NAZARBAYEV UNIVERSITY

ADAPTIVE NANOPORE SEQUENCING FOR INHERITED CARDIAC CONDITIONS

AIDANA GABDULKAYUM

(B.Sc. in Biotechnology, Northumbria University)

A THESIS SUBMITTED IN PARTIAL FULFILMENT OF THE REQUIREMENT OF NAZARBAYEV UNIVERSITY FOR THE DEGREE OF MASTER OF SCIENCE IN BIOLOGICAL SCIENCES AND TECHNOLOGIES

APRIL 2024

Student:	Aidana Gabdulkayum	
Supervisor:	Dr. Dos Sarbassov	
Co-Supervisor:	Dr. Ainur Akilzhanova	

Examiners

The M.Sc. thesis of Aidana Gabdulkayum has been approved by the examiners.

Dr. Gonzalo Hap Hortelano, PhD, School of Sciences and Humanities, Nazarbayev University

Dr. Elena Zholdybayeva, PhD, Laboratory of National Scientific Shared Laboratory of Biotechnology, National Center for Biotechnology

ABSTRACT

Inherited Cardiac Conditions pose a significant global health burden, described as heart structural or functional anomalies often inherited in an autosomal dominant pattern. It is also highly contributing to the mortality rates. Early diagnosis remains challenging, as most of the individuals may have asymptomatic initial stages subsequently leading to sudden cardiac death, therefore diagnostics is of high significance, especially in younger patients. Current genetic testing methods, such as Sanger sequencing and targeted panels, offer limited insight into structural genetic variations, necessitating the exploration of more comprehensive approaches. This study aims to design and study the adaptive sampling sequencing techniques, a method for targeting specific genomic regions without DNA modifications, for future use in genetic diagnosis for individuals with inherited cardiac conditions and compare the efficiency with currently used next-generation sequencing targeted panels. Through multiple sequencing runs and subsequent analysis, our findings revealed challenges in achieving consistent target region enrichment. The short sequence length observed in our study may contribute to the failure in target enrichment, highlighting the need for longer reads to improve coverage uniformity. Variant analysis using the Epi2me labs platform and annotations revealed the absence of pathogenic variants detected by Oxford Nanopore Technologies (ONT), in comparison with Illumina. This highlights the imperative for enhancing the variant calling algorithm's sensitivity and specificity. Adaptive sampling sequencing exhibits promise for diagnosis implementation. However, additional optimization is required to improve target region enrichment, particularly through addressing the limitations posed by sequence length.

Adaptive Nanopore Sequencing for Inherited Cardiac Conditions

Aidana Gabdulkayum

Table of Contents

Chapter 1 – Introduction

1.1 Inherited Cardiac Conditions

Cardiovascular diseases (CVDs) are the leading cause of death: they account for 1 in 4 of all deaths in the world. More than 75% of deaths from cardiovascular disease occur in low- and middle-income countries (WHO, 2019). Kazakhstan has one of the highest rates of CVD incidence (Mukasheva et al., 2022). Inherited cardiac conditions (ICCs) are a group of genetic disorders that affect the structure or function of the heart (Girolami et al., 2018). Arrhythmia syndromes, cardiomyopathies, aortapathies, channelopathies and hyperlipidemias are included in the ICCs (Pua et al., 2016). ICCs typically follows an autosomal dominant inheritance pattern, indicating that an individual affected by the condition has a 50% probability of transmitting the disease to each of their offspring. Due to no or only mild symptoms present in individuals with these conditions, diagnosis of ICCs can be challenging (Walsh & Cook, 2017). However, early diagnosis is important because ICCs can lead to sudden cardiac death (SCD), especially in young people. ICCs comprise a wide range of conditions from relatively common to extremely rare, including Long QT syndrome, Familial hypercholesterolemia, and Loeys-Dietz syndrome, etc.

Long QT (LQT) syndrome is defined as an autosomal dominant disorder characterized by the heart's electrical system abnormality. Affecting approximately 1 in 2000 people. This cardiac ion repolarization is characterized by a prolonged QT interval and T-wave abnormalities on the surface of an electrocardiogram (ECG) (Wallace et al., 2019). The QT intervals indicate the time span of action potential. The first description of QT interval prolongation was introduced by Jervell and Lange-Nielsen in 1957 (Krahn et al., 2022). Multiple factors are associated with the development of LQT, such as diabetes and epilepsy. Additionally, ion channel mutations could also become the reason for LQT (Shah et al., 2019). Based on the mutations, LQTS has been classified into 17 subtypes (Wallace et al., 2019). Moreover, 30-35% of LQTS patients are affected by LQT1 subtype. This subtype is defined by the loss of function of KCNQ1 gene, which encodes the α-subunit of the KV7.1 voltage-gated potassium channel. This channel is expressed within the cell membrane of cardiomyocytes. Approximately 30% of Long QT Syndrome (LQTS) cases, known as LQT2, are attributed to mutations in the KCNH2 gene, resulting in loss of function. LQT3, affecting around 10% of LQTS patients, is associated with gain-of-function mutations in the SCN5A gene. A small percentage, about 5% to 10% of individuals with LQTS, harbour multiple mutations in these genes, typically leading to a more severe phenotype and earlier onset. Most mutations are single nucleotide substitutions or small indels. However, a minority of cases

involve large gene rearrangements, such as single or multiple exon deletions/duplications.

Familial hypercholesterolemia (FH) is one the most common autosomal dominant disorders, affecting about 1 in 220 individuals globally (McGowan et al., 2019). This condition results in premature atherosclerotic cardiovascular disease, because of increased levels of low-density lipoprotein cholesterol (LDL-C) levels in the blood that deposits in tendons and skin xanthomas. If left undiagnosed and untreated from a young age, males face a 50% chance of experiencing a fatal or nonfatal coronary event by the age of 50, while females face a 30% risk by the age of 60 (Sturm et al., 2018). 85-90% of confirmed FH cases are caused by the pathogenic variants of LDL receptor gene. The fault of the functioning of LDLR is associated with an increased level of circulating LDL-C. Numerous LDLR variations have been identified as pathogenic, encompassing a range of genetic alterations such as large-scale changes in DNA copy number variations (CNV), insertions and deletions (Indel), as well as nonsense and missense mutations, along with mutations affecting splicing. (Benito-Vicente et al., 2018). While pathogenic variants of apolipoprotein B (ApoB) and gain of function for proprotein convertase subtilisin/kexin 9 (PCSK9) gene result in 5-15% of FH cases. Mutation in the ApoB gene is responsible for reduced attachment of LDL to the LDL receptor, and pathogenic variants of the PCSK9 gene result in enhanced breakdown of LDL receptors (McGowan et al., 2019).

Loeys-Dietz syndrome (LDS) is a disorder affecting connective tissues, defined as cardiovascular, neurocognitive, and craniofacial malfunctions, first described in 2005 by Loeys and Dietz. Moreover, connective tissue in the human body provides structure and support to the neighboring cells, additionally playing a role in cell adhesion and cell-to-cell communication. Some of the proteins are believed to act as a regulatory agent in sequestering cytokines like transforming growth factor-beta (TGFβ) (Meester et al., 2017). LDS often shows clinical overlaps with other connective tissue disorders like Marfan syndrome and Ehlers-Danlos syndrome. However, it has clinically distinguishing findings, including smaller diameter rupture of aortic aneurysms at a younger age, neck and head vessels becoming the main target for arterial tortuosity, and poor wound healing. Furthermore, genetically it is described as an autosomal dominant disease. There are two major genes associated with LDS: transforming growth factor beta receptors 1 and 2 (TGFBR1 and TGFBR2), classified as LDS type 1 and type 2, respectively (Velchev et al., 2021). Type 1 leads to severe craniofacial abnormalities, while type 2 is associated with minimal craniofacial abnormalities. Additionally, mutations in SMAD3 gene have been linked with LDS type 3, and types 4 and 5 are correlated with mutations in TGFB2 and TGFB3 ligands. Nevertheless, due to the

limited observational studies, the complete clinical range of LDS remains unclear (Gouda et al., 2022).

1.1.1 Diagnosis of Inherited Cardiac Conditions

Early diagnosis is crucial for patients with cardiovascular disorders, allowing healthcare providers to timely initiate treatment and reduce the chances of further disease progression and possible sudden cardiac death. Nevertheless, due to the presence of only mild symptoms in some of the ICCs, identification of a disorder in individuals can be challenging.

The clinical diagnosis for LQT is based on family history and ECG characteristics. Identification of a prolonged QT interval, measured in lead II of 12-lead ECG from the onset of QRS complex to the end of the T wave is crucial in diagnosis. Previous family history and sudden death in infants and young adults also confirm LQTS (Shah et al., 2019). Moreover, in the presence of family history, it is strongly suggested by the experts to undergo genetic studies, as one-fourth of patients with acquired LQTS have retained a pathogenic gene variant (Krahn et al., 2022).

FH is a remarkably underdiagnosed disorder, especially in children. Early lipid screening is strongly suggested for kids at the age of 2 years if a strong family history is present, it is additionally recommended for children 9-11 years with cardiac risk factors. Nevertheless, according to the survey data early diagnosis is limited because of limited public awareness regarding FH and its inherited nature. Therefore, the mean age of FH diagnosis is 50 years, by which most of the patients already experience atherosclerotic cardiovascular disease. Genetic testing is not a common recommendation during the diagnostic procedure. Nevertheless, it may show additional information concerning cardiac risks (McGowan et al., 2019).

The diagnostic procedure for Loeys-Dietz syndrome is a combination of clinical evaluation, imaging studies and genetic testing. The identification of arterial tortuosity, craniofacial anomalies, and joint laxity is performed by the physical examination with consideration of family history data. Individuals with LDS are required to perform echocardiography systematically to monitor the status of aortic root, to minimize the risk of aortic dissection and cerebral hemorrhages (MacCarrick et al., 2014). Even if vascular involvement is not yet confirmed, it is strongly suggested to perform genetic screening when the family history is positive (Loeys and Dietz, 2008).

Genetic testing is a valuable tool in the identification of genetic mutations underlying ICCs, being advised for individuals harboring determined genetic mutation (Stafford et al.,

2022). At present, there are several conventional methods of genetic testing for hereditary cardiovascular diseases. Including massively parallel sequencing, Sanger sequencing and various targeted panels for genes of interest (Hawkins, G., 2017). One of the targeted panels specifically designed for a comprehensive identification of causal variants in ICCs is TruSight Cardio Sequencing Panel implemented on Illumina (Pua et al., 2016). However, all these methods require careful preliminary sample preparation and DNA amplification steps and are quite time-consuming, often missing the detection of rare variants due to the lower coverage or limited sequencing depth.

1.2 Sequencing History

Research into human genetic variation using DNA sequencing has experienced significant changes since its introduction over four decades ago. Enabling the sequencing of the human genome in just a few days at a cost of around 1000 USD (Petersen et al., 2017). Sequencing is the process of precise nucleotide order identification in DNA or RNA molecules. The history of DNA sequencing development began with the pioneering work of Frederick Sanger, who developed the first sequencing method in 1977 (Sanger et al., 1977). The developed technique is now known as the Sanger sequencing. Sanger sequencing revolutionized the scientific community by enabling the identification of DNA and RNA base sequences. This breakthrough revolutionized the field of biology by furnishing the means to decode entire genes initially, and subsequently, whole genomes (van Dijk et al., 2014). The main principle involves chain-terminated nucleotides named dideoxy-nucleoside triphosphates (ddNTPs), which terminates the sequence elongation and produces fragments of varying lengths. Each ddNTP is tagged with a fluorescent dye, specific to each base, which allows further base identification, and the DNA fragments are then visualized by gel electrophoresis, which enables the identification of the base (van Dijk et al., 2018). Sanger sequencing quickly became the standard due to its accuracy and simplicity for the next 30 years (Heather & Chain, 2016). Although it was limited to analyzing one sequencing reaction at a time, and had a higher cost (Hu et al., 2021).

To address the cost and throughput concerns, in 2006 next-generation sequencing (NGS) technologies emerged in the market, initiated by the National Human Genome Research Institute. These also named "massively parallel" machines remarkably expanded the speed and output of sequencing. The difference from Sanger's sequencing is that instead of dNTPs, NGS methods employ luminescent techniques to quantify pyrophosphate synthesis (McCombie et al., 2018). The introduction of NGS sparked a significant transformation in

genomics research, making it feasible for numerous small laboratories to sequence entire genomes. Moreover, gene expression analyses shifted from microarray-based methods to NGS techniques, allowing for the identification and measurement of transcripts without prior gene-specific knowledge. This approach also yielded insights into alternative splicing and sequence variations. Another early application of NGS was in genome-wide mapping of protein-DNA interactions and epigenetic marks, accomplished through chromatin immunoprecipitation followed by sequencing (ChIP-seq) (Park, 2009). Although NGS technologies are extremely powerful and offer invaluable advantages, they also have some drawbacks. One prominent limitation is their inability to sequence longer reads, Illumina for example produces reads up to 350bp. Often genomes harbor numerous repeated sequences, which are difficult to assemble and fill the gaps with the use of short reads. Moreover, NGS is highly effective in the identification of smaller variants like single-nucleotide variations (SNVs) and shorter indels, nevertheless, detecting and characterization of larger structural variations (SVs) are of greater challenge (van Dijk et al., 2018). Therefore, the demand for technologies that can sequence longer reads at a higher speed has significantly increased.

Shortly after the introduction of NGS, third-generation sequencers (TGS) emerged. In contrast, TGS technologies directly focus on single DNA molecules, facilitating real-time sequencing. Making the reads available for immediate analysis as soon as they have passed through the sequencer (Lu et al., 2016). The first "true" TGS technologies released on the market are Pacific Biosciences (PacBio) in 2011, also termed as "single-molecule real-time" (SMRT) sequencing, and more recently introduced in 2014 Oxford Nanopore Technologies (ONT) (van Dijk et al., 2018). PacBio sequencing operates by capturing sequence data during the replication process of the target DNA molecule. This is facilitated by a template known as a SMRTbell, which is a closed, single-stranded circular DNA formed by attaching hairpin adaptors to both ends of a target double-stranded DNA molecule (Rhoads & Au, 2015).

1.3 Oxford Nanopore Technology

The principle of ONT is based on nanopores embedded in a thin membrane structure, which act as a molecular sensor, detecting and analyzing single-molecule amino acids, DNA, and RNA (Lin et al., 2021). Nanopore sequencing differs from other NGS devices by the ability to directly detect the nucleotides from a long stretch of single-stranded DNA passed through the nanopore by sensing the electrical changes during the nucleotide translocation (McCombie et al., 2018). At first ONTs initial device MinION generated approximately 0.5 Gb of data per one sequencing run. Over subsequent years, ONT successfully enhanced

throughput, by modifying the chemistry for faster translocation through the pores. Consequently, today up to 50Gb of data can be produced in a single run. There are currently three ONT platforms that can be used for sequencing long reads, MinION, GridION and PromethION, differing in the number of flow cells that can be held and a number of pores available in the flow cells. The MinION is a small pocket-size-like device that holds one flow cell. Furthermore, GridION can hold up to five flow cells, with 512 channels, in total having 2048 nanopores, as one channel contains four nanopores. Lastly, PromethION can hold 48 flow cells with 12000 nanopores in 3000 channels, which enables six-fold more generated long-read data (5-100 Gb) (Hu et al., 2021). The unique sequencing method of Nanopore provides several benefits over other technologies. First, unlike many existing technologies, it does not require imaging equipment to detect the nucleotides, hence, the size of the device is comparatively small. Secondly, as nanopore directly detects the input molecule, there is no limit to the length of DNA that could be sequenced, is able to generate reads from 1000 to over 1 million base pairs (Kono & Arakawa, 2019, Miller et al., 2021). Moreover, sequencing of long fragments resolves the challenges in identifying repeat expansion variants, homologous genes, pseudogenes, and rare variants (Conlin et al., 2022).

