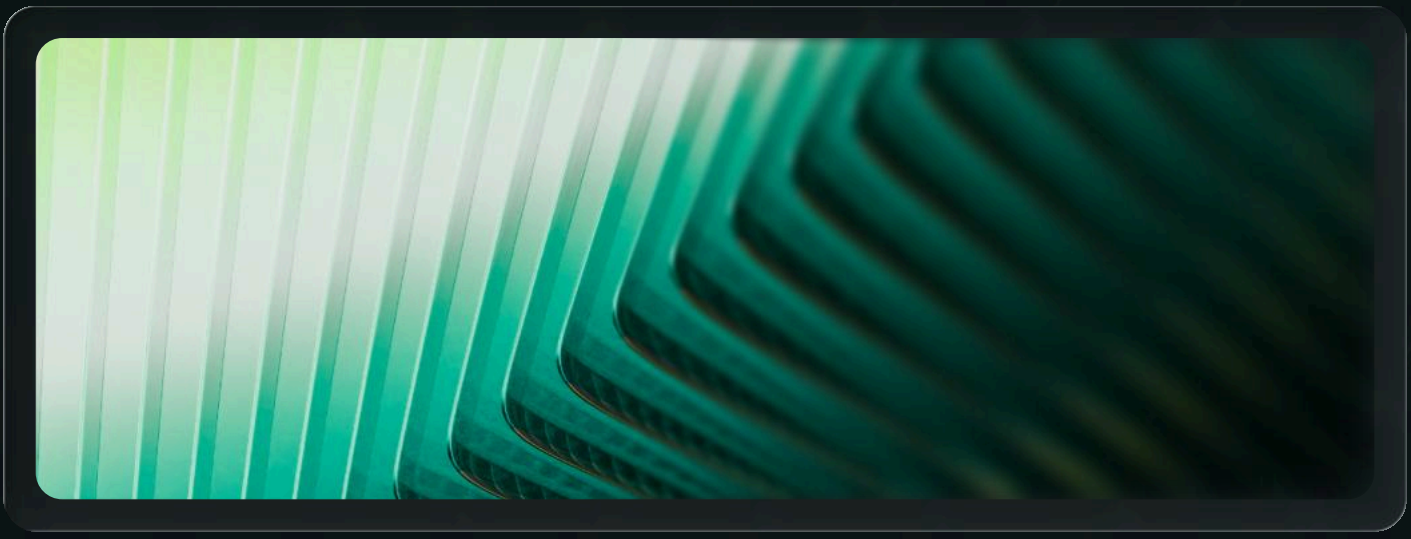
RENEW

Case Study



cDNA Sequencing for Isoform Detection

Long-Read Transcriptome Profiling for Rare Genetic Disease

Summary

Long-read cDNA sequencing enables isoform-level resolution of transcriptomic changes that are not accessible with short-read methods. In this study, Oxford Nanopore Technologies (ONT) long-read sequencing was applied to whole-blood samples to evaluate its ability to detect treatment-associated splice variants and characterize transcriptome-wide expression changes. Results demonstrate the utility of full-length transcript sequencing for identifying isoform-level biomarkers and supporting translational applications in therapeutic development.

Introduction

Short-read RNA-seq provides robust gene-level expression but is limited in resolving full-length transcripts and splice variants. Long-read sequencing addresses these limitations by capturing complete isoforms, enabling direct interrogation of transcript structure and isoform-specific changes.

To evaluate the impact of long-read sequencing in a translational setting, ONT cDNA sequencing was applied to pre-characterized patient samples and compared to a standard short-read RNA-seq approach.

