Designing an Assay to Evaluate NBS Results Using Targeted Long-Read Sequencing

Shenyi Ye

A thesis

submitted in partial fulfillment of the

requirements for the degree of

Master of Science

University of Washington

2025

Committee:

Danny Miller

Anna Scott

Program Authorized to Offer Degree:

Department of Laboratory Medicine and Pathology

University of Washington

Abstract

Designing an Assay to Evaluate NBS Results Using Targeted Long-Read Sequencing

Shenyi Ye

Chair of the Supervisory Committee:

Danny Miller

Department of Laboratory Medicine and Pathology

NBS (NBS) is a routine screening process that identifies selected genetic, metabolic, and endocrine disorders that can affect a newborn's health. Unfortunately, limitations in the follow-up process can create barriers to confirming screening results. Here, we demonstrate the ability of targeted long-read sequencing (T-LRS) in evaluating NBS results. Using adaptive sampling on the Oxford Nanopore platform on 8 positive control samples from Seattle Children's Hospital, we computationally targeted more than 500 genes relevant to metabolic and non-metabolic disorders and searched for pathogenic variants using a single data source. We detected all genomic variants identified by prior genetic testing, as well as additional variants not previously identified. T-LRS demonstrates to be an efficient and cost-effective method to evaluate individuals after a positive NBS result.

Table of Contents

LIST OF TABLES

LIST OF FIGURES

CHAPTER I: INTRODUCTION

A. Newborn Screening Overview

Newborn screening (NBS) is conducted after an infant is born and before they are discharged from the hospital. This routine screening identifies selected genetic, metabolic, and endocrine disorders that can affect a newborn's health. The conditions tested by NBS are treatable and actionable. In many cases, NBS results are validated by genetic testing. Each state differs in the selection of the NBS tests that are conducted. Typically, medical staff make a puncture to the infant's heel and collect drops of blood on a blood spot card. This card is then sent to a laboratory for testing.

Once the blood spot card arrives in the state laboratory, small samples from the dried blood spot are taken to conduct various tests. The process of NBS is crucial, as we want to quickly catch things that are actionable. The importance of NBS can be exemplified by medium-chain acyl-coenzyme a dehydrogenase (MCAD) deficiency, a condition which prevents fatty acid β-oxidation. MCAD deficiency results in hypoglycemia, and can lead to death in undiagnosed children.[1] Although management of the condition is as simple as giving carbohydrates, multiple cases of sudden neonatal death in those with MCAD deficiency have been reported.[2,3] Since then, MCAD deficiency has been included in Washington State's NBS panel since the early 2000s, the specific disorders included vary by state, as each has its own legislations governing NBS implementation. Accurate interpretation of NBS results depends on timely and proper sample collection, and in many cases, abnormal screens require follow-up testing to confirm the diagnosis. For families, the process of arriving at a definitive diagnosis,

especially when initial results are can be prolonged and complex, often described as a "diagnostic odyssey."[4]

Another reason that rapid and comprehensive follow-up of positive NBS results is critical is because, there are an increasing number of treatments for genetic conditions available. For example, the development of Spinraza to treat spinal muscular atrophy (SMA), a rare genetic condition that causes progressive muscle weakness and loss of motor function.[5] Through research efforts on identifying the genes associated with SMA and understanding the molecular basis of SMA, Spinraza became the first FDA-approved therapy for SMA in 2016.[6] The success of Spinraza highlights how genetic discoveries can lead to transformative, targeted therapies, however, detection of the disease through NBS remains a challenge. Approximately 5% of SMA cases are missed because current NBS assays fail to detect single allelic deletion with a point mutation.[7] More broadly, several studies suggest that with exome-based approaches, NBS miss as many as 12% of metabolic cases.[8] This underscores the critical need for follow-up testing strategies like LRS, which can help clarify ambiguous results or detect previously missed cases by identifying variants in complex genomic regions and capturing methylation patterns, ultimately improving diagnostic accuracy and patient outcomes.

As a final example, X-linked adrenoleukodystrophy (X-ALD) is a condition in which individuals do not have the ability to transport very long-chain fatty acids into peroxisomes for degradation. The buildup of very long-chain fatty acids will in turn damage the nervous system and the adrenal glands, leading to behavioral problems, muscle weakness, hearing loss, blindness, and possibly death.[9] X-ALD is a metabolic disorder defined by pathogenic mutations in the *ABCD1* gene and is estimated to affect 1 in 42,000 XY individuals or 1 in 17,000 births

worldwide.[10] Because this is an X-linked condition, XY individuals are more severely affected. However, it is estimated that 70% of 46, XX individuals who are carriers of a disease-causing variant in *ABCD1* develop symptoms at some point during their lives.[10] Adrenal insufficiency occurs in nearly all 46, XY individuals affected by X-ALD and is typically managed with steroid replacement therapy. For patients diagnosed with cerebral adrenoleukodystrophy, a stem cell transplant is needed to stop the progression of the disease.[10] The diagnosis of X-ALD is established by biochemical testing and confirmed through molecular genetic testing. If elevations in very long-chain fatty acids are observed in an individual, genetic testing results of *ABCD1* is used to confirm the diagnosis.

Molecular confirmation of variants in *ABCD1* often relies on PCR amplification and Sanger sequencing, which can be used to identify single nucleotide or small deletion/insertion (indels) variants but may be unable to detect larger structural changes, such as deletions and insertions. Further, the presence of four *ABCD1* pseudogenes in the human genome increases the rate of false positive and false negative results generated by short read sequencing (SRS).[11] The limitations posted by SRS have led to our interest in applying long read sequencing (LRS) methods to clarify clinical testing results and identify variants missed by standard clinical testing. Compared to SRS, LRS is better able to resolve structural variants (SVs) and disease-causing variants in repetitive regions of the genome.[12] More importantly, LRS can directly sequence native DNA, eliminating the step for PCR amplification while allowing us to detect methylation. The insights provided by LRS can enhance the evaluation of variants and thereby improve the confidence of testing results. Additionally, LRS is particularly powerful with phasing variants or assigning variants to the maternal or paternal chromosome. This is crucial in both diagnosing recessive disorders and improving carrier screening.

B. Sequencing Technology Review

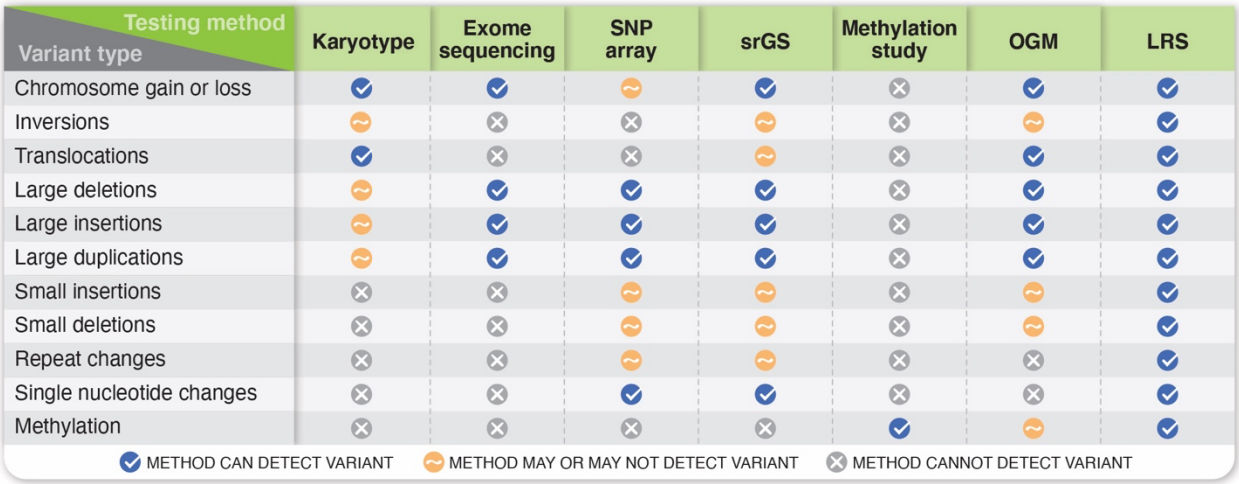
A variety of genomic technologies are available for identifying genetic variation, each with distinct advantages and limitations. Karyotyping provides a snapshot of all chromosomes and is useful for detecting mosaic events and chromosomal abnormalities such as aneuploidies or translocations, though it lacks the resolution to identify small variants like SNVs or indels and typically takes 1–2 weeks to complete.[12] Chromosomal microarray is often used to detect regions of homozygosity and copy number variants (CNVs) but cannot identify inversions or translocations. This method has a turnaround of approximately 7–10 days.[13] Sanger sequencing, while cost-efficient is limited in detecting large deletions or insertions and may take several weeks to yield results.[14] Targeted gene panels can capture regulatory regions not covered in exome sequencing and may return results in as little as 4 days, though both turnaround and content vary by laboratory.[15] Short-read sequencing approaches include exome sequencing, which evaluates most protein-coding regions and allows for data reanalysis, though interpretation can differ between labs and sensitivity for detecting CNVs is low. Results may be returned in weeks or months, with some rapid exome options available.[16] Whole genome sequencing (WGS) offers the most comprehensive analysis, but its complexity requires intensive data processing and presents interpretation challenges. Methylation analysis is valuable for detecting epigenetic alterations and mosaic patterns linked to phenotype but is restricted to certain conditions and does not reveal causative sequence variants. It is also time-consuming, often taking weeks to months.[17] Lastly, Optical Genome Mapping (OGM) provides high sensitivity for detecting structural variants and rearrangements but cannot detect SNVs or small indels, with an estimated turnaround of 1–2 weeks.[18]

Traditional genetic testing can occur in one of several ways **(Figure 1)**. With traditional genetic testing, the follow-up process of NBS can take up to six months. Challenges also arise when additional information is needed from both parents and one or both parents aren't available.

Our goal with long read sequencing (LRS) is to increase the diagnostic rate while shortening time to diagnosis. Because the technology is able to create read lengths that are longer and thereby easier to align to the genome, we can resolve SVs in complex regions of the genome. LRS not only outperforms short read sequencing in preparation time without polymerase chain reaction (PCR) amplification steps, but it also detects variants in repetitive regions, segmental duplications, and regions with high guanine-cytosine content.

Currently, there are two commercial LRS technologies available. In PacBio long-read sequencing, a DNA polymerase replicates a circular fragment of DNA multiple times. As the polymerase is working on replicating the strand, fluorescence is released and detected. The system then determines which DNA base was incorporated by the polymerase from the light emitted. PacBio offers two modes of sequencing depending on the goal. Continuous Long-Read Sequencing (CLR) generates very long read lengths at the expense of higher error, while Circular Consensus sequencing (CCS), also known as HiFi sequencing, is highly accurate but the read length is limited to about 20kb.[13] Both modes are able to detect DNA modifications. However, PacBio sequencing requires large, specialized instruments and high computational resources for data processing and storage. On the other hand, Oxford Nanopore Technologies (ONT), which sequences single strands of DNA based on characteristic disruptions in a current as a DNA or RNA molecule passes through a protein pore.[14] These pores are not limited by the length of the

DNA molecule and are capable of detecting a wide range of changes, including DNA modifications. One caveat of Nanopore long-read sequencing is that decoding the sequence from the signal may be computationally difficult, which is the primary source of error with the technology, especially in homopolymers or low-complexity repeats.[15]



Variant type \ Testing method	Karyotype	Exome sequencing	SNP array	srGS	Methylation study	OGM	LRS
Chromosome gain or loss	✓	✓	~	✓	✗	✓	✓
Inversions	~	✗	✗	~	✗	~	✓
Translocations	✓	✗	✗	~	✗	✓	✓
Large deletions	~	✓	✓	✓	✗	✓	✓
Large insertions	~	✓	✓	✓	✗	✓	✓
Large duplications	~	✓	✓	✓	✗	✓	✓
Small insertions	✗	✗	~	~	✗	~	✓
Small deletions	✗	✗	~	~	✗	~	✓
Repeat changes	✗	✗	~	~	✗	✗	✓
Single nucleotide changes	✗	✗	✓	✓	✗	✗	✓
Methylation	✗	✗	✗	✗	✓	~	✓

Figure 1: Sequencing Technology Overview

Comparison of variant detection capabilities across genomic testing methods. The table summarizes the ability of karyotyping, exome sequencing, SNP array, short-read genome sequencing (srGS), methylation studies, optical genome mapping (OGM), and long-read sequencing (LRS) to identify different variant types.

C. Targeted Long-Read Sequencing

Targeted Long-Read Sequencing (T-LRS) is a method that allows us to perform sequencing of a specific region of interest, allowing us to more efficiently perform in-depth analysis of disease-causing genes or regions. There are several T-LRS approaches. The simplest strategy is to use PCR to amplify region or regions of interest. Another T-LRS approach is

hybridization capture, where sheared DNA fragments containing regions of interest are selected using a hybridization-based kit.[16] However, with these PCR-based approaches, we see amplification biases related to PCR, as well as repeated optimization experiments and primer redesign. Another major limitation is the loss of methylation during the amplification process, which can hinder the study of epigenetic regulation in diseases such as cancer and imprinting disorders. Alternative approaches such as CRISPR/Cas9-based target enrichment and variations of this strategy have been developed to overcome the caveats of PCR-based approaches. Although there have been successes with CRISPR/Cas9-mediated protocols, difficulty in guide RNA design largely limits the adoption of such protocols.[16] Adaptive sampling (AS) using ONT achieves the same goal as other T-LRS approaches. One advantage is that this method is strictly computational and requires no additional experimental setup. With successful applications in human and nonhuman samples, AS has been used to characterize loci with clinically relevant repeat expansions, phasing of pathogenic variants, and evaluating complex structural rearrangements.[17,18] Generally, the decision to implement a T-LRS workflow is driven by costs. For all T-LRS approaches, it is important to note that prior knowledge of the disease-associated loci is required. In the context of NBS, this makes T-LRS the perfect candidate approach.

D. Literature Review of Relevant Genes

GALT

The *GALT* gene encodes galatose-1-phosphate uridylyltransferase, an enzyme responsible for converting galatose-1-phosphate into UDP-galactose. As GALT catalyzes a crucial step in galactose metabolism, its absence leads to accumulation of galactose-1-phsophate, which is toxic to liver, brain, and kidney cells.[19] Deficiency in GALT causes galactosemia, an autosomal

recessive disorder, symptoms may include jaundice, vomiting, and sepsis. Detection for galactosemia typically includes testing for total galactose and GALT activity. Genetic testing is essential in cases with ambiguous enzyme results. NBS and early diagnosis are critical to prevent the development of symptoms by starting a galactose-restricted diet.

