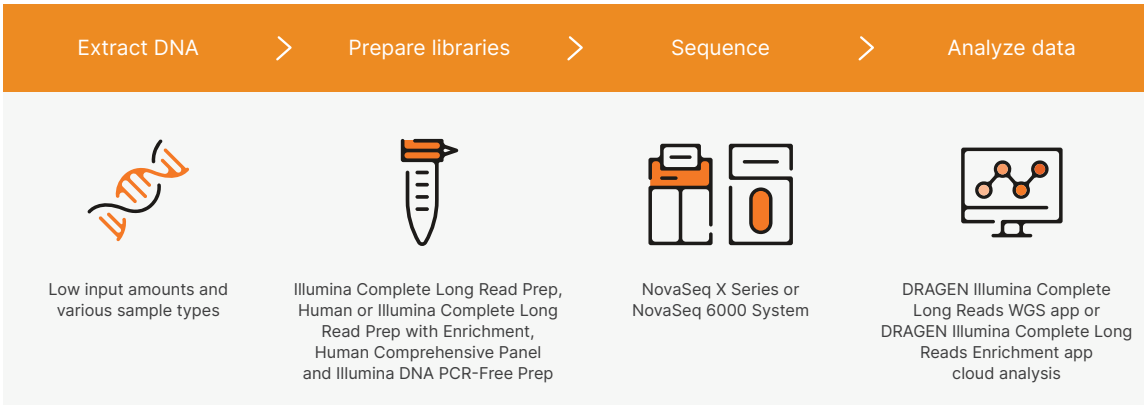

Figure 1: Illumina Complete Long Reads workflow—A standard next-generation sequencing (NGS) workflow with DNA extraction, library preparation, sequencing, and data analysis generates highly accurate long reads with robust and flexible performance for low DNA inputs and various sample types.

Illumina Complete Long Reads libraries were sequenced on the NovaSeq™ 6000 System using 2 × 151 bp read lengths and ~750 Gb per Illumina Complete Long Reads WGS sample and an average of 90–120 Gb per Illumina Complete Long Reads with Enrichment, Human Comprehensive Panel sample. Illumina DNA PCR-Free libraries were sequenced on the NovaSeq 6000 System using 2 × 151 bp read lengths and an average of 200 Gb per sample to achieve ≥ 30× coverage.

Secondary analysis of Illumina Complete Long Reads data combined with Illumina DNA PCR-Free data was performed by an automated DRAGEN Illumina Complete Long Reads WGS application* or DRAGEN Illumina Complete Long Reads Enrichment application† on BaseSpace™ Sequence Hub, resulting in variant calling merged files for both small variants and structural variants (Figure 2).[7]

* DRAGEN Illumina Complete Long Reads WGS app is called "DRAGEN ICLR WGS (Illumina Complete Long Reads)" on BaseSpace Sequence Hub.

† DRAGEN Illumina Complete Long Reads Enrichment app is called "DRAGEN ICLR Enrichment" on BaseSpace Sequence Hub.

Results

Illumina Complete Long Reads generated high-quality data with N50 of 6–7 kb (Table 1). The DRAGEN Illumina Complete Long Reads WGS and Enrichment apps used v4.2 of DRAGEN secondary analysis and merged data from Illumina Complete Long Reads libraries with matched Illumina DNA PCR-Free short-read WGS libraries. This analysis provided enhanced insights for three of the 10 patient samples, revealing variants associated with the hearing loss phenotype not identified by WES or standard short-read WGS alone.

- A ~58-kb deletion and hemizygous G>A stop variant in *STRC* in sample BCH-01 (Figure 3)
- A compound heterozygous 667-bp deletion and A>G missense variant in *OTOA* in sample BCH-02 (Figure 4)
- A 403-kb inversion interrupting *MITF* in sample BCH-03 (Figure 5)