Study Objectives

The current study assessed whether long-read cDNA sequencing improves the detection of treatment-associated transcriptomic changes with isoform-level resolution. Specifically, it aimed to:

1. Detect a therapy-associated splice variant hypothesized to mediate treatment response
2. Compare ONT long-read and Illumina short-read sequencing performance across standard QC and transcriptomic metrics
3. Replicate differential expression and pathway enrichment across key treatment groups (high dose vs. placebo, low dose vs. placebo, and age-stratified cohorts)

Study Subjects and Sample Preparation

Patients with a rare genetic disease were stratified into two age-based cohorts (Cohort 1, Cohort 2) and received high-dose therapy, low-dose therapy, or placebo. Whole blood samples were collected using standard clinical protocols. Total RNA was extracted, reverse- transcribed to cDNA, barcoded, and sequenced using Illumina NovaSeqX 10B and Oxford Nanopore Technologies PromethION platforms.

Figure 1A - Replicate Distributions (sample-level): Per-molecule telomere-length distributions for each replicate (N=3 per line), aggregated across all chromosome arms.

Illumina	Oxford Nanopore Technologies
Read Trimming: Raw FASTQ files were processed with cutadapt to remove adapter sequences and low-quality bases.	--
Read Alignment: Trimmed reads were aligned to the GRCh38 human reference genome using HISAT2.	Read Alignment: Long-read sequences were aligned to the human reference genome GrCh38 with Minimap2.
Transcript Quantification: Gene-level expression was quantified with featureCounts.	Transcript Quantification: Gene-level expression was quantified with featureCounts.
--	Transcript Isoform Analysis: Aligned BAM files were manually inspected in IGV to identify splice junctions consistent with the target isoform.
Differential Expression Analysis: Gene-level differential expression (DE) between treatment groups was computed with DESeq2.	Differential Expression Analysis: Gene-level differential expression (DE) between treatment groups was computed with DESeq2.
Gene Set Enrichment Analysis (GSEA): DE-ranked gene lists were analyzed using GSEA 4.3.3 on Windows 11 with MSigDB GO gene sets to evaluate enrichment of biological processes, cellular components, and molecular functions.	Gene Set Enrichment Analysis (GSEA): DE-ranked gene lists were analyzed using fgsea with MSigDB GO gene sets to evaluate enrichment of biological processes, cellular components, and molecular functions.

Illumina vs. ONT: QC and Read Performance

Both platforms generated high-quality cDNA data with distinct performance characteristics reflecting their underlying technologies (Figure 1, Table 1). Illumina produced more aligned reads (58.3M) due to short 150 bp paired-end sequencing and PCR amplification (Figure 1A). In contrast, ONT generated fewer reads (33.3M) but substantially longer fragments, yielding a higher total number of aligned bases (26.6 Gb vs. 17.4 Gb) (Figure 1B).

Alignment rates were similarly high for both platforms (Illumina 96.3%, ONT 96%), indicating strong mapping confidence despite increased read complexity in ONT data (Figure 1C). ONT achieved average read lengths >800 bp compared to ~150 bp for Illumina (Figure 1D), preserving full transcript structure and enabling improved isoform resolution.

In summary, Illumina provides higher read counts, while ONT captures more bases per read and supports more complete transcript reconstruction. This is critical for studies of alternative splicing, rare isoforms, and full-length transcripts.

