Arab American University – Palestine

Faculty of Graduate Studies

Identification of Novel Disease-Causing Gene Mutations Detected by Whole-Exome Sequencing and Whole-Genome Sequencing in Palestinian Families with Rare Diseases.

By

Shurouq Abdullah Yaghnam Alshaer

Supervisor

Dr. Nouar Mohammad Abdel Muti Qutob

Co-supervisor

Dr. Reham Abdelateef Mustafa Nazzal

This thesis was submitted in partial fulfillment of the requirements for the Master`s degree in

Molecular Genetics and Genetic Toxicology

March/2022

Identification of novel disease-causing gene mutations detected by whole-exome sequencing and whole-genome sequencing in Palestinian families with rare diseases.

By

Shurouq Abdullah Yaghnam Alshaer

This thesis was defended successfully on March 3rd, 2022 and approved by:

Committee members

1. Dr. Nouar Mohammad Abdel Muti Qutob
2. Dr. Reham Abdelateef Mustfa Nazzal
3. Dr. Siba Mohammad Yousef Shanak
4. Dr. Fawwaz Tarik Taher Awwad

DECLARATION

Hereby, I declare that this thesis has been composed by my own research work as a master's student in the molecular genetics and genetic toxicology program at the Arab American University of Palestine, and has not been previously submitted for any other degree or professional qualification, and every effort has been made to appropriately indicate, mention, and acknowledge others contributions (for data, arguments, words, and ideas).

Shurouq Abdullah Alshaer

Date: March 3rd, 2022

Signature:

DEDICATION

I’d like to dedicate this work to my husband, Malek, who has always believed in my capabilities and has supported me throughout my study journey. To my son, Zaid, who taught me to love without conditions. I also dedicate the work to my parents, Amal and Abdullah, who taught me to put pain before pleasure. I finally dedicate the work to my sincere sisters, Salam, Ahd and Aya, who have believed in my abilities to succeed and strive.

ACKNOWLEDGEMENT

First and foremost, I would like to sincerely thank my supervisors, Dr. Nouar Qutob and Dr. Reham Nazzal. I am deeply appreciative of their care and assistance, which has been immensely helpful to me in completing my thesis. Their expertise, knowledge, assiduity, and motivation have helped me to develop my research skills. I would also like to acknowledge the collaborations and contributions made by Exeter Genomics Laboratory. Thanks are also due to Husam Sallam and Rua' Thawabteh, the technical assistants in the molecular genetics laboratory of the AAUP, for providing professional support when needed in the lab. I would finally want to thank the patients and family members and staff who participated in the study.

LIST OF ABBREVIATIONS

ACMG: American College of Medical Genetics

BS: Bloom Syndrome

CF: Cystic Fibrosis

cHSP: Complex form of Hereditary Spastic Paraplegia

MECP2: Methyl-CpG Binding Protein 2

MPD: Microcephaly Primordial Dwarfism

RTT: Rett Syndrome

TOP3A or *TopIIIα*: DNA Topoisomerase 3-alpha

WES: Whole-Exome Sequencing

WGS: Whole-Genome Sequencing

ABSTRACT

Presence of isolated populations and communities in Palestine contributes to the rise of rare genetic diseases with founder pathogenic variants. These founder variants were found to be not common in the Palestinians genomic data or international controls. Therefore, it's essential to study the genetic makeup of such communities and build a map for the founder variants in a national screening theme. In this study, we used whole-exome and whole-genome sequencing approaches in order to decipher and segregate the molecular causes of rare genetic disorders among three Palestinian families living in Palestinian villages. Our study included a complex form of hereditary spastic paraplegia (cHSP), cystic fibrosis (CF), microcephaly primordial dwarfism (MPD), and Rett syndrome (RTT). The segregation analysis showed the presence of the two recessive disorders, cHSP and CF, in several genotypes in the same family. cHSP was found to be associated with novel biallelic *PTPN23* variant, p. (P1572T fs*12)], confirming that PTPN23 is a causal gene for hereditary spastic paraplegia. MPD was found to be caused by a novel frameshift-deletion biallelic *TOP3A* variant, [p.(Trp950Glyfs*58)]. Our sequencing results confirmed that the variants are inherited in an autosomal recessive manner. Furthermore, our segregation analysis supported the presence of a de *novo* inframe deletion variant in the *MECP2* gene, [p.(Ser411del)], that may have caused RTT in one identified case. Our study showed the genomic power to deliver a diagnosis and inform management plans in large highly consanguineous and isolated communities that have been without a diagnosis for several years. It also revealed the presence of different modes of inheritance for the rare diseases that can arise in these consanguineous settings and not exclusively the expected autosomal recessive mode.