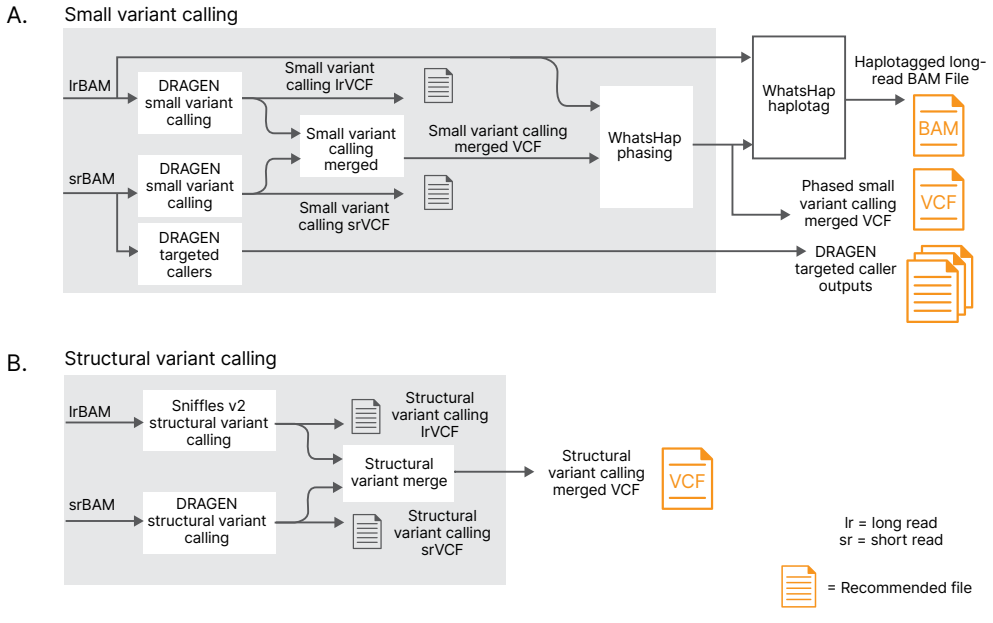

Figure 2: Illumina Complete Long Reads secondary analysis and file output—(A) For small variant calling, long and short reads are aligned separately and results are combined with machine learning logic to optimize variant calling. Long reads and merged small variants are phased using a phasing tool. (B) Long and short reads are separately used to perform structural variant (SV) calling with dedicated SV callers and results are merged using advanced logic to create a new, merged SV VCF file.[7]

Table 1: Sequencing metrics for Illumina Complete Long Reads for genetic hearing loss study

	Illumina Complete Long Read Prep, Human	Illumina Complete Long Read Prep with Enrichment, Human Comprehensive Panel
Input DNA	50 ng	50 ng
N50	6–7 kb	5–7 kb
Reads > 10 kb	5%–9%	Average 3%
Mean coverage	~30×	~37×
% Heterozygous SNVs phased[a]	> 98%	98%[b]
Mean SNVs identified	4.04M	4.01M
Mean indels identified[a]	960K	948K
Mean SVs identified[a]	24K	19K

a. SNVs, single nucleotide variants; indels, insertion–deletions; SVs, structural variants.
b. On-target.