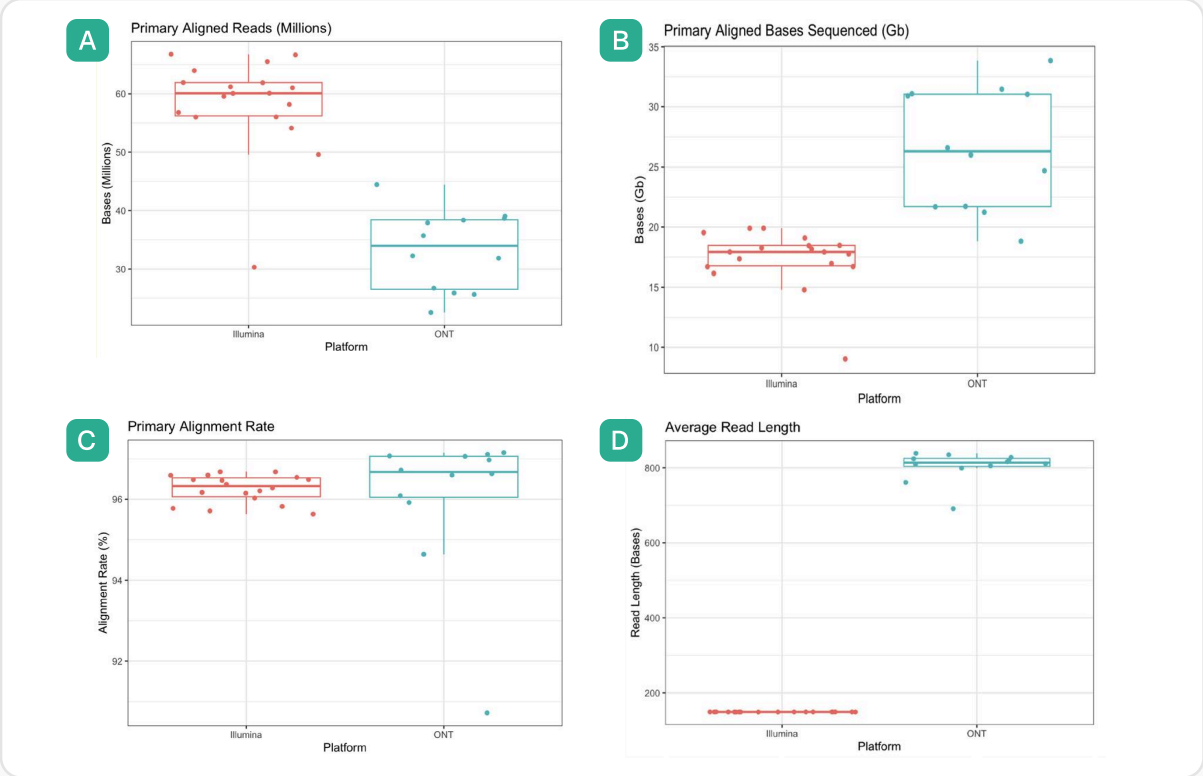


Figure 1. Illumina vs. ONT cDNA QC and read performance metrics, including primary aligned reads (A), primary aligned bases (B), primary alignment rate (C), and average read length (D). mRNA was extracted from whole blood samples from patients, reverse transcribed to cDNA, and sequenced on Illumina's NovaSeqX 10B (n = 18) or ONT's PromethION platform (n = 12).

Metric	Illumina	ONT	p-value
Primary Aligned Reads	58.3 M	33.3 M	1.97E-09
Primary Aligned Bases Sequenced	17.4 Gb	26.6 Gb	2.82E-07
Primary Alignment Rate	96.3%	96%	0.65
Avg Read Length	149.2 bp	803.2 bp	8.77E-33
Isoform Resolution	Limited	High	-

Isoform Transcript Analysis

No evidence of the hypothesized isoform was detected in ONT long-read datasets from either high-dose or placebo samples (data not shown). This absence may reflect low expression in whole blood or context-dependent splicing restricted to specific tissues, cell types, or timepoints not captured in this study.

Despite this, ONT enabled high-confidence splice junction mapping and full-length transcript resolution, supporting validation of complex isoforms beyond the capabilities of short-read sequencing. A representative IGV plot of a ubiquitously expressed transcript demonstrates complete exon coverage (Figure 2).

Long-read transcriptome profiling also supported robust downstream analyses of treatment- associated expression changes across cohorts (Figure 3), enabling differential expression and gene set enrichment analyses of therapeutic response.

