كلية الطب والصيدلة وطب الأسنان
FACULTÉ DE MÉDECINE, DE PHARMACIE ET DE MÉDECINE DENTAIRE

جامعة سيدي محمد بن عبد الله - فاس
UNIVERSITÉ SIDI MOHAMED BEN ABDELLAH DE FES

Evaluation of nanopore sequencing as a diagnostic tool for Wilson disease in Morocco
(About 4 cases)

MEMOIRE PRESENTE PAR :

Docteur ELMAKHZEN BADREDDINE
Né le 05/09/1992 à FES

Pour l'obtention du diplôme de spécialité en médecine

Option : Génétique Médicale

Rapporteur : Pr BOUGUENOUCH LAILA

Sous la direction de : Pr OULDIM KARIM

Session Juin 2023

Pr. BOUGUENOUCH LAILA
Professeur Agrégé
Génétique Médicale

كلية الطب والصيدلة وطب الأسنان
ⵜⴰⵖⵔⵉⵙⵜ ⵏ ⵜⵓⵙⵙⵔⵜ ⴷ ⵜⵓⴳⵓⵔⵉⵜ ⴷ ⵜⵓⵙⵙⵔⵜ ⵏ ⵜⵉⵏⵙⴰⵏ
FACULTÉ DE MÉDECINE, DE PHARMACIE ET DE MÉDECINE DENTAIRE

جامعة سيدي محمد بن عبد الله - فاس
ⵜⴰⵙⴷⴰⵡⵉⵜ ⵙⵉⴷⵉ ⵎⵓⵃⴰⵎⴷ ⴱⵏ ⵄⴱⴷⵍⵍⴰⵀ - ⴼⴰⵙ
UNIVERSITÉ SIDI MOHAMED BEN ABDELLAH DE FES

LE VALUATION OF NANOPORE SEQUENCING AS A DIAGNOSTIC TOOL FOR WILSON DISEASE IN MOROCCO (About 4 cases)

MEMOIRE PRESENTE PAR :

Docteur ELMAKHZEN BADREDDINE
Né le 05/09/1992 à FES (MAROC)

POUR L'OBTENTION DU DIPLOME DE SPECIALITE EN MEDECINE

OPTION : GÉNÉTIQUE MÉDICALE

Rapporteur PROFESSEUR BOUGUENOUCH LAILA

Sous la direction de : PROFESSEUR OULDIM KARIM

Session juin 2023

REMERCIEMENTS

A dieu

Je remercie le bon dieu tout puissant qui m'a donné la force et la volontéd'achever ce travail

A ma famille

Je remercie toute ma famille pour son soutien inconditionnel et sesencouragements continus

A mes maitres

Pr BOUGUENOUCH Laila, *je la remercie de m'avoir encouragé et de m'avoiraccordé toute sa confiance et son attention tout au long de ma formation. aussi pour son support inestimable et pour la formation que j'aieu l'honneur de suivre sous sa responsabilité durant ces quatre années de résidanat*

Pr. Dr. HUGUES Abriel *thanks you; without your support and guidance, this work would never been achieved.*

You have entrusted me this rich work of interest and guided me in every

step of its implementation.

Your seriousness, your competence and sense of duty have greatly impressed me.

Your tireless encouragement. Your kindness deserves all of admiration.

Please find here the expression of my, in the name of the departement and every moroccan, my respectful consideration and deep admiration for all of your scientific and human qualities.

Pr OULDIM Karim *pour son support inestimable et pour la formation que j'aieu l'honneur de suivre sous sa responsabilité durant ces quatre années de résidanat.*

SOMMAIRE

Table des matières

LIST OF ABBREVIATIONS

AR	autosomal recessive
ATP7B	ATPase copper transporting beta
BAM	Binary Alignment Map
CNV	Copy-number variation
IGV	Integrative Genomics Viewer
NGS	Next generation sequencing
ONT	Oxford nanopore technology
PCR	Polymerase chain reaction
SNV	Single nucleotid variation
SV	structural variation

LIST OF FIGURES

LIST OF TABLES

INTRODUCTION

Wilson's disease is an autosomal recessive disorder related to a dysfunction of copper transport expressed by the gene ATP7B. It can present with hepatic, neurologic, or psychiatric disturbances or a combination of these symptoms. In most instances, the diagnosis is established by biochemical findings. The presence of Kayser-Fleisher rings in the eyes is almost pathognomonic. As a recessive disease, it requires the presence of biallelic pathogenic variants in ATP7B. The diagnosis by itself, in many cases, is not complicated; however, genetic counselling requires molecular confirmation. According to the literature, deletion and duplication are rare in Wilson's disease. On the other hand, many studies report a high frequency of composite heterozygotes. To our knowledge, no molecular study has been conducted in the Moroccan population using nanopore sequencing.

Our study aimed to test the feasibility of long-range PCR and nanopore sequencing of the whole ATP7B gene. In particular, our aims were as follows:

1. to identify the mutational profile of Wilson's disease in the Moroccan population,
2. to demonstrate the interest and feasibility of applying long-read sequencing methods using Oxford Nanopore Technology (ONT) to identify variants in the ATP7B gene,
3. to test the economical and biological impact of nanopore sequencing in populations with limited access to genetic testing.

We designed primer pairs to amplify an average of 10 kb-long PCR fragments covering the entire ATP7B gene. Following PCR amplifications, barcoded libraries were prepared and sequenced on ONT flow cells using the Mk1C device. In this pilot study, we included 4 patients clinically diagnosed with Wilson's disease from the Department of Gastroenterology (CHU Fez).

MATERIALS AND METHODS

A. Inclusion criteria

Our work is a retrospective study that included 4 patients with Wilson disease recruited during the period between March 2022 and March 2023. Patients are followed at the gastroenterology department of CHU Hassan II. The diagnosis was made on the basis of clinical signs, biological data and ultrasound. The majority of patients benefited from genetic consultation. Patients had no previous genetic testing.

B. Blood sampling

Before any sample is taken, a form is filled in with the patient's details and consent or that of his or her guardian if he or she is a minor. Two EDTA tubes of blood are taken for DNA extraction, which can be done by kit or by salt. Tubes can be stored at +4°C or -20°C for later use.

C. Nucleic acid preparation and purification techniques

DNA (deoxyribonucleic acid) extraction is a technique that isolates DNA from a cell in sufficient quantity and quality to enable its analysis. In the Medical Genetics and Oncogenetics laboratory, extraction is carried out using two methods: either SEL or a commercial KIT (purgene). We used the kit method for DNA extraction to obtain better quality and to extract long molecular weight DNA. This technique is necessary for long-range PCR to preserve long DNA fragments.

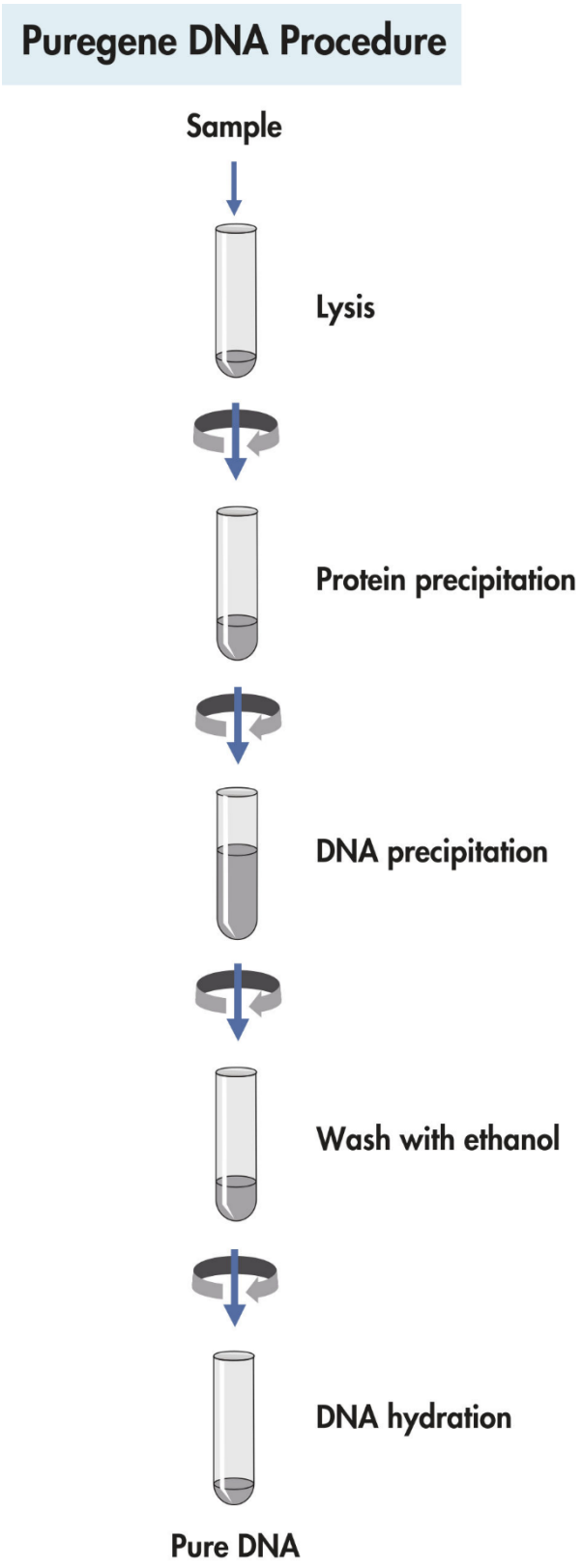

Figure 1:hight molecular weight dna extraction

1- Principle

This is a simple, fast and advantageous method when the volume of blood is low. We used the Puregene Blood Kit.

2- Procedure:

- Three milliliters of RBC lysis solution was dispensed into a centrifuge tube.
- One milliliter of 1ML whole blood was added and mixed by inverting 10 times.
- Incubate for 5 min at room temperature (15–25°C). Invert at least once during the incubation.
- Centrifuge for 2 min at 2000 x g to pellet the white blood cells.
- Carefully discard the supernatant by pipetting or pouring, leaving approximately 70 µl of the residual liquid and the white blood cell pellet.
- Vortex the tube vigorously to resuspend the pellet in the residual liquid.

  -Vortexing greatly facilitates cell lysis in the next step.

  -The pellet should be completely dispersed after vortexing.
- Then, 1 ml of Cell Lysis Solution was added, and the cells were pipetted up and down to lyse the cells or vortexed vigorously for 10 s.