1.3.1 Adaptive Sampling for Nanopore Enrichment

Long-read sequencing pulled a lot of attention and interest in the scientific community, due to the ability to resolve the problems associated with short-read technologies. It allows us to deeper understand the nature of genetic mutations, by the possibility of carrying out phasing of reads by haplotype. Additionally, long-reads supply DNA modification information, from signals generated during the sequence without secondary manipulation (Nakamura et al., 2024). Nevertheless, the clinical diagnosis rates are still less than required using whole genome sequencing. Therefore, targeted long-read sequencing that has been developed by ONT leads to reducing the cost and increasing the diagnosis rates. For example, Polymerase Chain Reaction (PCR) enrichment and Cas9-based enrichment on ONT, however, require time-consuming upfront sample preparations (Yamaguchi et al., 2021).

ONT has recently announced a novel Adaptive Sampling technique, which enables the performance of whole genome sequencing while having the capability to enhance specific regions of interest within the genome (Miyatake et al., 2022). Target regions are specified in the Browser Extensible Data (BED) file before sequencing allows you to determine whether the DNA being sequenced is included in the target, by sequencing the first 400-500 base pairs (bp) (Mariya et al., 2022). In case the DNA sample does not contain the target region, the

software will eject the sample from the pore by reversing the voltage (Kipp et al., 2023). Currently, interest in nanopore sequencing is growing in the field of molecular pathology, especially due to its ability for precise resolution of structural variations. Moreover, adaptive sampling could be implemented for phasing and detection of clinically relevant variants like single nucleotide variants (SNVs), structural variants (SVs), copy number variations (CNVs) and methylation differences. Thus, overcoming the challenges present in existing technologies (Miller et al., 2021). Some studies have shown the ability of adaptive nanopore sequencing to better characterize and classify structural variants in germline tandem duplication of exon 18-20 of BRCA1 (Filser et al., 2023).

1.4 Hypothesis

Considering the above-mentioned information, it is hypothesized that adaptive sample sequencing would achieve high enrichment of targeted regions and therefore more precise detection of SVs and SNVs and be comparable with other short-read sequencing techniques. The main aim of this project is to study and design an adaptive sampling sequencing method for the enrichment of 174 genes associated with ICCs using the ONT GridION sequencing machine.

1.5 Aims

Aim 1: Design a BED file. Sequence the genetic material obtained from patients with cardiovascular disorders and perform a molecular genetic assessment of target regions.

Aim 2: Carry out a comparative analysis of the depth of coverage and genetic variants of Adaptive Sampling with the Illumina TruSight Cardio panel.

Chapter 2 – Materials and Methods

2.1 Rationale

The selection of methods was based according to the aims of the study mentioned above. The whole blood samples were collected from study participants and extracted DNA. The genomic DNA (gDNA) was further used for the adaptive sample ONT sequencing together with the generated BED file as mentioned in aim 1. To accomplish the second aim the comparative analysis of the depth sequencing coverage and genetic variants between NGS and TGS was carried out using bioinformatic tools (Figure 1).

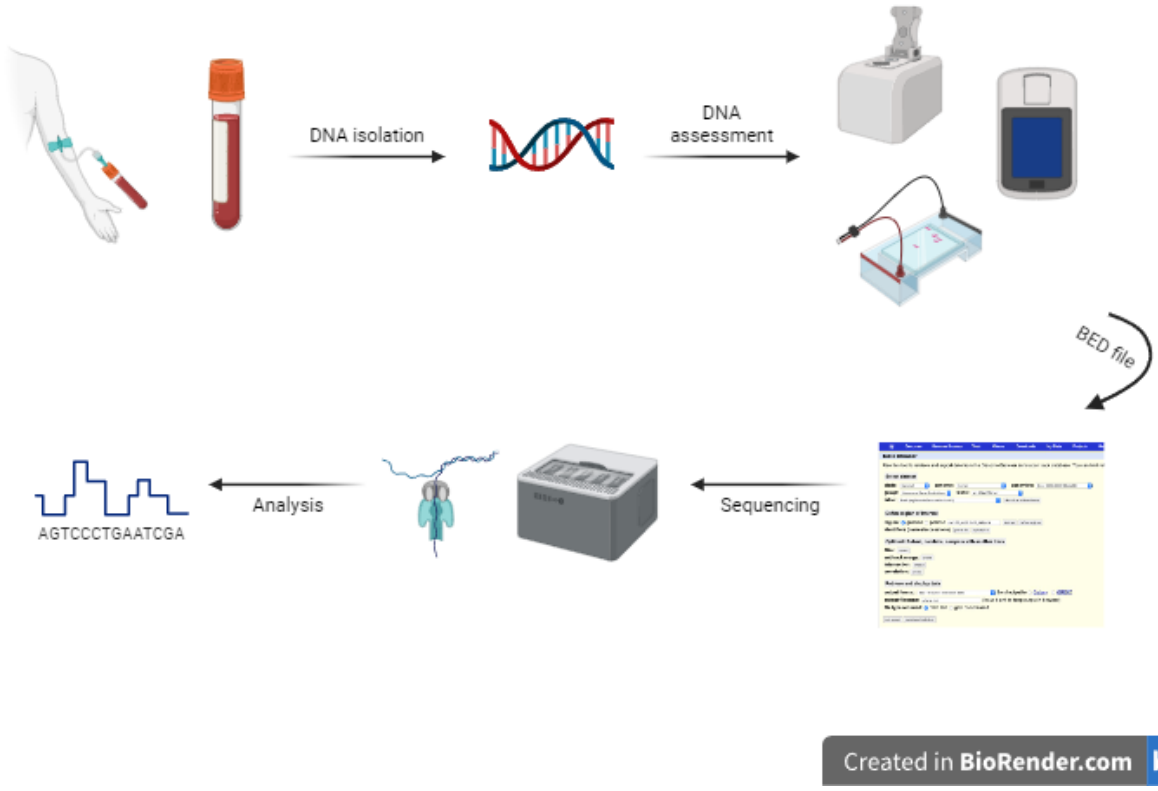


Figure 1. The overview of the experiment steps

2.2 Recruitment of study participants

The recruitment of study participants has been carried out based on the National Research Cardiac Surgery Center, Astana. Inclusion criteria for participants: patients with a confirmed or suspected diagnosis of hereditary cardiac syndromes (FH, LDS, LQTS etc.) and collection of informed consent for the study participation. Additionally, age, gender, occupation, lifestyle, sports activity, and information from family members were collected. The blood samples were collected into two tubes containing K2EDTA (blood cells and blood

plasma) and one coagulation activator gel from the ulnar vein for patients with confirmed or suspected cardiac conditions requiring genetic testing. Plasma and blood serum were aliquoted by 500µl into 1.5 ml tubes and frozen in the laboratory biorepository.

2.3 DNA isolation

DNA samples were isolated from 250µl of whole blood samples, by using commercially available DNA isolation kits like Illustra Blood genomic prep spin kit (Cytiva, USA) or Qiagen QIAmp DNA mini and Blood mini (Qiagen, Germany) according to the manufacturer's recommendation (Figure 2). The above-mentioned DNA isolation kits are based on the column spin isolation methods. The choice of the current extraction method is justified by fast protocol and good quality of final material, suitable for further experiments. Additionally, column-based extraction allows us to skip the step of shearing gDNA for later library preparation.

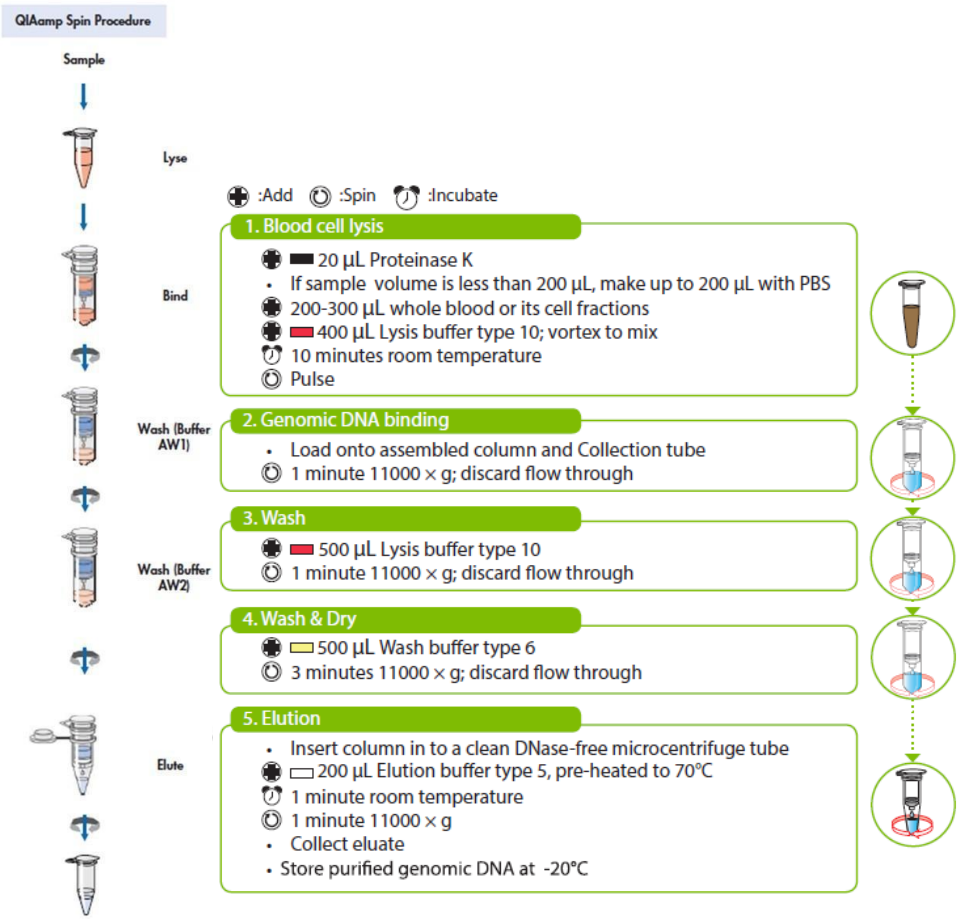


Figure 2. QIAamp and Illustra DNA isolation protocols (Qiagen, Germany; Cytiva, USA).

2.4 DNA assessment

The assessment of quality and quantity of extracted gDNA was carried out by the spectrophotometric method using the NanoDrop-2000 and the fluorometric method using the Qubit 2.0. DNA that did not meet the quality and quantity criteria was re-isolated until the required quality and quantity parameters are achieved. The 260nm/280nm and 260nm/230nm absorbance ratios are commonly used in estimating the purity of nucleic acids in NanoDrop spectrophotometer (García-Alegría et al., 2020). Furthermore, nucleotides normally absorb at 260nm, therefore, to assess the purity of DNA 260nm/280nm is used, and the accepted range is 1.8~2.0. The ratio below 1.6 is contaminated by proteins. The 260nm/230nm ratio is often used as a secondary measure of DNA purity, indicating the contaminations with phenol and carbohydrates, which are absorbed at 230nm. The expected value for the 260/230 ratio is 2.0~2.2 (Thermo Fisher Scientific, 2024). An example of a NanoDrop slope meeting the quality threshold is shown in Figure 3. The qualitative analysis was performed using gel electrophoresis on agarose gel. Electrophoresis on agarose gel is a commonly used technique to identify, separate, and analyze DNA samples. Moreover, agarose gels have a greater range of separation, best resolving 50-20000bp DNA fragments (Green & Sambrook, 2019). The locations of fragments were identified by the addition of ethidium bromide fluorescent dye, then detected by ultraviolet light on a gel documentation system (BioRad). Later storage of DNA is carried out in the Biobank of the Laboratory for further need.

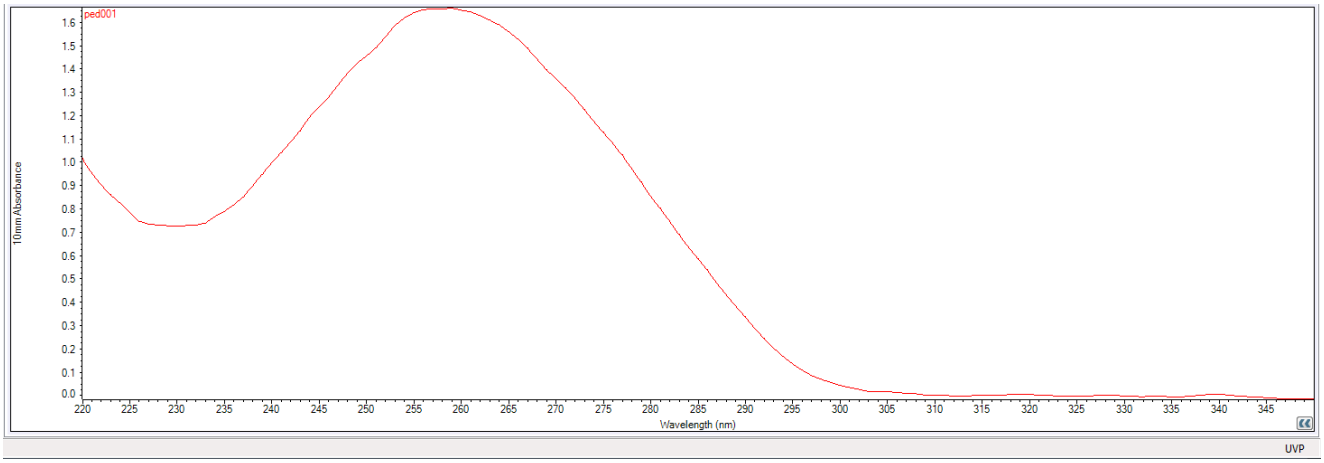


Figure 3. An example of a NanoDrop 2000 slope of samples gDNA assessment.

2.5 The specification of target regions

The specification of target regions for adaptive sampling enrichment includes the selection of genes of interest and the identification of chromosome locations. According to the ONT recommendations, the most suitable format for identification of target regions is BED file. The BED file is a flexible format for the display of data lines. The line fields include the name of a chromosome, the starting and ending position in a chromosome, and the name of the gene. Furthermore, genes of interest were identical to the TruSight Cardio panel used for targeted sequencing on Illumina. It includes 174 genes associated with ICCs, including core and emerging genes for 17 ICCs (Table 1). According to the recommendations, the list of genes was converted to gene ID through the Ensembl genome database. The BED file was retrieved by using the Table Browser in the UCSC genome browser. The selected dataset included: Mammal (clade), Human (genome), GRCh38 (assembly), and NCBIRefseq (track) shown in Figure 4. Additionally, to increase the chances of acceptance by the pore of our target regions, flanking regions "Buffer" were added at both sides of every single target indicated in the BED file. The size of the Buffer region has constituted 10% of the target region, following the adaptive sampling best practice recommendations (Oxford Nanopore Technologies, 2023). The genomic regions were manipulated using the `flank` function from the BEDTools suite (Quinlan and Hall, 2010). Specifically, we employed BEDTools version 2.31.0 to extend each region by 10% upstream and downstream.

Table 1. Cardiac conditions and number of genes included into the experiment design (Illumina).