TABLE OF CONTENTS

PAGE

LIST OF TABLES

TABLE **PAGE**

LIST OF FIGURES

FIGURE **PAGE**

CHAPTER 1: INTRODUCTION

Rare diseases have become a special concern in the genomics field due to their commonality among certain populations especially the small isolated ones. Advancement in molecular genetics techniques holds potential for discovering the molecular basis of these diseases. They also ensure translating these discoveries into improved clinical health through better diagnosis and management.

1.1 Rare Diseases: Prevalence, Inheritance, and Challenges

Rare diseases (often referred to as "orphan diseases") are generally defined as conditions that affect a very limited number of individuals compared to the overall population (Pogue et al. 2018). To be precise, diseases with a prevalence of less than 5 in 10,000 people are rare (Dawkins et al. 2018). The prevalence threshold for which a disease is considered rare varies across countries. For example, in the USA, the disease which affects less than 200,000 patients is considered as a rare disease, while in Japan the disease which affects less than 50,000 is a rare disease (Mukherjee 2019). Rare diseases are many, geographically scattered, and have a heterogeneous nature (Nguengang Wakap et al. 2020). The Orphanet portal (the portal for rare diseases and orphan drugs) (http://www.orpha.net) lists more than six thousand unique rare diseases in its database (Nov, 2020). Individually, these diseases are rare, but collectively, they impact 4-8% of the population (Bick et al. 2019). According to the Global Genes Project (https://globalgenes.org/rare-facts/), 400 million people globally suffer from a rare disease (Aug, 2021). Most of these diseases are genetically inherited and many are solely pediatric-onset (Nguengang Wakap et al. 2020). The fact that most of rare diseases are genetically inherited makes the diagnoses difficult and expensive, and limits treatment options (Pogue et al. 2018). Challenges stemming from rare

diseases' low prevalence including lack of knowledge and expertise, in addition to chronic, incurable, and life-threatening nature, shall raise awareness to prioritize research and follow up of emerging rare diseases (Nguengang Wakap et al. 2020). However, research investigating rare diseases started with attempts to detect genetic variants and mutations that are causative of specific diseases. The accurate molecular genetic investigation of rare diseases can be critical to support disease classification, risk prediction among relatives, give access to clinical screening, inform of prevention strategies, and probably point out the most convenient treatment (Ellard et al. 2020). There is a serious need for these investigations in Arab communities as they are at a high risk for rare diseases (Al-Gazali, Hamamy, and Al-Arrayad 2006).

1.2 Rare Diseases in Arab Communities

Current data suggest that in Arab countries, rare genetic disorders are more prevalent than in developed countries (Al-Gazali, Hamamy, and Al-Arrayad 2006). The main factors which may contribute to the high prevalence of genetically inherited disorders include: (i) High consanguinity rates, childbearing at older ages, and high birth rates. (ii) Failure to take public health measures to control and avoid congenital and genetically defined disorders (Tadmouri 2006). Consanguineous marriage draws significant interest as a causative factor in the prevalence of rare genetic disorders , in particular, among the rural Arab communities (Zlotogora et al. 2006). First cousin marriages are particularly common in many Arab countries and comprise nearly one quarter of all marriages (Al-Gazali and Hamamy 2014). Autosomal recessive disorders are reported as the most frequent disorders due to consanguinity (Shalev 2019). Up to 45 % of the Palestinian Arab couples who live in Israel are consanguineous. More than 20% of these spouses are related as first cousins (Zlotogora et al. 2006). The incidence of malformations / genetic diseases is 6,8% in children born to consanguineous Arab spouses in north Israel (Shalev 2019). In Palestine, the high rate of

consanguinity does not only result in relatively high prevalence of many recessive disorders, it also results in the presence of two different autosomal recessive disorders within the same sibship in one family, and this must be taken into consideration (Zlotogora 1997). Particular homozygosity in dominant genes is one of the other consequences of the high rate consanguinity. It involves the pathogenesis of multifactorial disorders, and is probably one of the main causes of an increased malformation rate among the population being identified (Zlotogora 1997). Data from the Israeli National Genetic Database (INGD) show that many of the variants for autosomal recessive disorders, reported among Israeli Arabs are novel. Most of them were identified in a single community only (Zlotogora 2019). One of the consequences of this is that many of the variants are classified as variants of uncertain significance especially for missense variants.