  Note: Usually, no incubation is needed; however, if cell clumps are visible, incubate at 37°C until the solution is homogeneous.
- Optional: If RNA-free DNA is needed, add 5 µl of RNase A Solution, and mix by inverting 25 times. Incubate for 15 min at 37°C.

Then, incubate for 3 min on ice to quickly cool the sample.

- Add 300 µl of Protein Precipitation Solution, and vortex vigorously for 20 s at high speed.
- Centrifuge for 5 min at 2000 x g.

- Pipette 1 ml isopropanol into a clean 15 ml tube and add the supernatant from the previous step by pouring carefully. (Be sure that the protein pellet is not dislodged during pouring).
- Mix by inverting gently 50 times until the DNA is visible as threads or a clump.
- Centrifuge for 3 min at 2000 x g. (The DNA may be visible as a small white pellet)
- The supernatant was carefully discarded, and the tube was drained by inverting on a clean piece of absorbent paper, taking care that the pellet remained in the tube.
- Add 1 ml of 70% ethanol and invert several times to wash the DNA pellet.
- Centrifuge for 1 min at 2000 x g.
- The supernatant was carefully discarded. Drain the tube on a clean piece of absorbent paper, taking care that the pellet remains in the tube. The pellet was air dried for 5-10 min.
- Add 100 µl DNA Hydration Solution and vortex for 5 s at medium speed to mix.
- Incubate at 65°C for 1 h to dissolve the DNA.
- Incubate at room temperature overnight with gentle shaking. Ensure that the tube cap is tightly closed to avoid leakage. Samples can then be centrifuged briefly and transferred to a storage tube.

3-Dosing DNA:

The DNA concentration of the sample was estimated spectrophotometrically. Nucleic acids have an ultraviolet absorption peak, with a maximum at 260 nm. Interference by contaminants can be recognized by

calculating a "ratio". As proteins absorb at 280 nm, the A260/A280 ratio is used to estimate nucleic acid purity. Pure DNA should have a ratio of approximately 1.8, while pure RNA should have a value of approximately 2.0. Absorbance at 230 nm reflects contamination of the sample by substances such as carbohydrates, peptides, phenols or aromatic compounds. In the case of pure samples, the A260/A230 ratio should be approximately 2.2.

In the laboratory, "NANODROP" (Figure 2) was used to perform the assay. Simply place 2 µl of extracted DNA in the device to determine its concentration and obtain the purity curve.

Figure 2:spectrophometre

D- Longrange PCR:

PCR (polymerase chain reaction) is used to amplify a specific region of a given nucleic acid in vitro to obtain sufficient quantities to detect and study it. To achieve this, a series of reactions is repeated in a loop, replicating a double-stranded DNA template. During the PCR, the products obtained at the end of each cycle serve as a template for the following cycle, so amplification is exponential.

Double-stranded DNA replication involves three steps:

1. DNA denaturation to obtain single-stranded templates;
2. Hybridization of specific primers
3. Polymerization of the complementary strand by the polymerase enzyme.

At the end of each cycle, the products are in the form of double-stranded DNA.

Long-range PCR has been commonly used to prepare specific high-molecular-weight DNA fragments for a variety of applications, including cloning and genome mapping sequencing. Generally, to successfully amplify all amplicons in an experiment, we need to change the annealing temperature and extension time, which are specific to each amplicon because the primers may have very different Tm values.

In addition to long-range PCR, a variety of other methods, such as solution-based capture, microarray-based capture, molecular inversion probes (MIPs) and multiplex PCR, have been used in target enrichment applications. Target enrichment is a highly effective way of reducing costs and saving time when only specific genomic regions (such as all exons in a gene or a genomic region spanning a few GWAS loci) are of interest. Approaches based on capture, such as solution-based capture and microarray-based capture, achieve high performance and have advantages for medium to large target regions (10–50 Mb)22. However, microarray-based methods, such as Agilent SureSelect and HaloPlex, require large amounts of

input DNA to be successful as well as expensive hardware working with microarray slides. Solution-based capture, such as NimbleGen SeqCap, is less extensively used because of performance issues, but solution-based capture techniques are constantly improving. Generally, GC-rich segments were not well represented in capture samples. This may be attributed to sequencing bias, as well as difficulty in capturing high-GC templates. This is less a concern for the long-range PCR that ‘‘capture’’ large regions at once, especially for specific enzymes (such as PrimeSTAR GXL and QIAGEN LongRange PCR polymerase) that were optimized for amplifying GC-rich segments.

MIPs are generally believed to be superior in terms of specificity but far less amenable to multiple sample coprocessing in a single reaction. Moreover, its design has to consider the uniqueness of each target region fragment and the most suitable hybridization conditions. Long-range PCR has a unique niche: it does not require customized design by commercial vendors and can be afforded by small laboratories when a small number of samples and continuous regions (such as full gene regions including introns) are of interest.

1- First PCR: target amplification

The first PCR was designed to amplify the target region of the ATP7b gene, 3 utr and 5 utr.

Primers were designed to generate amplicons ranging from 7 kb to 10 kb with overlapping amplicons.

We designed 10 pairs of primers dividing the entire gene into 10 amplicons. Each amplicon was designed to contain a tail that is necessary for barcoding. This part in red (table 2): tail that acts as a priming site for the “outer” primers.

To perform our PCR, we used a very specific taq polymerase designed for long amplicons. In our study, we used the LongAmp Hot Start Taq 2x Master mix.

Reagent	**Volume**
Template DNA	100 ng
Forward primer at 10 µM (840)	1 µl
Reverse primer at 10 µM (841)	1 µl
LongAmp Hot Start Taq 2x Master mix	12.5 µl
Nuclease-free water	Up to 25 µl
TOTAL	25 µl

table 1: PCR-1 protocol

ampl 1 f	TTTCTGTTGGTGCTGATATTGCTTTTCCAGGTGCGGAGTTCA
ampl1r	ACTTGCCTGTCGCTCTATCTTCCTCATGCAAAAGCTGGCTCC
ampl2 f	TTTCTGTTGGTGCTGATATTGCCTCCCTCCCTCTCAGGTCTC
ampl2 r	ACTTGCCTGTCGCTCTATCTTCCCCAGCCCTCCTTTCTGAAG
ampl3f	TTTCTGTTGGTGCTGATATTGCAGCAGTGTTACAGGCTGACC
ampl3r	ACTTGCCTGTCGCTCTATCTTCACATCAATCCTCCCTGCTGC
ampl4f	TTTCTGTTGGTGCTGATATTGCCCCTTCTCTAGCCCCTCCTT
ampl4r	ACTTGCCTGTCGCTCTATCTTCAGGCTTGAGAGCTAATGGGC
ampl5 f	TTTCTGTTGGTGCTGATATTGCCTGAGGCCTGGATTTGCTCA
ampl5 r	ACTTGCCTGTCGCTCTATCTTCACGTGGAGGATTCTGAGGGA
ampl6f	TTTCTGTTGGTGCTGATATTGCTGCATTTGGAGCGTGTACCT
ampl6r	ACTTGCCTGTCGCTCTATCTTCCGGAGAGAACACCTGGAACC
ampl7f	TTTCTGTTGGTGCTGATATTGCGGTGTTCTCTCCGTGTTGGT
ampl7r	ACTTGCCTGTCGCTCTATCTTCGAGGCCCCAAAGCTGAGAAT
ampl 8f	TTTCTGTTGGTGCTGATATTGCCCATGACCCCCGATTCCAAA
ampl8r	ACTTGCCTGTCGCTCTATCTTCGCACAGCAAAGGCCCTACTA
ampl9f	TTTCTGTTGGTGCTGATATTGCTTGTCTCTGTTCCTGCCGTC
ampl9r	ACTTGCCTGTCGCTCTATCTTCAGCTGAGGTGTCACTGAACG
ampl10f	TTTCTGTTGGTGCTGATATTGCTCATGGCTAAGGCACCCAAG
ampl10r	ACTTGCCTGTCGCTCTATCTTCAGGCCAAGGAACCCAAAACA

Table 2: Primer design for PCR

We amplified using the following program in a thermocycler: We optimized the PCR with the aim of having a significant annealing temperature for all amplicons. The extension time ranged from 8 to 10 mn depending on the length of the amplicon.

Cycle step	Temperature	Time	Number of cycles
Initial denaturation	94°C	30 secs	1
Denaturation	94°C	20 secs	35 (Denaturation, Annealing, Extension)
Annealing	59°C	30 secs	
Extension	65°C	8-10 min	
Final extension	65°C	7 min	1
Hold	4°C	∞	

table 3: PCR -1 CYCLING TEMPERATURE

1-1 PCR controls

The controls used for each PCR include the following:

- Positive controls, to ensure that PCR conditions are correct, represented by DNA from a normal healthy subject.
- Negative control to ensure the absence of contamination, in which there is no DNA.

1-2 Control of PCR products

PCR products were checked on a 2% agarose gel in the presence of a size marker to verify the size and specificity of the amplified product.

1-3 Purify the remaining DNA using Ampure beads 0.8X:

For better purification, we purified each amplicon separately.

We prepared AMPure XP beads for use (warmed up for approximately 5 minutes at room temperature) and resuspended them by vortexing.

The PCR products were transferred to a 1.5 ml DNA LoBind tube, and 0.8x AMPure beads were added. Mix by flicking the tube.

Incubate on a rotator mixer for 5' at RT

Spin down the sample (1000 × g for 15'') and pellet on a magnet. Keep the tube on the magnet and pipette off the supernatant. Keep on magnet and add 5x volume of freshly prepared 70% EtOH (with nuclease-free water)

Without disturbing the pellet. Remove the EtOH using a pipet.

Repeat the previous step.

Spin down and place the tube back on the magnet. Pipette off any residual EtOH and briefly allow to dry (15 to 30'' max).

Remove the tube from the magnet and resuspend in 12 µl of 10 mm Tris-HCl pH 8.0 with 50 mM NaCl. Incubate for 2 minutes at RT.

Pellet beads on the magnet until the eluate is clear and colourless (approximately 2 to 5 minutes).

Remove and retain the eluate (keep on ice).

1-4 Dose the purified PCR products using Nanodrop

After purifying the PCR products, we took 1.5 µl of the products and assessed the yield and quality of DNA using a nanodrop eader.