Cardiac Conditions	**№ of Genes Covered**
Aortic Valve Disease	3
Marfan Syndrome	3
Loeys – Dietz Syndrome	4
Short QT Syndrome	4
Catecholaminergic Polymorphic Ventricular Tachycardia	6
Familial Hypercholesterolemia	7
Restrictive Cardiomyopathy	9
Non – Compaction Cardiomyopathy	10
Noonan Syndrome	11
Arrhythmogenic Right Ventricular Cardiomyopathy	11
Brugada Syndrome	13
Structural Heart Disease	15
Long QT Syndrome	15
Familial Aortic Aneurysm	16
Familial Atrial Fibrillation	21
Hypertrophic Cardiomyopathy	47
Dilated Cardiomyopathy	59

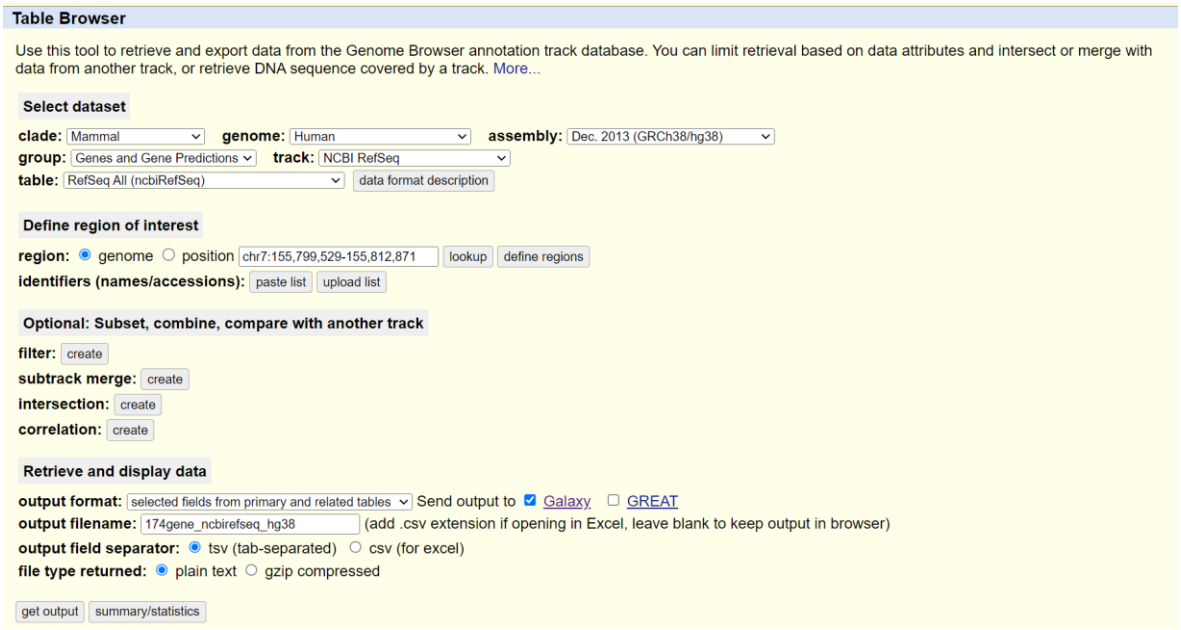


Figure 4. Example of UCSC genome browser.

2.6 Sequencing

The sequencing of adaptive sampling was carried out using the Oxford Nanopore GridION device and specific GridION R9.4.1 flongle flow cells. The sequencing libraries were prepared using the ligation sequencing kit (SQK-LSK 109) according to the manufacturer's recommendations. The library preparation protocol consists of 3 main steps: DNA repair and end-prep; Adapter ligation and clean-up; Priming and loading. The recommended library input for the adaptive sampling method is 50fmol in each loading, therefore, 3-4ug of gDNA was adjusted to the volume of 48ul in nuclease-free water, skipping the fragmentation step due to the choice of DNA extraction method. Furthermore, the first step of library preparation involved the repairment of DNA ends with the use of NEBnext enzymes to generate blunt ends and the addition of polyA tails at the 3' end. Next adapters with the motor enzyme at the end are ligated to the dA–tailed fragments, enabling the attachment of the library to the nanopore. Following the ligation step, all unligated adapters are removed through the purification step. By aiming the longer fragment sequencing size selection step involved the usage of a long fragment buffer and AMPure XP bead-based purification. Finally, the library is loaded onto the flongle flow cell, with sequencing buffer and loading beads, aiming to distribute the library evenly across the nanopore surface. The flow cells are checked for the number of available pores before the start of sequencing by running the "Flow cell check" command on the sequencer machine, using MinKnow software.

Minknow is a software introduced by ONT, which provides all the interactions with the sequencer. The flow cells must exceed the number of 1000 available pores, which is considered to be the “low pore” condition of a flow cell. The key features taken into account after the sequencing run are total data produced in Gb; average N50 (sequence length); and number of reads generated.

2.7 Bioinformatic analysis

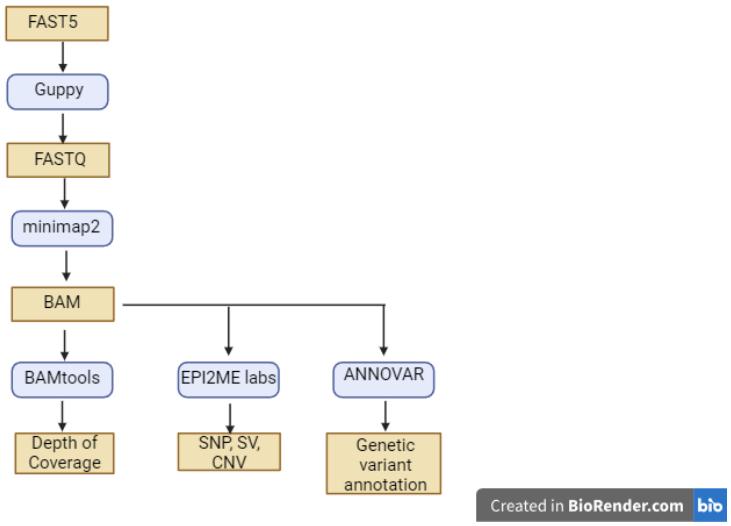


Figure 5. A systematic overview of bioinformatic flow.

The sequencing data was analyzed by bioinformatic methods. Software packages of the Laboratory of Bioinformatics and Systems Biology, NLA for bioinformatic analysis of sequencing data were used. Basecalling of FAST5 files was carried out using Guppy basecaller. To align the reads to GRCh 38 human reference genome Minimap2 was used, and further converted into BAM files using SAMtools. The comparative analysis of the depth of coverage between Illumina and ONT Adaptive Sampling was performed using the `coverage` function from the BEDTools suite (Quinlan & Hall, 2010). BEDTools version 2.31.0 across genomic regions of interest. Calling of SNVs, SVs and CNVs and STR genotyping was carried out using the Epi2me labs bioinformatic platform designed by Oxford Nanopore Technologies. The analysis pipeline for human variation detection was implemented using the "wf-human-variation" workflow available on GitHub (https://github.com/epi2me-labs/wf-human-variation), developed by Epi2me Labs. The version used in this study was 2.0.0, and the pipeline configuration and parameters are described in the repository documentation (Figure 5). Moreover, called variants through workflow were filtered based

on ClinVar annotations, and were further annotated using ANNOVAR (ver.2019Oct24) to obtain additional information about their genomic features and functional consequences (Wang etc., 2010).

2.8 Permits and Approval

All human specimens were collected legally following the ethical and biosafety procedure at Nazarbayev University and NLA, which meets or exceeds the US regulatory standards for the human care and treatment of humans in research. The research protocol, informed consent forms and other related documents have been submitted to the Local Ethical Commission of the Center for Life Sciences, National Laboratory Astana, Nazarbayev University (Protocol № 02-2022, dated 01.04.2022) in the framework of "Comparative evaluation of genetic variants in sudden cardiac death victims and cardiac patients: genetic markers of risk and diagnosis" project.

Chapter 3 – Results

3.1 Information regarding the patients enrolled in the study.

Throughout the study, 9 participants were recruited by the National Cardiac Surgery Centre. The age of recruited individuals was between 2-47 years old. Among the recruited participants, 7 patients have been diagnosed with Long QT syndrome, and 2 other participants have been diagnosed with Familial hypercholesterolemia and Loeys-Dietz syndrome. Diagnosed disorders are included in the ICCs term. The summary of recruited participants is presented in Table 2.

Table 2. Summary of recruited participants

ID name	Disease	Date of Birth	Sex
FH001	FH	09.09.2009	Female
LDS001	LDS	23.04.2019	Female
PED001	LQTS	05.02.1999	Male
PED007	LQTS	09.04.2006	Female
PED011	LQTS	10.02.1975	Female
PED014	LQTS	01.01.1999	Female
PED015	LQTS	24.08.2021	Male
PED016	LQTS	24.05.1989	Female
PED017	LQTS	30.06.1991	Male
FH: Familial Hypercholesterolemia; LDS: Loeys-Dietz syndrome; LQTS: Long QT syndrome			

3.2 gDNA fluorometric and spectrophotometric assessment

gDNA was isolated from whole blood samples using commercial kits with the column method for DNA isolation. The assessment of gDNA was performed using Nanodrop 2000, Qubit2.0, and gel electrophoresis. Spectrophotometric and fluorometric assessment of gDNA is summarized in Table 3. All the samples have been identified to be of suitable purity and concentration for further DNA library preparation steps, according to the recommendations of ONT. The results of DNA integrity testing by electrophoresis are shown in Figure 6. gDNA samples showed no contamination by proteins and RNA, the range of the used ladder is 100 bp to 15000 bp, indicating that the size of isolated DNA samples is bigger than 15000 bp.

Table 3. gDNA assessment results

ID	Concentration (ng/ul)	260/280	260/230	Qubit dsDNA BR assay kit, (ng/ul)
PED001	82.5	1.94	2.28	71.4
PED011	105.8	1.88	2.22	83.8
PED014	73	1.96	2.31	59
PED015	120.9	1.86	2.34	117
PED016	159.4	1.86	2.34	139
PED017	134.5	1.87	2.37	120
PED007	115.5	1.83	2.15	75.8
LDS001	133.5	1.86	2.34	127
FH001	136.5	1.87	2.34	132

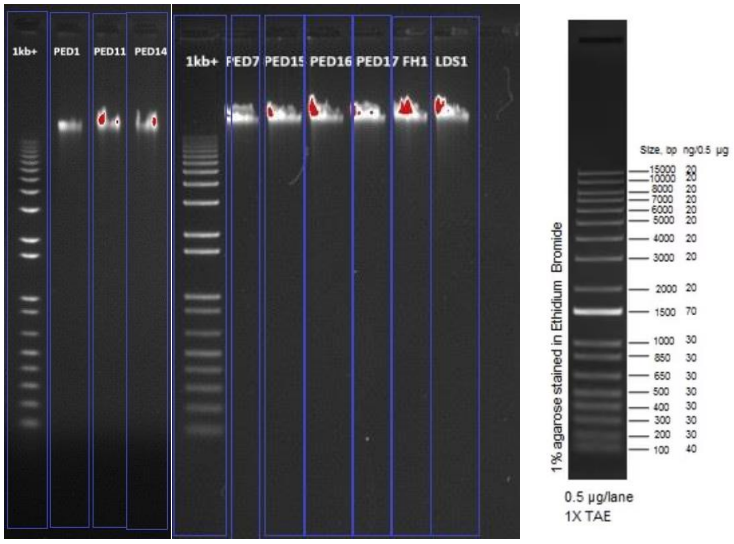


Figure 6. Gel electrophoresis assessment, represents the DNA assessment using gel electrophoresis (1% agarose gel in ethidium bromide), 1kb+ ladder (Invitrogen, USA) has been used to measure the length of gDNA.

3.3 Sequencing library preparation summary

4 ug of gDNA was taken as an input for library preparation. The final concentration of the library was in the range of 5 -55 fmol, as recommended by ONT. Additionally, the number of available pores in the flow cells ranged between 1047 and 1615. Libraries were loaded 2-3 times each 24 hours. The overview of DNA libraries is summarised in Table 4.

Table 4. Library for sequencing, final library concentrations and flow cell IDs

ID	DNA used for library (ug)	Final library concentration (fmol)	Number of library loads	Flow Cell ID	Number of pores
FH001	2ug	31	3	FAS49142	1394
LDS001	4ug	81	3	FAV65425	1301
PED001	4ug	57	3	FAS92410	1344
PED007	4ug	68	2	FAV69863	1261
PED011	4ug	82	2	FAS96946	1615
PED014	4ug	48	2	FAS88107	1149
PED015	4ug	55	3	FAS87728	1321
PED016	4ug	51	3	FAS92290	1375
PED017	4ug	50	3	FAS88675	1047

3.4 Adaptive sampling sequence summary

Figure 2 represents the final report of sequencing outputs. The 72-hour sequence resulted in 90.76 Gb of produced data, with 592 base pairs estimated N50 number for FH001 sample. LDS001, PED001, PED007, PED011, PED014, PED015, PED016 and PED017 resulted in 36.3 Gb, 64 Gb, 118.76 Gb, 57.14 Gb, 61.88 GB, 49.97 GB, 49.05 Gb and 39.84 Gb respectively. A comparison of the average sequence length (N50) for samples sequenced on GridION with adaptive sampling is shown in Figure 7. N50 is on average 824bp ranging between 592-1000bp.

Table 5. Total data produced by Adaptive sampling sequence

ID	Reads generated (M)	Total data produced (Gb)
FH001	5.81	90.76
LDS001	1.88	36.3
PED001	3.55	64
PED007	6.17	118.76
PED011	2.92	57.14
PED014	3.35	61.88
PED015	2.54	49.97
PED016	2.42	49.05
PED017	1.77	39.84

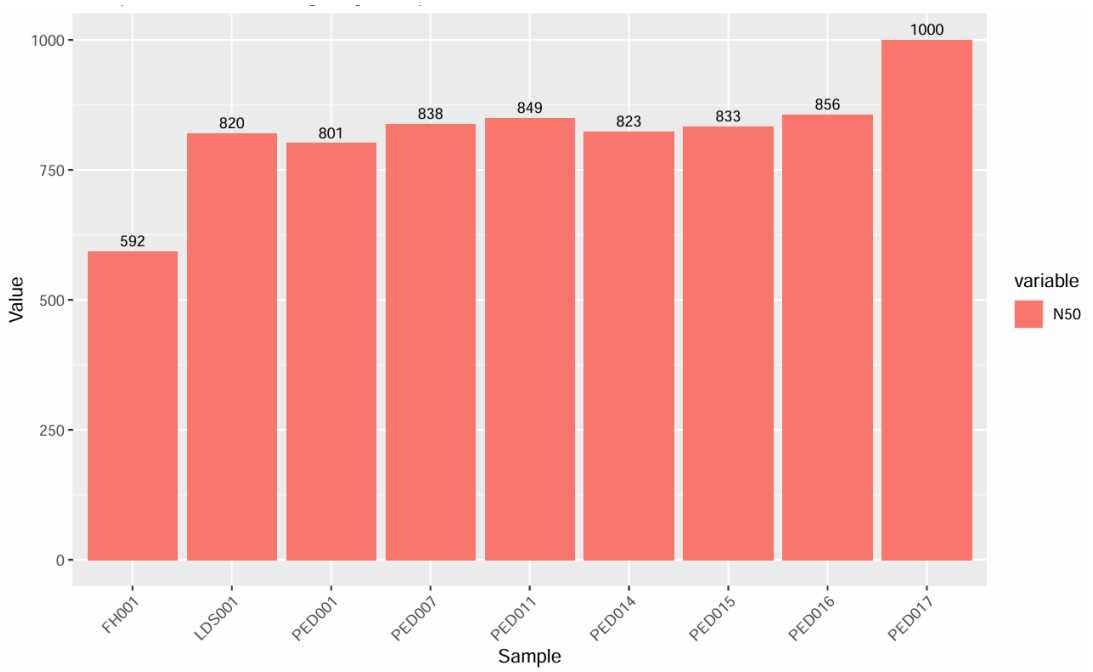


Figure 7. Comparison of sequence length (N50) between samples sequenced on ONT with adaptive sampling

3.5 Bioinformatic analysis

3.5.1 Depth of coverage comparison using BAM tools

174 predisposition genes were obtained and listed in the BED file with 10% flanking regions on both ends. The average depth of coverage of on-target regions indicated in a BED file and all other off-target regions were 4441 (range: 2372-6336) and 5447 (range: 3043-10443), respectively (Figure 8). Two samples had the lowest average depth of coverage.