GAA

The *GAA* gene encodes acid alpha-glucosidase, a lysosomal enzyme essential for glycogen degradation. Mutations in *GAA* cause Pompe disease, a rare autosomal recessive disorder leading to glycogen accumulation in cardiac and skeletal muscles.[20] Deficiency of the enzyme alpha-glucosidase results in glycogen build up. In muscle cells, accumulated glycogen disrupts contractile function, leading to cell death.

Pompe disease exists in two forms. Infantile-onset Pompe disease (IOPD) presents within the first few days to months of life and is characterized by muscle weakness, feeding difficulties, and hypertrophic cardiomyopathy. Without enzyme replacement therapy, IOPD results in death by age two. Late-onset Pompe disease (LOPD) manifests later, with a slower progression and less cardiac involvement.[21] Diagnosis of Pompe is confirmed using enzymatic assay of GAA activity and molecular testing. NBS programs have increased efforts to detect Pompe disease to allow timely initiation of treatment.

CHAPTER II: METHODS

A. Quality Control Steps Before Library Preparation

All extracted DNA samples go through a series of quality control (QC) steps before sequencing. Samples are quantified using Nanodrop and Qubit instruments, followed by the Femto pulse system.

To test the validity of the pipeline, a total of eight samples provided by the Seattle Children's Hospital were used as positive controls. Extracted DNA was first measured by Qubit and Nanodrop, followed by the Femto pulse. If Femto pulse results show multiple peaks near the largest peak, additional treatment steps prior to library preparation are required. Samples were sheared to obtain the optimal length of 15-20kb for the purpose of AS. Prior to shearing, Femto peaks were at approximately 60kb (**Figure 2**), after shearing, the Femto peaks were at 15-20kb (**Figure 3**). Shearing DNA into fragments allows the pore to read through the strand quicker.

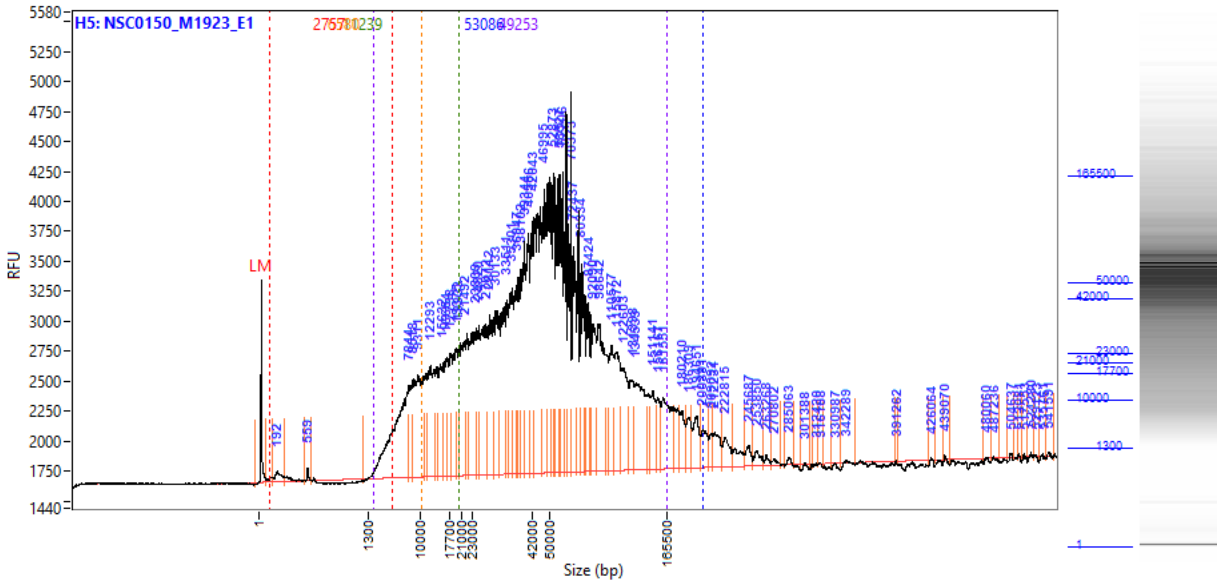


Figure 2: Femto Pulse Trace Prior to Treatment

Trace image of genomic DNA analyzed using the Agilent Femto Pulse system. The trace displays multiple distinct peaks centered around 60,000 bp, indicating the presence of high molecular weight DNA fragments of different sizes.

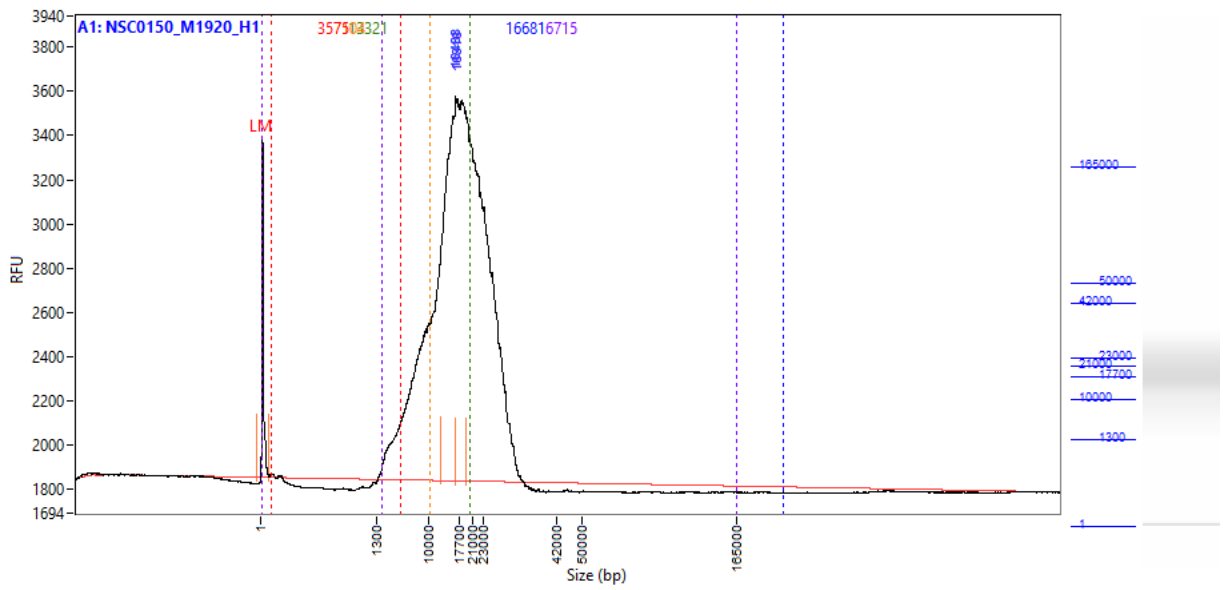


Figure 3: Femto Pulse Trace After Treatment

Trace image of genomic DNA analyzed using the Agilent Femto Pulse system. A single, sharp peak is observed around 20,000 bp.

B. Library Preparation for Adaptive Sampling

The Ligation Sequencing Kit V14 from nanopore is used to prepare a library for AS. Before loading the library for sequencing, ends of the DNA need to be repaired and prepped for ligation. Proteinase K solution cleans up any protein residue that might interfere with the ligation. After Ampure clean up, the library is then ligated with adapters and goes through one more round of Ampure clean up before loading onto the Promethion.

Table 1: Sequencing Statistics

Sample #	NBS1	NBS3	NBS4	NBS6	NBS7	NBS8	NBS9	NBS10
Initial concentration (ng/µL)	21.1	17.6	16.4	29.6	17.1	48.3	21.2	57.8
Input DNA (ng)	1371	1144	1066	1924	1111	3139	1378	3757
Final library yield (ng)	300	452.2	680	500	788	918	748	1115
Conversion efficiency (%)	24.1	39.5	63.8	26.1	70.9	29.2	54.3	30.6
Total libraries	1	1	1	1	2	2	1	3
Library load (ng)	300	452.2	1066	500	788	918	748	1115

Previously extracted DNA from blood samples was used as input for ligation-based library preparation, with 330–950 ng of DNA per sample. Libraries were sequenced using the Oxford Nanopore PromethION platform with R10.4.1 flow cells for 24–72 hours. Standard downstream analysis included base calling, alignment, variant calling, haplotype phasing, and methylation profiling.

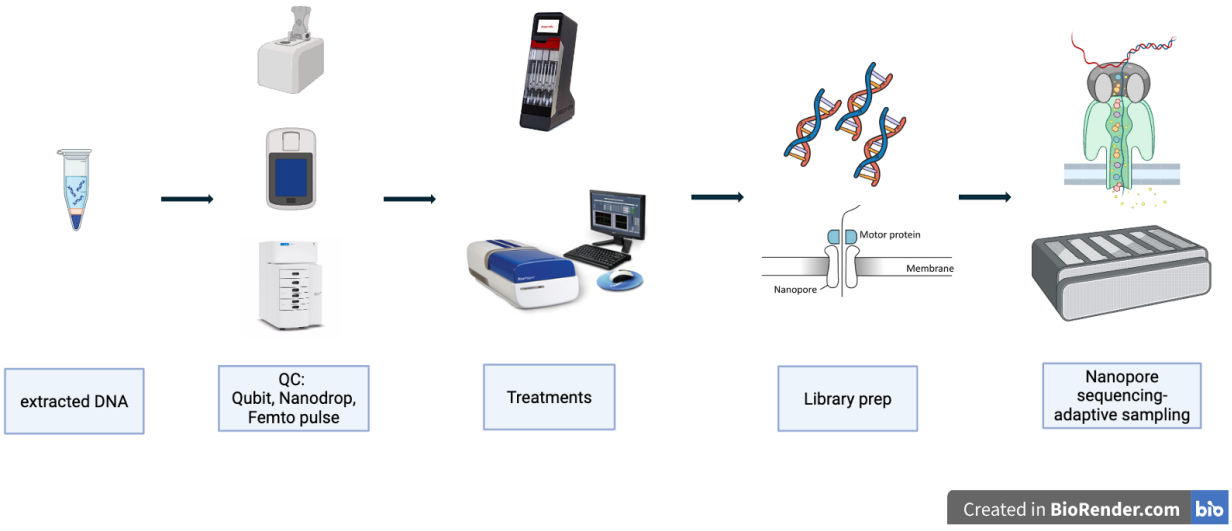


Figure 4: Standard Workflow

Extracted DNA go through a series of quality control steps including the Nanodrop, Femto Pulse, and Qubit system. Sometimes samples are treated either by shearing or size selection. After that, we prep the library using the ligation kit. Lastly, samples are loaded onto the sequencer for sequencing.

CHAPTER III: RESULTS

Table 2: Summary of Sequencing Results

Sample #	NBS1	NBS3	NBS4	NBS6	NBS7	NBS8	NBS9	NBS10
Sequencing method	T-LRS	T-LRS	T-LRS	T-LRS	T-LRS	T-LRS	T-LRS	T-LRS
Data generated (GB)	30	43.78	16.53	38.86	44.3	59.95	41.7	58.48
Depth of coverage (avg of target region)	58x	56x	37x	65x	64x	111x	75x	121x
Average N50 (kb)	14.5	10.07	20.9	13.1	15.1	14.4	15.9	17.1
No. of reads (M)	20.88	39.77	10.24	29.53	30.33	41	32.2	37.5
Run time (hrs)	20.85	24	23	72	63	63.28	22	65.86

Eight T-LRS runs are shown with key performance metrics, including total data generated (in gigabases), average depth of coverage across the target region, average read N50 (in kilobases), total number of reads (in millions), and run time (in hours). All runs achieved sufficient coverage for downstream analysis, with average depth ranging from 37x to 121x.

A. Pilot Experiment

To test whether the target genome size would exceed the threshold for AS, we first piloted a sequencing run with two samples NBS8 and NBS10. Two libraries were prepared for sample NBS8, and three libraries were prepared for sample NBS10 the difference in the number of libraries loaded was because we were trying to maximize the amount of data from each sample. Sample NBS8 had a lower DNA yield, and we were not able to load a third library.

For the purpose of this project, we are most interested in the target region statistics. The total target size is 73.45 mb, and 2.37% of the human genome was targeted (**Appendix A**). For

NBS8 our sequencing results indicate that 8.7 gigabytes (GB) of data and 1,002,113 reads were generated in the target region. The N50 for NBS8 was 14.4 kb, the average read depth across the targeted region was 114x. For NBS10, 9.5 GB of data and 935,960 reads were generated in the target region. The N50 was 17.1 kb, the average read depth across the targeted region was 124x. All genes were covered.

After looking at the sequencing results of our first two sample results, we decided to increase the target genome size to include non-metabolic disorders. The updated total target size is 79.60 Mb, and 2.57% of the human genome was targeted (**Appendix B**). To evaluate whether our panel would be compatible with AS on Nanopore, we first sequenced samples NBS1 and NBS6. The total target genome size exceeded the recommended threshold of 2% for adaptive sampling for optimal performance. Despite this, we successfully completed the sequencing runs and achieved adequate depth of coverage across the regions of interest, demonstrating the feasibility of our approach.

B. Variant Analysis

Raw data generated goes through a standard analysis pipeline (**Figure 5**). For each sample, unmapped BAM files were aligned to the GRCh38 human reference genome. Following alignment, variant callers were used to ensure the detection of different variant types. SNVs and indels are called with Clair3,[22] while SVs were called using Sniffles2,[23] CuteSV,[24] and qDNAseq[25] Detected variants were assigned to haplotypes using LongPhase,[26] enabling phasing of SNVs and SVs across reads. Variant annotation was performed using the Variant Effect Predictor (VEP) for SNVs and indels, while structural variants were annotated and integrated with SNV data at later stages of the pipeline. At multiple points throughout the workflow, quality

control metrics such as alignment statistics and variant counts were generated to monitor and validate the sequencing and variant calling performance per sample.