2- Barcoding - Long-range PCR2

In a 0.2 ml PCR tube, prepare the PCR mix (on ice) as follows: Use one different barcoding primer/sample.

Reagent	Volume
Template DNA	50 ng
Barcoding primer at 10 µM (from PBK004)	2 µl
LongAmp Hot Start Taq 2x Master mix	25 µl
Nuclease-free water	Up to 50 µl
TOTAL	50 µl

table 4: PCR-2 protocol

Cycle step	Temperature	Time	Number of cycles
Initial denaturation	94°C	30 secs	1
Denaturation	94°C	20 secs	15
Annealing	62°C	30 secs	
Extension	65°C	7 min	
Final extension	65°C	7 min	1
Hold	4°C	∞	

table 5: PCR-2 cycling protocol

Second purification and dosage is performed.

3- Nanopore library preparation using SQK-PBK004 and a FONGLE Flow cell

We diluted 100 fmol (as little as 50 fmol) of PCR product in 5 µl in 10 mM Tris-HCl pH 8.0 with 50 mM NaCl in a 1.5 ml DNA LoBind tube. At this step, we can pool the different amplicons if needed.

Add 0.5 µl RAP to the 5 µl amplified DNA library mix gently by pipetting.

Incubate the reaction for 5 minutes at RT and then store the library on ice until ready to load.

Then, proceed to the priming and loading of the Flongle cell.

4- Loading of the Flongle cell:

Sequencing buffer (SQB), loading beads (LB) and flush buffer (FB) were thawed from the Flongle Sequencing Expansion and Flush Tether (FLT) from the Flow Cell Priming Kit (EXP-FLP002) at room temperature.

The sequencing buffer (SQB), flush tether (FLT) and flush buffer (FB) tubes were mixed by vortexing and spun down at room temperature.

Place the Flongle adapter into the MinION Mk1B or one of the five GridION positions.

The adapter should sit evenly and flat on the MinION Mk1B or GridION platform. This ensures that the flow cell assembly is flat during the next stage.

Place the flow cell into the Flongle adapter, and press the flow cell down until you hear a click.

In a fresh 1.5 ml Eppendorf DNA LoBind tube, mix 117 µl of Flush Buffer (FB) with 3 µl of Flush Tether (FLT) and mix by pipetting.

Peel back the seal tab from the Flongle flow cell, up to a point where the sample port is exposed.

To prime your flow cell with the mix of flush buffer (FB) and flush tether (FLT) that was prepared earlier, ensure that there is no air gap in the sample port or the pipette tip. Place the P200 pipette tip inside the sample port and slowly dispense the priming fluid into the Flongle flow cell. To avoid flushing the flow cell too vigorously, load the priming mix by twisting the pipette plunger down.

Vortex the vial of loading beads (LB). Please note that the beads settle quickly, so immediately prepare the Sequencing Mix in a fresh 1.5 ml Eppendorf DNA LoBind tube for loading the Flongle, as follows:

Reagents	Volume
Sequencing Buffer (SQB)	15 µl
Loading Beads (LB), mixed immediately before use	10 µl
DNA library	5 µl
Total	30 µl

table 6: Sequencing mix preparation

To add the Sequencing Mix to the flow cell, ensure that there is no air gap in the sample port or the pipette tip. Place the P100 tip inside the sample port and slowly dispense the Sequencing Mix into the flow cell by twisting the pipette plunger down.

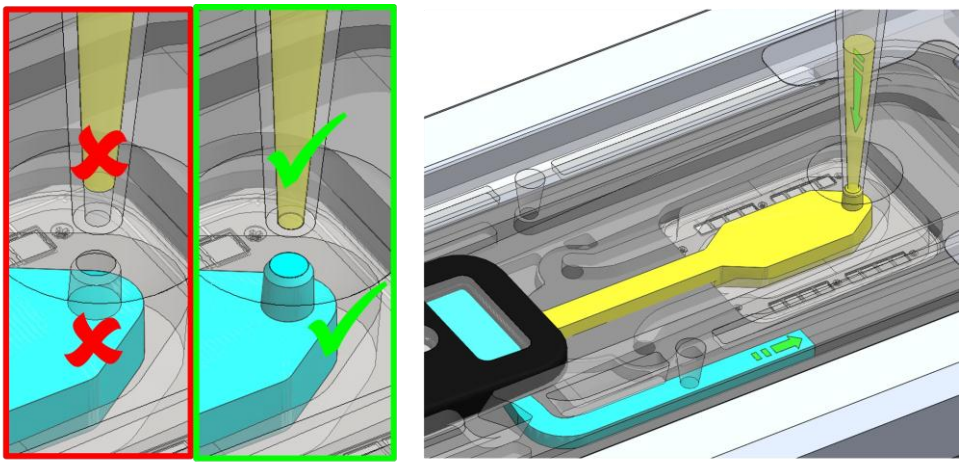

Figure 3: presentation of the mix loading in the nanopore flow cell

E- bioinformatic analysis

The Poretools toolkit was used to extract FASTQ files from FAST5 files.

Base-calling was performed using Guppy software from ONT (v6.0.6) with a super accuracy model (dna_r9.4.1_450 bps_sup.cfg).

Galaxy, a web-based platform for processing NGS data, was employed for variant analysis (https://usegalaxy.org/). Reads were aligned to the GRCh38 human reference genome using Minimap2 (v2.24) (Li, 2018) and visualized with the Integrative Genomics Viewer (IGV) browser. Variant calling on each partition of candidate positions was performed using Clair. The structural variations were called using Sniffles (version 2.0.3) (Sedlazeck et al., 2018).[1]

RESULTS

A- Observation:

We included 4 patients clinically diagnosed with Wilson disease in our study.

The table below represents the patients' clinical data.

	REPORTED CASES IN THE FAMILY	*AGE OF DIAGNOSIS*	*CONSANGUINITY*	*CLINICAL SYMPTOMS*
P1	NA	17 yd	yes	Hepatocellular Insufficiency, slight cytolysis
P2	NA	35 yd	yes	Cholestatic icterus
P3	Affeted brother	18yd	yes	Hepatocellular Insufficiency, jaundice
P4	Affected sister	15yd	yes	Cholestatic icterus, homogeneous hepatomegaly

All patients underwent a ceruloplasmin test and were clinically diagnosed on the basis of the clinical and biochemical findings (low serum ceruloplasmin concentration, low serum concentration of total copper, and increased urinary copper excretion).

B- DNA extraction results:

DNA extraction was performed using a kit. The absorbance ratio

DO260/DO280, was between 1.8 and 2, leading to the conclusion that these extracts are not contaminated by proteins and can be qualified as pure.

C- Results of ATP7B gene amplification by PCR:

PCR of the ATP7B gene amplicons was performed in each of the 4 patients. followed by agarose gel electrophoresis to verify amplifications after each PCR (including barcoding PCR).

The results found in the electrophoretic profiles of all patients showed that the product amplified for sequencing did not contain any contaminants, and the sizes of the amplicons were concurdant, so sequencing could proceed.

D- Sequencing results:

The results of sequencing are shown in the figures below, the IGV image with BAM files. We used the canonical transcript NM_000053.4 as a reference transcript.

After analysis of the results, it was possible to identify three different types of single nucleotide variation that were classified according to ACMG recommendations. In addition to these mutations, we were able due to the large size of the sequenced fragment to identify an intronic deletion in two patients, an intronic

duplication in one patient, and, most interestingly, a 3 kb deletion in exon 5 in patient 2. The patient also harbored a homozygous (NM_000053.4(ATP7B): c.3694del) SNV frameshift classified as propably pathogenic.

For this inversion, a pcr confirmation will be performed. In the study of the ATP7B gene, we identified many polymorphisms that were not taken into account in the data analysis.

The table below summarizes the sequencing results of our patient representing the identified disease or probably disease-causing variations.

	Amplicons	resultats	clinvar
P2	1-10	Amplicon 7 : Dup-Inversion 3 kb c.3419del	Not reported
P1	1-10	Ampl 10 : exon 17 c.3694 A>C + c.3694del	Reported Not reported
P3	1-10	Ampl 10: exon 16 c.3419del	Not reported
P4	1-10	Ampl 10: exon 16 c.3419del	Not reported

table 7: Sequencing results of our cases

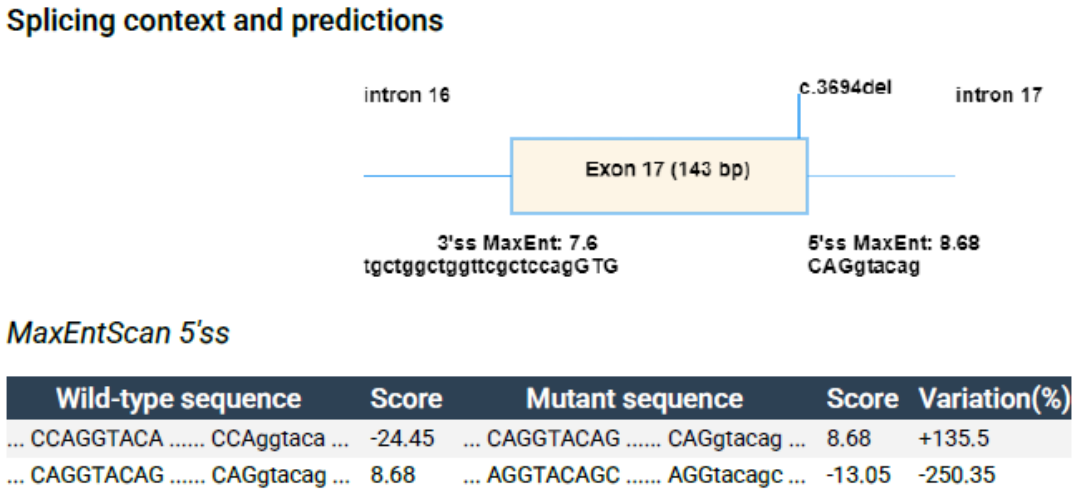

Figure 4: Schematic presentation of localization of the variation with splice score for deletion c.3694 i patient 1 amplicon 10-exon16

For patient 1, we identified two pathogenic variations:

- The variation detected in exon 17 (c.3694del) was responsible for a single base deletion with a frameshift. p.(Thr1232ProfsTer98). The variation is absent from the population database, mainly in the gnomAD exome, genome, and V3, and has never been reported in ClinVar and Clingen. This variant was predicted to be deleterious by SIFT and MutationTaster and is responsible for a premature stop codon. It is located in a region with a high frequency of pathogenic missense variations. Therefore, the variation was classified as likely pathogenic (PVS1-PM2).
- A second missense variation c.3694A>C has been reported in many other Wilson disease patients and is classified as pathogenic (figure 10).