In comparison with Illumina, adaptive sample sequencing has shown a lower average depth of coverage, with a mean of 5868 (range: 2442-8795), while Illumina sequencing generated a higher average depth of coverage, with a mean of 19743 (range: 4308-28060). Among the three samples analysed on Illumina, PED007 exhibited the lowest coverage, while on ONT adaptive sampling, LDS001 samples resulted in the lowest coverage (Figure 9).

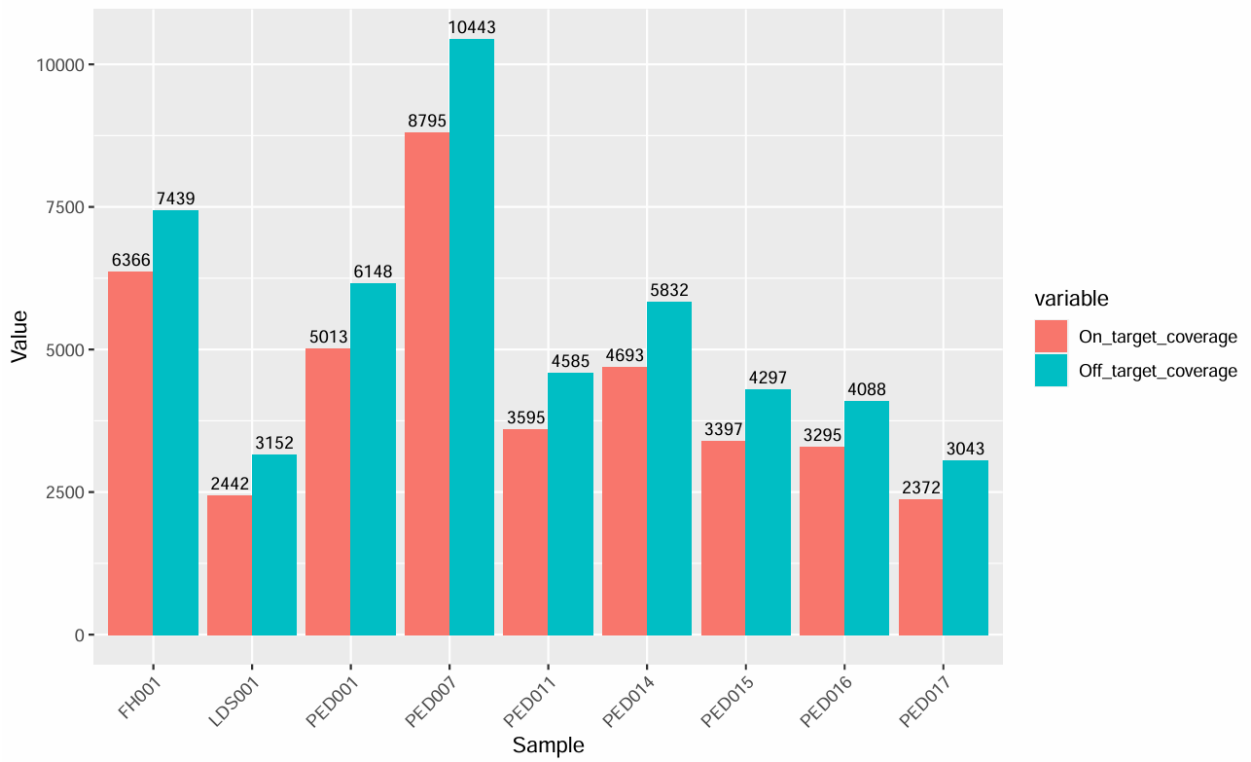


Figure 8. Comparison of average depth of coverage of on-target and off-target regions

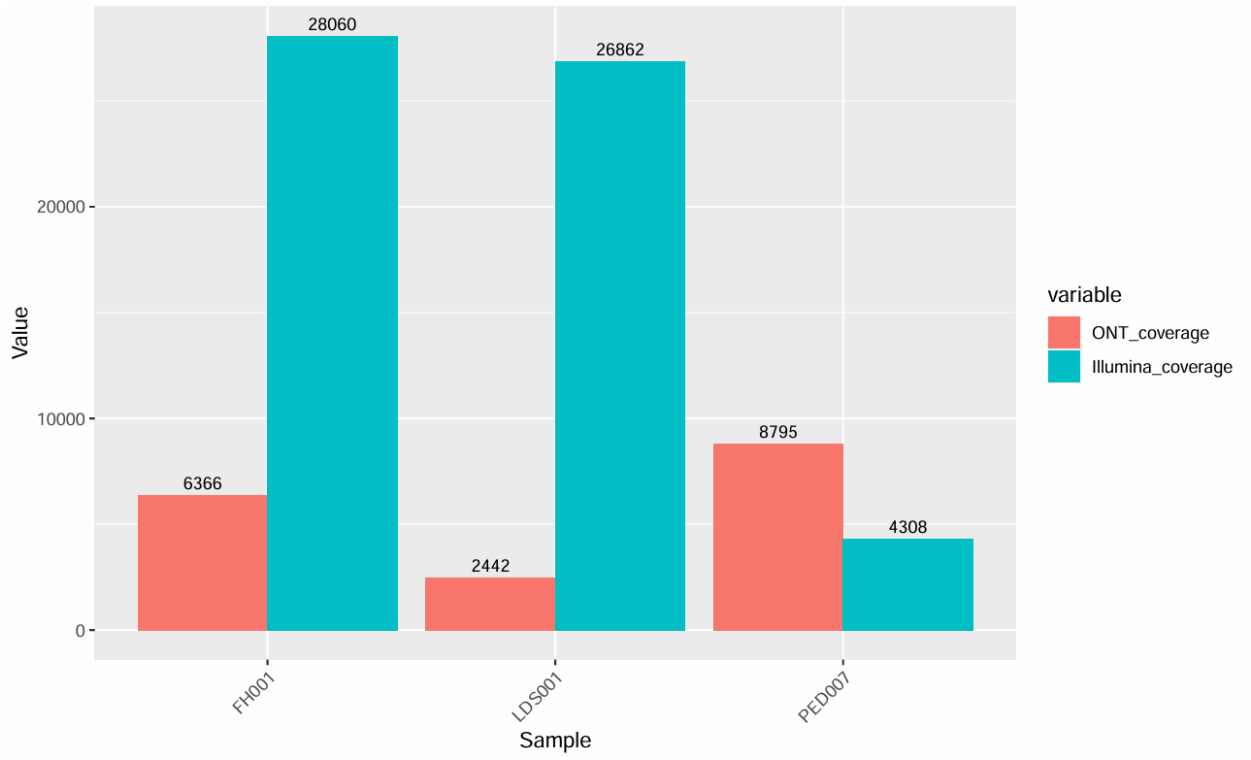


Figure 9. Comparison of average depth of coverage between Illumina and ONT adaptive sampling

3.5.2 Extraction of SNVs/Indels

The analysis of genetic variants using the Epi2me labs human variation workflow revealed varying numbers of variants across the samples examined. Sample PED007 exhibited the highest number of genetic variants, with a total of 8202 identified variants, followed by sample PED014 (3892 variants) and PED016 (4006 variants). In contrast, sample FH001 had the lowest number of identified variants, totalling 1321 (Figure 10). Notably, the number of variants identified varied considerably among the samples, suggesting differences in genetic diversity or sequencing depth. These findings underscore the importance of considering sample-specific characteristics and sequencing methodologies when interpreting genetic variant data.

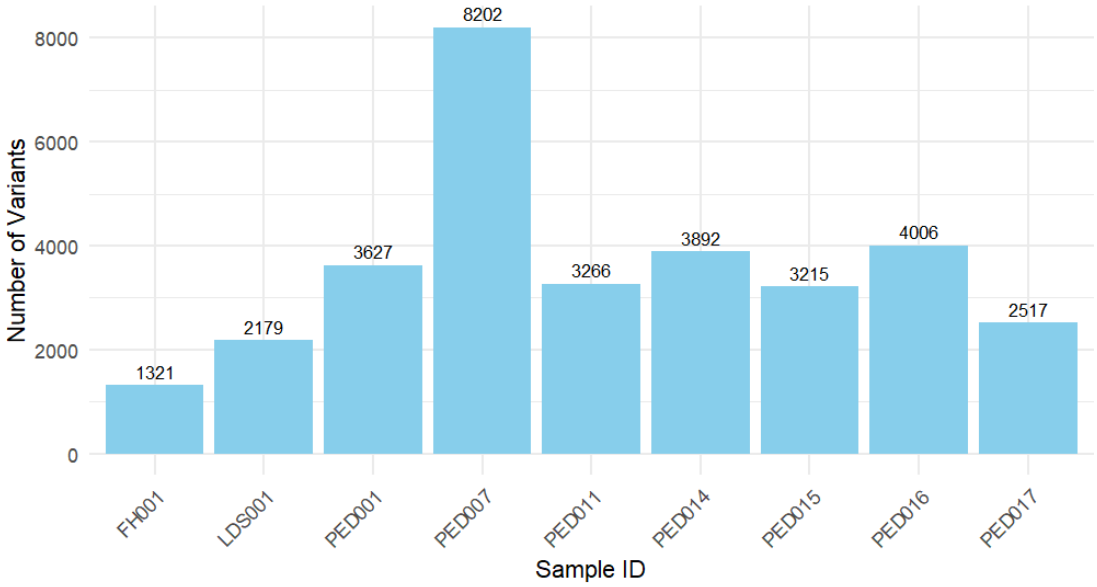


Figure 10. The number of variants is called through human variation workflow.

In Table 6, the variants identified in each sample have been annotated with ClinVar using the SNPEff annotation tool included into human variation workflow. The table summarizes the number of effects by type and region for each sample. It indicates that the majority of variants are categorized as Intergenic and Intron variant types, while Missense variants are the least identified type. In Figure 11, the distribution of variant types in FH001, LDS001 and PED007 sample is represented as a pie chart. Distribution of variant types for other samples are listed in Figure S3.

The identified genetic variants were additionally annotated using ANNOVAR annotation tool and revealed no pathogenic or likely pathogenic variants in the data produced

by adaptive sampling. While Illumina produced data has identified few pathogenic and likely pathogenic variants through annotation tool (Table S2- 12).

Table 6. Type and affecting region of variants indicated by annotation and their amount in each sample.

ID	Intergenic variant	Intron variant	Missense variant	Non-coding transcript exon variant	Synonymous variant	Upstream gene variant	Downstream gene variant
FH001	3	7161	19	71	52	124	525
LDS001	2090	220	0	2	2	3	15
PED001	3106	548	1	2	4	43	59
PED007	7223	2450	13	76	22	652	644
PED011	2964	689	5	12	11	179	136
PED014	3514	874	9	17	15	225	152
PED015	2902	750	9	38	15	254	160
PED016	3117	1946	6	18	24	157	180
PED017	3106	548	1	2	4	43	59

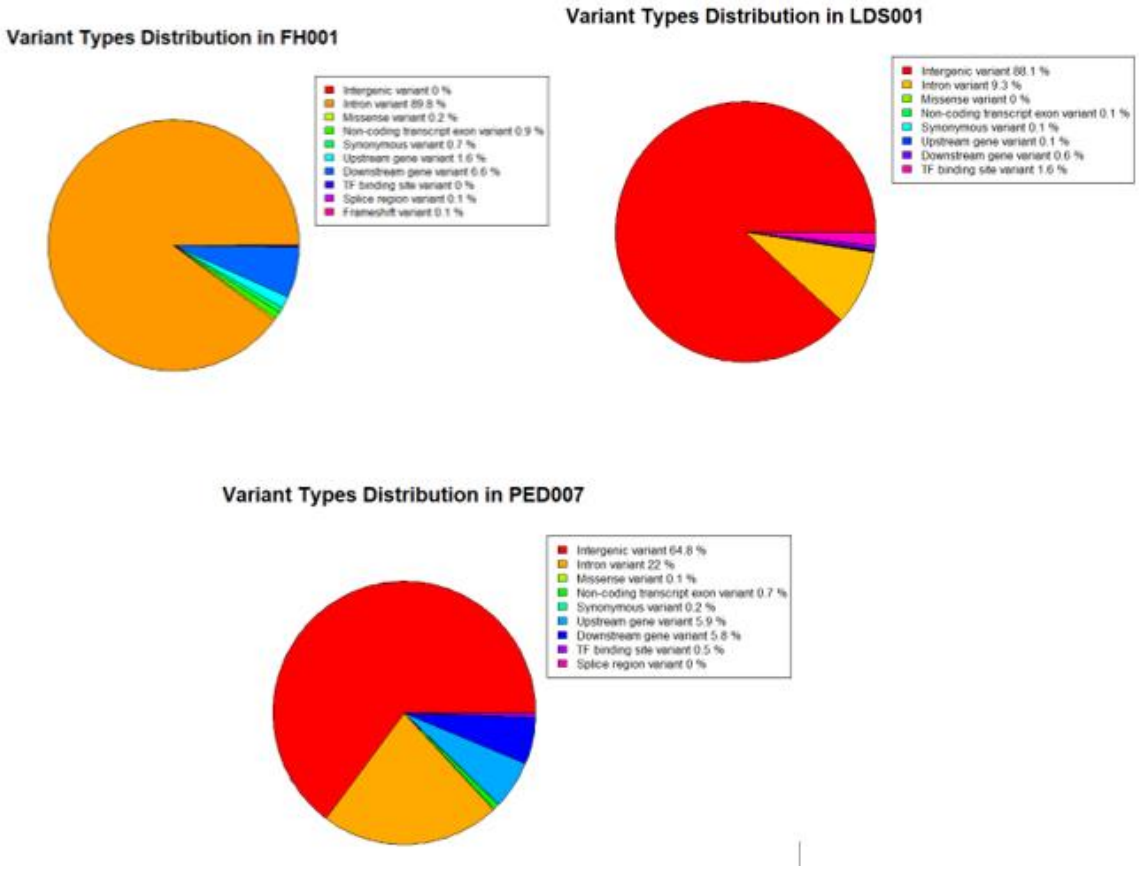


Figure 11. Distribution of the variant types among some samples annotated by SNPEff.

Chapter 4 – Discussion

The present study investigated the potential of ONT Adaptive Sampling as a novel approach for efficiently enriching regions of interest in whole-genome sequencing, particularly in comparison to the established targeted panels on Illumina. Additionally delving into the genetic landscape of ICCs. Understanding sequencing methods is crucial for future clinical implementations, and for improving the current diagnostics pace, when existing technologies fall short in providing precise genetic diagnostics. The primary focus of this study was to elucidate the workflow principles, develop an optimized protocol for target enrichment, and compare the performance of ONT adaptive sampling with Illumina TruSight Cardio Panel for enriching target regions associated with ICCs. The selection of enrolled patients was based on the diagnosis of the disorder included in the term ICCs. Moreover, 9 enrolled patients were predominantly diagnosed with LQTS and 2 patients with FH and LDS. All the patients have registered a previous family history of having CVDs, and only one patient PED001 has no prevalent history of CVDs in the family. That could have indicated the autosomal dominance inheritance pattern of the disorder.