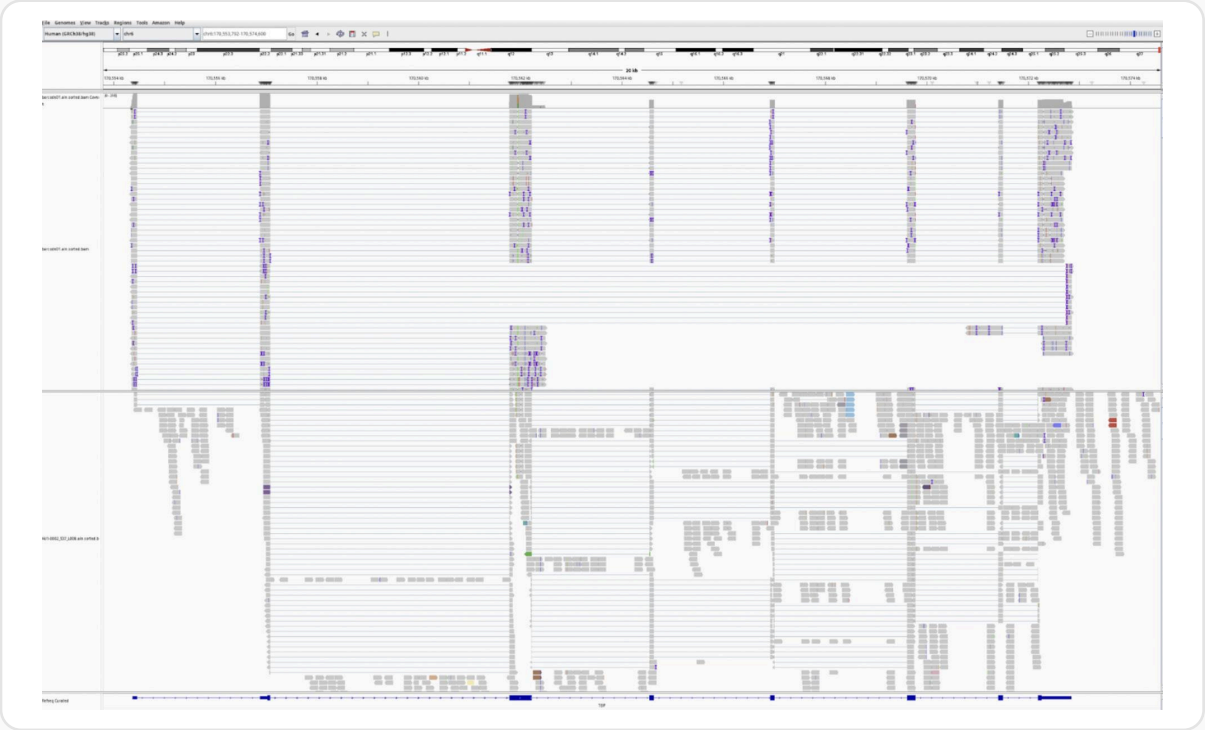


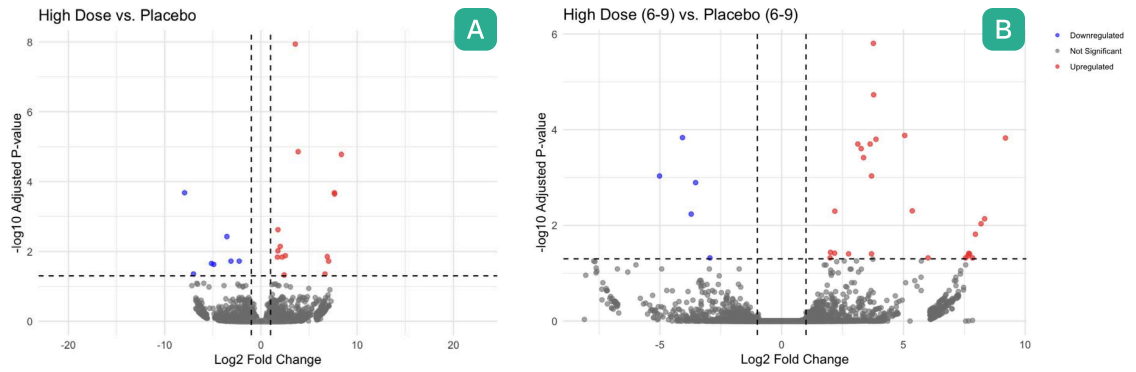
Figure 2. Representative Integrative Genome Viewer (IGV) visualization of ONT long-read cDNA sequencing. This representative transcript, unrelated to the therapy-targeted isoform, illustrates how ONT long reads span multiple exons in a single read to enable direct isoform reconstruction and splice junction resolution that is not possible with traditional short-read RNA-seq platforms.