Figure 8 shows amplicon 10 of patient 1, where we identified a misalignment region in intron 16. The misalignment is related to an intronic inversion.

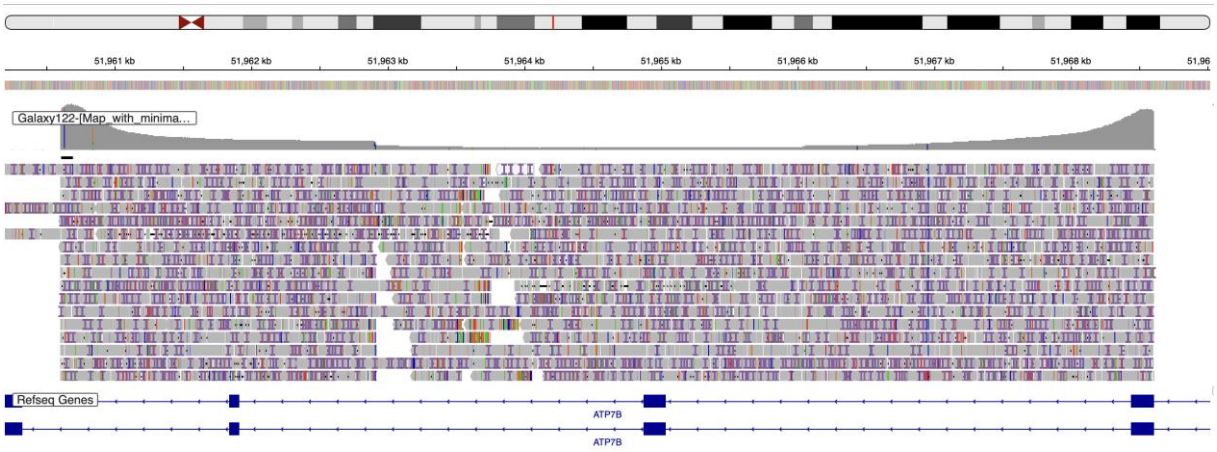

Figure 5: IGV presentation of amplicon 7

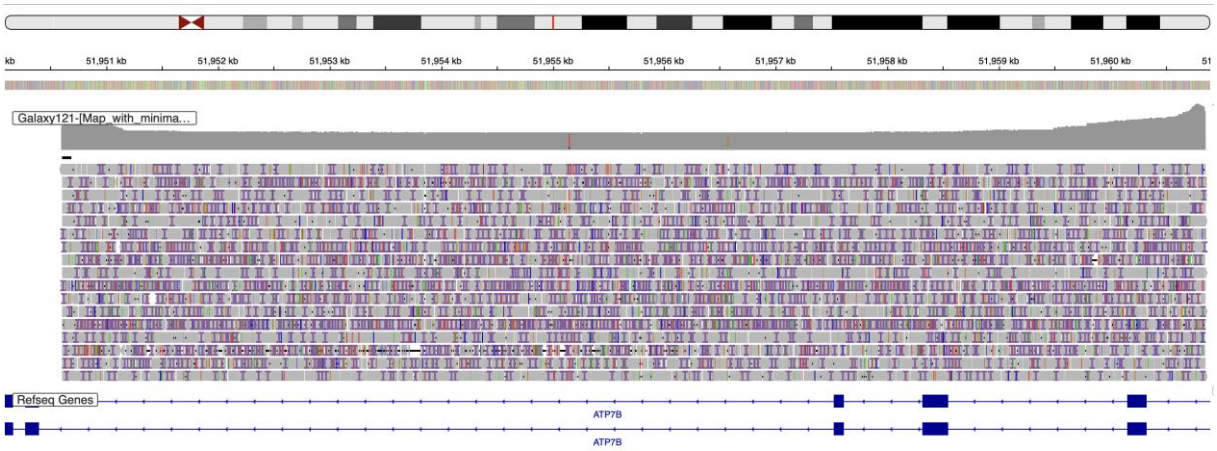

Figure 6: IGV presentation of amplicon 8

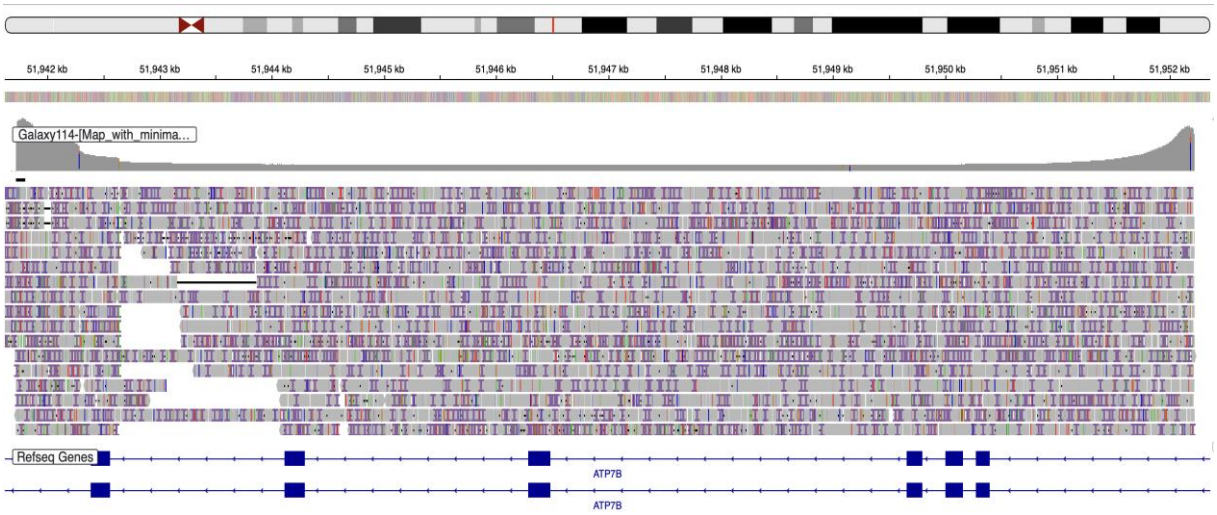

Figure 7: IGV presentation of amplicon 9

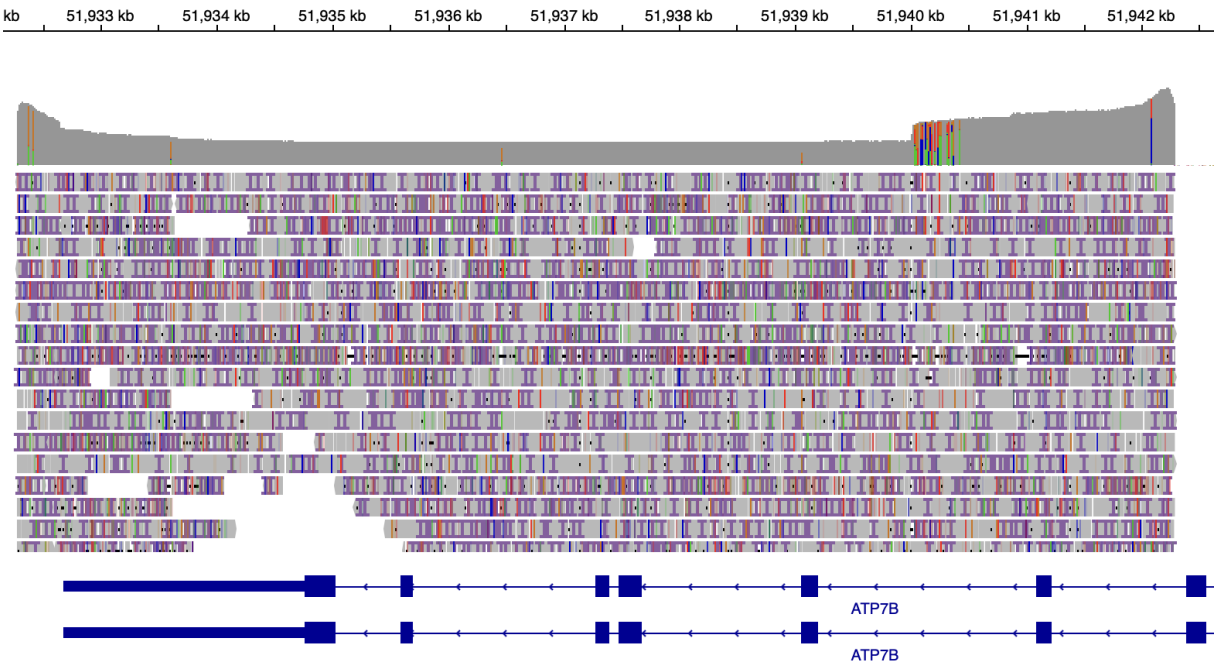

Figure 8: IGV presentation of amplicon 10 for patient 1: Intronic inversion

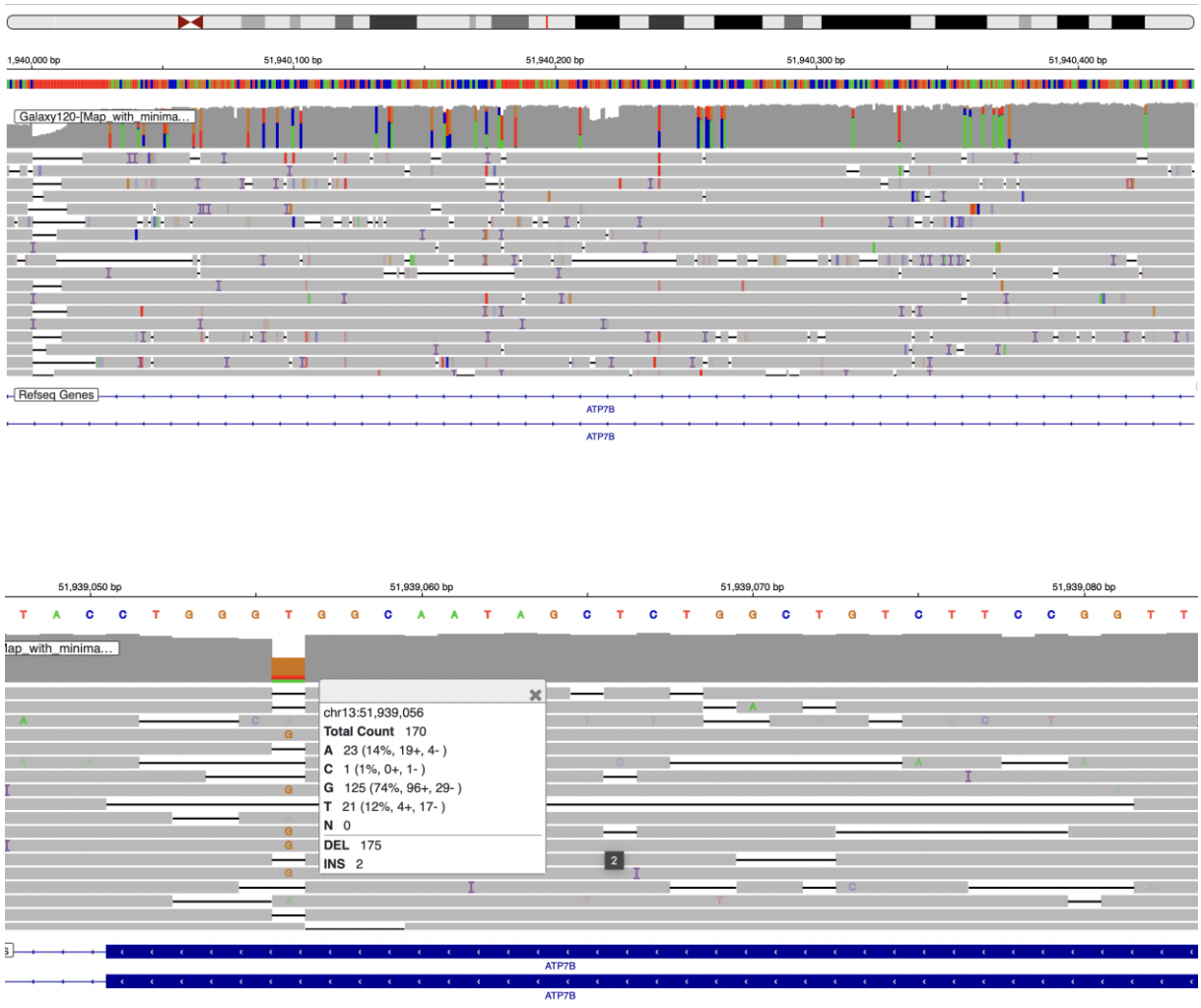

Figure 9: IGV/PATIENT 1 have 2 variations at the same nucleotidique position. The presence of a deletion in 50%

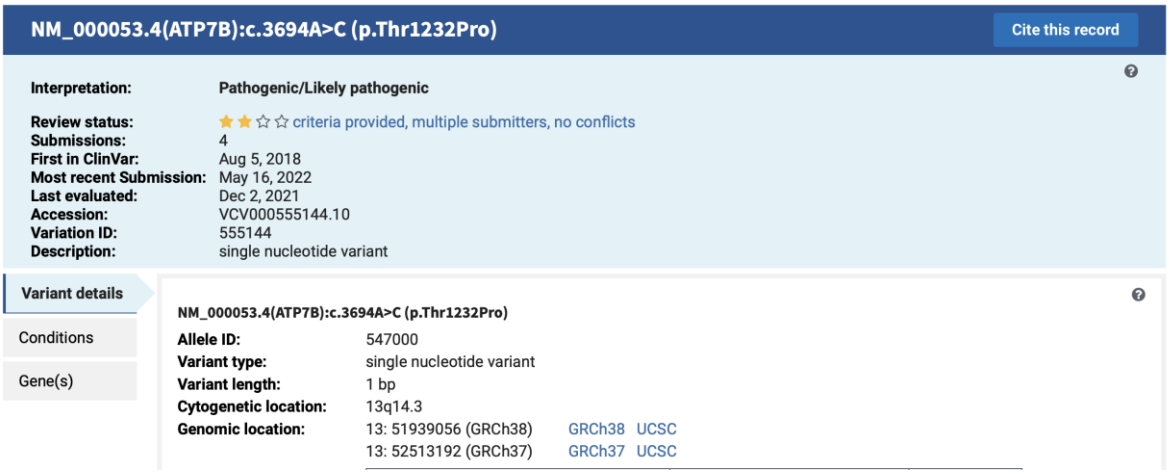

Figure 10: ClinVar reported pathogenic mutation

For P2-P3-P4, we identified a single base homozygous deletion in exon 16. The variation was not reported in gnomad v3 and was never described in clinvar or in clingen. It is classified as likely pathogenic by VARSOME. (PS1-PM1). The variation is responsible for the NP_000044.2:p.(Val1140AlafsTer8) frameshift.