Central to the success of targeted sequencing is the accurate definition of target regions, as they are computationally defined for sequencing (Weilguny et al., 2023). To achieve an on-target region enrichment, firstly a BED file has been designed to indicate the target regions. According to the ONT guideline recommendations target region should range from 0.1 to 10% of the whole genome. In this case, for human genome should be between 3-300 Mb (Miller et al., 2021). The careful selection and delineation of target regions ensure comprehensive coverage and minimize off-target effects, optimizing the sensitivity and specificity of variant detection. Additionally, the target regions have been increased by 10% of the gene size, in a BED file with sequences that do not initially map to our on-target regions, named flanking regions. Flanking regions increase the number of accepted reads hitting the target region (Oxford Nanopore Technologies, 2020). By using the 174 genes mostly associated with ICCs and the UCSC genome table browser, a BED file covering 0.5% of the whole human genome has been created.

To improve the efficiency of adaptive sampling enrichment, it is suggested by ONT to take 4µg of gDNA as a starting material for library preparation to load 50 fmol and more into the flow cell (Table 4). This helps in decreasing the open pore state and minimizing wasted sequencing capacity, as most of the strands that enter the pores are not from the target regions. All the samples have been sequenced for 72 hours with 2-3 loadings of DNA library. The adaptive sampling sequence resulted in a different amount of total data produced from each

sample, with the lowest being 36.3 Gb for LDS001 and the highest 118.76 Gb for PED007, even though samples have been adjusted under the same conditions (Table 5). This may depend on different factors, like the active number of available pores in a flow cell and the read length of each sample (Yamaguchi et al., 2021). The average read length is approximately 824bp between all samples, falling below the typical range expected for Oxford Nanopore sequencing. (Figure 7) (Loose et al., 2016). The DNA enrichment or depletion decision time for adaptive sampling on the GridION sequencing machine is 450bp. Considering that it raises concerns regarding the ability to enrich the target regions with the obtained N50 (Payne et al., 2020). The Minknow software is configured to enable simulations for the development of tools and analysis methods. One of the reasons for the shorter sequence length generated is that MinKnow simulations have limitations, and rejected reads are not fully removed from the pore, instead, the original read is fragmented at the point where it could have been unblocked (Munro et al., 2024). Therefore, it is plausible that the adaptive sampling algorithm may not effectively capture and enrich the regions of interest within the specified parameters.

Moreover, the comparison of coverage for on-target and off-target regions reveals a noteworthy observation: the average depth of coverage for off-target regions is higher than that for on-target regions (Figure 8). This discrepancy may be attributed to the relatively short N50 length, which limits the effective capture and enrichment of target regions while allowing for greater coverage of non-specific, off-target sequences. Additionally, the larger size of off-target regions compared to the 174 genes in the on-target BED file further exacerbates this disparity, potentially diluting the sequencing effort and resulting in suboptimal enrichment of the intended targets.

The comparison of target region enrichment between nanopore sequencing and the Illumina TruSight Cardio Panel provides valuable insights into the performance and utility of each platform for genetic analysis of ICCs (Figure 9). By evaluating the depth of coverage, uniformity, and specificity of target region enrichment, this study sheds light on the potential of nanopore sequencing as a complementary or alternative approach to Illumina sequencing in clinical and research settings. Above mentioned issue with shorter N50 generated may contribute to a lower average depth of coverage compared to Illumina sequencing, which utilizes even shorter reads and employs amplification steps to enhance coverage and enrichment.

The utilization of the Epi2me labs platform, particularly the human variation workflow, represents a valuable tool for identifying structural variants (SVs) in the sequenced samples (Figure 10) (Parker, 2023). According to the workflow, most of the identified variants in all

samples fall under the intergenic and intronic variant types. Intronic and intergenic variants are non-coding and may not alter protein sequences (Table 6 and Figure 11). Therefore, the identification of a lower number of synonymous and frameshift variants might explain the annotation results by ANNOVAR, identifying mostly benign and likely benign variants (Table S5- 12). The absence of pathogenic or likely pathogenic variants identified through this platform, especially in comparison to Illumina sequencing results, raises important considerations regarding the sensitivity and accuracy of variant detection methodologies.

The discrepancy in variant detection between nanopore sequencing using Epi2me labs and Illumina sequencing underscores the inherent differences in sequencing technologies and analysis pipelines. One possible explanation for the discrepancy could be the inherent error profiles of nanopore sequencing, which may lead to challenges in accurately identifying and characterizing variants, particularly those of low frequency or in regions of high complexity (Jain et al., 2016). Additionally, differences in sequencing coverage and depth between the two platforms may contribute to variations in variant calling sensitivity and specificity.

The lack of pathogenic or likely pathogenic variants identified through nanopore sequencing highlights the importance of corroborating sequencing results with orthogonal validation methods, such as Sanger sequencing or targeted PCR assays. Furthermore, it underscores the need for careful interpretation of sequencing data, considering the strengths and limitations of the sequencing platform and analysis pipeline employed.

While nanopore sequencing offers advantages such as real-time data analysis and long-read capabilities, the observed limitations underscore the importance of careful protocol optimization and consideration of sequencing parameters for targeted applications. Moving forward, efforts to improve variant detection sensitivity and accuracy in nanopore sequencing could involve optimization of DNA isolation and shearing to achieve longer fragments, refining bioinformatics algorithms, optimizing sequencing protocols, and integrating complementary sequencing technologies or validation approaches. By addressing these challenges, nanopore sequencing has the potential to become a powerful tool for comprehensive genomic analysis, clinical diagnostics, and personalized medicine in the context of ICCs and other genetic disorders. Due to the ability of sequencing not only targeted regions like Illumina panels, but it also allows to expand the diagnostic field by looking at off-target regions, therefore, increasing the chances of correct diagnosis and treatment plan creation.

In conclusion, while the adaptive sampling sequence holds promise for targeted sequencing of ICCs, the observed variability in total data produced and coverage metrics highlights the need for further optimization and refinement of the sequencing protocol. By

addressing these challenges and leveraging the strengths of nanopore sequencing technology, future studies can contribute to advancing our understanding of the genetic basis of cardiac diseases and other disorders and pave the way for personalized approaches to diagnosis and treatment.

LIST OF TABLES

LIST OF FIGURES AND ILLUSTRATIONS

ABBREVIATIONS

ApoB	Apolipoprotein B
BED	Browser Extensible Data
bp	base pairs
CNV	copy number variation
CVDs	Cardiovascular diseases
ddNTPs	dideoxy-nucleoside triphosphates
DNA	Deoxyribonucleic acid
ECG	Electrocardiogram
FH	Familial Hypercholesterolemia
fmol	Femtomole
gDNA	Genomic DNA
ICCs	Inherited Cardiac Conditions
Indel	Insertions and Deletions
K2EDTA	Dipotassium ethylenediaminetetraacetic acid
kbp	kilo base pairs
LDL-C	Low density lipoprotein cholesterol
LDS	Loeys-Dietz syndrome
LQTS	Long QT Syndrome
N50	Median value for sequence length
NGS	Next-Generation Sequencing
NLA	National Laboratory Astana
nm	nanometer
ONT	Oxford Nanopore Technologies
PCSK9	Proprotein convertase subtilisin/kexin 9
RNA	Ribonucleic acid
SMRT	Single-molecule real-time
SNVs	Single nucleotide variants
SVs	Structural variants
TGFβ	Transforming growth factor
TGS	Third-Generation Sequencing
UCSC	University of California, Santa Cruz
USD	United States Dollar

Acknowledgments

I would like to express my deepest gratitude to my supervisor, Dr. Dos Sarbassov, and my co-supervisor, Dr. Ainur Akilzhanova, for their invaluable guidance, support, and encouragement throughout the course of my master's thesis. Their knowledge, guidance, and support have played a crucial role in molding my path of research.

I am also indebted to my laboratory colleagues at the National Laboratory of Astana for their collaboration, camaraderie, and insightful discussions, which have enriched my academic experience and contributed to the success of my research endeavours.

I extend my heartfelt thanks to all the professors from the School of Sciences and Humanities, biology department whose lectures, seminars, and guidance have broadened my knowledge and inspired my scientific curiosity. Their dedication to teaching and research has been a constant source of inspiration.

Special gratitude goes to my groupmates for their friendship and unwavering support throughout this academic journey. Their encouragement and companionship have made the challenges more manageable and the successes more meaningful.

Last but not least, I would like to thank my family for their unconditional love, support, and understanding. Their unwavering belief in me has been my greatest motivation and source of strength throughout this journey.

Declaration

I hereby declare that the thesis is my original work, and it has been written by me in its entirety. I have duly acknowledged all the sources of information which have been used in the thesis. This thesis has also not been submitted for any degree in any university previously.

Aidana Gabdulkayum

12 April 2024

Chapter 5 – References

Adaptive sampling. Oxford Nanopore Technologies. (2023, July 5). https://community.nanoporetech.com/docs/plan/best_practice/adaptive-sampling/v/ads_s1016_v1_revj_12nov2020

Benito-Vicente, A., Uribe, K., Jebari, S., Galicia-Garcia, U., Ostolaza, H., Martin, C. (2018). Familial hypercholesterolemia: The most frequent cholesterol metabolism disorder caused disease. *International Journal of Molecular Sciences, 19*(11), 3426. https://doi.org/10.3390/ijms19113426

Burgess, C. (2023). How can I determine the number of reads that are being rejected due to adaptive sampling? Oxford Nanopore Technologies Help Center. https://help.nanoporetech.com/en/articles/6627843-how-i-can-determine-the-number-of-reads-that-are-being-rejected-due-to-adaptive-sampling

Conlin, L. K., Aref-Eshghi, E., McEldrew, D. A., Luo, M., Rajagopalan, R. (2022). Long-read sequencing for Molecular Diagnostics in constitutional genetic disorders. *Human Mutation,* 43(11), 1531–1544. https://doi.org/10.1002/humu.24465

Filser, M., Schwartz, M., Merchadou, K., Hamza, A., Villy, M.-C., Decees, A., Frouin, E., Girard, E., Caputo, S. M., Renault, V., Becette, V., Golmard, L., Servant, N., Stoppa-Lyonnet, D., Delattre, O., Colas, C., Masliah-Planchon, J. (2023). Adaptive Nanopore sequencing to determine pathogenicity ofbrca1exonic duplication. *Journal of Medical Genetics,* 1–4. https://doi.org/10.1136/jmg-2023-109155

García-Alegría, A. M., Anduro-Corona, I., Pérez-Martínez, C. J., Guadalupe Corella-Madueño, M. A., Rascón-Durán, M. L., Astiazaran-Garcia, H. (2020). Quantification of DNA through the NanoDrop spectrophotometer: Methodological validation using standard reference material and Sprague Dawley rat and human DNA. *International Journal of Analytical Chemistry, 2020,* 1–9. https://doi.org/10.1155/2020/8896738

Girolami, F., Frisso, G., Benelli, M., Crotti, L., Iascone, M., Mango, R., Mazzaccara, C., Pilichou, K., Arbustini, E., Tomberli, B., Limongelli, G., Basso, C., Olivotto, I. (2018). Contemporary genetic testing in inherited cardiac disease. *Journal of Cardiovascular Medicine,* 19(1), 1–11. https://doi.org/10.2459/jcm.0000000000000589

Gouda, P., Kay, R., Habib, M., Aziz, A., Aziza, E., Welsh, R. (2022). Clinical features and complications of Loeys-Dietz Syndrome: A systematic review. International Journal of Cardiology, 362, 158–167. https://doi.org/10.1016/j.ijcard.2022.05.065

Green, M. R., Sambrook, J. (2019). Analysis of DNA by Agarose Gel electrophoresis. *Cold Spring Harbor Protocols*, *2019*(1). https://doi.org/10.1101/pdb.top100388

Guelly, C., Abilova, Z., Nuralinov, O., Panzitt, K., Akhmetova, A., Rakhimova, S., Kozhamkulov, U., Kairov, U., Molkenov, A., Seisenova, A., Trajanoski, S., Abildinova (Rashbayeva), G., Kaussova, G., Windpassinger, C., Lee, J. H., Zhumadilov, Z., Bekbossynova, M., Akilzhanova, A. (2021). Patients with coronary heart disease, dilated cardiomyopathy and idiopathic ventricular tachycardia share overlapping patterns of pathogenic variation in cardiac risk genes. *PeerJ,* 9. https://doi.org/10.7717/peerj.10711

Hawkins, G. (2017). Chapter 5 - Analysis of Human Genetic Variations Using DNA Sequencing. Basic Science Methods for Clinical Researchers. *Academic Press*, 77-98. https://doi.org/10.1016/B978-0-12-803077-6.00005-9.

Hu, T., Chitnis, N., Monos, D., & Dinh, A. (2021). Next-generation sequencing technologies: An overview. *Human Immunology*, *82*(11), 801–811. https://doi.org/10.1016/j.humimm.2021.02.012

Jain, M., Olsen, H. E., Paten, B., Akeson, M. (2016). The Oxford Nanopore Minion: Delivery of Nanopore sequencing to the Genomics Community. Genome Biology, 17(1). https://doi.org/10.1186/s13059-016-1103-0

Kipp, E. J., Lindsey, L. L., Milstein, M. S., Blanco, C. M., Baker, J. P., Faulk, C., Oliver, J. D., Larsen, P. A. (2023). Nanopore adaptive sampling for targeted mitochondrial genome sequencing and bloodmeal identification in hematophagous insects. *Parasites & Vectors*, 16(68). https://doi.org/10.1186/s13071-023-05679-3

Kono, N., Arakawa, K. (2019). Nanopore sequencing: Review of potential applications in functional genomics. Development, *Growth & Differentiation*, 61(5), 316–326. https://doi.org/10.1111/dgd.12608

Krahn, A. D., Laksman, Z., Sy, R. W., Postema, P. G., Ackerman, M. J., Wilde, A. A. M., Han, H.-C. (2022). Congenital long QT syndrome. *JACC: Clinical Electrophysiology*, *8*(5), 687–706. https://doi.org/10.1016/j.jacep.2022.02.017

Lin, B., Hui, J., Mao, H. (2021). Nanopore technology and its applications in Gene Sequencing. *Biosensors*, 11(7), 214–230. https://doi.org/10.3390/bios11070214

Loeys, BL., Dietz, HC. (2008). Loeys-Dietz Syndrome. *Gene Reviews*. https://www.ncbi.nlm.nih.gov/books/NBK1133/

Loose, M., Malla, S., Stout, M. (2016). Real-time selective sequencing using Nanopore Technology. Nature Methods, 13(9), 751–754. https://doi.org/10.1038/nmeth.3930

Lu, H., Giordano, F., Ning, Z. (2016). Oxford Nanopore Minion Sequencing and Genome Assembly. Genomics, Proteomics & Bioinformatics, 14(5), 265–279. https://doi.org/10.1016/j.gpb.2016.05.004

Mariya, T., Kato, T., Sugimoto, T., Miyai, S., Inagaki, H., Ohye, T., Sugihara, E., Muramatsu, Y., Mizuno, S., Kurahashi, H. (2022). Target enrichment long-read sequencing with adaptive sampling can determine the structure of the small supernumerary marker chromosomes. *Journal of Human Genetics*, 67(6), 363–368. https://doi.org/10.1038/s10038-021-01004-x

McCombie, W. R., McPherson, J. D., Mardis, E. R. (2018). Next-generation Sequencing Technologies. *Cold Spring Harbor Perspectives in Medicine*, *9*(11), a036798. https://doi.org/10.1101/cshperspect.a036798

McGowan, M. P., Hosseini Dehkordi, S. H., Moriarty, P. M., Duell, P. B. (2019). Diagnosis and treatment of heterozygous familial hypercholesterolemia. *Journal of the American Heart Association*, *8*(24), e013225. https://doi.org/10.1161/jaha.119.013225

Meester, J. A., Verstraeten, A., Schepers, D., Alaerts, M., Van Laer, L., Loeys, B. L. (2017). Differences in manifestations of Marfan syndrome, Ehlers-Danlos syndrome, and