Differential Expression & Pathway Analyses

ONT-based analysis replicated core differential expression and GSEA results observed with Illumina, confirming cross-platform reproducibility while providing enhanced transcript- level resolution. Given the consistency, ONT results are highlighted to demonstrate the added value of long-read sequencing.

Volcano plots show significant gene expression differences across treatment groups (Figure 3A–B). A greater number of differentially expressed genes were observed in Cohort 1 (Figure 3B–C), suggesting a potential age-dependent treatment response.

GSEA did not identify significantly enriched pathways, likely due to limited sample size (Figure 4). However, the data demonstrate that long-read cDNA sequencing supports robust pathway-level analysis while preserving isoform-level resolution and capturing broader, cohort-specific transcriptomic trends.



B

Cohort 1 High Dose v Placebo Differential Expression

ONT Top 20 Upregulated Genes

gene_name <chr>	baseMean <dbl>	log2FoldChange <dbl>	lfcSE <dbl>	stat <dbl>	pvalue <dbl>	padj <dbl>
BRPF3-AS1	62.56104	9.191382	1.6912304	5.434731	5.487925e-08	1.492167e-04
ENSG00000302062	34.41068	8.335088	1.8783478	4.437457	9.102785e-06	7.279550e-03
CNN1	30.97584	8.188698	1.8723478	4.373492	1.222749e-05	9.235148e-03
ENSG00000249771	26.59378	7.955681	1.8720579	4.249698	2.140591e-05	1.531649e-02
SLC2A1-DT	24.63781	7.840363	2.0111615	3.898425	9.682031e-05	4.783118e-02
ENSG00000291250	22.53066	7.716207	1.9388294	3.979828	6.896528e-05	3.937865e-02
LINC02917	21.93162	7.683967	1.9188691	4.004425	6.216863e-05	3.841739e-02
ENSG00000294869	21.37140	7.648863	1.9433140	3.935989	8.285460e-05	4.332340e-02
ENSG00000308028	20.02655	7.552235	1.9517434	3.869481	1.090671e-04	4.783118e-02
ENSG00000296262	38.54218	6.011617	1.5462054	3.887981	1.010814e-04	4.783118e-02
CEP295	48.68898	5.364061	1.1763722	4.559833	5.119428e-06	4.971331e-03
TMEM176A	300.21214	5.053907	0.9112021	5.546417	2.915831e-08	1.321357e-04
ENSG00000289271	132.67775	3.872758	0.7183154	5.391445	6.989322e-08	1.583664e-04
ENSG00000267174	3545.71882	3.776473	0.6350935	5.946327	2.742267e-09	1.864056e-05
ENSG00000301694	392.55943	3.766437	0.5843721	6.445272	1.153932e-10	1.568771e-06
ENSG00000284931	2042.90440	3.690126	0.7483857	4.930781	8.190163e-07	9.278772e-04
RPH3A	853.54214	3.689151	0.9275052	3.977499	6.964381e-05	3.937865e-02
PF4V1	1064.80488	3.633067	0.6857923	5.297621	1.173213e-07	1.993728e-04
HBG1	25015.84925	3.363192	0.6549863	5.134751	2.825174e-07	3.840824e-04
C1D	302.92842	3.264440	0.6235254	5.235457	1.645772e-07	2.486030e-04

ONT Top Downregulated Genes

gene_name <chr>	baseMean <dbl>	log2FoldChange <dbl>	lfcSE <dbl>	stat <dbl>	pvalue <dbl>	padj <dbl>
ENSG00000256020	76.9051	-5.015827	1.0163592	-4.935093	8.011277e-07	0.0009278772
IGKV3-11	1097.3726	-4.072523	0.7434620	-5.477782	4.306910e-08	0.0001463811
IGHA1	8860.0994	-3.715335	0.8258359	-4.498878	6.831314e-06	0.0058044817
IGKV3D-11	534.8812	-3.535100	0.7285726	-4.852091	1.221665e-06	0.0012775798
IGKC	25027.6108	-2.945899	0.7601540	-3.875398	1.064507e-04	0.0478311839