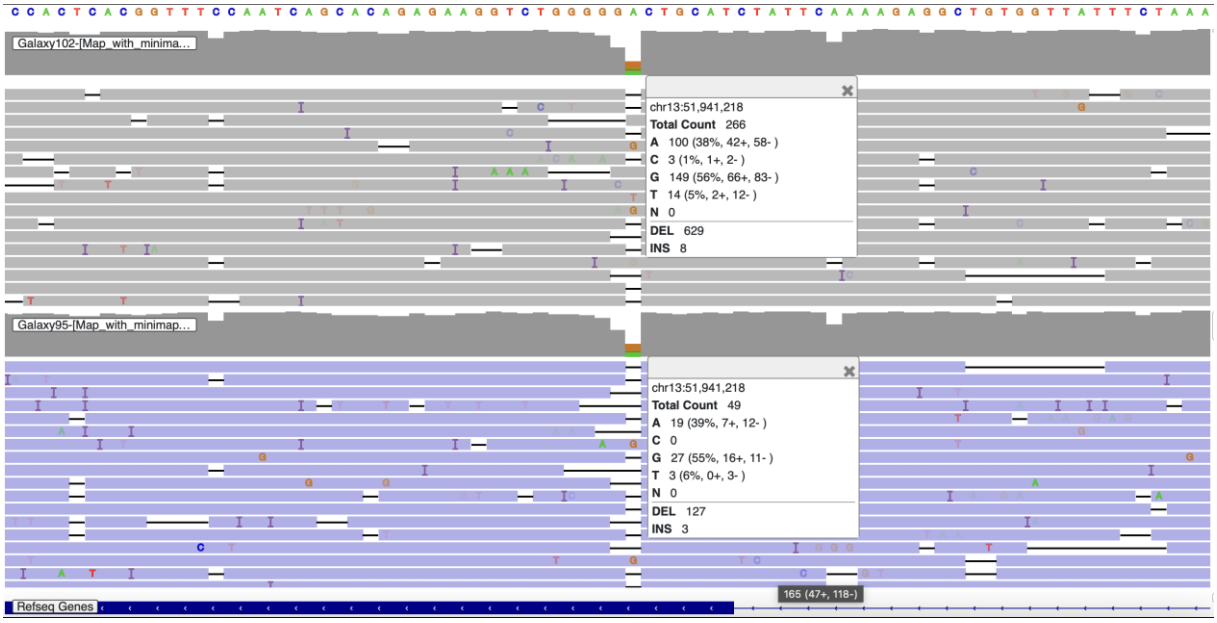

Figure 11: Patient 3 and 4 IGV presentation of a single deletion. We noticed the presence of an interfiring G base related to alignment quality.

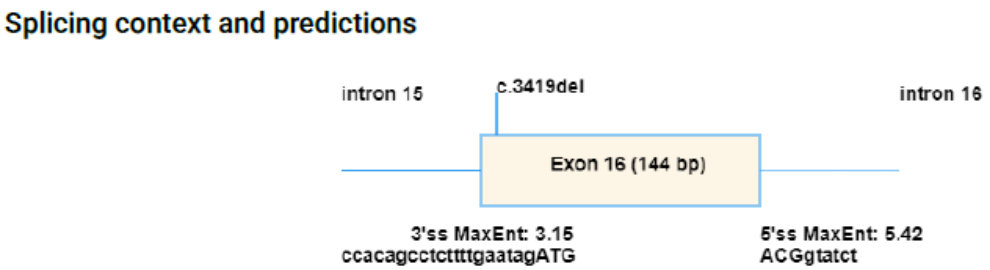

Figure 12: Chematic presentation of localization of the variation in exon 16 with splice score

In addition to the reported SNV, we also identified a duplication-inversion with deletion of exon 5 associated with an intronic deletion in intron 5 for patient 2, and the latter was reported in 2 patients in our study population.

Confirmation of these variants by specific PCR is recommended.

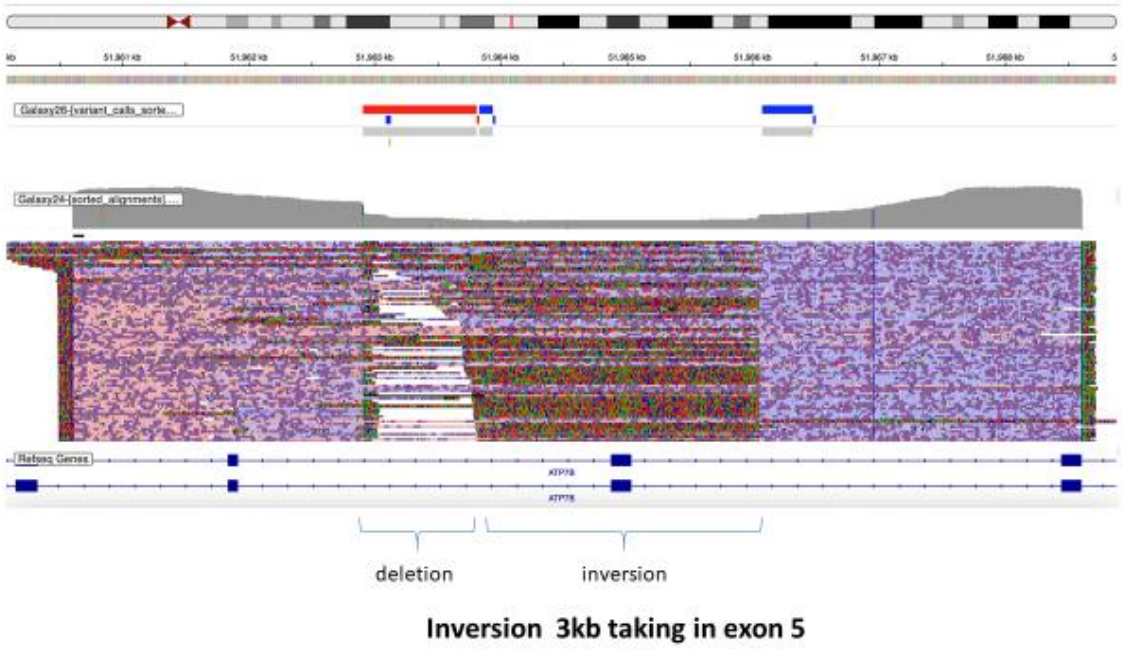

Figure 13: IGV presentation of amplicon 7 for patient 2 showing an intronic deletion (in red) and 3 kb duplication-inversion with deletion of exon 5.