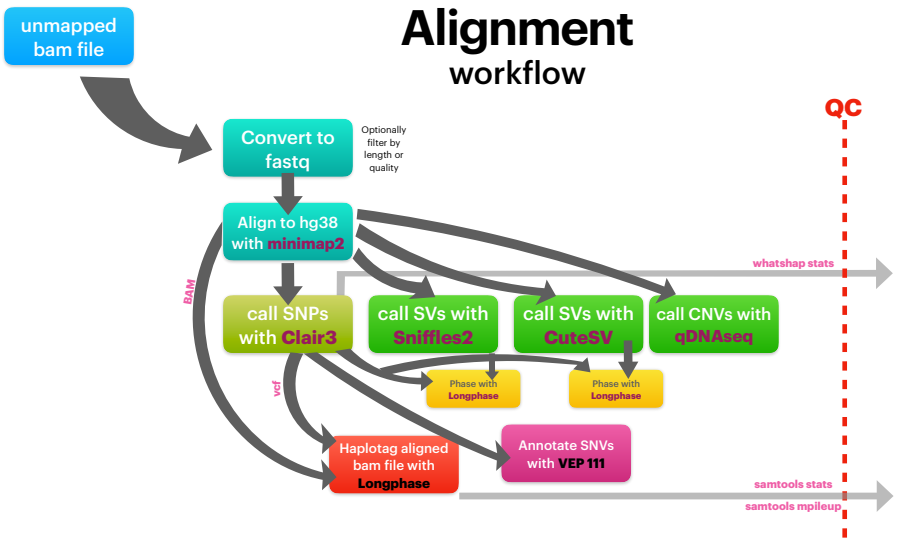


Figure 5: Alignment Workflow

This pipeline starts with unmapped bam files. Input file is aligned to the reference genome, GRCh38. After alignment, we call different classes of variants using callers shown in green. Individual reads and variants are assigned to a haplotype using LongPhase (seen in red). QC information is generated at several steps.

For each sample, we conducted variant analysis to confirm if the results match the known variants reported by Seattle Children's Hospital. The presence of pseudodeficiency alleles in most of the samples made it difficult to carry out the standard analysis pipeline. We first examined filtered variants in the VEP file. These variants were prioritized based on predicted protein impact, phenotype relevance, and population frequency.

High priority variants identified by the VEP file were further inspected using the Integrative genomics viewer (IGV) to confirm the accuracy of variant calls. Individual bam files

were loaded and aligned to the GRCh38 reference genome, a minimum of 10x read depth was used as a threshold for reliability. IGV's base coloring and quality score features provided an additional layer of validation.

Table 3: Summary of Variants Identified

		Variant 1		Variant 2		Variant 3		Variant 4		Total variants
Study ID	**Gene**	**Variant 1**	**Inheritance**	**Variant 2**	**Inheritance**	**Variant 3**	**Inheritance**	**Variant 4**	**Inheritance**	**Total Variants**
NBS1	*GALT*	c.563A>G (p.Gln188Arg)		c.855G>T (p.Lys285Asn)						2
NBS3	*GALT*	NM_000155.3:c.563A>G (p.Gln188Arg)		c.940A>G (p.Asn314Asp)						2
NBS4	*GALT*	c.563A>G (p.Gln188Arg)		c.563A>G (p.Gln188Arg)						2
NBS6	*GAA*	c.1933G>T (p.Asn645Tyr)	Not Pat	c.2015G>A (p.Arg672Gln)	Not Pat	c.2065G>A (p.Glu689Lys)	Not Pat	c.2238G>C (p.Trp746Cys)	Pat	4
NBS7	*GAA*	c.1477C>T (p.Pro493Ser)	Mat	c.1726G>A (p.Gly576Ser)	Pat	c.2065G>A (p.Glu689Lys)	Pat	c.2221G>A (p.Asp741Asn)	Mat	4
NBS8	*GAA*	NM_000152.3:c.2051C>T (p.Pro684Leu)	Mat	Deletion of exon 18	Pat					2
NBS9	*GAA*	NM_000152.3:c.752C>T (p.Ser251Leu)	Mat	c.761C>T (p.Ser254Leu)	Mat	c.1726G>A (p.Gly576Ser)	Pat	c.2065G>A (p.Glu689Lys)	Pat	4
NBS10	*GAA*	c.1927G>A (p.Gly643Arg)	Pat	c.2560C>T (p.Arg854Ter)	Pat	c.1726G>A (p.Gly576Ser)	Mat	c.2065G>A (p.Glu689Lys)	Mat	4

Table 3 shows variants identified in *GALT* and *GAA* genes across eight NBS (NBS) cases. For each study ID, up to four variants are listed with corresponding inheritance information when available (Pat = paternal, Mat = maternal, Not Pat = not paternal). Total variants per case is indicated in the final column.

IGV Screenshots of Variants Per Sample

IGV screenshot showing long-read alignment at the *GAA* and *GALT* loci for samples NBS1–NBS10. Each panel displays sequencing reads aligned to GRCh38 in the IGV browser. The coverage track above each alignment panel shows read depth at each genomic position, with peaks corresponding to higher coverage. Reads are colored by mismatched base, A = green, C = blue, G = orange, T = red, and gray for matches to the reference. Variants appear as colored mismatches in the read stack.

Where applicable, haplotypes assigned are visualized by IGV's grouping and coloring options, with distinct blocks of reads indicating phased alleles. Phased reads are grouped by haplotype, allowing visual separation of maternally and paternally inherited variants when haplotyping is successful.

M1919/NBS1

This case was flagged for low GALT enzyme activity and suspected galactosemia. Two variants were identified in the *GALT* gene: c.563A>G (p.Gln188Arg) and c.855G>T (p.Lys285Asn). Both variants were classified as missense. Variants were confirmed using IGV, screenshots of the variants (**Figures 6-7**) confirm that they are true and do exist.

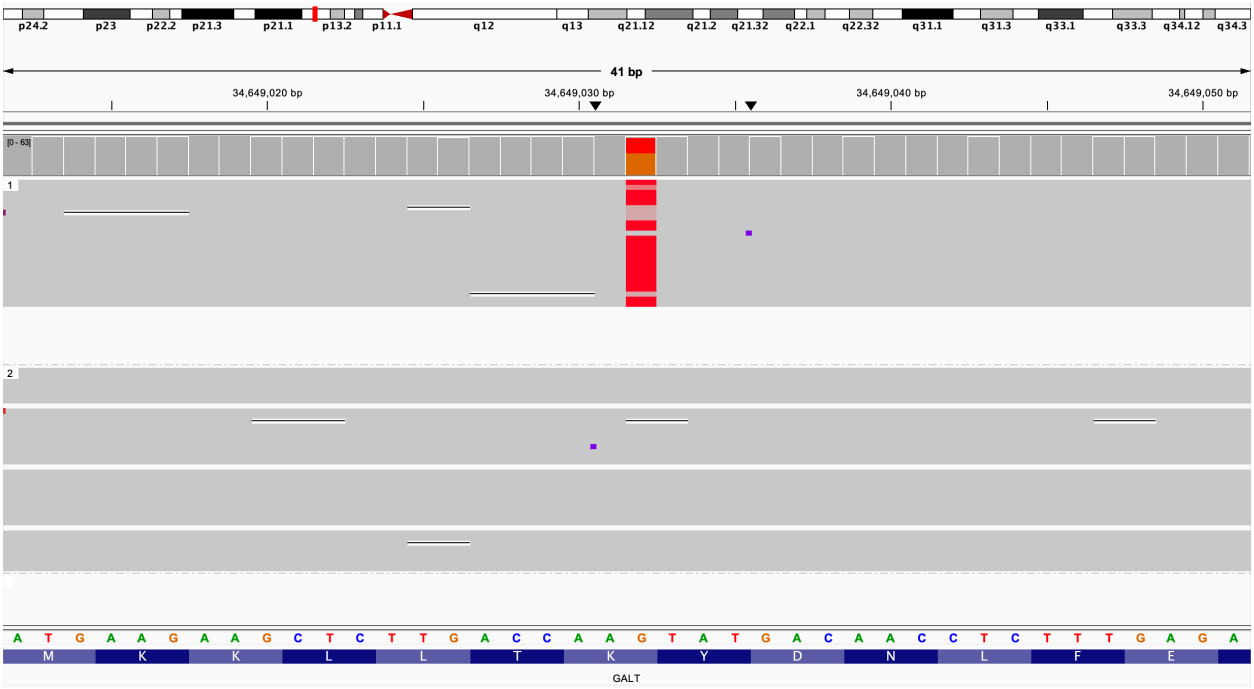


Figure 6: IGV View of NBS1 Variant 1 c.855G>T

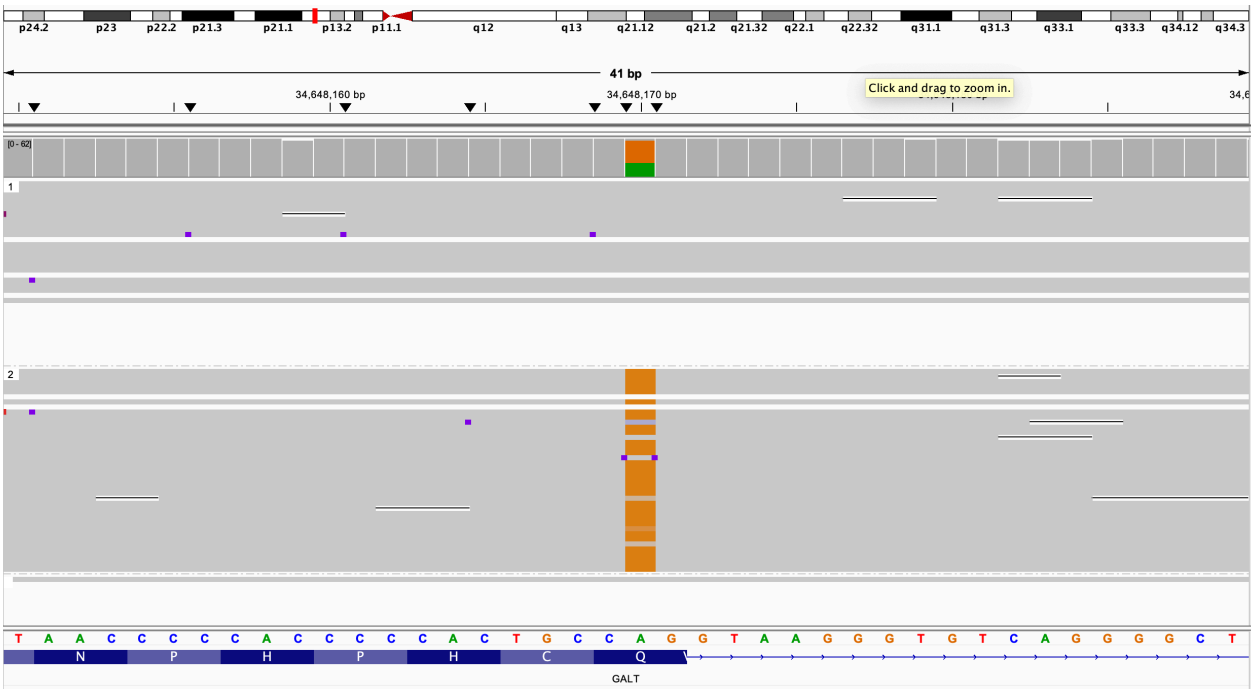


Figure 7: IGV View of NBS1 Variant 2 c.563A>G

M1920/NBS3

This case was flagged for low GALT enzyme activity and suspected to have galactosemia. Two variants c.563A>G (p.Gln188Arg) and c.940A>G (p.Ile314Val) were detected in *GALT*. Variants were confirmed using IGV, screenshots of the variants (**Figures 8-9**) support compound heterozygosity.