Loeys-Dietz syndrome. Annals of Cardiothoracic Surgery, 6(6), 582–594. https://doi.org/10.21037/acs.2017.11.03

Miller, D. E., Sulovari, A., Wang, T., Loucks, H., Hoekzema, K., Munson, K. M., Lewis, A. P., Fuerte, E. P., Paschal, C. R., Walsh, T., Thies, J., Bennett, J. T., Glass, I., Dipple, K. M., Patterson, K., Bonkowski, E. S., Nelson, Z., Squire, A., Sikes, M., Eichler, E. E. (2021). Targeted long-read sequencing identifies missing disease-causing variation. *The American Journal of Human Genetics*, 108(8), 1436–1449. https://doi.org/10.1016/j.ajhg.2021.06.006

Miyatake, S., Koshimizu, E., Fujita, A., Doi, H., Okubo, M., Wada, T., Hamanaka, K., Ueda, N., Kishida, H., Minase, G., Matsuno, A., Kodaira, M., Ogata, K., Kato, R., Sugiyama, A., Sasaki, A., Miyama, T., Satoh, M., Uchiyama, Y., Matsumoto, N. (2022). Rapid and comprehensive diagnostic method for repeat expansion diseases using Nanopore sequencing. *Npj Genomic Medicine*, 7(1). https://doi.org/10.1038/s41525-022-00331-y

Mukasheva, G., Abenova, M., Shaltynov, A., Tsigengage, O., Mussabekova, Z., Bulegenov, T., Shalgumbaeva, G., Semenova, Y. (2022). Incidence and mortality of cardiovascular disease in the Republic of Kazakhstan: 2004-2017. *Iranian Journal of Public Health*, 51(4), 821–830. https://doi.org/10.18502/ijph.v51i4.9243

Munro, R., Wibowo, S., Payne, A., Loose, M. (2024). Icarust, a real-time simulator for Oxford Nanopore Adaptive Sampling. Bioinformatics. https://doi.org/10.1093/bioinformatics/btae141

Nakamura, W., Hirata, M., Oda, S., Chiba, K., Okada, A., Mateos, R, N., Sugawa, M., Iida, N., Ushiama, M., Tanabe, N., Sakamoto, H., Kawai, Y., Tokunaga, K., Tsujimoto, S., Shiba, N., Ito, S., Yoshida, T., Shiraishi, Y. (2023). A comprehensive workflow for target adaptive sampling long-read sequencing applied to hereditary cancer patient genomes. *medRxiv The Preprint Serve for Health Sciences.* 2023.05.30.23289318. https://doi.org/10.1101/2023.05.30.23289318

Oxford Nanopore Technologies. (2020). Adaptive sampling – considerations for experimental design. https://community.nanoporetech.com/info_sheets/adaptive-sampling/v/ads_s1016_v1_revj_12nov2020/considerations-for-experimental-design

Park, P. J. (2009). Chip–SEQ: Advantages and challenges of a maturing technology. Nature Reviews Genetics, 10(10), 669–680. https://doi.org/10.1038/nrg2641

Parker, M. (2023, June 19). Targeted BRCA gene analysis with Oxford Nanopore. EPI2ME Labs. https://labs.epi2me.io/human-targeted-analysis/#appendix-1---bam-file-generation

Payne, A., Holmes, N., Clarke, T., Munro, R., Debebe, B. J., & Loose, M. (2020). Readfish enables targeted nanopore sequencing of gigabase-sized genomes. Nature Biotechnology, 39(4), 442–450. https://doi.org/10.1038/s41587-020-00746-x

Petersen, B.-S., Fredrich, B., Hoeppner, M. P., Ellinghaus, D., Franke, A. (2017). Opportunities and challenges of whole-genome and -exome sequencing. BMC Genetics, 18(1). https://doi.org/10.1186/s12863-017-0479-5

Pua, C. J., Bhalshankar, J., Miao, K., Walsh, R., John, S., Lim, S. Q., Chow, K., Buchan, R., Soh, B. Y., Lio, P. M., Lim, J., Schafer, S., Lim, J. Q., Tan, P., Whiffin, N., Barton, P. J., Ware, J. S., & Cook, S. A. (2016). Development of a comprehensive sequencing

assay for inherited cardiac condition genes. *Journal of Cardiovascular Translational Research, 9*(1), 3–11. https://doi.org/10.1007/s12265-016-9673-5

Quinlan, A. R., & Hall, I. M. (2010). BEDTools: a flexible suite of utilities for comparing genomic features. Bioinformatics, 26(6), 841–842. https://doi.org/10.1093/bioinformatics/btq033

Rhoads, A., & Au, K. F. (2015). PacBio sequencing and its applications. Genomics, Proteomics & Bioinformatics, 13(5), 278–289. https://doi.org/10.1016/j.gpb.2015.08.002

Sanger, F., Nicklen, S., Coulson, A. R. (1977). DNA sequencing with chain-terminating inhibitors. Proceedings of the National Academy of Sciences, 74(12), 5463–5467. https://doi.org/10.1073/pnas.74.12.5463

Schudel, A., Chapman, A. P., Yau, M.-K., Higginson, C. J., Francis, D. M., Manspeaker, M. P., Avecilla, A. R., Rohner, N. A., Finn, M. G., Thomas, S. N. (2020). Programmable multistage drug delivery to lymph nodes. Nature Nanotechnology, 15(6), 491–499. https://doi.org/10.1038/s41565-020-0679-4

Shah, S. R., Park, K., Alweis, R. (2019). Long QT Syndrome: A comprehensive review of the literature and current evidence. *Current Problems in Cardiology*, *44*(3), 92–106. https://doi.org/10.1016/j.cpcardiol.2018.04.002

Stafford, F., Krishnan, N., Richardson, E., Butters, A., Hespe, S., Burns, C., Gray, B., Medi, C., Nowak, N., Isbister, J. C., Raju, H., Richmond, D., Ryan, M. P., Singer, E. S., Sy, R. W., Yeates, L., Bagnall, R. D., Semsarian, C., Ingles, J. (2022). The role of genetic testing in diagnosis and care of inherited cardiac conditions in a specialised multidisciplinary clinic. *Genome Medicine*, 14(1), 145. https://doi.org/10.1186/s13073-022-01149-0

Sturm, A. C., Knowles, J. W., Gidding, S. S., Ahmad, Z. S., Ahmed, C. D., Ballantyne, C. M., Baum, S. J., Bourbon, M., Carrié, A., Cuchel, M., de Ferranti, S. D., Defesche, J. C., Freiberger, T., Hershberger, R. E., Hovingh, G. K., Karayan, L., Kastelein, J. J., Kindt, I., Lane, S. R., Rader, D. J. (2018). Clinical genetic testing for familial hypercholesterolemia. *Journal of the American College of Cardiology*, *72*(6), 662–680. https://doi.org/10.1016/j.jacc.2018.05.044

Thermo Fisher Scientific. (n.d.). *T042-TECHNICAL BULLETIN NanoDrop Spectrophotometers*. Note-on-the-260_280-and-260_230-Ratios. https://dna.uga.edu/wp-content/uploads/sites/51/2019/02/Note-on-the-260_280-and-260_230-Ratios.pdf

van Dijk, E. L., Auger, H., Jaszczyszyn, Y., & Thermes, C. (2014). Ten years of next-generation sequencing technology. Trends in Genetics, 30(9), 418–426. https://doi.org/10.1016/j.tig.2014.07.001

van Dijk, E. L., Jaszczyszyn, Y., Naquin, D., & Thermes, C. (2018). The Third Revolution in Sequencing Technology. Trends in Genetics, 34(9), 666–681. https://doi.org/10.1016/j.tig.2018.05.008

Velchev, J. D., Van Laer, L., Luyckx, I., Dietz, H., & Loeys, B. (2021). Loeys-Dietz syndrome. Advances in Experimental Medicine and Biology, 251–264. https://doi.org/10.1007/978-3-030-80614-9_11

Wallace, E., Howard, L., Liu, M., O'Brien, T., Ward, D., Shen, S., Prendiville, T. (2019). Long QT Syndrome: Genetics and Future Perspective. *Pediatric Cardiology*, *40*(7), 1419–1430. https://doi.org/10.1007/s00246-019-02151-x

Walsh, R., Cook, S. A. (2017). Issues and challenges in diagnostic sequencing for inherited cardiac conditions. *Clinical Chemistry*, 63(1), 116–128. https://doi.org/10.1373/clinchem.2016.254698

Wan, Y., Yang, L., Wang, J. X., Tham, C.-Y., de Sessions, P. F., Cheng, A. (2023). Direct RNA sequencing coupled with adaptive sampling enriches RNAS of interest in the transcriptome. *Research Square*. https://doi.org/10.21203/rs.3.rs-2567801/v1

Wang, K., Li, M., Hakonarson, H. (2010). ANNOVAR: Functional annotation of genetic variants from next-generation sequencing data. *Nucleic Acids Research*, 38:e164.

Weilguny, L., De Maio, N., Munro, R., Manser, C., Birney, E., Loose, M., & Goldman, N. (2023). Dynamic, adaptive sampling during nanopore sequencing using Bayesian Experimental Design. Nature Biotechnology, 41(7), 1018–1025. https://doi.org/10.1038/s41587-022-01580-z

WHO. (2019). Health Topics. Cardiovascular diseases. https://www.who.int/health-topics/cardiovascular-diseases/#tab=tab_1

Yamaguchi, K., Kasajima, R., Takane, K., Hatakeyama, S., Shimizu, E., Yamaguchi, R., Katayama, K., Arai, M., Ishioka, C., Iwama, T., Kaneko, S., Matsubara, N., Moriya, Y., Nomizu, T., Sugano, K., Tamura, K., Tomita, N., Yoshida, T., Sugihara, K., Ikenoue, T. (2021). Application of targeted nanopore sequencing for the screening and determination of structural variants in patients with Lynch syndrome. *Journal of Human Genetics*, 66(11), 1053–1060. https://doi.org/10.1038/s10038-021-00927-9

Chapter 6 – Appendices
Appendix 1 – Protocol links

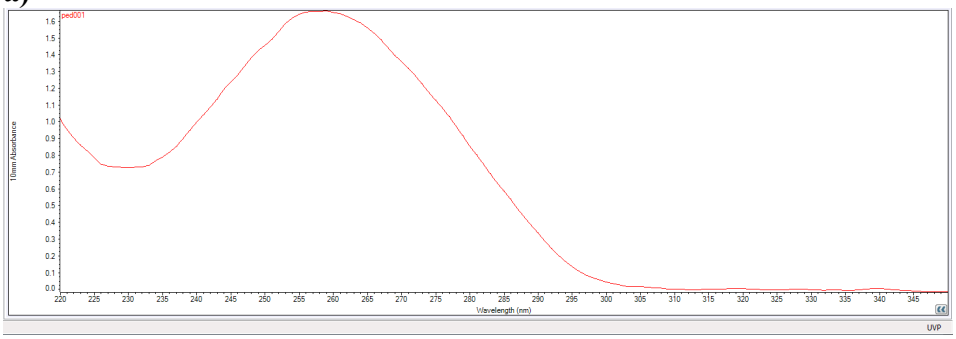
Method name:	Protocol link:
Column DNA extraction	https://cdn.cytivalifesciences.com/api/public/content/digi-14876-pdf https://www.qiagen.com/ja-us/products/discovery-and-translational-research/dna-rna-purification/dna-purification/genomic-dna/qiaamp-dna-blood-kits
Library preparation	https://community.nanoporetech.com/protocols/gDNA-sqk-lsk109/v
Adaptive sampling guide	https://community.nanoporetech.com/info_sheets/adaptive-sampling/v/ads_s1016_v1_revj_12nov2020/considerations-for-experimental-design
Sequencing	https://community.nanoporetech.com/docs/prepare/library_prep_protocols/experiment-companion-minknow/v/mke_1013_v1_revdb_11apr2016/minion-mk1c-quick-start-guide
TruSight Cardio panel	https://www.illumina.com/content/dam/illumina/gcs/assembled-assets/marketing-literature/trusight-cardio-data-sheet-m-gl-00688/trusight-cardio-data-sheet-m-gl-00688.pdf

Appendix 2 – Summary of reagents and resources

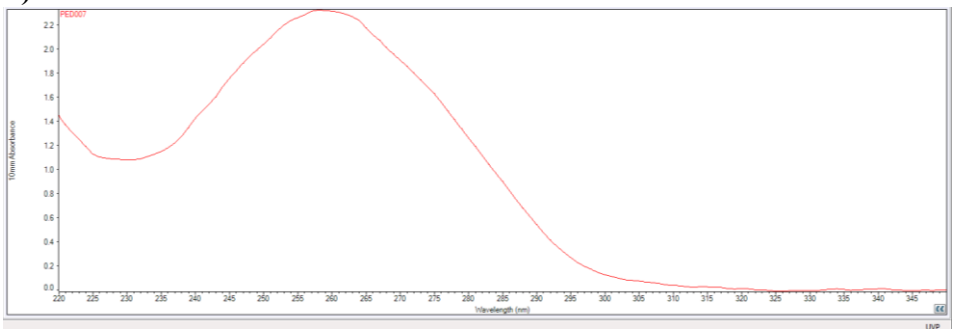
№	Type of equipment	Instrument	Purpose of equipment
1	Benchtop Centrifuge	Eppendorf Centrifuge 5810R	Rapid Spinning for Tubes and Plates
2	Microcentrifuge	Eppendorf Centrifuge 5418R	Rapid Spinning for Microcentrifuge Tube
3	Spectrophotometer	Nanodrop 2000	DNA Quantitative Assessment, Spectrophotometric Analysis
4	Fluorometer	Qubit 2.0	DNA Quantitative Assessment, Fluorometric analysis
5	Gel Documentation System	BioRad GelDoc	DNA Qualitative Assessment, Gel Analysis
6	Third Generation Sequencing Platform, Nanopore Sequencer	GridION, Oxford Nanopore	Genome and Targeted Sequencing
7	Flow Cell	Oxford Nanopore Technologies	Genome and Targeted Sequencing
8	QIAmp DNA Mini and Blood Mini Kit	Qiagen, 51106	DNA Isolation
9	Illustra Blood Genomic Mini Prep Kit	Cytiva, 28-9042-65	DNA Isolation
10	DNA and RNA free UltraPureTM Water	Thermo Scientific, 10977035	DNA Rehydration
11	Qubit dsDNA Broad Range	Thermo Scientific, Q32851	Fluorometric DNA Assessment
12	UltraPure Agarose Powder, 500g	Thermo Fisher Scientific, 16500-500	Agarose Gel for Electrophoresis
13	DNA Ladder (1 kb+)	Invitrogen, 10488-085	Assessment of DNA Weight
14	Ethidium Bromide (0.5mg/ml)	Sigma-Aldrich, E1510	Qualitative DNA Assessment on Gel Doc
15	Ligation Sequencing Kit	Oxford Nanopore, SQK-LSK 109	DNA Library Preparation
16	Flow Cell Priming Kit	Oxford Nanopore, EXP-FLP 002	Priming of Sequencing Flow Cells
17	Flow Cell Wash Kit	Oxford Nanopore, EXP-WSH 004	Washing of Sequencing Flow Cells

Appendix 3 - Supplementary figures

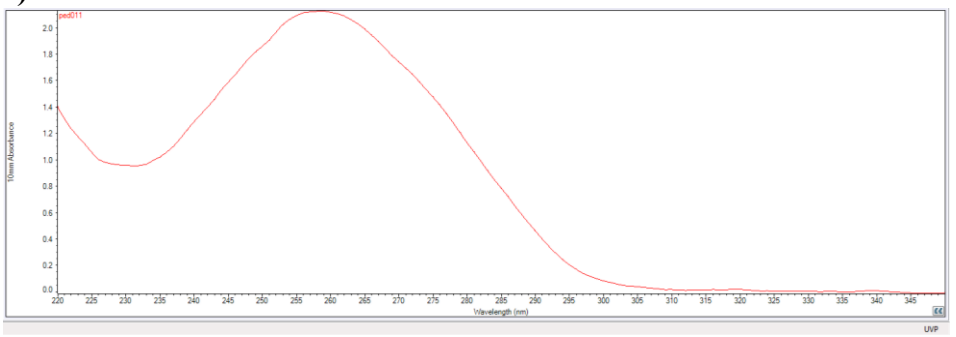
a)