With this remarkable increase in rare genetic diseases among Palestinians, it was decided to set up a community-based, translational genomics project, called "Stories of Hope, Stories from Palestine" headed by Dr.Reham Nazzal. This project focuses on characterizing the clinical and molecular basis of rare diseases in Palestinian communities and translating research findings into improved diagnostic provision and clinical care for affected patients, families, and their communities. The preliminary data collected in this project indicated a promising potential for novel gene discoveries and led to the identification of biomolecular disease mechanisms while providing impacting diagnostic and clinical management benefits for participating families. This project seeks to empower collaborations with renowned international scientists and institutions to develop a solid scientific platform to enable a long-term patient and family recruitment program for continued discoveries of neurodevelopmental molecules and biological pathways after the identification of diseases-causative variants using whole-genome sequencing (WGS) and whole-exome sequencing (WES) techniques.

1.3 Emergence of WGS and WES in the Diagnosis of Rare Diseases

Fifteen years after the human genome was first assembled and decoded, which took over a decade and costed around $3 billion, next-generation sequencing (NGS) techniques have emerged bringing personalized genomic medicine into reality (Muzzey, Evans, and Lieber 2015). NGS or massively parallel sequencing is a technique of simultaneously sequencing millions of DNA fragments that can analyze multiple gene regions simultaneously at once. Due to this featured ability, NGS has been quickly implemented in the clinical laboratory to test both germline and somatic mutations (Yohe and Thyagarajan 2017). NGS technology demonstrated accuracy in the detection of variations between the genome of a patient and the reference genome. Recently, NGS-based tests are increasingly utilized to detect the presence of rare diseases, evaluate both the likelihood and occurrence of cancer, and check for aneuploidies among fetuses (Muzzey, Evans, and Lieber 2015). The most frequently used NGS techniques for research and diagnosis purposes are WES and selected genes sequencing (gene-panels) (Klebe, Stevanin, and Depienne 2015). Over the last 10 years, whole-exome sequencing confirmed its potency in both clinical and research fields as a tool used for revealing novel disease-causing genes for many rare diseases that were troublesome to previous approaches (Boycott et al. 2019). WES has helped in the diagnosis of 30%–50% of rare Mendelian diseases (Frésard and Montgomery 2018). WES is currently used as a diagnostic tool in specific settings such as mapping different autosomal recessive disorders and detecting novel causative genes and pathogenic mutations in consanguineous families (Harripaul et al. 2017, M. et al. 2016). However, WES has the potential to miss important types and regions of disease-causing genomic variation, such as structural variants, indels, and intronic single nucleotide variants (SNVs) (Lionel et al. 2018). WGS, however, has the potential to capture almost every genomic variation in an accurate, rapid, and an unbiased manner with a single test in

patients with complex presentations and different clinical indications (Bick et al. 2019), therefore it exhibits diagnostic superiority over WES (Lionel et al. 2018). Furthermore, 9% of the first exons of reference sequenced (RefSeq) genes were not found to be covered by WES. On the other hand, WGS gave a more complete coverage of exons compared to WES (Bick et al. 2019). WGS is regarded as a first-tier test in the diagnosis of acutely ill inpatient infants as the time-to-diagnosis using rapid WGS is significantly less than WES (Bick et al. 2019) and this results in improved outcomes and net healthcare savings (Farnaes et al. 2018). Indications of using WGS rather than WES are many, and of them (i) Clinical diagnosis of a disease that has been linked to many genes, (ii) Clinical symptoms that are insufficient to make a clinical diagnosis on their own but are known to be related with a variety of genetic disorders, (iii) Atypical clinical abnormalities not previously explained in the aspect of the molecular diagnosis, and (iv) Early onset of disorder typically seen in adulthood (Bick et al. 2019). In all cases, it is imperative to correctly analyze WGS and WES data using different tools to make an accurate diagnosis.

1.3.1 Analysis of WGS and WES Data

WGS and WES data of rare diseases are analyzed using different bioinformatics tools and databases. One of the known databases is genome Aggregation Database (gnomAD) which is a searchable framework for genes and variants in different populations (Prokop et al. 2018). It presently encompasses about 17 million variants obtained by exome sequencing of 125, 748 individuals and about 262 million variants detected by genome sequencing of 15,708 individuals who participated in different disease-specific and population genetic studies [https://gnomad.broadinstitute.org/] (Ellard et al. 2020). Among the other databases, gene-phenotype catalog (OMIM) [www.omim.org/], ClinVar (database of archived human variations and phenotypes supported by evidence)[https://www.ncbi.nlm.nih.gov/clinvar/]

,Orphanet portal, and HGMD [http://www.hgmd.cf.ac.uk/ac/index.php] are used to give higher yield of diagnosis (Boycott et al. 2019). There are also many upgraded algorithms and software tools that are used to detect, filter, and classify variants in the generated WGS and WES data. Such algorithms allow for the detection and classification of the more likely deleterious variants (Ulintz, Wu, and Gates 2019).