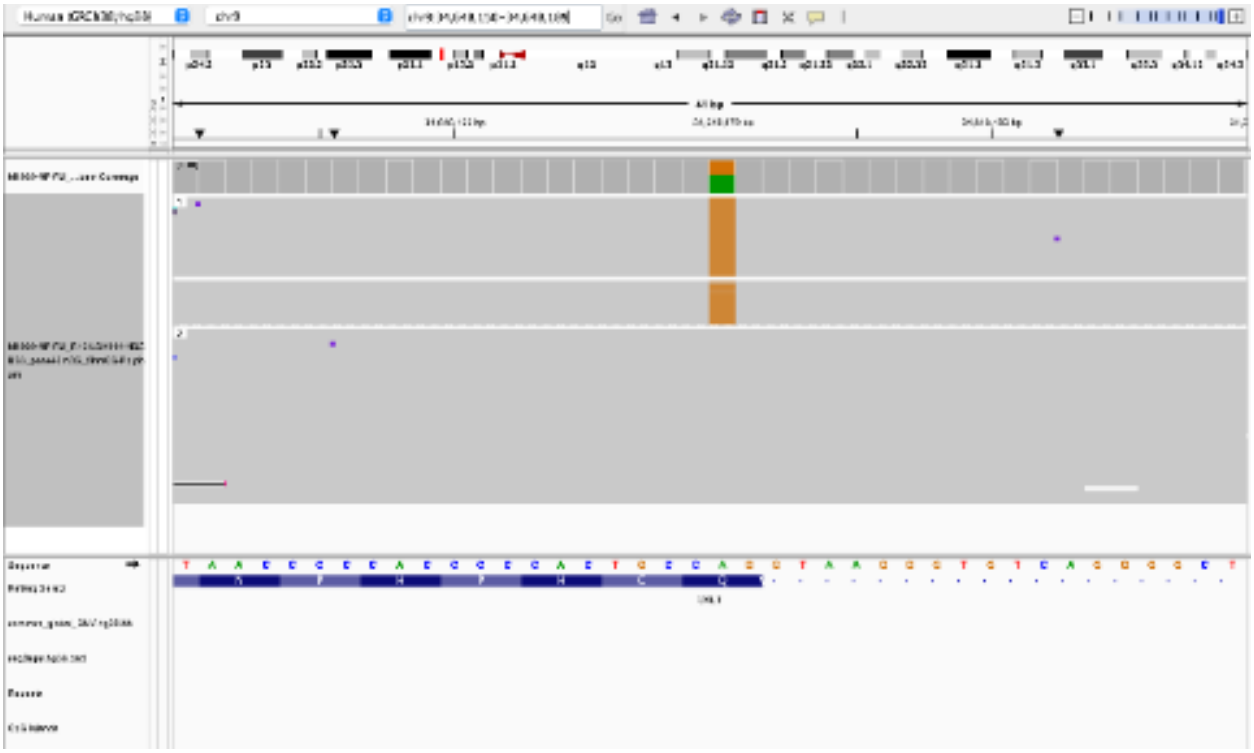


Figure 8: NBS3 Variant 1 c.563A>G

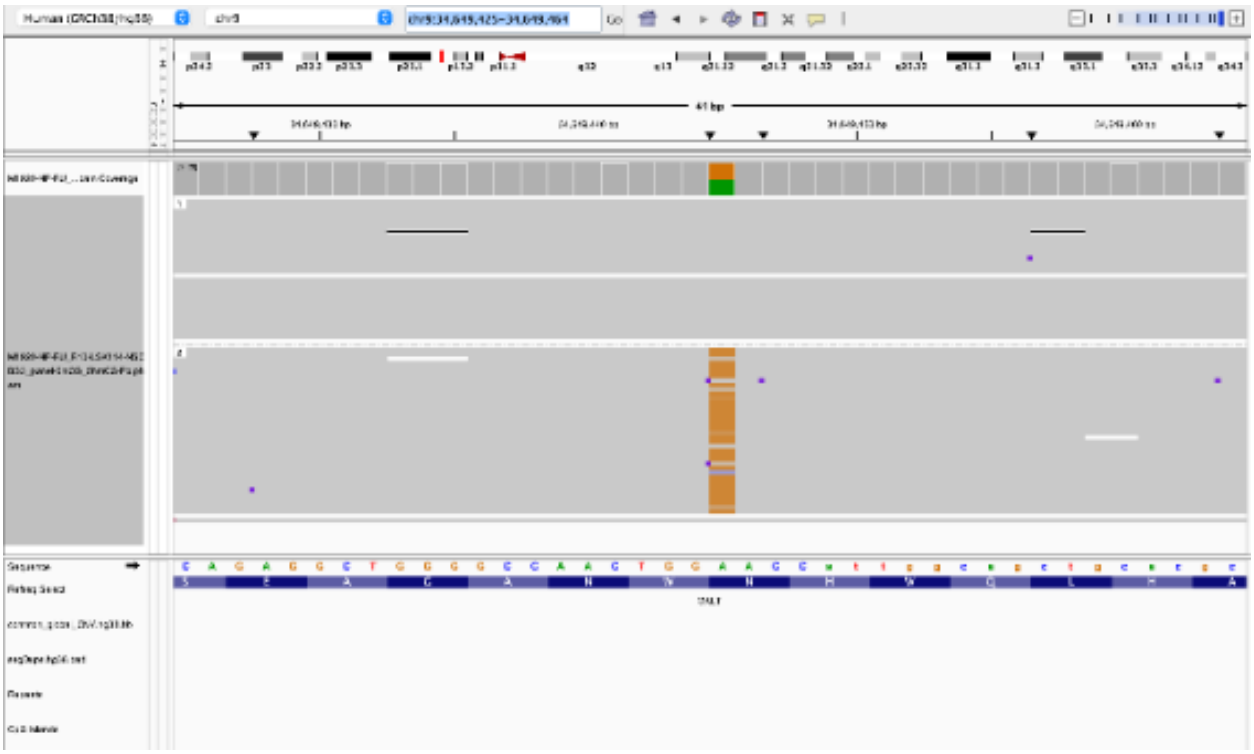


Figure 9: NBS3 Variant 2 c.940A>G

M1921/NBS4

This case was flagged for low GALT enzyme activity and suspected galactosemia. A homozygous pathogenic *GALT* variant (c.563A>G; p.Gln188Arg) was identified, consistent with classic galactosemia. Haplotype phasing seen in IGV (**Figure 10**) confirmed both alleles carried the same variant.

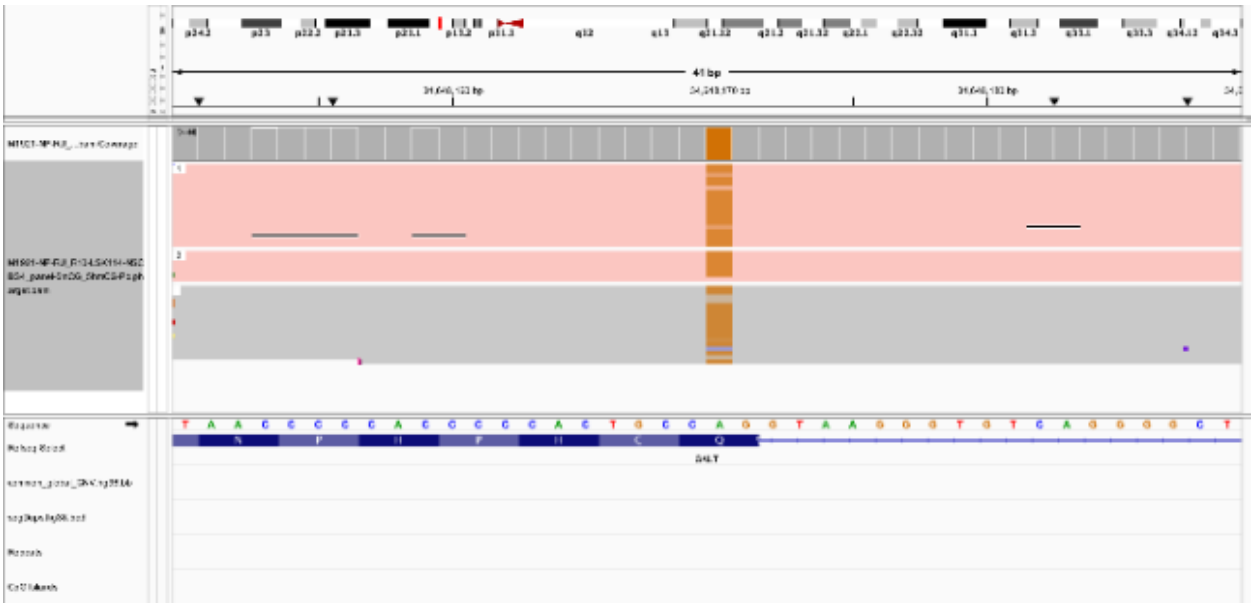


Figure 10: NBS4 Variants c.563A>G

M1922/NBS6

This case was flagged for deficiency in the GAA enzyme and suspected for Pompe disease. Four missense variants were detected in *GAA* c.1933G>T (p.Asn645Tyr), c.2015G>A (p.Arg672Gln), c.2065G>A (p.Glu689Lys), and c.2238G>C (p.Trp746Cys). IGV phasing (**Figures 11-14**) showed that the first three variants were inherited from the same parent, while c.2238G>C was on the opposite haplotype.

Figure 11 NBS6 Variant 1 c.193G>T

Figure 12: NBS6 Variant 2 c.2015G>A

Figure 13: NBS6 Variant 3 c.2065G>A

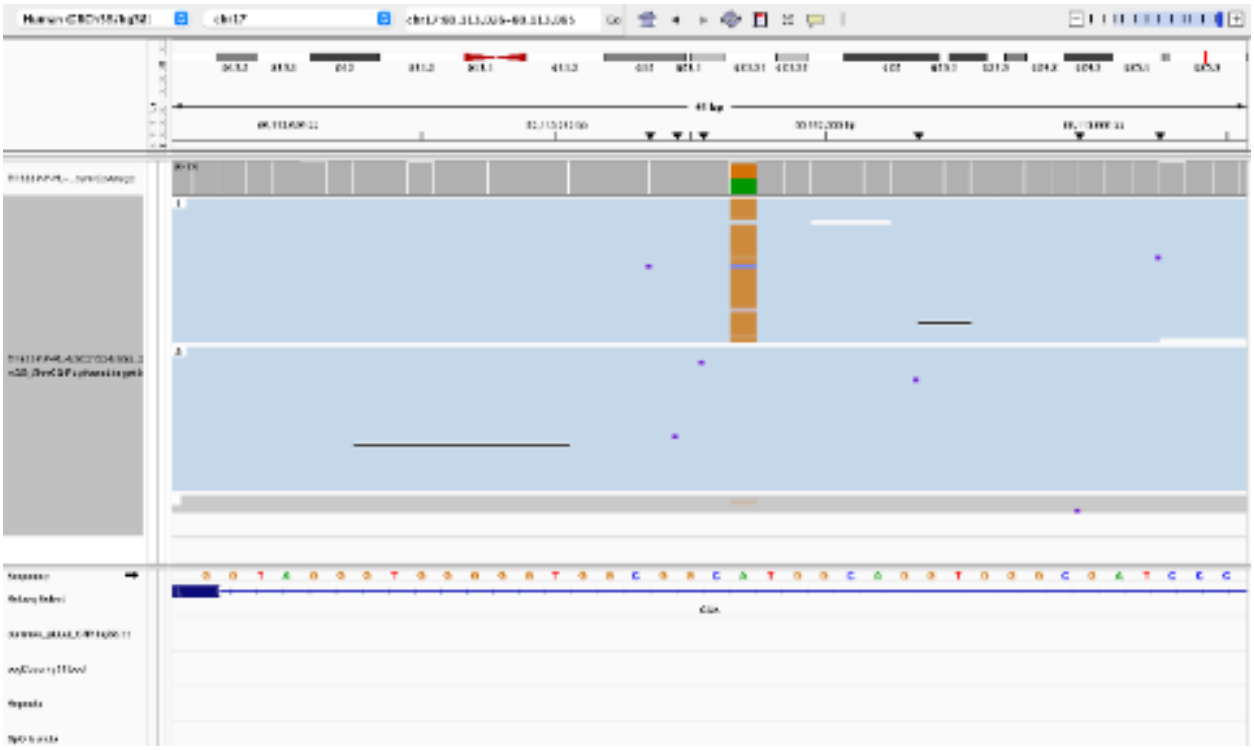


Figure 14: NBS6 Variant 4 c.2238G>C

M1923/NBS7

This case was flagged for deficiency in the GAA enzyme and suspected for Pompe disease. Four *GAA* variants were identified and phased into two haplotypes: c.1477C>T (p.Pro493Ser) and c.2221G>A (p.Asp741Asn) were maternally inherited, while c.1726G>A (p.Gly576Ser) and c.2065G>A (p.Glu689Lys) were paternally inherited. This finding is significant as it demonstrates that the pseudo-deficiency alleles are inherited on a single allele. As a result, the individual is at most a carrier and not expected to be affected by the condition, highlighting the value of phasing in variant interpretation.