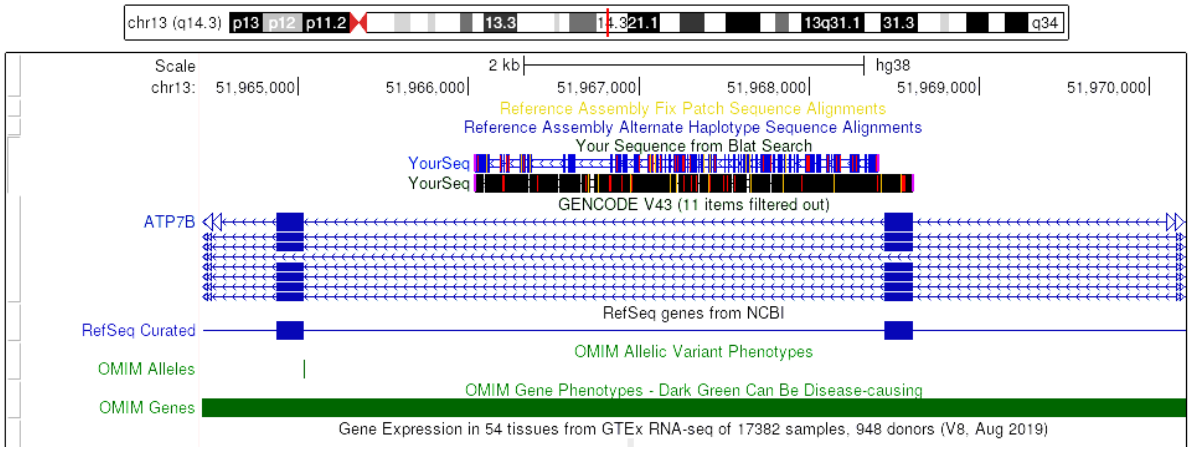

Figure 14: ucsc presentation of amplicon 7 for patient 2 showing duplicated-inversed region of the gene: in black the 5'-3' fragment, in blue the same fragment in 3'-5' (inverted).

DISCUSSION

A- Next generation sequencing:

Next-generation sequencing (NGS), also known as massively parallel sequencing or high-throughput sequencing, is a technology enabling the simultaneous sequencing of millions of DNA or RNA sequences. [16]

It offers several advantages over traditional sequencing methods, including higher throughput with the ability to multiplex samples, greater sensitivity in variant detection, faster turnaround time and lower cost. NGS represents a genuine technological revolution in sequencing, following Sanger sequencing.

Second-generation sequencing methods can be grouped into two broad categories: hybridization sequencing and sequencing by synthesis (SBS). SBS methods represent the advanced version of Sanger sequencing, without the dideoxy terminators, in combination with repeated cycles of synthesis, imaging and methods for incorporating additional nucleotides. At first glance, these new methods may appear expensive, but reactions are run in parallel, often in nanoliter, picoliter or zeptoliter volumes, in small chambers, so the cost per base pair sequenced is lower. Ongoing refinements and miniaturization further reduce costs. [17]

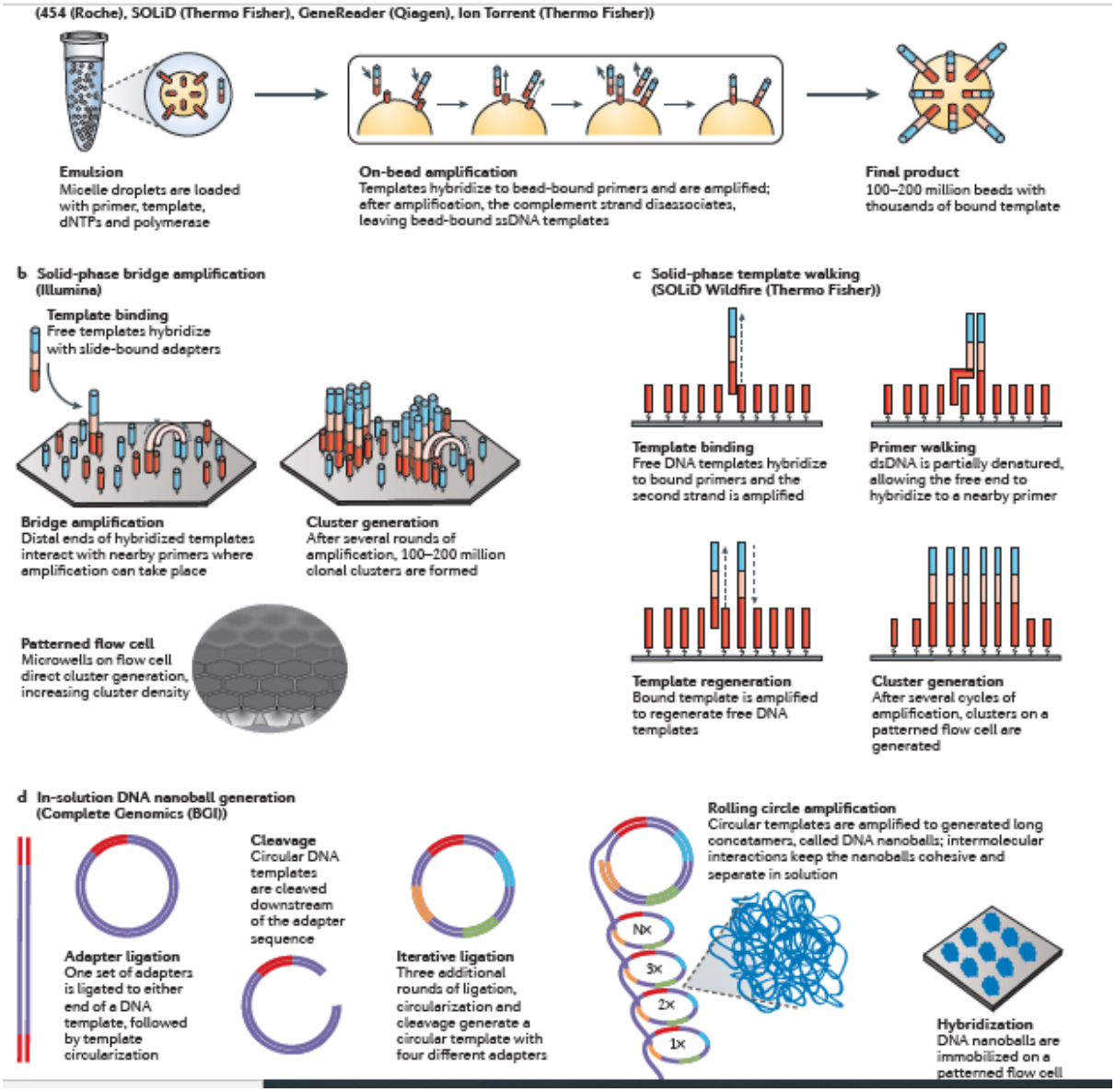

Figure 14: Different NGS second-generation techniques17

B- Nanopore

1- Introduction:

The idea of nanopore sequencing was proposed by Deamer and Branton and independently by Church. The concept is based on the detection of an electric current as nucleotides pass through a tiny channel. In 1993, Deamer, Branton and Kasiannowicz used α-hemolysin (α-HL), a toxic poreforming protein secreted by Staphylococcus aureus, to detect the passage of DNA through the α-HL nanopore according to Song et al. In 1996, their results of DNA translocation through the α-HL nanopore were published (Kasianowicz et al., 1996).[2]

A nanopore is simply a small hole on the order of 1 nm in internal diameter that consists of certain cellular transmembrane proteins. Nanopore sequencing works on the principle of minute variations in electrical current through the nanopore, which is immersed in a conductive fluid to which a voltage is applied when a moving nucleotide (or DNA strand) passes through it. Each nucleotide of the DNA molecule, as it passes through the nanopore, obstructs the nanopore to a different and characteristic degree, and the amount of current change is characteristic for each different nucleotide. DNA can be forced through the hole one base at a time, similar to a thread through the eye of a needle. The change in current can be read directly, and the sequence of the passing DNA can be determined by detecting variations in the generated current specific to the base passed.[3]

Although the technology has not been used routinely for sequencing, according to the scientists who developed it, it can sequence the entire 5.4-kilobase genome of the virus in a single continuous read. However, the initial target is to provide 100-kilobase reads. This is still much longer than the fragments processed by other technologies. The initial system will be equipped with a function to read

DNA at a rate of several hundred kilobases per second. [4]

The category of nanopore used can be divided into two parts, solid-state nanopores and biological nanopores:

1-1 Biological nanopores:

Poreforming proteins can form nanopores; typically, the protein is mushroom-shaped with a hollow core running through it. Phi 29 connector, MspA porin and α-hemolysin are some poreforming proteins. α-Hemolysin (α-HL, also known as a-toxin) is the first and most widely used biological nanopore, holding enormous value in the field of DNA sequencing. α-HL is an exotoxin secreted by the bacterium Staphylococcus aureus, a human pathogen. This mushroom-shaped heptamer is a 232.4 kDa transmembrane channel consisting of a 3.6 nm diameter cap and a 2.6 nm diameter transmembrane β-barrel. In laboratory work, a nanopore is inserted into a lipid bilayer film, followed by manipulations of single-channel proteins and measurements.[5]

Biological nanopores are derived from natural protein molecules or from genetically engineered artificial nanopores. However, biological nanopores are fragile and have characteristics such as a short lifespan, intrinsic instability and a strict requirement for a specific environment, making them unsuitable for long-term biosensor operation.[6]

1-2 Solid-state nanopores:

Solid-state nanopores are essentially created in films of silicon compounds, with silicon nitride being the most commonly used. A number of techniques are used to fabricate solid-state nanopores, including "ion-beam unfolding and sculpting" and "electron-beam fabrication". The diameter of solid-state nanopores can be controlled within a precise range from subnanometers to hundreds of nanometers,

depending on experimental parameters. In general, compared with lipid membranes, dielectric materials such as SiN, which offer superior chemical and thermal stability, are used in solid-state nanopores. However, this stability depends on the parameters that lead to the formation of these pores. Regarding graphene-based nanopores, although their chemical and thermal stability has not been demonstrated, they possess unique chemical properties that are highly advantageous compared with their biological complements. Solid-state nanopores have paved the way for a wide range of research, particularly in the fields of DNA sequencing, protein interaction identification, molecular transport and environmental identification.[7 3]

Si3N4 and SiO2 nanopores are the most widely used, and their manufacture is compatible with industrial integrated circuit processes based on complementary metal oxide semiconductors. These nanopores can be ion-etched into independent Si3N4 and SiO2 films using an argon ion beam or an electron beam.