Figure 3. Differential expression analysis in patients treated with therapy vs. placebo. Volcano plots show transcript-level differences for high-dose vs. placebo (A) and high-dose vs. placebo in Cohort 1 (B), with more differentially expressed genes observed in Cohort 1 (n = 3 per group). Gene lists of top up- and down-regulated genes are included in (C).

Cohort 1 High Dose v Placebo Gene Set Enrichment Analysis

ONT Upregulated Pathways (Top 20, Not Statistically Significant)

pathway	pval	padj	ES	NES	nMoreExtreme	size
GOMF_INOSITOL_1_4_5_TRISPHOSPHATE_BINDING	0.0150508210	0.9730848	0.9977321	1.480633	76	5
GOMF_TYPE_5_METABOTROPIC_GLUTAMATE_RECEPT...	0.0193510555	1.0000000	0.9972096	1.479857	98	5
GOCC_PLATELET_DENSE_GRANULE_LUMEN	0.0383111806	1.0000000	0.9941356	1.475296	195	5
GOBP_REGULATION_OF_MICROVILLUS_ORGANIZATION	0.0457388585	1.0000000	0.9931520	1.473836	233	5
GOMF_MYOSIN_HEAVY_CHAIN_BINDING	0.0467161845	1.0000000	0.9928346	1.473365	238	5
GOBP_AMMONIUM_TRANSMEMBRANE_TRANSPORT	0.0486708366	1.0000000	0.9924762	1.472833	248	5
GOMF_AMMONIUM_TRANSMEMBRANE_TRANSPORTER...	0.0486708366	1.0000000	0.9924762	1.472833	248	5
GOBP_CELLULAR_RESPONSE_TO_POTASSIUM_ION	0.0494526974	1.0000000	0.9923686	1.472673	252	5
GOBP_TRANSEPITHELIAL_CHLORIDE_TRANSPORT	0.0498436278	1.0000000	0.9923412	1.472633	254	5
GOBP_POSITIVE_REGULATION_OF_CENTRIOLE_REPLICA...	0.0048365254	0.6534986	0.9993102	1.456288	24	6
GOBP_POSITIVE_REGULATION_OF_SYNAPTIC_VESICLE_...	0.0203134068	1.0000000	0.9972486	1.453283	104	6
GOBP_CDP_CHOLINE_PATHWAY	0.0220545560	1.0000000	0.9969606	1.452864	113	6
GOBP_PYRIMIDINE_RIBONUCLEOTIDE_CATABOLIC_PRO...	0.0266976204	1.0000000	0.9963831	1.452022	137	6
GOBP_CARDIAC_MUSCLE_CELL_MYOBLAST_DIFFERENT...	0.0363706713	1.0000000	0.9950380	1.450062	187	6
GOMF_TRANSMEMBRANE_RECEPTOR_PROTEIN_TYROSI...	0.0379183595	1.0000000	0.9948165	1.449739	195	6
GOCC_CLATHRIN_COMPLEX	0.0448829561	1.0000000	0.9936404	1.448025	231	6
GOMF_PHOSPHATE_ION_BINDING	0.0185752585	1.0000000	0.9977023	1.426988	96	7
GOMF_G_PROTEIN_COUPLED_GLUTAMATE_RECEPTOR...	0.0231711988	1.0000000	0.9971705	1.426228	120	7
GOBP_CMP_METABOLIC_PROCESS	0.0294906166	1.0000000	0.9963404	1.425041	153	7
GOBP_RNA_GUANINE_N7_METHYLATION	0.0317885867	1.0000000	0.9961106	1.424712	165	7