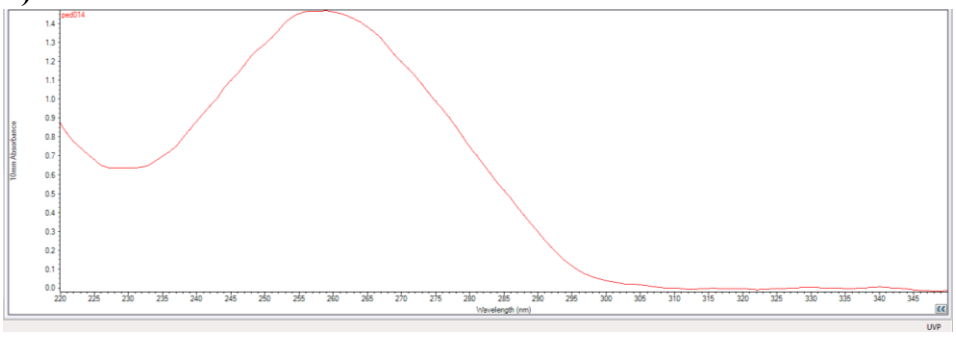
b)



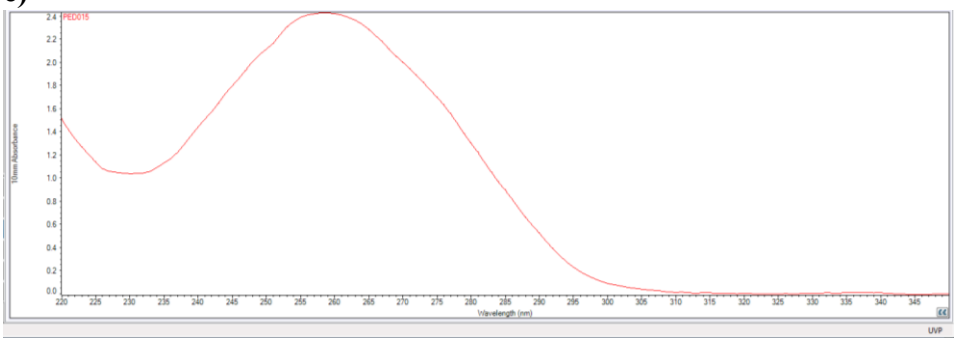
c)



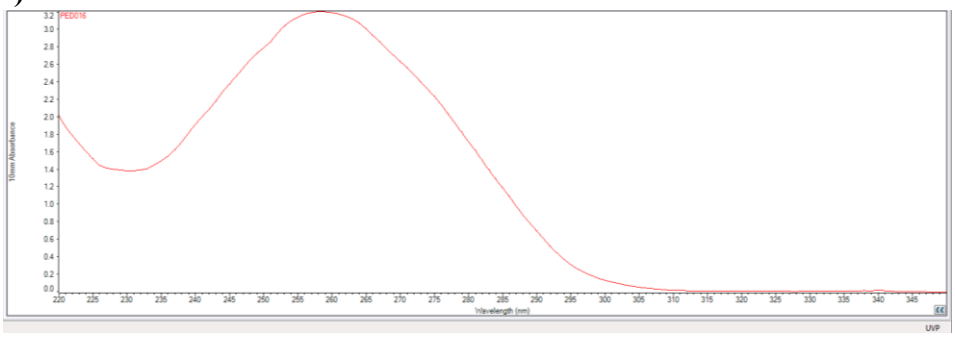
d)



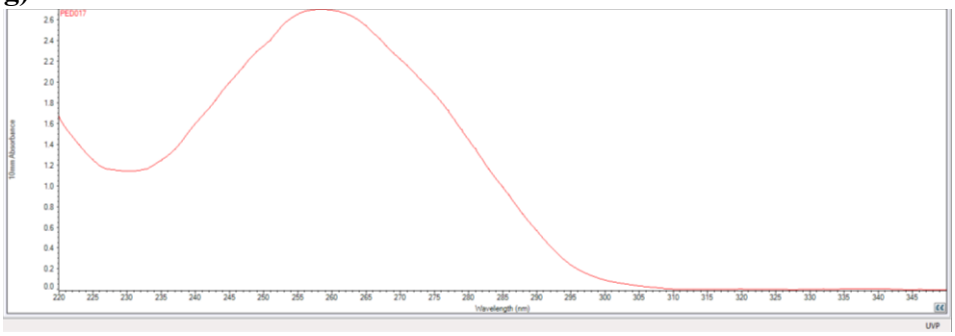
e)


PED015
10mm Absorbance
Wavelength (nm)
UVP

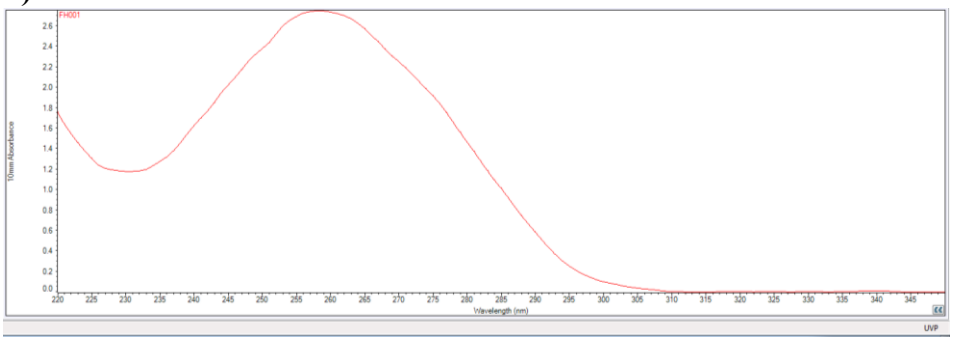
f)


PED016
10mm Absorbance
Wavelength (nm)
UVP

g)


PED017
10mm Absorbance
Wavelength (nm)
UVP

h)


PH001
10mm Absorbance
Wavelength (nm)
UVP

i)

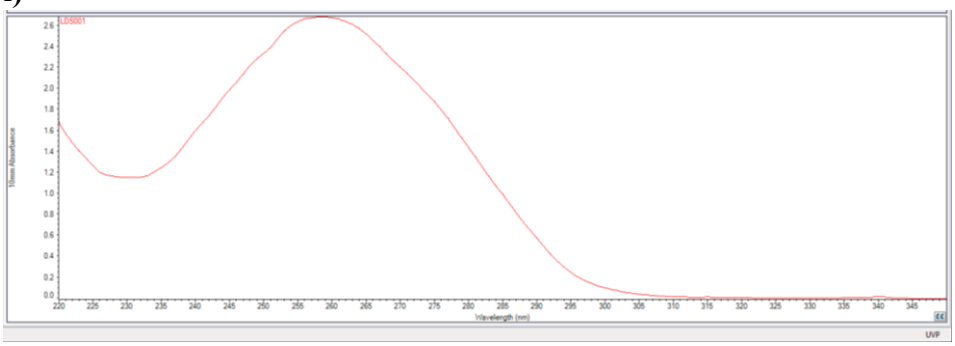
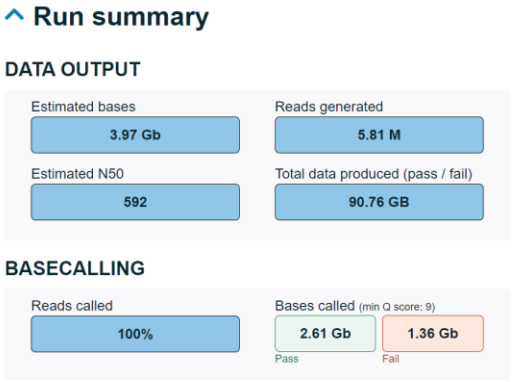
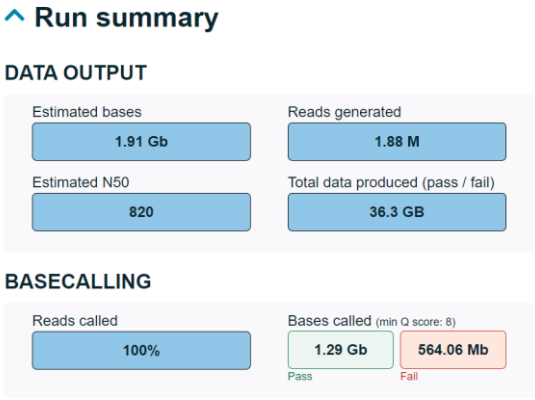


Figure S1. Overview of Nanodrop slope results, for the assessment of gDNA purity. a) Nanodrop results for PED001; b) Nanodrop results for PED007; c) Nanodrop results for PED011; d) Nanodrop results for PED014; e) Nanodrop results for PED015; f) Nanodrop results for PED016; g) Nanodrop results for PED017; h) Nanodrop results for FH001; i) Nanodrop results for LDS001.

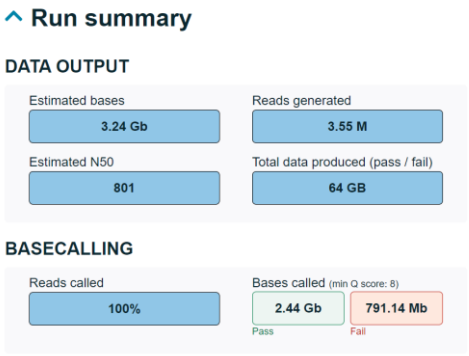
a)


Run summary
DATA OUTPUT
Estimated bases
3.97 Gb
Reads generated
5.81 M
Estimated N50
592
Total data produced (pass / fail)
90.76 GB
BASECALLING
Reads called
100%
Bases called (min Q score: 9)
2.61 Gb
1.36 Gb
Pass
Fail

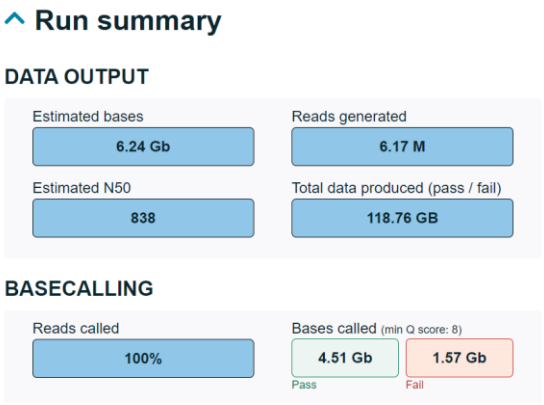
b)


Run summary
DATA OUTPUT
Estimated bases
1.91 Gb
Reads generated
1.88 M
Estimated N50
820
Total data produced (pass / fail)
36.3 GB
BASECALLING
Reads called
100%
Bases called (min Q score: 8)
1.29 Gb
564.06 Mb
Pass
Fail

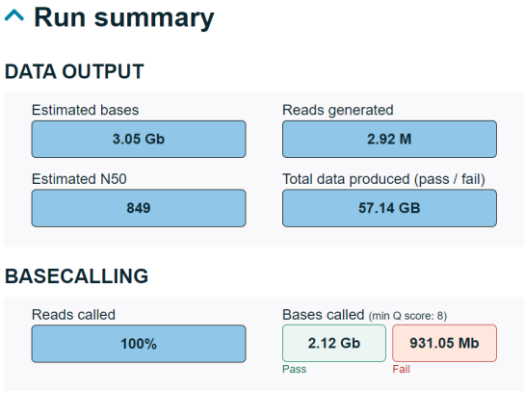
c)


Run summary
DATA OUTPUT
Estimated bases
3.24 Gb
Reads generated
3.55 M
Estimated N50
801
Total data produced (pass / fail)
64 GB
BASECALLING
Reads called
100%
Bases called (min Q score: 8)
2.44 Gb
791.14 Mb
Pass
Fail

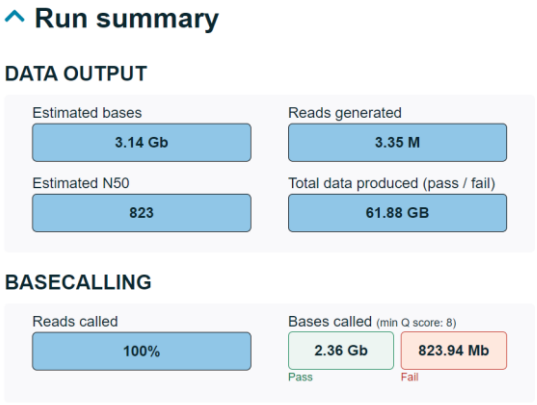
d)


Run summary
DATA OUTPUT
Estimated bases
6.24 Gb
Reads generated
6.17 M
Estimated N50
838
Total data produced (pass / fail)
118.76 GB
BASECALLING
Reads called
100%
Bases called (min Q score: 8)
4.51 Gb
1.57 Gb
Pass
Fail

e)


Run summary
DATA OUTPUT
Estimated bases
3.05 Gb
Reads generated
2.92 M
Estimated N50
849
Total data produced (pass / fail)
57.14 GB
BASECALLING
Reads called
100%
Bases called (min Q score: 8)
2.12 Gb
931.05 Mb
Pass
Fail

f)


Run summary
DATA OUTPUT
Estimated bases
3.14 Gb
Reads generated
3.35 M
Estimated N50
823
Total data produced (pass / fail)
61.88 GB
BASECALLING
Reads called
100%
Bases called (min Q score: 8)
2.36 Gb
823.94 Mb
Pass
Fail

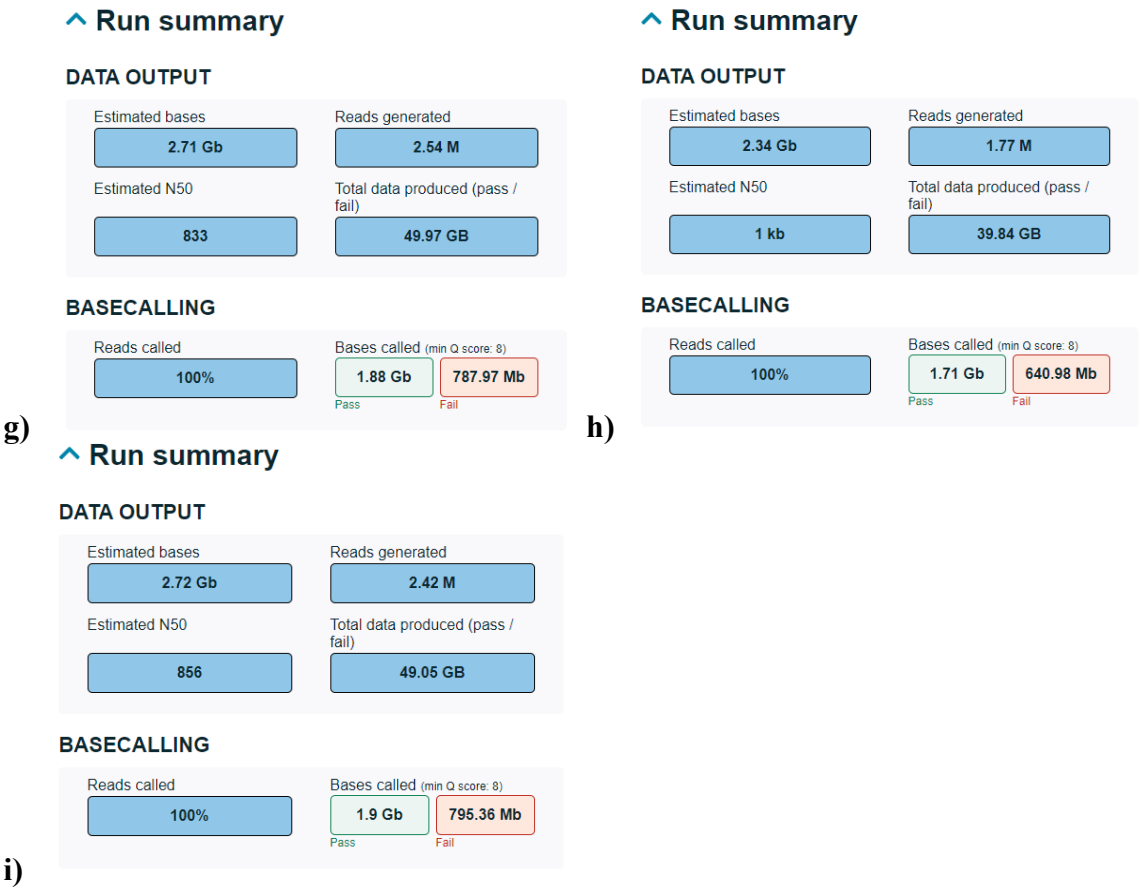


Figure S2. Overview on sequencing run summary. a) FH001 sequencing summary; b) LDS001 sequencing summary; c) PED001 sequencing summary; d) PED007 sequencing summary; e) PED011 sequencing summary; f) PED014 sequencing summary; g) PED015 sequencing summary; h) PED016 sequencing summary; i) PED017 sequencing summary