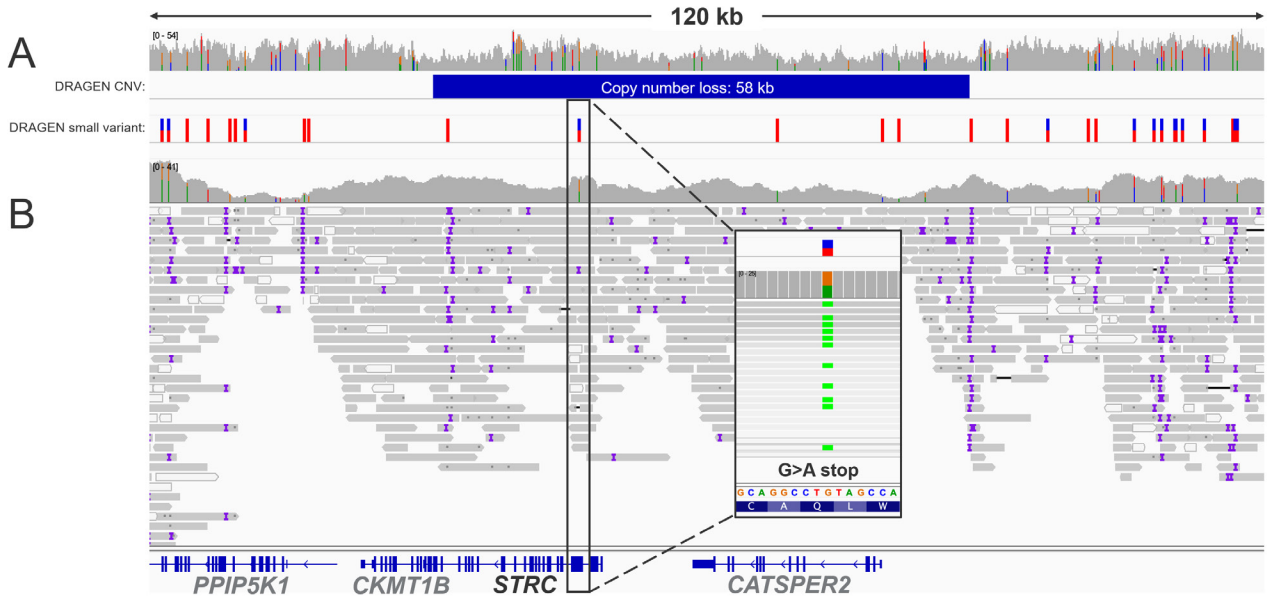

Figure 3: Identification of pathogenic deletion and hemizygous variant in *STRC* in sample BCH-01—A ~58 kb deletion and G>A stop variant are shown in *trans*. (A) The depth-based DRAGEN copy number variant (CNV) caller (variant called in the file <*>.cnv.vcf.gz) using the short-read data detected a ~58 kb deletion at chr15:43,600,595-43,658,379 covering *STRC*. (B) The DRAGEN small variant caller (variant called in the file <*>.combined_iclr_sbs.phased.vcf.gz) identified a heterozygous G>A mutation at chr15:43,616,338. This exon is challenging to align due to a highly similar paralog. The zoomed-in box on the G>A variant illustrates the read alignments shaded by mapping quality, high to low. The reads with low MAPQ (no variant) should map to the paralog. The copy number loss supports the zygosity of G>A stop variant as hemizygous.