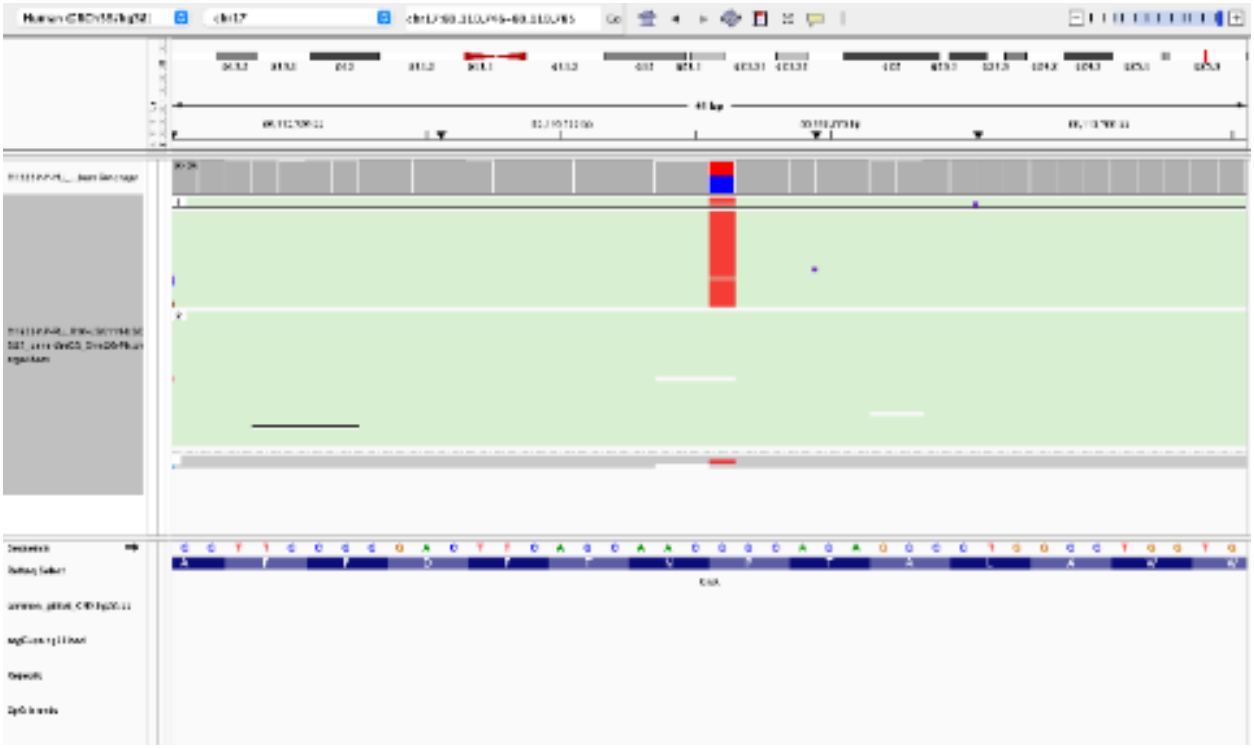


Figure 15: NBS7 Variant 1 c.1477C>T

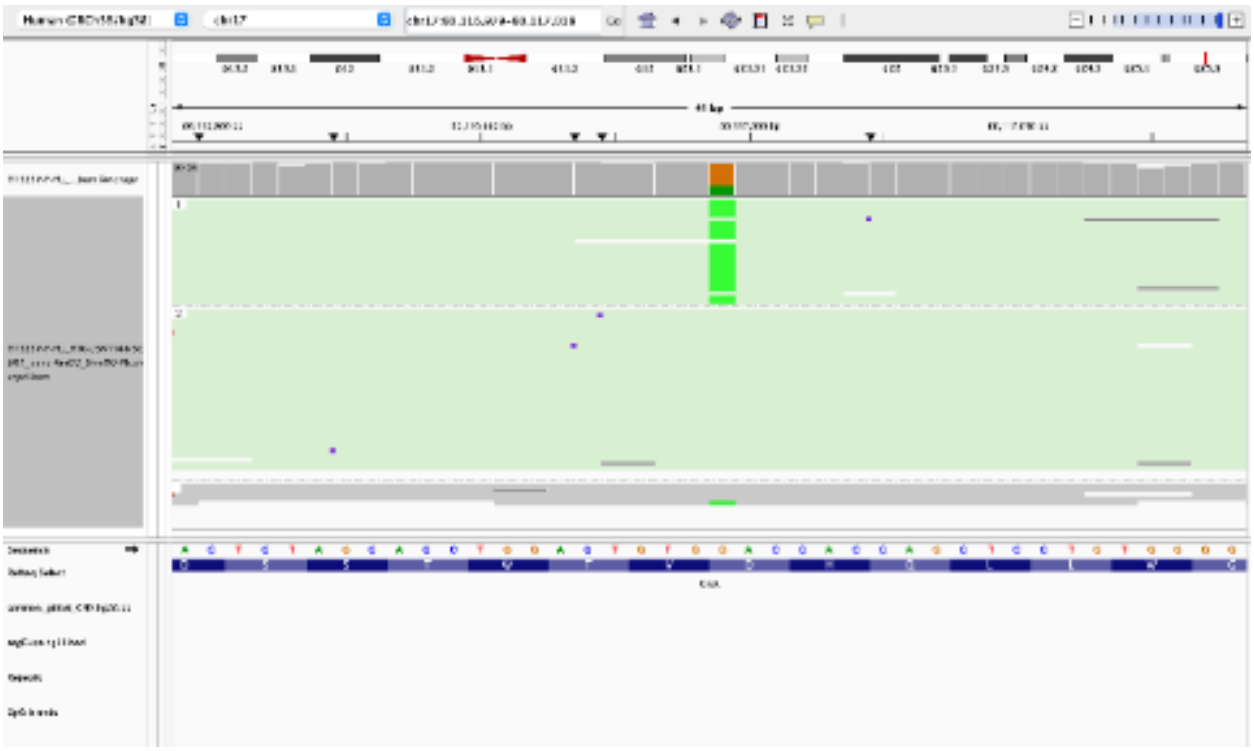


Figure 16: NBS7 Variant 2 c.1726G>A

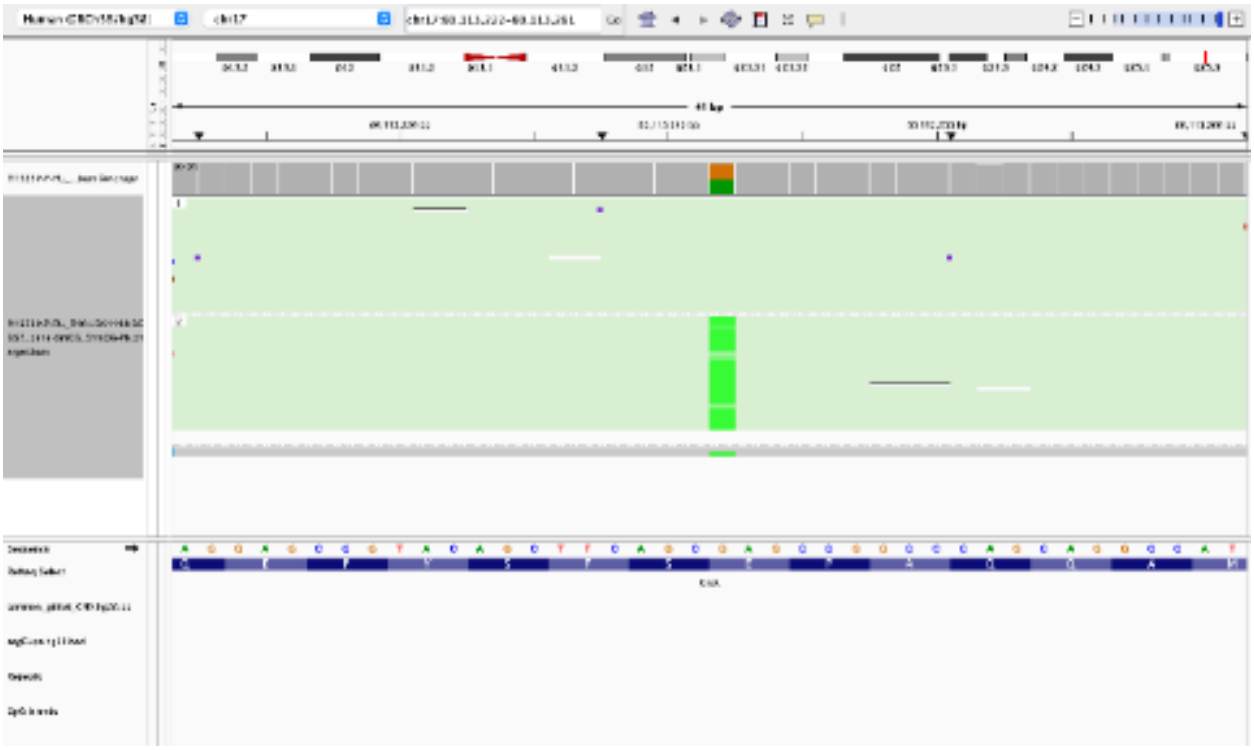


Figure 17: NBS7 Variant 3 c.2065G>A

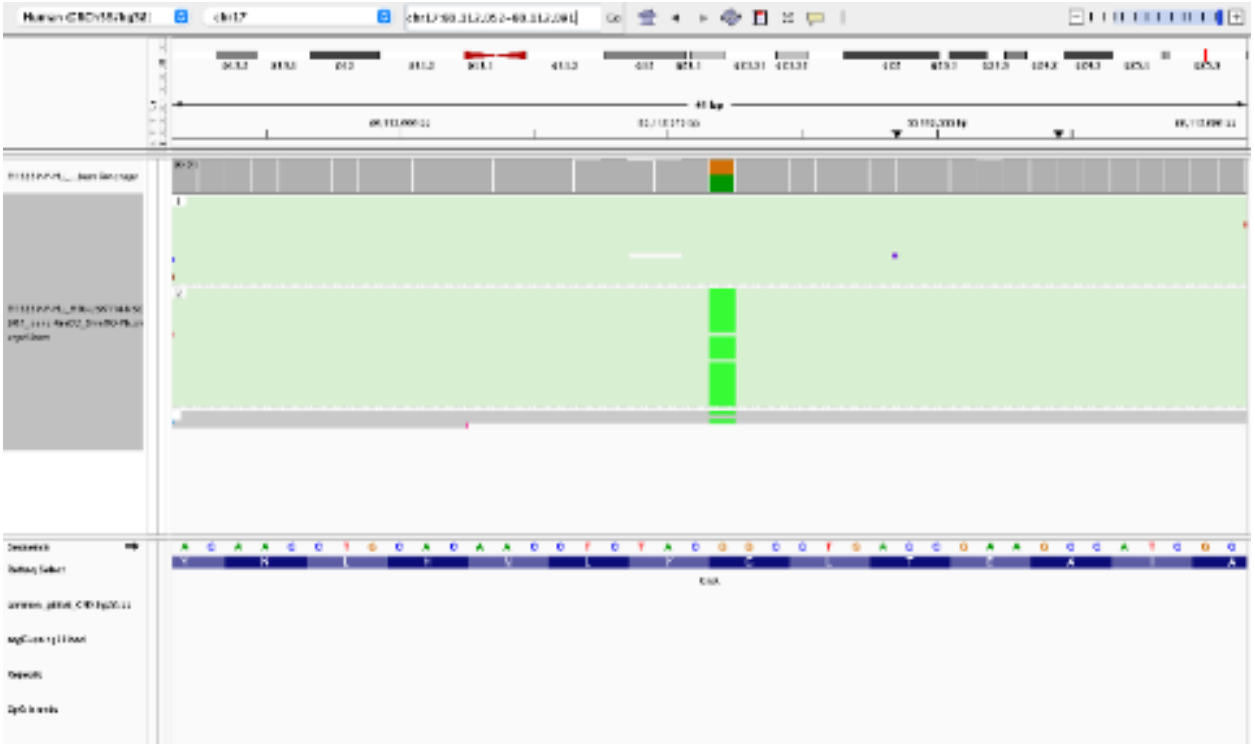


Figure 18: NBS7 Variant 4 c.2221G>A

M1924/NBS8

This case was flagged for deficiency in the GAA enzyme and suspected for Pompe disease. Two pathogenic *GAA* variants were detected: a missense variant (c.2051C>T) maternally inherited, and a deletion of exon 18 on the paternal allele.

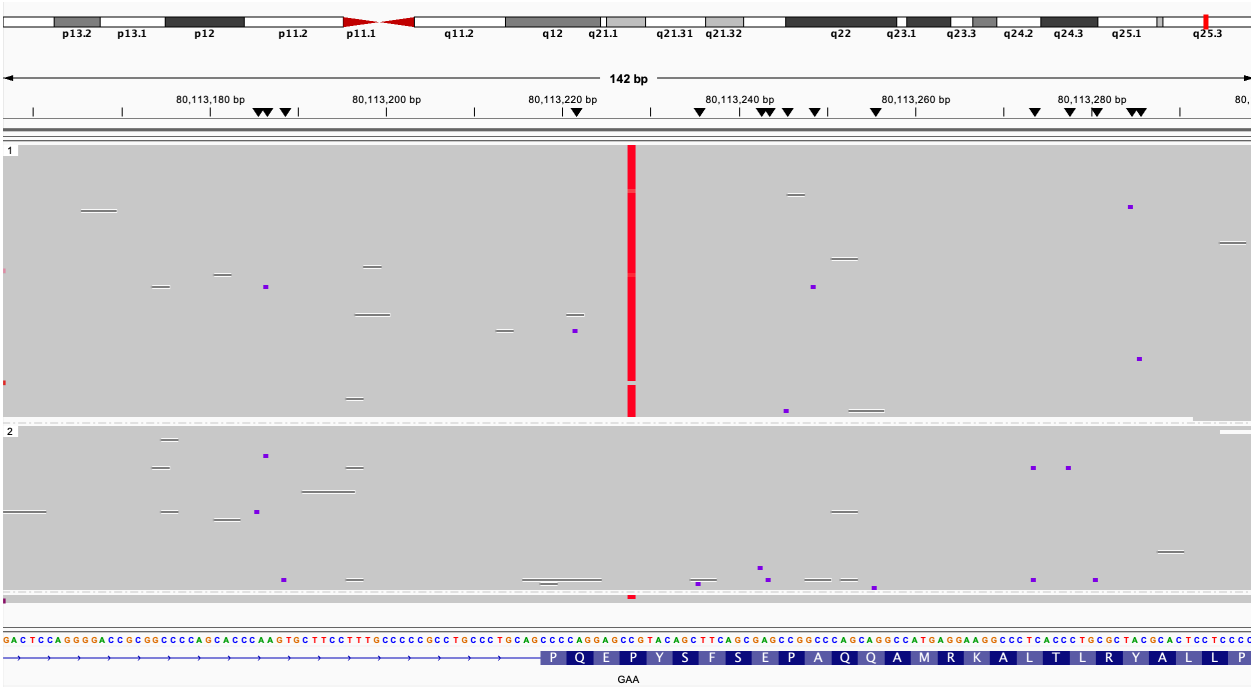


Figure 19: NBS8 Variant 1 c.2051C>T

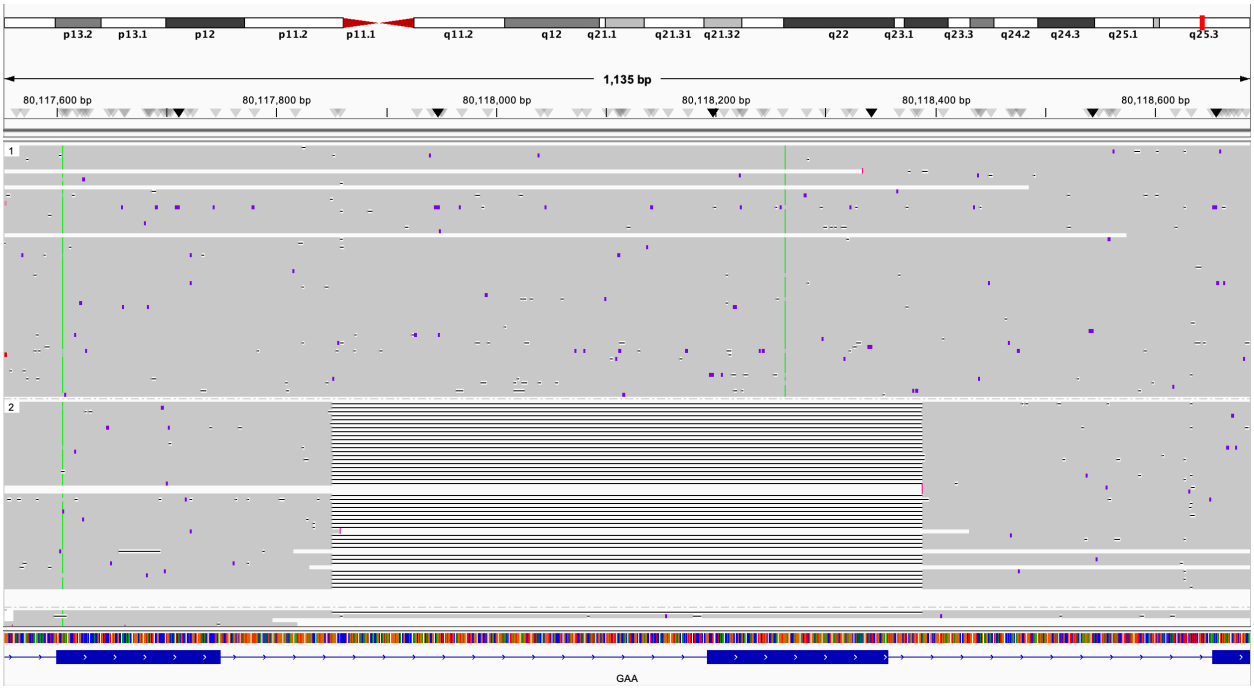


Figure 20: NBS8 Variant 2 Deletion of exon 18

M1925/NBS9

This case was flagged for deficiency in the GAA enzyme and suspected for Pompe disease. Four *GAA* variants were observed: c.752C>T and c.761C>T, both maternally inherited and closely located within exon 4, and c.1726G>A and c.2065G>A, both paternally inherited.