2- Nanopore sequencing technology:

ONT is a single-molecule sequencing technology based on nanopores. The first prototype, MinION, was released in 2014. The updated platform, PromethION, was released in 2015 with improved throughput. Two versions of PromethION, named ProtheION 24 and 48, incorporate 24 and 48 flow tanks, respectively. Thanks to the explosion in the number of flow tanks compared with MinION, the PromethION system was able to generate up to 7.6 TB of data, whereas MinION could only generate 50 GB in 72 hours of operation. [18]

There are three forms of nanopore sequencing: 1D, 2D and 1D2. The 1D kit uses a nanopore where only one strand of DNA is sequenced. The 2D kit was first used in ONT. A hairpin structure was used at one end of the double-stranded DNA to connect two strands. Once sequencing of one strand is complete, sequencing of

the other strand begins immediately. In this way, it is equivalent to repeating the sequencing twice, which can be used for base correction.[19] 1D2 is similar to 2D but does not require hairpins to physically link the two DNA strands together. The reaction system for nanopore sequencing is implemented in a flow cell, in which two compartments filled with ionic solution have been separated by membranes containing either 2048 (MinION) or 12,000 (PromethION) nanopores. The nanoporebased gene sequencing process can be divided into three parts: library preparation, sequencing and basecalling.[5]

2-1 Sequencing process:

The figure below shows the schematic sequencing process. The DNA carrier to be sequenced is mixed with copies of the processive enzyme. As the DNA-protein complex approaches the nanopore, the enzyme binds to a single-stranded leader at the end of the double-stranded DNA template, opens the double strand and passes a single strand through the nanopore. A single molecule with high specificity can interfere with the current as the open long DNA strand passes through the nanopore one base at a time. These current signals can be used to determine the base type. [8]

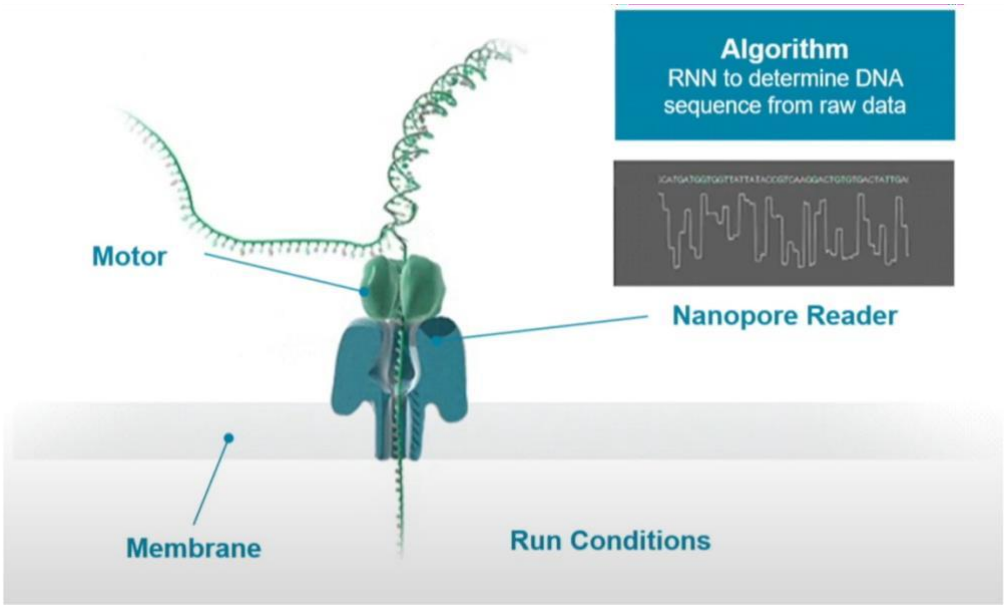

Figure 15: Schematic presentation of a single nanopore[5]

2-2 Basecalling:

During the base reading process, due to the difference in charge and structure of the nucleotides as they cross the nanopore, the measured current causes small perturbations. These electrical signals can then be translated into DNA sequences using deep learning algorithms. However, the readout signals are noisy and random because they originate from several molecules in the nanopores, which is difficult for the flip-flop.[9]

In addition, the strength of the hole is determined by the bases of several nucleotides located at the narrowest point of the hole. As the final step in interpreting the entire DNA sequence, data analysis using deep learning is a challenge, requiring efficient algorithms and a large amount of data for computational training.

2-3 Advantages:

The nanopore system has many potential advantages. In fact, it could enable real-time sequencing of single molecules at low cost; it should also make it possible to read very long DNA molecules in a single read. Nanopore technology should enable sequencing at a very low cost, from $25 to $40 per gigabase of sequence. This means that sequencing a human genome with standard 30-fold coverage would only cost a few thousand dollars. As this is a very fast sequencing process, it could be a future sequencing choice for almost any application.[10 4]

2-4 Limitations:

At this stage, the technology presents a number of problems that need to be addressedmainly accuracy. however nanopore had commercialized a new cells R10 which thery promised that they will increase the sequencing accuracy . besides that nanopore propose a 2D sequencing technology, allowing the sequencing in consecutive time the forwrs and the reverse brin strand , by that error related to

sequencing could be corrected at the base calling step

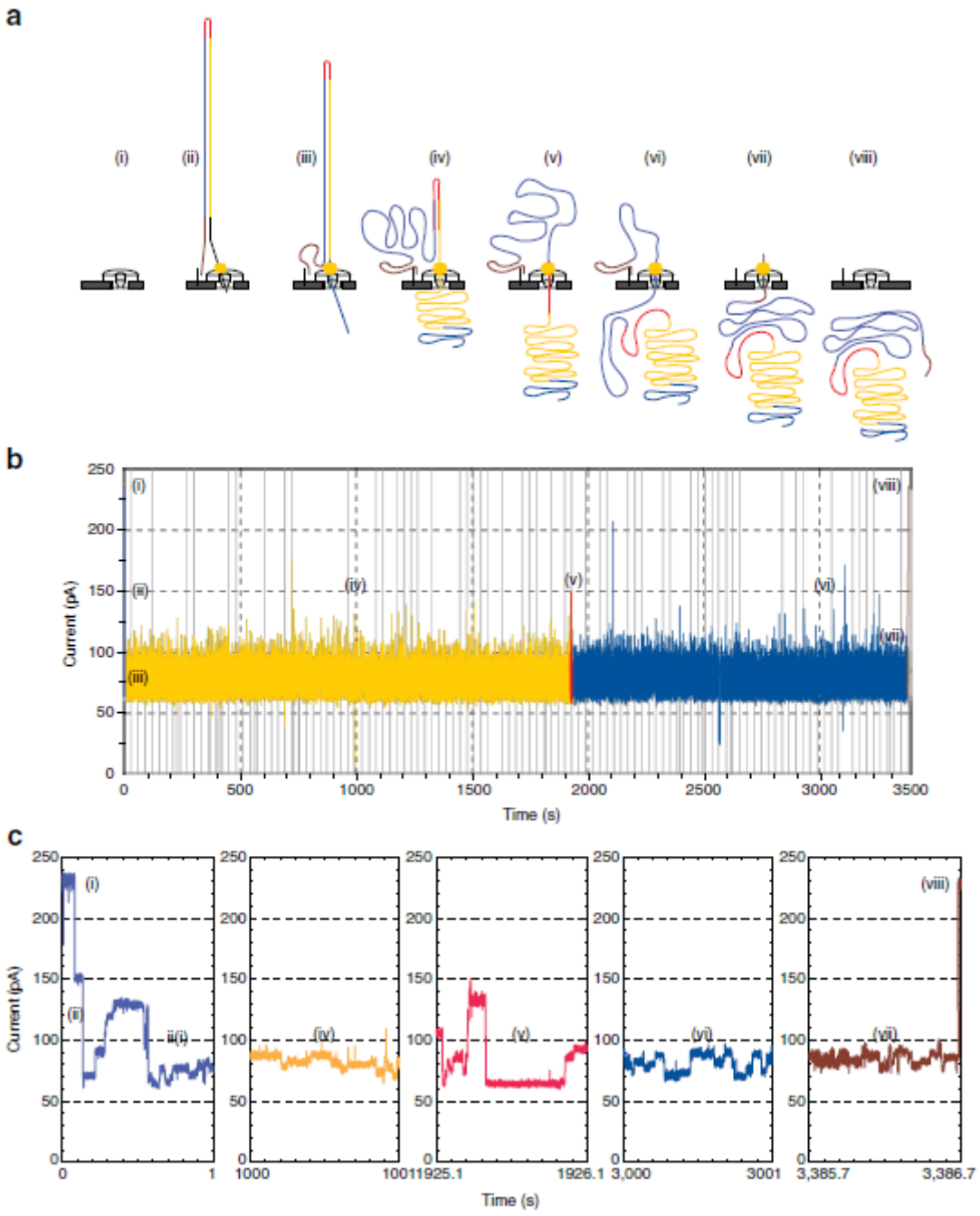

Figure 16: Data for a 2D read of a full-length λ phage dsDNA from the MinION nanopore sequencer[8]