1.3.2 Variants Classification Using ACMG Guidelines

The increasing use of NGS approaches in genetic testing in the clinical laboratories has been accompanied by challenges and errors in sequence variants interpretation. Such errors summoned establishing united interpretative categories of sequence variants and interpretation algorithm (Richards et al. 2015). The American College of Medical Genetics (ACMG) contributed in solving this issue by establishing and improving guidelines for interpretation and classification of sequence variants (Nykamp et al. 2017). ACMG guidelines depend on integration of the clinical (For example; patient phenotype, population data) and biological scientific data (For example; protein functional and structural data, computational data, and allelic data) together while interpreting WES/WGS variants in order to make a classification (Richards et al. 2015). These guidelines draw a skeleton to classify variants as “pathogenic”, “likely pathogenic”, “uncertain significance”, “likely benign” or “benign” based on a set of series criteria with levels of evidence stated as very strong, strong, moderate, or supporting **(Table 1)** (Ellard et al. 2020). ACMG guidelines also confirmed that all of the classifications have to be done with respect to a disease and inheritance pattern (Ellard et al. 2020).

Table 1. ACMG criteria for classifying pathogenic and likely pathogenic variants (adapted from Richards et al. 2015).

Evidence	Category	Description
Very strong	PVS1	Null variant (nonsense, frameshift, canonical ±1 or 2 splice sites, initiation codon, single or multiexon deletion) in a gene where LOF is a known mechanism of disease
Strong	PS1	Same amino acid change as a previously established pathogenic variant regardless of nucleotide change
	PS2	De novo (both maternity and paternity confirmed) in a patient with the disease and no family history
	PS3	Well-established in vitro or in vivo functional studies supportive of a damaging effect on the gene or gene product
	PS4	The prevalence of the variant in affected individuals is significantly increased compared with the prevalence in controls
Moderate	PM1	Located in a mutational hot spot and/or critical and well-established functional domain (e.g., active site of an enzyme) without benign variation
	PM2	Absent from controls (or at extremely low frequency if recessive) in Exome Sequencing Project, 1000 Genomes Project, or Exome Aggregation Consortium
	PM3	For recessive disorders, detected in *trans* with a pathogenic variant
	PM4	Protein length changes as a result of in-frame deletions/insertions in a nonrepeat region or stop-loss variants
	PM5	Novel missense change at an amino acid residue where a different missense change determined to be pathogenic has been seen before
	PM6	Assumed de novo, but without confirmation of paternity and maternity
Supporting	PP1	Cosegregation with disease in multiple affected family members in a gene definitively known to cause the disease
	PP2	Missense variant in a gene that has a low rate of benign missense variation and in which missense variants are a common mechanism of disease
	PP3	Multiple lines of computational evidence support a deleterious effect on the gene or gene product (conservation, evolutionary, splicing impact, etc.)
	PP4	Patient's phenotype or family history is highly specific for a disease with a single genetic etiology
	PP5	Reputable source recently reports variant as pathogenic, but the evidence is not available to the laboratory to perform an independent evaluation

1.4 Equity and Diversity in Genomics Field

Despite the rapid advancements in the genomics field, it has so far failed to equitably involve diverse and under-represented populations in human genomic studies that have developed the evidence base and translated findings to better health outcomes (Jooma et al. 2019). Several actions must be taken in order to achieve health equity in genomic medicine application, to assure: (i) supporting the representation of under-represented groups, (ii) ensuring an equal access to genomic services, and (iii) investing in research that enhances infrastructure outside of academic medical centers (Jooma et al. 2019). Ensuring the equity and diversity will contribute to raising the accuracy, acceptability, and usability of genomic data in clinical care (Hindorff et al. 2018). Genomic data obtained from studies on isolated populations (population isolates) are helpful in simplifying diagnosis and discovering of disease causing variants of different complex diseases.