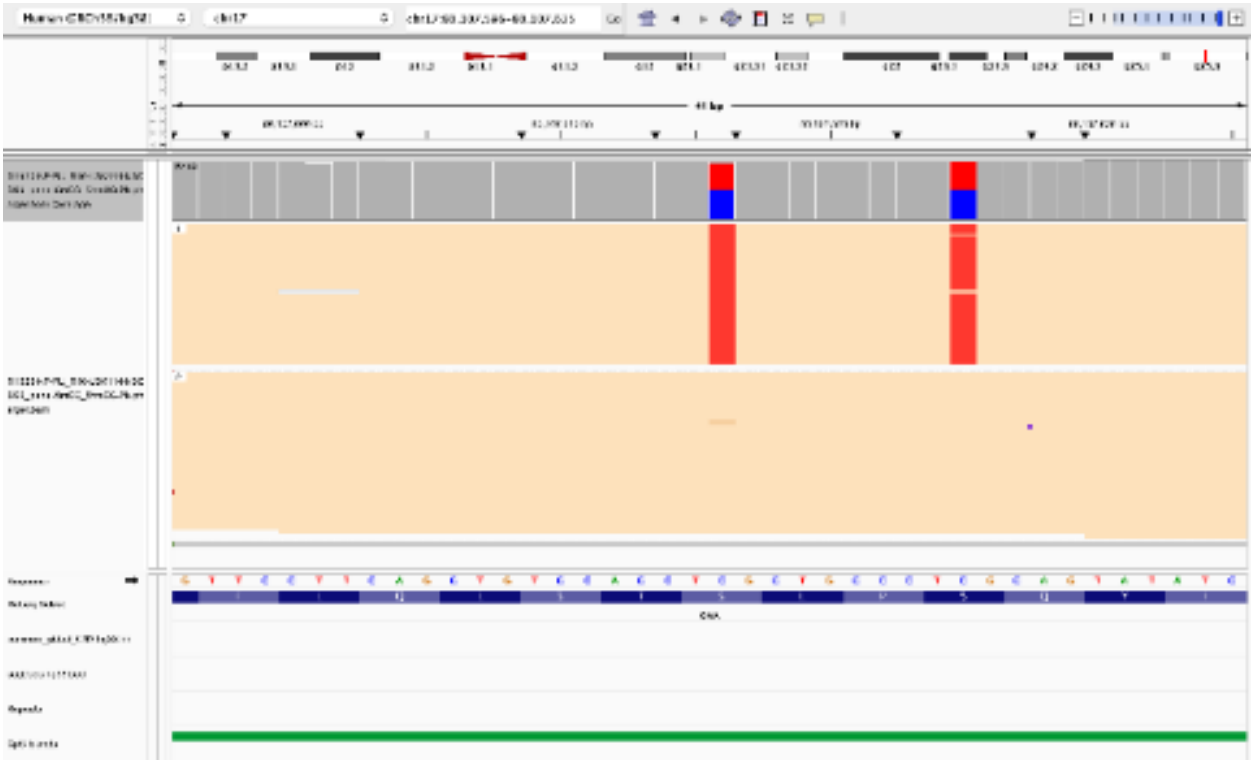


Figure 21: NBS9 Variant 1 and 2 c.752C>T, c.761C>T

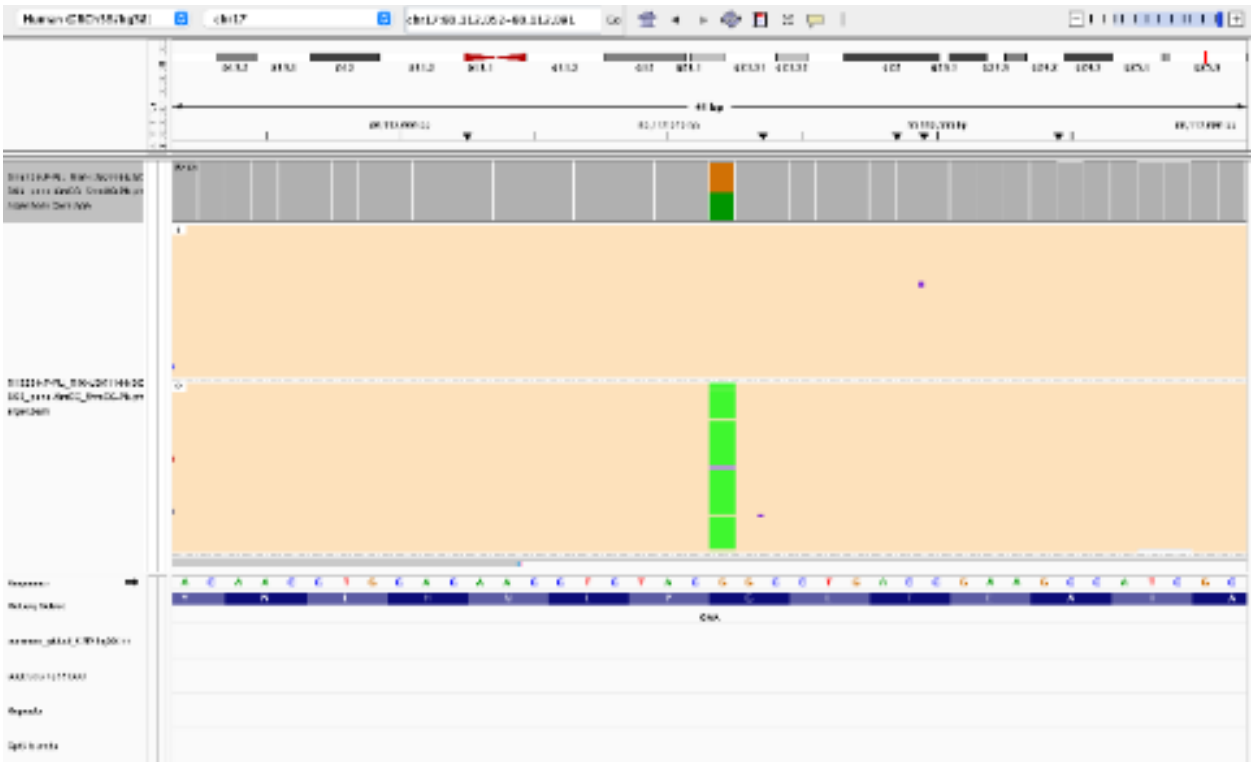


Figure 22: NBS9 Variant 3 c.1726G>A

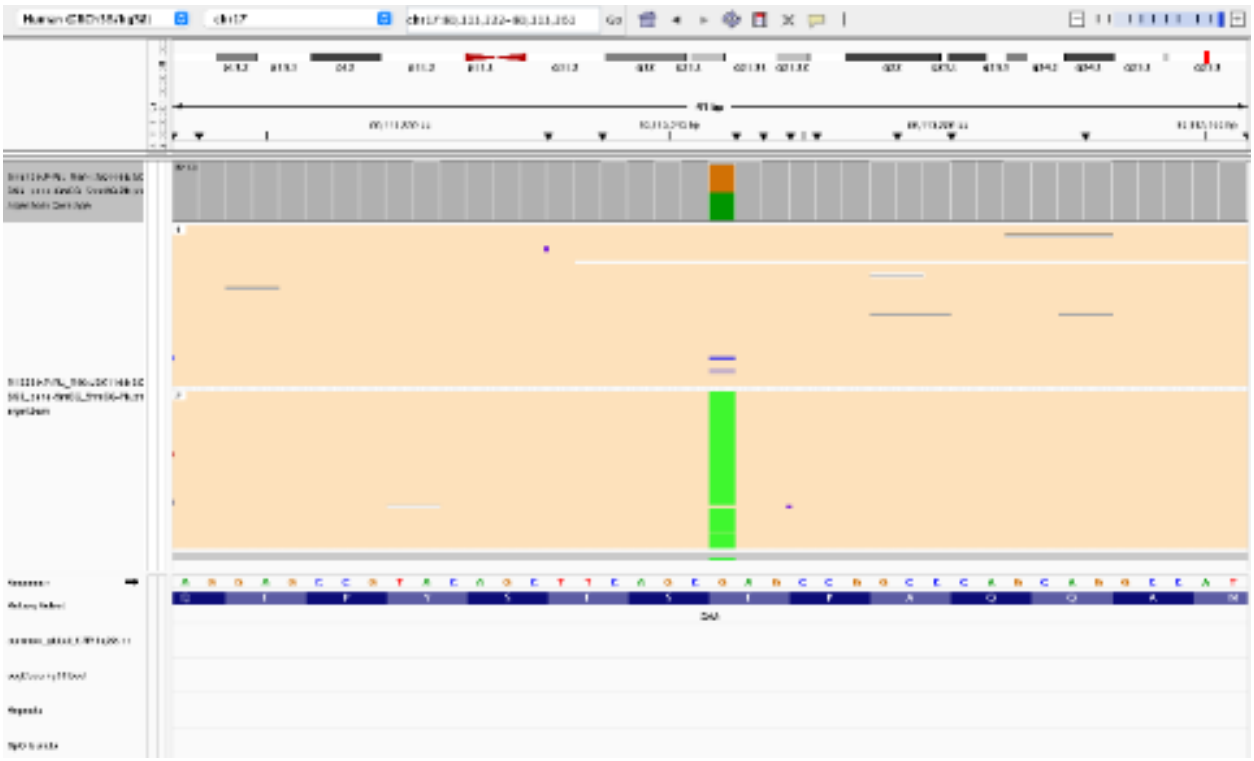


Figure 23: NBS9 Variant 4 c.2065G>A

M1926/NBS10

This case was flagged for deficiency in the GAA enzyme and suspected for Pompe disease. Four *GAA* variants were identified and phased into maternal and paternal haplotypes: c.1927G>A (p.Gly643Arg) and c.2560C>T (p.Arg854) were on the paternal allele, while c.1726G>A (p.Gly576Ser) and c.2065G>A (p.Glu689Lys) were maternally inherited. This case provides another compelling value of phasing. Analysis showed that both variants are located on the same allele, which indicates that the infant is a carrier and not at risk of disease. This level of resolution could not have been achieved through Sanger sequencing of the infant alone, which further demonstrates the added clinical utility of long-read sequencing for variant interpretation.