ONT Upregulated Pathways (Top 20, Not Statistically Significant)

pathway	pval	padj	ES	NES	nMoreExtreme	size
GOBP_REGULATION_OF_ANIMAL_ORGAN_FORMATION	0.0022513303	0.6534986	-0.9996232	-1.601042	10	5
GOBP_REGULATION_OF_COMPLEMENT_ACTIVATION_C...	0.0079819894	0.8021394	-0.9989110	-1.599901	38	5
GOBP_N_TERMINAL_PROTEIN_LIPIDATION	0.0094146541	0.8402521	-0.9986455	-1.599476	45	5
GOBP_CELLULAR_RESPONSE_TO_ABIOTIC_STIMULUS	0.0009212345	0.6534986	-0.9700468	-1.598407	1	221
GOBP_REGULATION_OF_NMDA_RECEPTOR_ACTIVITY	0.0159639787	0.9970649	-0.9978987	-1.598279	77	5
GOBP_NEGATIVE_REGULATION_OF_PROTEIN_CATABOL...	0.0165779779	0.9970649	-0.9978083	-1.598135	80	5
GOCC_INTRACILIARY_TRANSPORT_PARTICLE_A	0.0208759722	1.0000000	-0.9971947	-1.597152	101	5
GOCC_MICROVESICLE	0.0227179697	1.0000000	-0.9969401	-1.596744	110	5
GOMF_ESTROGEN_RESPONSE_ELEMENT_BINDING	0.0028967515	0.6534986	-0.9996662	-1.594146	13	6
GOBP_EPITHELIAL_CELL_DIFFERENTIATION_INVOLVED...	0.0028967515	0.6534986	-0.9996573	-1.594131	13	6
GOBP_LEYDIG_CELL_DIFFERENTIATION	0.0033105731	0.6534986	-0.9996210	-1.594073	15	6
GOBP_OPSONIZATION	0.0093109870	0.8402521	-0.9988665	-1.592870	44	6
GOBP_CELLULAR_RESPONSE_TO_ACIDIC_PH	0.0376586165	1.0000000	-0.9944827	-1.592808	183	5
GOCC_RIPOPTOSOME	0.0444126074	1.0000000	-0.9936221	-1.591430	216	5
GOBP_SPERM_CAPACITATION	0.0262776743	1.0000000	-0.9969807	-1.589863	126	6
GOMF_DELAYED_RECTIFIER_POTASSIUM_CHANNEL_AC...	0.0403476102	1.0000000	-0.9947539	-1.586312	194	6
GOBP_REGULATION_OF_ADIPONECTIN_SECRETION	0.0475894889	1.0000000	-0.9938966	-1.584945	229	6
GOCC_RIBONUCLEOPROTEIN_GRANULE	0.0009041591	0.6534986	-0.9669862	-1.577117	1	208
GOBP_ADAPTIVE_IMMUNE_RESPONSE_BASED_ON_SOM...	0.0025432350	0.6534986	-0.9250426	-1.573664	4	281
GOBP_LYMPHOCYTE_MEDIATED_IMMUNITY	0.0025303644	0.6534986	-0.9258518	-1.570864	4	277

Figure 4. Gene Set Enrichment Analysis (GSEA) of Cohort 1 (Treatment vs. Placebo). GSEA was performed using fgsea with MSigDB gene ontology sets to evaluate pathway-level transcriptional differences in Cohort 1 (n = 3 per group). While no pathways reached statistical significance in this pilot dataset, the GSEA framework successfully demonstrated analytical feasibility, revealing transcriptomic trends that may be further explored in larger cohorts.

Conclusion

This pilot study demonstrates the advantages of long-read cDNA sequencing for transcript-level analysis beyond the capabilities of short-read platforms. Despite lower read counts, ONT generated longer, more information-rich reads, enabling more complete transcript reconstruction and isoform-level resolution.

Although the targeted splice variant was not detected, full-length coverage supported confident negative reporting and accurate splice junction mapping. Downstream analyses further showed that ONT data enables robust transcriptomic profiling, including detection of potential age-dependent treatment responses.