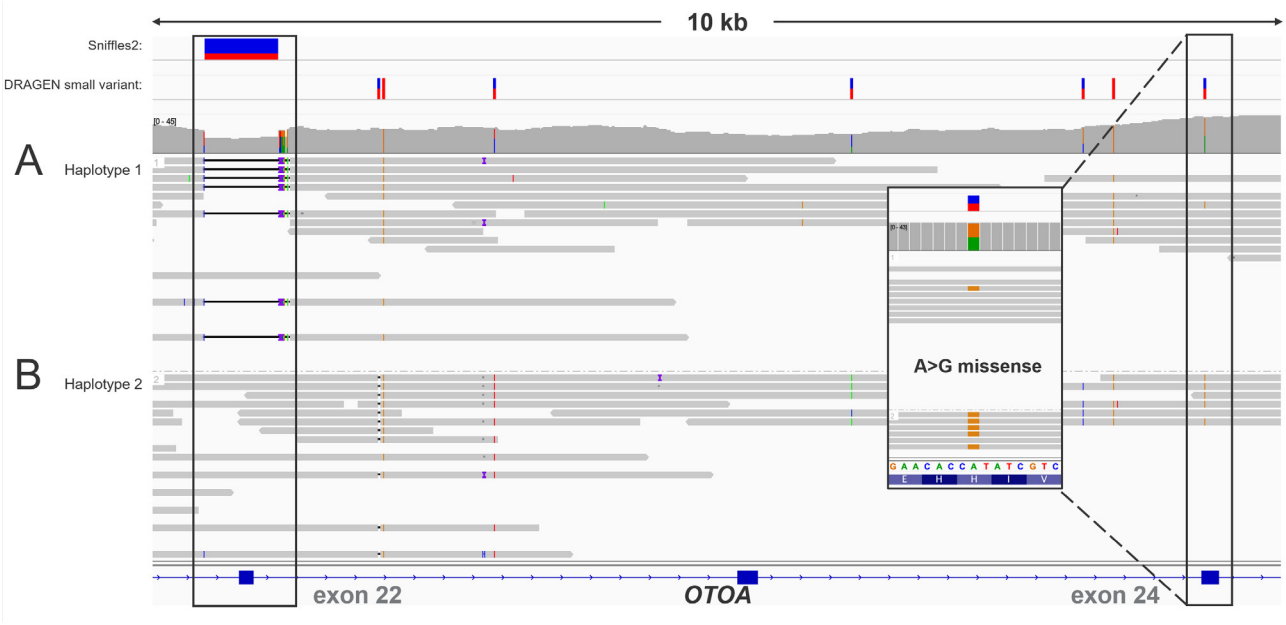

Figure 4: Identification of compound heterozygous pathogenic variants in *OTOA* in sample BCH-02—A 667-bp deletion and A>G missense variant are shown in *trans*. (A) In haplotype 1, Illumina Complete Long Reads structural variant caller, Sniffles2, (variant called in the output file <*>.combined_iclr_sbs.sv.vcf.gz) detected a 667-bp deletion at chr16:21,735,952-21,736,619 covering exon 22 of *OTOA*. (B) In haplotype 2, DRAGEN small variant caller (variant called in the output file <*>.combined_iclr_sbs.phased.vcf.gz) identified a heterozygous A>G mutation at chr16:21,744,915, a missense variant in an exon that is challenging to map with short reads. Haplotype unassigned reads not shown.

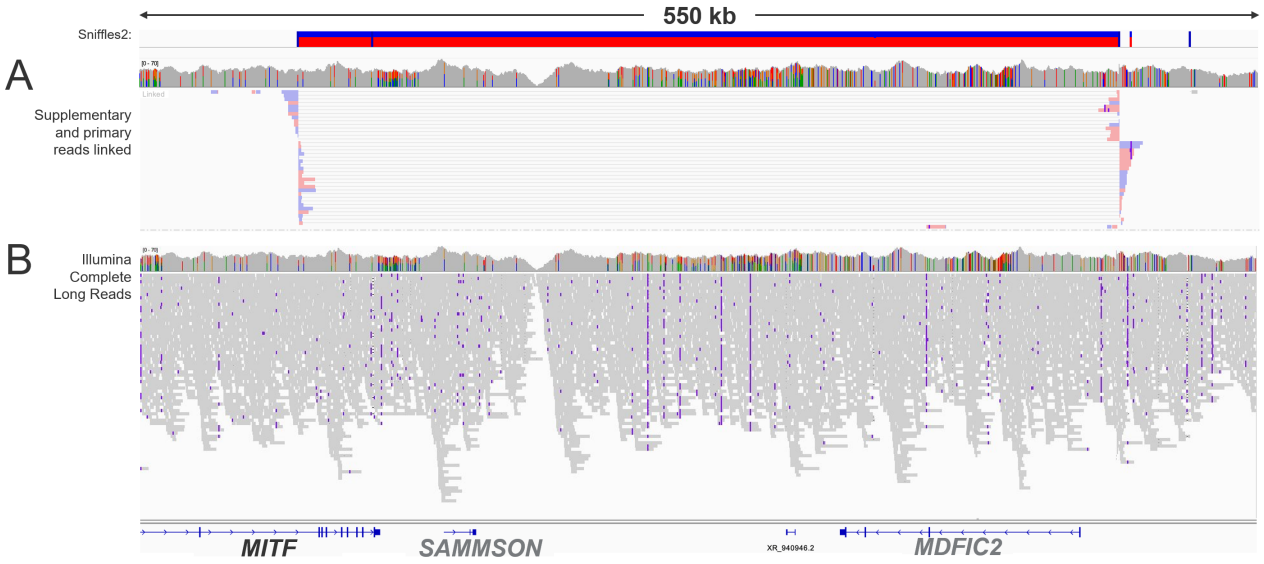

Figure 5: Identification of inversion interrupting *MITF* in sample BCH-03—The DRAGEN Illumina Complete Long Reads structural variant caller, Sniffles2, (variant called in the output file <*>.combined_iclr_sbs.sv.vcf.gz) detected a 403-kb heterozygous inversion at chr3:69,927,919-70,331,154 interrupting *MITF*. (A) Supplementary alignments linked to primary reads, (B) ungrouped read alignment.

Discussion

STRC variants identified for sample BCH-01

The comprehensive suite of DRAGEN secondary analysis tools was used to analyze Illumina Complete Long Reads and enhance standard short-read WGS data, enabling detection of compound pathogenic variants in the challenging-to-map *STRC* gene region for sample BCH-01 (Figure 3). The depth-based DRAGEN copy number variant (CNV) caller using the short-read data detected a ~58 kb deletion. Sequence data from Illumina Complete Long Reads, from both whole-genome and enriched using the Human Comprehensive Panel, detected a heterozygous G>A stop variant on the other allele. However, the copy number loss and closer inspection of mapping quality support the zygosity of the call as hemizygous.