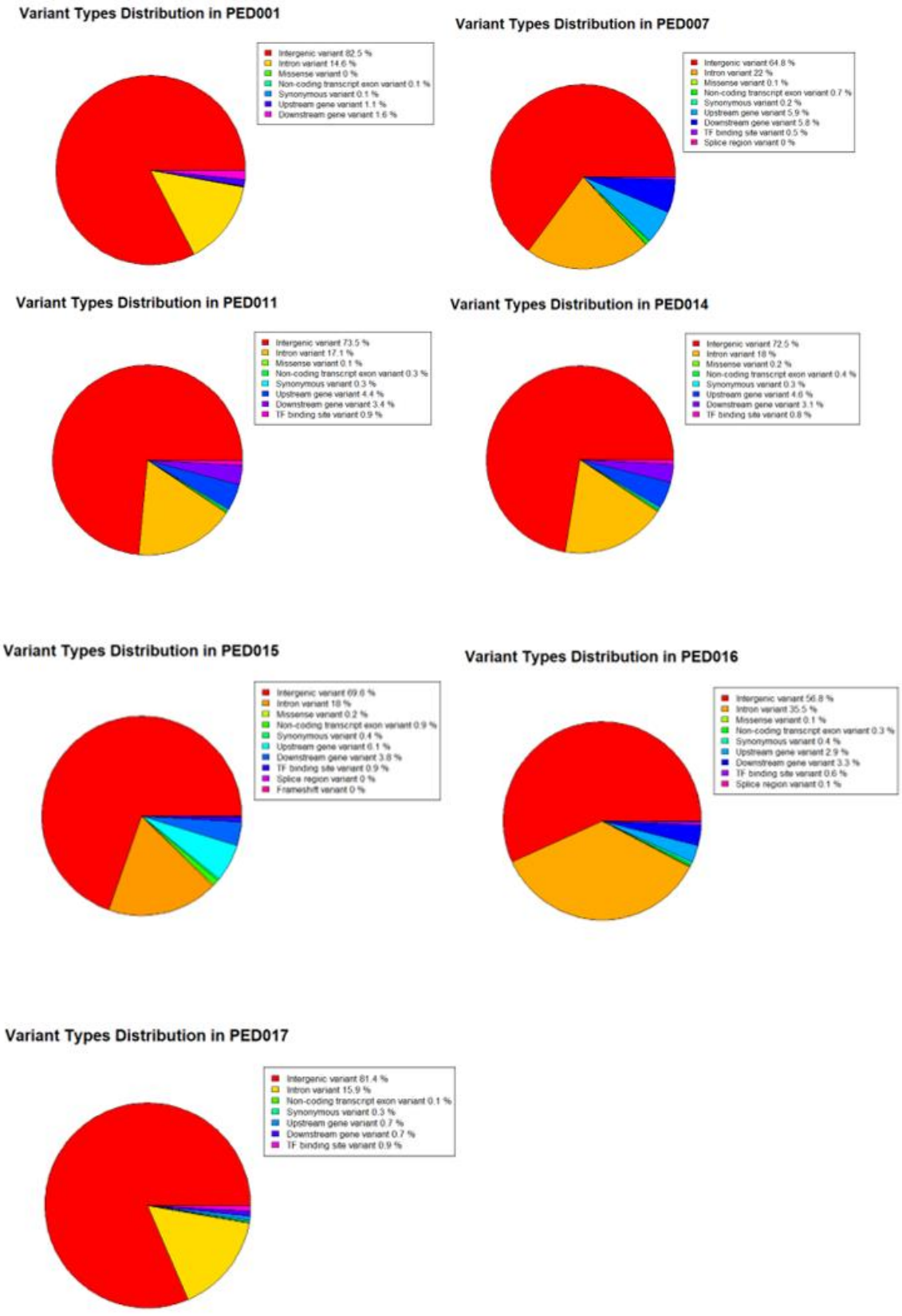


Figure S3. Distribution of the variant types among PED001-PED017 samples annotated by SNPEff.

Appendix 4 - Supplementary tables

Table S1. Raw data of BAM coverage summary of the average depth of target on-target and off-target regions, N50 resulted and number of variants identified through Epi2me labs human variation workflow.

ID	N50 (bp)	Average target depth of Coverage	Average off-target depth of Coverage	Variants identified
FH001	592	6366	7439	1321
LDS001	820	2442	3152	2179
PED001	801	8795	10443	3627
PED007	838	5013	6148	8202
PED011	849	3595	4585	3266
PED014	823	4693	5832	3892
PED015	833	3397	4297	3215
PED016	856	3295	4088	4006
PED017	1000	2372	3043	2517

Table S2. Identified pathogenic variants via Illumina TruSight Cardio for FH001

Gene APOE	Position 45412079	Transcript NM_000041:exon4	Exonic effect nonsynonymous SNV	Genotype het
ID rs7412	OMIM 107741	Nomenclature c.C526T:p.R176C	ClinVar classification Pathogenic	ACMG/AMP classification **Pathogenic** (upd: 3.10.23)
Phenotype Hyperlipoproteinemia, type III	1000G 0.075	ExAC 0.072	MutationTaster 0.930	gnomAD 0.0833
Gene LDLR	Position 11216000	Transcript NM_001195799:exon3	Exonic effect nonsynonymous SNV	Genotype hom
ID n/a	OMIM 606945	Nomenclature c.G295A:p.E99K	ClinVar classification Pathogenic/Likely_pathogenic	ACMG/AMP classification Likely_pathogenic
Phenotype Familial_hypercholesterolemia	1000G n/a	ExAC n/a	MutationTaster 1.0	gnomAD n/a

Table S3. Identified pathogenic variants via Illumina TruSight Cardio for PED001

Gene SCN5A	Position 38592513	Transcript NM_001099405:exon27	Exonic effect nonsynonymous SNV	Genotype het
ID rs137854601	OMIM 600163	Nomenclature :c.G5296A:p.E1766K	ClinVar classification Pathogenic	ACMG/AMP classification **Likely pathogenic**
Phenotype Long_QT_syndrome_1|Sinus_node_disease|Congenital_long_QT_syndrome|Brugada_syndrome|Long_QT_syndrome_3|	1000G n/a	SIFT 0.001	MutationTaster 1.0	gnomAD n/a

Table S4. Identified pathogenic variants via Illumina TruSight Cardio for PED007

Gene KCNH2	Position 150648799	Transcript NM_001204798:exon3	Exonic effect nonsynonymous SNV	Genotype het
ID rs121912504	OMIM 152427	Nomenclature c.C662T:p.A221V	ClinVar classification Pathogenic/Likely pathogenic	ACMG/AMP classification **Pathogenic**
Phenotype Prolonged_QT_interval|Long_QT_syndrome|Congenital_long_QT_syndrome|Long_QT_syndrome_2|Cardiovascular_phenotype|not_provided	1000G n/a	ExAC n/a	MutationTaster 1.0	gnomAD n/a

Table S5. Identified variants via ONT adaptive sampling for LDS001

Gene RYR2	Position 237727137	Transcript NM_001035:exon76	Exonic effect synonymous SNV	ID rs2685301
OMIM 180902	Nomenclature c.C10776T:p.S3592S	gnomAD n/a	ClinVar Benign/Likely_benign	ACMG/AMP Benign
Phenotype Hyperlipoproteinemia, type Ventricular arrhythmias	1000G n/a	ExAC 0.9885	MutationTaster n/a	

Table S6. Identified variants via ONT adaptive sampling for PED001

Gene RYR2	Position 237700418 237718470 237727137	Transcript NM_001035:exon65 NM_001035:exon73 NM_001035:exon76	Exonic effect synonymous SNV synonymous SNV synonymous SNV	ID rs2797436 rs2797441 rs2685301
OMIM 180902	Nomenclature c.T9318G:p.S3106S c.C10503T:p.T3501T c.C10776T:p.S3592S	gnomAD n/a	ClinVar Benign/Likely_benign Benign/Likely_benign Benign/Likely_benign	ACMG/AMP Benign Benign Benign
Phenotype Hyperlipoproteinemia, type Ventricular arrhythmias	1000G n/a	ExAC 0.9917 0.9886 0.9885	MutationTaster n/a n/a n/a	

Table S7. Identified variants via ONT adaptive sampling for PED007

Gene RYR2	Position 237651483 237700418 237718470 237727137	Transcript NM_001035:exon51 NM_001035:exon65 NM_001035:exon73 NM_001035:exon76	Exonic effect synonymous SNV synonymous SNV synonymous SNV synonymous SNV	ID rs684923 rs2797436 rs2797441 rs2685301
OMIM 180902	Nomenclature c.C7806T:p.H2602H c.T9318G:p.S3106S c.C10503T:p.T3501T c.C10776T:p.S3592S	gnomAD n/a	ClinVar Benign/Likely_benign Benign/Likely_benign Benign/Likely_benign Benign/Likely_benign	ACMG/AMP Benign Benign Benign Benign
Phenotype Hyperlipoproteinemia, type Ventricular arrhythmias	1000G n/a	ExAC 0.5198 0.9917 0.9886 0.9885	MutationTaster n/a n/a n/a n/a	

Table S8. Identified variants via ONT adaptive sampling for PED011

Gene RYR2	Position 237718470 237727137	Transcript NM_001035:exon73 NM_001035:exon76	Exonic effect synonymous SNV synonymous SNV	ID rs2797441 rs2685301
OMIM 180902	Nomenclature c.C10503T:p.T3501T c.C10776T:p.S3592S	gnomAD n/a	ClinVar Benign/Likely_benign Benign/Likely_benign	ACMG/AMP Benign Benign
Phenotype Hyperlipoproteinemia, type Ventricular arrhythmias	1000G n/a	ExAC 0.9886 0.9885	MutationTaster n/a n/a	

Table S9. Identified variants via ONT adaptive sampling for PED014

Gene RYR2	Position 237651483 237700418 237718470 237727137	Transcript NM_001035:exon51 NM_001035:exon65 NM_001035:exon73 NM_001035:exon76	Exonic effect synonymous SNV synonymous SNV synonymous SNV synonymous SNV	ID rs684923 rs2797436 rs2797441 rs2685301
OMIM 180902	Nomenclature c.C7806T:p.H2602H c.T9318G:p.S3106S c.C10503T:p.T3501T c.C10776T:p.S3592S	gnomAD n/a	ClinVar Benign/Likely_benign Benign/Likely_benign Benign/Likely_benign Benign/Likely_benign	ACMG/AMP Benign Benign Benign Benign
Phenotype Hyperlipoproteinemia, type Ventricular arrhythmias	1000G n/a	ExAC 0.5198 0.9917 0.9886 0.9885	MutationTaster n/a n/a n/a n/a	

Table S10. Identified variants via ONT adaptive sampling for PED015

Gene RYR2	Position 237651483 237700418 237718470 237727137	Transcript NM_001035:exon51 NM_001035:exon65 NM_001035:exon73 NM_001035:exon76	Exonic effect synonymous SNV synonymous SNV synonymous SNV synonymous SNV	ID rs684923 rs2797436 rs2797441 rs2685301
OMIM 180902	Nomenclature c.C7806T:p.H2602H c.T9318G:p.S3106S c.C10503T:p.T3501T c.C10776T:p.S3592S	gnomAD n/a	ClinVar Benign/Likely_benign Benign/Likely_benign Benign/Likely_benign Benign/Likely_benign	ACMG/AMP Benign Benign Benign Benign
Phenotype Hyperlipoproteinemia, type Ventricular arrhythmias	1000G n/a	ExAC 0.5198 0.9917 0.9886 0.9885	MutationTaster n/a n/a n/a n/a	

Table S11. Identified variants via ONT adaptive sampling for PED016

Gene LMNA	Position 156137743	Transcript NM_001257374:exon10	Exonic effect synonymous SNV	ID rs4641
OMIM 150330	Nomenclature c.C1362T:p.H454H	gnomAD n/a	ClinVar Benign	ACMG/AMP Benign
Phenotype Cardiomyopathy, dilated	1000G n/a	ExAC 0.2655	MutationTaster n/a	
Gene SEMA4A	Position 156158748	Transcript NM_001193302:exon4	Exonic effect synonymous SNV	ID rs562037528
OMIM 607292	Nomenclature c.C96T:p.I32I	gnomAD n/a	ClinVar Conflicting_interpretations_of_pathogenicity	ACMG/AMP Likely benign
Phenotype Cone-rod dystrophy 10	1000G n/a	ExAC 0.0011	MutationTaster n/a	
Gene RYR2	Position 237651483 237700418 237718470 237727137	Transcript NM_001035:exon51 NM_001035:exon65 NM_001035:exon73 NM_001035:exon76	Exonic effect synonymous SNV synonymous SNV synonymous SNV synonymous SNV	ID rs684923 rs2797436 rs2797441 rs2685301
OMIM 180902	Nomenclature c.C7806T:p.H2602H c.T9318G:p.S3106S c.C10503T:p.T3501T c.C10776T:p.S3592S	gnomAD n/a	ClinVar Benign/Likely_benign Benign/Likely_benign Benign/Likely_benign Benign/Likely_benign	ACMG/AMP Benign Benign Benign Benign
Phenotype Ventricular arrhythmias	1000G n/a	ExAC 0.5198 0.9917 0.9886 0.9885	MutationTaster n/a n/a n/a n/a	

Table S12. Identified variants via ONT adaptive sampling for PED017

Gene RYR2	Position 237700418 237718470 237727137	Transcript NM_001035:exon65 NM_001035:exon73 NM_001035:exon76	Exonic effect synonymous SNV synonymous SNV synonymous SNV	ID rs2797436 rs2797441 rs2685301
OMIM 180902	Nomenclature c.T9318G:p.S3106S c.C10503T:p.T3501T c.C10776T:p.S3592S	gnomAD n/a	ClinVar Benign/Likely_benign Benign/Likely_benign Benign/Likely_benign	ACMG/AMP Benign Benign Benign
Phenotype Hyperlipoproteinemia, type Ventricular arrhythmias	1000G n/a	ExAC 0.9917 0.9886 0.9885	MutationTaster n/a n/a n/a	