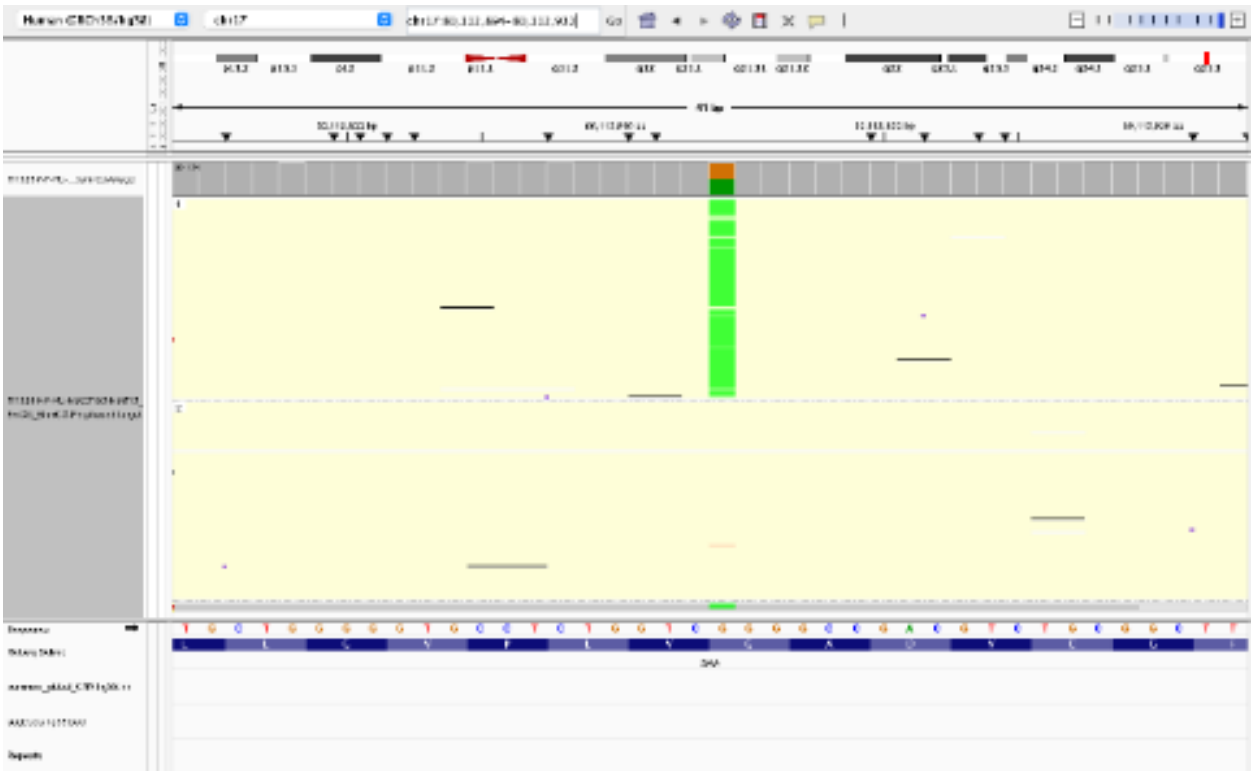


Figure 24: NBS10 Variant 1 c.1927G>A

Figure 25: NBS10 Variant 2 c.2560C>T

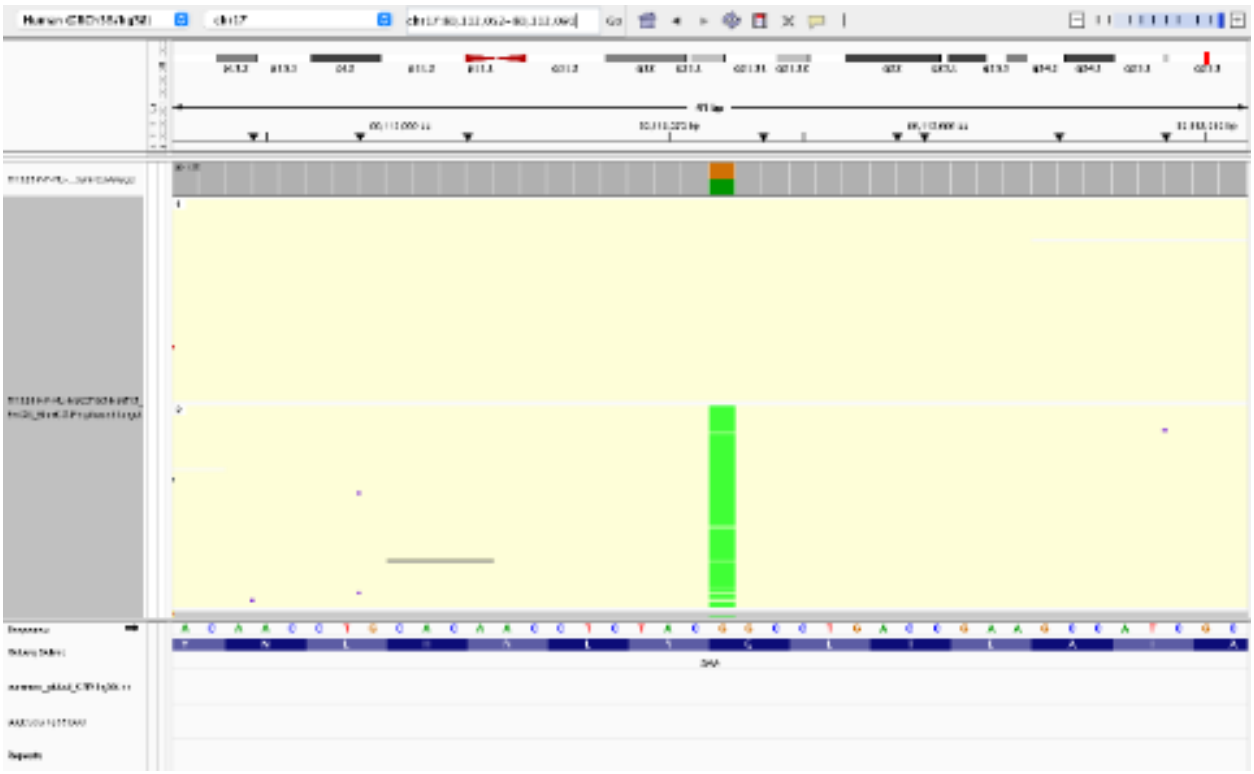


Figure 26: NBS10 Variant 3 c.1726G>A

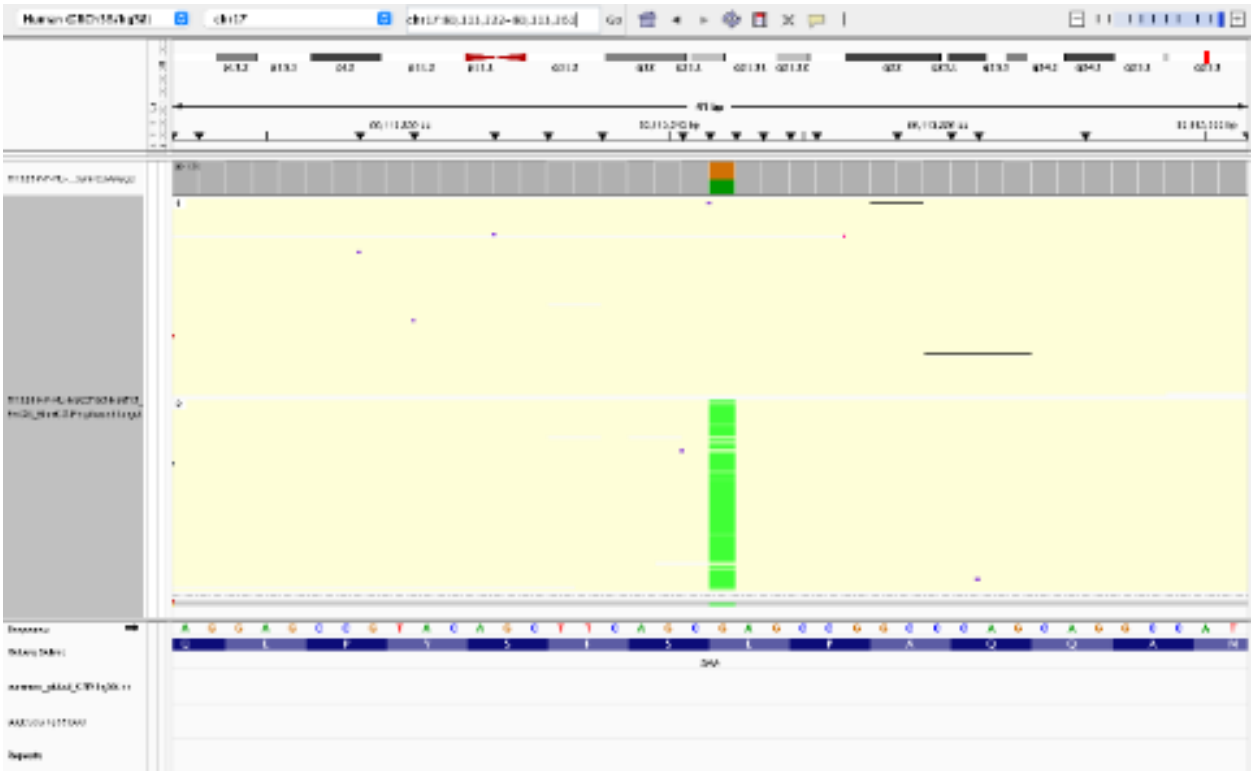


Figure 27: NBS10 Variant 4 c.2065G>A

CHAPTER IV: DISCUSSION

We demonstrate the ability of using the Oxford Nanopore long-read sequencing platform for validating NBS results. Of the eight positive control samples provided by Seattle Children's Hospital, we were able to identify both SNVs and indels across clinically relevant genes, with a focus on regions that are often challenging for short-read technologies. Our results show T-LRS can accurately detect and phase pathogenic variants in genes such as *GALT* and *GAA*, both associated with conditions screened for in state-mandated NBS programs. These findings highlight the potential of T-LRS to enhance the diagnostic rate of NBS follow-up, especially when traditional testing methods are inconclusive.

One of the biggest advantages of LRS is its ability to capture complex variants along with phasing. For example, we detected an exon 18 deletion in *GAA* in NBS8, which may have been missed or misinterpreted by conventional methods like SRS. These results support prior studies that have underscored the ability of long-read technologies in characterizing regions that are difficult to sequence. Our eventual goal is to perform whole-genome sequencing using LRS to follow up on NBS results, however, T-LRS remains a cost-effective approach scalable for clinical workflows. By focusing on specific regions of interest, we achieved sufficient coverage and depth, allowing us to perform comprehensive analysis on each sample.

CHAPTER V: LIMITATIONS

Despite these strengths, several limitations must be acknowledged. First, the sample size was relatively small, limiting the generalizability of our findings. Although our results are promising, additional validation across larger sample size and a more diverse cohort with genes other than *GAA* and *GALT* are necessary to determine the utility of T-LRS in NBS results follow

up. Second, while T-LRS offers excellent variant detection capabilities, it requires relatively high DNA quality. Depending on the DNA extraction method and sample source, some samples may require further treatment or cleanup to improve read quality and yield, which may not be feasible in a state laboratory. Third, the use of our analysis pipeline, followed by manual inspection in IGV allowed us to confirm the accuracy of variants based on read depth and base quality. However, the proposed workflow remains computationally intensive, requiring manual curation for variant interpretation and phasing, which may be a barrier for clinical adoption without automation. Lastly, it's important to reiterate that while T-LRS offers a great way to detect complex variants and accurate phasing, it may not fully replace existing biochemical assays. For disorders such as classic galactosemia or Pompe disease, enzymatic activity measurements remain critical for diagnosis.

CHAPTER VI: FUTURE DIRECTIONS

Future studies should focus on testing this workflow on DNA directly extracted from Guthrie cards to evaluate its feasibility in a public health setting. To further assess the reproducibility of this approach, collaborations with state laboratories to test the workflow is recommended. Additionally, it would be beneficial to explore other difficult to sequence genes such as the SMN region that may benefit from long-read technologies. Ultimately, as sequencing costs decrease and with continued improvements in variant calling algorithms, we are confident that T-LRS will become a valuable resource to NBS follow up, by enabling more accurate diagnoses, reducing the number of false positives and negatives, and supporting timely treatments for affected newborns.

ACKNOWLEDGEMENTS

I would like to thank my committee, Danny and Anna, for their mentorship and guidance throughout my thesis project. Thank you for believing in me when I said I couldn't do it. Thank you for giving me feedback and always asking me hard questions. To everyone in the Miller Lab, thank you for creating an inclusive environment. I've been so lucky to be surrounded by people who are not only great at research, but also fun and welcoming colleagues. I'll always remember the laughs we shared learning a new protocol, troubleshooting experiments, and eating lunch at that round table. I'm truly grateful to have been part of the Miller Lab.

To my friends and family, even though you guys never remembered exactly what I am studying, your constant love, encouragement, and understanding have kept me going when things felt overwhelming. I couldn't have done this without you guys by my side. And lastly, to my dad, thank you for making so many sacrifices so that I could have better educational opportunities. None of this would have been possible without your strength and hard work. This one is for us.

REFERENCES

1. Merritt JL, Chang IJ. Medium-Chain Acyl-Coenzyme A Dehydrogenase Deficiency. In: Adam MP, Feldman J, Mirzaa GM, et al., eds. *GeneReviews®*. University of Washington, Seattle; 1993. Accessed April 21, 2024. http://www.ncbi.nlm.nih.gov/books/NBK1424/

2. Ahrens-Nicklas RC, Pyle LC, Ficicioglu C. Morbidity and mortality among exclusively breastfed neonates with medium-chain acyl-CoA dehydrogenase deficiency. *Genet Med Off J Am Coll Med Genet*. 2016;18(12):1315-1319. doi:10.1038/gim.2016.49

3. Wilcken B, Hammond J, Silink M. Morbidity and mortality in medium chain acyl coenzyme A dehydrogenase deficiency. *Arch Dis Child*. 1994;70(5):410-412.

4. Clark MM, Stark Z, Farnaes L, et al. Meta-analysis of the diagnostic and clinical utility of genome and exome sequencing and chromosomal microarray in children with suspected genetic diseases. *Npj Genomic Med*. 2018;3(1):1-10. doi:10.1038/s41525-018-0053-8

5. Lunn MR, Wang CH. Spinal muscular atrophy. *Lancet Lond Engl*. 2008;371(9630):2120-2133. doi:10.1016/S0140-6736(08)60921-6

6. Nusinersen (Spinraza®) – Spinal Muscular Atrophy (SMA) | National Institute of Neurological Disorders and Stroke. Accessed April 29, 2025. https://www.ninds.nih.gov/about-ninds/what-we-do/impact/ninds-contributions-approved-therapies/nusinersen-spinrazar-spinal-muscular-atrophy-sma

7. Vrščaj E, Dangouloff T, Osredkar D, Servais L. Newborn screening programs for spinal muscular atrophy worldwide in 2023. *J Neuromuscul Dis*. 2024;11(6):1180-1189. doi:10.1177/22143602241288095

8. Bick SL, Nathan A, Park H, Green RC, Wojcik MH, Gold NB. Estimating the sensitivity of genomic newborn screening for treatable inherited metabolic disorders. *Genet Med*. 2025;27(1). doi:10.1016/j.gim.2024.101284

9. Knapkova M, Hall K, Loeber G. Reliability of Neonatal Screening Results. *Int J Neonatal Screen*. 2018;4(3):28. doi:10.3390/ijns4030028

10. Raymond GV, Moser AB, Fatemi A. X-Linked Adrenoleukodystrophy. In: Adam MP, Feldman J, Mirzaa GM, et al., eds. *GeneReviews®*. University of Washington, Seattle; 1993. Accessed March 24, 2024. http://www.ncbi.nlm.nih.gov/books/NBK1315/

11. Liu S, Li L, Wu H, et al. Genetic analysis and prenatal diagnosis of 76 Chinese families with X-linked adrenoleukodystrophy. *Mol Genet Genomic Med*. 2022;10(1):e1844. doi:10.1002/mgg3.1844

12. Damaraju N, Miller AL, Miller DE. Long-Read DNA and RNA Sequencing to Streamline Clinical Genetic Testing and Reduce Barriers to Comprehensive Genetic Testing. *J Appl Lab Med*. 2024;9(1):138-150. doi:10.1093/jalm/jfad107

13. Blood M. Sequencing 101: long-read sequencing. PacBio. March 2, 2023. Accessed May 5, 2025. https://www.pacb.com/blog/long-read-sequencing/

14. How nanopore sequencing works. Oxford Nanopore Technologies. Accessed May 5, 2025. https://nanoporetech.com/platform/technology

15. Delahaye C, Nicolas J. Sequencing DNA with nanopores: Troubles and biases. *PLoS ONE*. 2021;16(10):e0257521. doi:10.1371/journal.pone.0257521

16. Mastrorosa FK, Miller DE, Eichler EE. Applications of long-read sequencing to Mendelian genetics. *Genome Med*. 2023;15:42. doi:10.1186/s13073-023-01194-3

17. Miller DE, Sulovari A, Wang T, et al. Targeted long-read sequencing identifies missing disease-causing variation. *Am J Hum Genet*. 2021;108(8):1436-1449. doi:10.1016/j.ajhg.2021.06.006

18. Miller DE, Lee L, Galey M, et al. Targeted long-read sequencing identifies missing pathogenic variants in unsolved Werner syndrome cases. *J Med Genet*. 2022;59(11):1087-1094. doi:10.1136/jmedgenet-2022-108485

19. Elsas LJ, Lai K. The molecular biology of galactosemia. *Genet Med Off J Am Coll Med Genet*. 1998;1(1):40-48. doi:10.1097/00125817-199811000-00009

20. Kroos M, Pomponio RJ, van Vliet L, et al. Update of the Pompe disease mutation database with 107 sequence variants and a format for severity rating. *Hum Mutat*. 2008;29(6):E13-26. doi:10.1002/humu.20745

21. Leslie N, Bailey L. Pompe Disease. In: Adam MP, Feldman J, Mirzaa GM, Pagon RA, Wallace SE, Amemiya A, eds. *GeneReviews®*. University of Washington, Seattle; 1993. Accessed June 6, 2025. http://www.ncbi.nlm.nih.gov/books/NBK1261/

22. Zheng Z, Li S, Su J, Leung AWS, Lam TW, Luo R. Symphonizing pileup and full-alignment for deep learning-based long-read variant calling. *Nat Comput Sci*. 2022;2(12):797-803. doi:10.1038/s43588-022-00387-x

23. Smolka M, Paulin LF, Grochowski CM, et al. Detection of mosaic and population-level structural variants with Sniffles2. *Nat Biotechnol*. 2024;42(10):1571-1580. doi:10.1038/s41587-023-02024-y

24. Jiang T, Liu Y, Jiang Y, et al. Long-read-based human genomic structural variation detection with cuteSV. *Genome Biol*. 2020;21(1):189. doi:10.1186/s13059-020-02107-y