OTOA variants identified for sample BCH-02

The results for sample BCH-02 illustrate the benefits of Illumina Complete Long Reads to enhance variant calling in standard short-read WGS data and the importance of haplotype phasing for resolution of variants (Figure 4). The heterozygous 667-bp deletion in the *OTOA* gene was not detected in the PCR-free short-read WGS data, but this deletion was called using the Illumina Complete Long Reads SV caller, Sniffles2, in sequence data from Illumina Complete Long Reads WGS and Illumina Complete Long Reads with Enrichment, Human Comprehensive Panel libraries. On the other *OTOA* allele, the heterozygous A>G missense variant was also detected by the DRAGEN small variant caller in all three data sets: PCR-free short-read WGS data, Illumina Complete Long Read WGS data, and Illumina Complete Long Reads with Enrichment, Human Comprehensive Panel data.

MITF variant identified for sample BCH-03

The *MITF* inversion in sample BCH-03 was not detected by previous analysis with an alternative short-read WGS pipeline. However, the improved resolution of DRAGEN secondary analysis tools identified the *MITF* inversion in the PCR-free short-read WGS data and the event was detected by the Illumina Complete Long Reads SV caller, Sniffles2, in the Illumina Complete Long Reads WGS data (Figure 5). Note that the Human Comprehensive Panel only targets 0.65% of the *MITF* gene and consequently does not target the regions needed to enable detection of this variant by Illumina Complete Long Reads with Enrichment.

Summary

Illumina Complete Long Reads enables highly accurate sequencing of challenging regions of the genome from low input DNA. This study of genomic contributions to hearing loss demonstrates how Illumina Complete Long Reads and the comprehensive suite of DRAGEN analysis tools can enhance standard short-read WGS data to help identify disease-associated variants.

Learn more

Illumina Complete Long Reads

Illumina Complete Long Read Prep, Human

Illumina Complete Long Read Prep with Enrichment, Human

Illumina Human Comprehensive Panel

Human whole-genome sequencing

Long-read sequencing technology

References

1. Shearer AE, Hildebrand MS, Schaefer AM, et al. Genetic Hearing Loss Overview. 1999 [Updated 2023]. In: Adam MP, Feldman J, Mirzaa GM, et al., editors. GeneReviews® [Internet]. Seattle (WA): University of Washington, Seattle; 1993-2024. Available from: ncbi.nlm.nih.gov/books/NBK1434/
2. Perry J, Redfield S, Oza A, et al. Exome Sequencing Expands the Genetic Diagnostic Spectrum for Pediatric Hearing Loss. *Laryngoscope*. 2023;133(9):2417-2424. doi:10.1002/lary.30507
3. Illumina. Illumina Complete Long Read Prep, Human data sheet. illumina.com/content/dam/illumina/gcs/assembled-assets/marketing-literature/illumina-long-read-prep-human-data-sheet-m-gl-01420/illumina-long-read-prep-data-sheet-m-gl-01420.pdf. Published 2023. Accessed February 14, 2024.
4. Illumina. Illumina Complete Long Read Prep with Enrichment, Human data sheet. illumina.com/content/dam/illumina/gcs/assembled-assets/marketing-literature/illumina-long-read-enrich-data-sheet-m-gl-02188/ilmn-long-read-enrich-data-sheet-m-gl-02188.pdf. Published 2024. Accessed February 14, 2024.
5. Illumina. Illumina Human Comprehensive Panel data sheet. illumina.com/content/dam/illumina/gcs/assembled-assets/marketing-literature/illumina-long-read-enrich-hu-comp-panel-data-sheet-m-gl-02191/long-read-hu-comp-panel-data-sheet-m-gl-02191.pdf. Published 2024. Accessed February 14, 2024.
6. Illumina. Comprehensive whole-genome sequencing with Illumina Complete Long Read Prep, Human technical note. illumina.com/content/dam/illumina/gcs/assembled-assets/marketing-literature/illumina-long-read-prep-human-tech-note-m-gl-01421/ilmn-long-read-hu-tech-note-m-gl-01421.pdf. Published 2023. Accessed January 22, 2024.
7. Roessler K. Illumina Complete Long Reads software analysis workflow for human WGS. illumina.com/science/genomics-research/articles/complete-long-read-software-analysis.html. Published 2023. Accessed January 22, 2024.

illumina®

1.800.809.4566 toll-free (US) | +1.858.202.4566 tel
techsupport@illumina.com | www.illumina.com

M-GL-02523 v1.0