Data for a 2D read of a full-length λ phage dsDNA from the MinION nanopore sequencer. a Steps in DNA translocation through the nanopore: (i) open channel; (ii) dsDNA with lead adaptor (blue), bound molecular motor (orange) and hairpin adaptor (red) is captured by the nanopore; capture is followed by translocation of the (iii) lead adaptor, (iv) template strand (gold), (v) hairpin adaptor, (vi) complement strand (dark blue) and (vii) trailing adaptor (brown); and (viii) status returns to open channel. b Raw current trace for the passage of the single 48-kb λ dsDNA construct through the nanopore. Regions of the trace corresponding to steps i–viii are labelled. (c) Expanded time and current scale for raw current traces corresponding to steps i–viii. Each adaptor generates a unique current signal used to aid base calling[8]

C- Wilson disease:

1- Review of wilson disease :

Wilson disease is a disorder of copper metabolism that can present with hepatic, neurologic, or psychiatric disturbances, or a combination of these findings, in individuals ranging from age three years to older than 50 years; symptoms vary among and within families. Liver disease includes recurrent jaundice, simple acute self-limited hepatitis-like illness, autoimmune-type hepatitis, fulminant hepatic failure, or chronic liver disease. Neurologic presentations include movement disorders (tremors, poor coordination, loss of fine-motor control, chorea, choreoathetosis) or rigid dystonia (mask-like facies, rigidity, gait disturbance, pseudobulbar involvement). Psychiatric disturbance includes depression, neurotic behaviors, disorganization of personality, and, occasionally, intellectual deterioration. Kayser-Fleischer rings, frequently present, result from copper deposition in Descemet's membrane of the cornea and reflect a high degree of copper storage in the body.[11]

The prevalence of Wilson disease is estimated at one in 30,000 in most populations, with a corresponding carrier frequency in the general population of one in 90 [Sandahl et al 2020].

The understanding of the phenotypic spectrum has further expanded through the widespread use of molecular genetic testing, which has confirmed the diagnosis in individuals with atypical clinical and biochemical findings.[11]

Because Wilson disease is a treatable condition, it is appropriate to offer predictive testing to asymptomatic at-risk adults and children. The optimal time for determination of genetic risk and discussion of the availability of prenatal/preimplantation genetic testing is before pregnancy. • It is appropriate to offer genetic counselling (including discussion of potential risks to offspring and reproductive options) to young adults who are affected, are carriers, or are at risk of being carriers. • Carrier testing for the reproductive partners of affected individuals and known carriers should be considered, particularly if consanguinity is likely and/or if both partners are of the same ethnic background.[12 11]

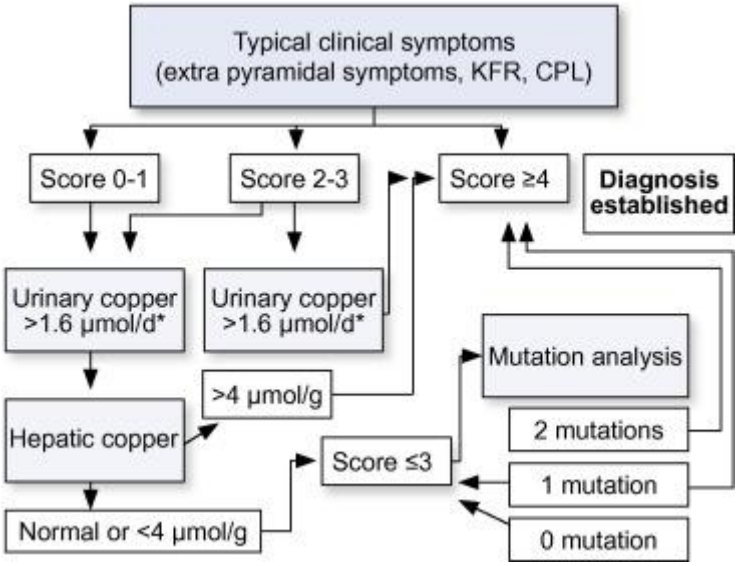

Figure 17: Diagnostic algorithms for Wilson's disease based on the Leipzig score[13]

2- Genetics of wilson disease in moroccan population :

To our knowledge, a genetic study has never been conducted in Moroccan patients diagnosed with Wilson disease. This finding would mainly be attached to the limitations accessibility for genetic testing in Morocco and the high price of testing disease. In addition to the high reported frequency of composite heterozygosity for Wilson disease and the absence of a mutational hot spot, all of these factors contributed to the test implementation delay. [14]

However, the situation is changing with the emergence of very encouraging pharmaceutical therapies and the possibility of transplantation in many medical centers.

All of these recent developments encourage us to review the necessity of genetic testing, and among many techniques, we propose that nanopore testing might be the most cost-effective, in parallele with NGS second generation testing.

In a recent study conducted in the CHU Mohamed 6 in Marrakech, they sequenced the 21 exons of the ATP7b gene and flanking regions in 12 patients clinically diagnosed with Wilson disease.[15]

They identified six different homozygous mutations of the ATP7B gene in ten patients, while in two patients, no mutation was detected in the promoter and exonic regions.

In different populations, individuals with no mutations have been reported in different proportions. Failure to detect any mutations may be explained by unknown mutations that may be located outside of the exons and flanking regions, such as the promoter, introns, or other DNA control regions, thus highlighting the limitations in genetic testing for WD diagnosis. [15]

We were able to use long-range PCR with nanopore sequencing to cover the entire gene in contrast to the reported study.

On the other hand, the missense mutation p. T1232P was detected in two patients (P7, P10) but only in one patient in our study, who had a heterozygous mutation at the same position, with the second variation being a deletion (c.3694delA).

The missense substitution in exon 17 involves the ATP loop of the protein. Both patients were diagnosed at a young age (7 and 8 years), and both showed similar hepatic symptoms. This mutation was previously described. Advanced protein functional studies indicate that this missense mutation is expected to disrupt protein function. The same effect would be expected for the deletion variation, and a functional study will be necessary.[16]

In their reported study, they also identified many other variations, all of which have been reported in other populations, with the exception of the missense

variation p.G836E. It can probably be considered a rare mutation in the worldwide population, since only one case was previously reported in Italy in a patient of Moroccan origin. This variation occurs in exon 10. Some studies have suggested that the A-domain regulates the release of the ion at the luminal site by triggering movements of transmembrane helices. The TGE motif in the A-domain is responsible for dephosphorylation by removing γ-phosphate from the DKTGT motif in the P-domain.

The study also used MLPA for negative patients to detect structural duplications or deletions. No structural variations were reported; however, we clearly reported the presence of deletions and a 3 kb inversion involving exon 5. It is worth mentioning that MLPA cannot detect inversions. All of these conditions in the utility, price and rapidity of nanopore sequencing in combination with the detection of single nucleotide variation and structural variation would enhance the use of this target sequencing in some diseases.

mutation	LAFHAL et al[15] (12 cases)	Our study (4 cases)
c.3694A>C	2 cases	1 case
c.3694delA	Not reported	1 case
c.2507G>A	1 case	Not reported
c.865C>T	2 cases	Not reported
c.51+4A>T	2 cases	Not reported
c.1746dup	2 cases	Not reported
c.3310T>C	1 case	Not reported
c.3419del	No reported	3 cases

table 8: comparison of our sequencing results with the study of lafhal et al

CONCLUSION

The analysis of sequencing data allowed us to identify two variants for one patient: a missense variant p. Thr1232Pro (rs568009639), already reported as pathogenic, and a single base deletion that has never been described at the same position (c.3694 delA) and one structural variation, an inversion, and presume the presence of a founder effect of a single frameshift deletion that was identified in 3 of our population study.

These preliminary data highlight the economic and biological importance of sequencing the whole ATP7b gene to identify variants responsible for Wilsons disease. This is particularly important in diseases with a high frequency of heterozygote composite in genetically poorly explored populations. It is anticipated that with this method, we may find more variants in Wilson disease patients.

The advantages of using nanopore sequencing were clearly elucidated in target sequencing, as it is a cost-effective method to diagnose SNVs and CNVs.

In the present study of the Department of Genetics in Marrakech, 2 patients were negative for Sanger sequencing. We further aim to collaborate with them and use nanopore sequencing for these 2 patients to explore intronique regions and resolve, if possible, the molecular diagnosis.

REFERENCES

ABSTRACT

Background – Wilson's disease is an autosomal recessive disorder related to a dysfunction of copper transport expressed by the gene ATP7B. It can present with hepatic, neurologic, or psychiatric disturbances or a combination of these findings. In most instances, the diagnosis is established by biochemical findings. The presence of Kayser-Fleisher rings in the eyes is almost pathognomonic. As a recessive disease, it requires the presence of biallelic pathogenic variants in ATP7B. The diagnosis by itself, in many cases, is not complicated; however, genetic counselling requires molecular confirmation. According to the literature, deletion and duplication are rare in Wilsons disease. On the other hand, many studies report a high frequency of composite heterozygotes. To our knowledge, no molecular study has been conducted in the Moroccan population using nanopore sequencing.

Aim - Our study aimed to test the feasibility of long-range PCR and nanopore sequencing of the whole ATP7B gene. In particular, our aims were 1) to identify the mutational profile of Wilson's disease in the Moroccan population, 2) to demonstrate the interest and feasibility of applying long-read sequencing methods using Oxford Nanopore Technology (ONT) to identify variants in the ATP7B gene, and 3) to test the economical and biological impact of nanopore sequencing in populations with limited access to genetic testing.

Methods and Results - We designed primer pairs to amplify an average of 10 kb-long PCR fragments covering the entire ATP7B gene. Following PCR amplifications, barcoded libraries were prepared and sequenced on ONT flow cells using the Mk1C device. In this pilot study, we included 4 patients clinically diagnosed with Wilson's disease from the Department of Gastroenterology (CHU Fez).

The analysis of sequencing data allowed us to identify two variants for one patient: a missense variant p. Thr1232Pro (rs568009639), already reported as pathogenic, and a single base deletion that has never been described at the same position (c.3694 delA).

One structural variation, an inversion, and presumed the presence of a founder effect of a single frameshift deletion that was identified in 3 of our poopulation studies.

Conclusion - These preliminary data highlight the economic and biological importance of sequencing the whole ATP7b gene to identify variants responsible for Wilsons disease. This is particularly important in diseases with a high frequency of heterozygote composite in genetically poorly explored populations. It is anticipated that with this method, we may find more variants in Wilson disease patients.

We further aim to collaborate with the Department of Genetics in Marrakech and try to use nanopore sequencing for patients in whom Sanger sequencing failed to resolve the diagnosis.