25. Scheinin I, Sie D, Bengtsson H, et al. DNA copy number analysis of fresh and formalin-fixed specimens by shallow whole-genome sequencing with identification and exclusion of problematic regions in the genome assembly. *Genome Res*. 2014;24(12):2022-2032. doi:10.1101/gr.175141.114

26. Lin JH, Chen LC, Yu SC, Huang YT. LongPhase: an ultra-fast chromosome-scale phasing algorithm for small and large variants. *Bioinformatics*. 2022;38(7):1816-1822. doi:10.1093/bioinformatics/btac058

APPENDIX A

Targeted Gene List- first version

Total target size is 73.45 MB

Total percent of diploid human genome is 2.37%

ABCC8, ABCD1, ABCD3, ABCD4, ACAD8, ACAD9, ACADL, ACADM, ACADS, ACADSB, ACADVL, ACAT1, ACOX1, ACSF3, ACY1, ADA, ADAMTSL2, ADAR, ADCK3, ADK, ADSL, AGA, AGK, AGL, AGPAT2, AGPS, AGXT, AHCY, AKT2, ALAD, ALAS2, ALDH5A1, ALDH7A1, ALDOA, ALDOB, ALG1, ALG11, ALG12, ALG13, ALG2, ALG3, ALG6, ALG8, ALG9, AMACR, AMN, AMPD1, AMT, ANO10, ANO5, ANTXR2, APRT, APTX, ARG1, ARSA, ARSB, ASAH1, ASL, ASPA, ASS1, ATIC, ATP13A2, ATP2A1, ATP6V0A2, ATP7B, AUH, B3GALNT2, B3GLCT, B4GALT1, B4GAT1, BCKDHA, BCKDHB, BCS1L, BOLA3, BSCL2, BSND, BTD, C10ORF2, C12ORF65, CA5A, CACNA1S, CAPN3, CASQ1, CASR, CAV1, CAV3, CBS, CD320, CHKB, CIDEC, CLCN1, CLCNKB, CLDN16, CLDN19, CLN3, CLN5, CLN6, CLN8, CLPB, CNNM2, CNNM4, COG1, COG4, COG5, COG6, COG7, COG8, COL11A2, COL2A1, COQ2, COQ4, COQ5, COQ6, COQ7, COQ9, CPOX, CPS1, CPT1A, CPT2, CTDP1, CTH, CTNS, CTSA, CTSC, CTSD, CTSK, CUBN, D2HGDH, DAG1, DBT, DDOST, DGUOK, DHCR7, DHDDS, DHODH, DLD, DMD, DNAJB6, DNAJC12, DNM1L, DOLK, DPAGT1, DPM1, DPM2, DPM3, DPYD, DPYS, DYM, DYSF, EBP, ECHS1, EGF, EMD, ENO3, EPM2A, ETFA, ETFB, ETFDH, FAH, FAM111A, FBP1, FBXL4, FDX1L, FECH, FH, FHL1, FKRP, FKTN, FLAD1, FLNA, FLNB, FMO3, FOLR1, FOXRED1, FUCA1, FUT8, FXYD2, G6PC, GAA, GALC, GALE, GALK1, GALNS, GALT, GAMT, GATM, GBA, GBE1, GCDH, GCH1, GCK, GCSH, GFM1, GIF, GLA, GLB1, GLDC, GLRX5, GLUD1, GLUL, GM2A,

GMPPA, GMPPB, GNE, GNMT, GNPAT, GNPTAB, GNPTG, GNS, GPC3, GPHN, GUSB, GYG1, GYS1, GYS2, HADH, HADHA, HADHB, HAMP, HCFC1, HEXA, HEXB, HFE, HFE2, HGD, HGSNAT, HIBCH, HLCS, HMBS, HMGCL, HMGCS2, HNF1A, HNF1B, HNF4A, HPD, HPRT1, HRAS, HSD17B10, HSD17B4, HYAL1, IDH2, IDS, IDUA, IFIH1, INSR, ISCU, IVD, KCNA1, KCNJ10, KCNJ11, KCNJ2, L2HGDH, LAMA2, LAMP2, LARGE, LCT, LDB3, LDHA, LIAS, LIPA, LIPE, LIPT1, LMBRD1, LMNA, LPIN1, MAGT1, MAN1B1, MAN2B1, MANBA, MCCC1, MCCC2, MCEE, MCOLN1, MFN2, MFSD8, MGAT2, MLYCD, MMAA, MMAB, MMACHC, MMADHC, MOCOS, MOCS1, MOCS2, MOGS, MPDU1, MPI, MPV17, MTHFR, MTR, MTRR, MUT, MYH3, MYOT, NAGA, NAGLU, NAGS, NBAS, NDUFAF2, NDUFS1, NEU1, NFU1, NGLY1, NHLRC1, NIPA2, NPC1, NPC2, NT5C3A, OAT, OPA1, OPA3, OTC, OXCT1, PAH, PC, PCBD1, PCCA, PCCB, PCK1, PDHA1, PDHB, PDHX, PDSS1, PDSS2, PDX1, PEPD, PEX1, PEX10, PEX11B, PEX12, PEX13, PEX14, PEX16, PEX19, PEX2, PEX26, PEX3, PEX5, PEX6, PEX7, PFKM, PGAM2, PGK1, PGM1, PHKA1, PHKA2, PHKB, PHKG2, PHYH, PLIN1, PMM2, PNP, PNPLA2, POLG, POLG2, POMGNT1, POMGNT2, POMK, POMT1, POMT2, PPARG, PPOX, PPT1, PRKAG2, PRKAG3, PRODH, PRPS1, PSAP, PTF1A, PTRF, PTS, PYGL, PYGM, QDPR, RAI1, RBCK1, REN, RFT1, RNASEH2A, RNASEH2B, RNASEH2C, RRM2B, RYR1, SAMHD1, SARS2, SCN4A, SEC23B, SERAC1, SERPINA1, SGCA, SGCB, SGCD, SGCG, SGSH, SI, SIL1, SLC12A3, SLC16A1, SLC17A5, SLC22A5, SLC25A1, SLC25A13, SLC25A15, SLC25A20, SLC25A26, SLC25A3, SLC25A4, SLC2A1, SLC2A2, SLC30A10, SLC35A1, SLC35A2, SLC35C1, SLC37A4, SLC39A4, SLC3A1, SLC40A1, SLC46A1, SLC5A1, SLC6A19, SLC6A8, SLC6A9, SLC7A7, SLC7A9, SMPD1, SPG7, SSR4, STAC3, STT3A, STT3B, SUCLA2, SUCLG1, SUGCT, SUMF1, SUOX, TALDO1, TANGO2, TAT, TAZ, TBC1D4, TCAP, TCF4, TCN2, TFR2, TIMM8A, TK2, TMEM126A,

TMEM165, TMEM70, TNPO3, TPMT, TPP1, TREX1, TRIM32, TRIM37, TRPM6, TUSC3, TYMP, UCP2, UMOD, UMPS, UPB1, UROD, UROS, WFS1, XDH, ZMPSTE24

APPENDIX B

Targeted Gene List- updated version

Total target size is 79.60 MB

Total percent of diploid human genome is 2.57%

ABCC8, ABCD1, ABCD3, ABCD4, ACAD8, ACAD9, ACADL, ACADM, ACADS, ACADSB, ACADVL, ACAT1, ACOX1, ACSF3, ACY1, ADA, ADAMTSL2, ADAR, ADCK3, ADK, ADSL, AGA, AGK, AGL, AGPAT2, AGPS, AGXT, AHCY, AKT2, ALAD, ALAS2, ALDH5A1, ALDH7A1, ALDOA, ALDOB, ALG1, ALG11, ALG12, ALG13, ALG2, ALG3, ALG6, ALG8, ALG9, AMACR, AMN, AMPD1, AMT, ANO10, ANO5, ANTXR2, APRT, APTX, ARG1, ARSA, ARSB, ASAH1, ASL, ASPA, ASS1, ATIC, ATP13A2, ATP2A1, ATP6V0A2, ATP7B, AUH, B3GALNT2, B3GLCT, B4GALT1, B4GAT1, BCKDHA, BCKDHB, BCS1L, BOLA3, BSCL2, BSND, BTD, C10ORF2, C12ORF65, CA5A, CACNA1S, CAPN3, CASQ1, CASR, CAV1, CAV3, CBS, CD320, CHKB, CIDEC, CLCN1, CLCNKB, CLDN16, CLDN19, CLN3, CLN5, CLN6, CLN8, CLPB, CNNM2, CNNM4, COG1, COG4, COG5, COG6, COG7, COG8, COL11A2, COL2A1, COQ2, COQ4, COQ5, COQ6, COQ7, COQ9, CPOX, CPS1, CPT1A, CPT2, CTDP1, CTH, CTNS, CTSA, CTSC, CTSD, CTSK, CUBN, D2HGDH, DAG1, DBT, DDOST, DGUOK, DHCR7, DHDDS, DHODH, DLD, DMD, DNAJB6, DNAJC12, DNM1L, DOLK, DPAGT1, DPM1, DPM2, DPM3, DPYD, DPYS, DYM, DYSF, EBP, ECHS1, EGF, EMD, ENO3, EPM2A, ETFA, ETFB, ETFDH, FAH, FAM111A, FBP1, FBXL4, FDX1L, FECH, FH, FHL1, FKRP, FKTN, FLAD1, FLNA, FLNB, FMO3, FOLR1, FOXRED1, FUCA1, FUT8, FXYD2, G6PC, GAA, GALC, GALE, GALK1, GALNS,

GALT, GAMT, GATM, GBA, GBE1, GCDH, GCH1, GCK, GCSH, GFM1, GIF, GLA, GLB1, GLDC, GLRX5, GLUD1, GLUL, GM2A, GMPPA, GMPPB, GNE, GNMT, GNPAT, GNPTAB, GNPTG, GNS, GPC3, GPHN, GUSB, GYG1, GYS1, GYS2, HADH, HADHA, HADHB, HAMP, HCFC1, HEXA, HEXB, HFE, HFE2, HGD, HGSNAT, HIBCH, HLCS, HMBS, HMGCL, HMGCS2, HNF1A, HNF1B, HNF4A, HPD, HPRT1, HRAS, HSD17B10, HSD17B4, HYAL1, IDH2, IDS, IDUA, IFIH1, INSR, ISCU, IVD, KCNA1, KCNJ10, KCNJ11, KCNJ2, L2HGDH, LAMA2, LAMP2, LARGE, LCT, LDB3, LDHA, LIAS, LIPA, LIPE, LIPT1, LMBRD1, LMNA, LPIN1, MAGT1, MAN1B1, MAN2B1, MANBA, MCCC1, MCCC2, MCEE, MCOLN1, MFN2, MFSD8, MGAT2, MLYCD, MMAA, MMAB, MMACHC, MMADHC, MOCOS, MOCS1, MOCS2, MOGS, MPDU1, MPI, MPV17, MT, ATP6, MT, ATP8, MT, CO1, MT, CO2, MT, CO3, MT, CYB, MT, ND1, MT, ND2, MT, ND3, MT, ND4, MT, ND4L, MT, ND5, MT, ND6, MT, RNR1, MT, RNR2, MT, TA, MT, TC, MT, TD, MT, TE, MT, TF, MT, TG, MT, TH, MT, TI, MT, TK, MT, TL1, MT, TL2, MT, TM, MT, TN, MT, TP, MT, TQ, MT, TR, MT, TS1, MT, TS2, MT, TT, MT, TV, MT, TW, MT, TY, MTHFR, MTR, MTRR, MUT, MYH3, MYOT, NAGA, NAGLU, NAGS, NBAS, NDUFAF2, NDUFS1, NEU1, NFU1, NGLY1, NHLRC1, NIPA2, NPC1, NPC2, NT5C3A, OAT, OPA1, OPA3, OTC, OXCT1, PAH, PC, PCBD1, PCCA, PCCB, PCK1, PDHA1, PDHB, PDHX, PDSS1, PDSS2, PDX1, PEPD, PEX1, PEX10, PEX11B, PEX12, PEX13, PEX14, PEX16, PEX19, PEX2, PEX26, PEX3, PEX5, PEX6, PEX7, PFKM, PGAM2, PGK1, PGM1, PHKA1, PHKA2, PHKB, PHKG2, PHYH, PLIN1, PMM2, PNP, PNPLA2, POLG, POLG2, POMGNT1, POMGNT2, POMK, POMT1, POMT2, PPARG, PPOX, PPT1, PRKAG2, PRKAG3, PRODH, PRPS1, PSAP, PTF1A, PTRF, PTS, PYGL, PYGM, QDPR, RAI1, RBCK1, REN, RFT1, RNASEH2A, RNASEH2B, RNASEH2C, RRM2B, RYR1, SAMHD1, SARS2, SCN4A, SEC23B, SERAC1,

SERPINA1, SGCA, SGCB, SGCD, SGCG, SGSH, SI, SIL1, SLC12A3, SLC16A1, SLC17A5, SLC22A5, SLC25A1, SLC25A13, SLC25A15, SLC25A20, SLC25A26, SLC25A3, SLC25A4, SLC2A1, SLC2A2, SLC30A10, SLC35A1, SLC35A2, SLC35C1, SLC37A4, SLC39A4, SLC3A1, SLC40A1, SLC46A1, SLC5A1, SLC6A19, SLC6A8, SLC6A9, SLC7A7, SLC7A9, SMPD1, SPG7, SRD5A3, SSR4, STAC3, STT3A, STT3B, SUCLA2, SUCLG1, SUGCT, SUMF1, SUOX, TALDO1, TANGO2, TAT, TAZ, TBC1D4, TCAP, TCF4, TCN2, TFR2, TIMM8A, TK2, TMEM126A, TMEM165, TMEM70, TNPO3, TPMT, TPP1, TREX1, TRIM32, TRIM37, TRPM6, TUSC3, TYMP, UCP2, UMOD, UMPS, UPB1, UROD, UROS, WFS1, XDH, ZMPSTE24, CFTR, GALT, IL2RG, JAK3, HBB, HBA1, HBA2, SMN1, SMN2, CYP21A2, ASL, SLC22A5, HADHA, ACADVL, PC, DUOX2, DUOXA2, FOXE1, GLIS3, IGSF1, IYD, KDM6A, KMT2D, NKX2, 1, PAX8, POU1F1, PROP1, SLC16A2, SLC26A4, SLC5A5, TG, THRA, THRB, TPO, TSHB, TSHR, UBR1