1.5 Bottleneck Events and Genetic Diseases

Bottleneck events such as migration and isolation result in a drastic reduction in population size, resulting in the presence of the founder effect (Major et al. 2021). Genetic drift and inbreeding are common to occur in small populations (Cheptou et al. 2017), and lead to the presence of specific alleles in higher prevalence (Major et al. 2021). When population isolation is combined with geographic isolation, genetic drift reduces heterogeneity through randomly driving some alleles to fixation while driving others to extinction (Kristiansson, Naukkarinen, and Peltonen 2008). Founder alleles may increase the likelihood of more prevalent complex disorders (Major et al. 2021). The enrichment of numerous recessive disorders in founder populations is one example of such drift and the founder effect (Kristiansson, Naukkarinen, and Peltonen 2008). The identified genetic variants among isolated populations are largely replicated among individuals which aided

in discovering new pathways behind rare disease processes (Kristiansson, Naukkarinen, and Peltonen 2008).

1.6 Hereditary Spastic Paraplegia Forms, Symptoms, and Inheritance

Hereditary Spastic Paraplegia (HSP) is a syndromic hereditary group of clinically and genetically heterogeneous neurologic disorders with major signs in lower limb spasticity or weakness (Blackstone 2018, de Souza et al. 2017) in its pure form, while other neurological signs appear in its complex form (Toft et al. 2019). Usually, the progressive spastic paraplegia in the HSP pure form is accompanied with hyperreflexia, ankle clonus, Babinski sign, and increased tone in the lower limbs. Whereas spastic paraplegia in the complex form is related to other neurological or extra-neurological symptoms, including intellectual disability, ataxia, deafness, optic atrophy, epileptic seizures, dysmorphic features (microcephaly, macrocephaly, short stature, facial dimorphisms, etc.) (de Souza et al. 2017, Klebe, Stevanin, and Depienne 2015,Ortega and Rosemberg 2019). HSP can be inherited in all of the Mendelian and non Mendelian inheritance patterns including autosomal recessive, dominant, X-linked, or maternal pattern traits (de Souza et al. 2017). The pure form of HSP is more common in families with the autosomal pattern of inheritance. In contrast, the HSP complex form is frequently found in autosomal recessive or X-linked HSP (Faber et al. 2017). The major pathological hallmark of HSP is retrograde distal axonopathy of the longest descending motor fibers of the corticospinal tract and posterior columns (de Souza et al. 2017). HSP prevalence is estimated to range from 1.3 to 9.6 in 100,000 people (Ortega and Rosemberg 2019). HSP age of symptoms onset ranges widely from infancy to late adulthood (Polymeris et al. 2016).

1.6.1 The Heritable Component of Hereditary Spastic Paraplegia

The presence of founder mutations in certain isolated populations and levels of consanguine marriages can increase autosomal recessive HSP prevalence in certain geographic areas (Klebe, Stevanin, and Depienne 2015). NGS techniques have greatly increased the number of known HSP related genes and both the HSP phenotype and genotype have been expanded (Novarino et al. 2014). More than 70 loci and 50 genes have been reported to be associated with HSP and referred as spastic paraplegia genes (SPGs) (Polymeris et al. 2016). The heritable component of HSP is complex and includes over 70 identified genetic subtypes (de Souza et al. 2017). HSP genetic subtypes were described and sequentially numerated, according to the gene discovery order (SPG1, SPG2, SPG3, etc.) (Faber et al. 2017). HSP can clinically and genetically overlap with other neurodegenerative disease including Charcot-Marie-Tooth (CMT), Spastic Ataxias (SPAX), Hypomyelinating-Leukodystrophy Disorders (HLD), Amyotrophic Lateral Sclerosis (ALS), and others among them (Souza et al. 2017). HSP is diagnosed based on certain clinical, technical, and genetic findings including: First, the presence of the HSP clinical symptoms (non-progressive or slowly progressive bilateral lower limb spasticity and weakness). Second, the neurological examination of the corticospinal tract involvement represented by spasticity, hyperreflexia, and other responses. Third, family history shows similar neurological disorders, with taking into account the probability of sporadic cases occurring and the incomplete genetic penetrance. Fourth, a clearly defined molecular genetic test revealing a previously reported gene related to the HSP or a novel mutation in non-disease-related genes that involves in molecular mechanisms that can lead to corticospinal tract impairment (Souza et al. 2017). The autosomal recessive HSP includes more than 40 subtypes. SPG11 is the primary prototype of thin-corpus callosum autosomal recessive HSP (up to 40% among all cases). It is caused by compound heterozygous or homozygous

mutations in the *SPG11* gene (located at 15q21.1). *SPG11* gene produce spatacsin protein which involved in neuronal axonal development, function, and intracellular cargo trafficking. SPG15 subtype is the second most frequent autosomal recessive HSP. It is juvenile or adult onset neurodegenerative disorder and caused by compound heterozygous and homozygous mutations in the *ZFYVE26* gene (14q24.1) producing the spastizin protein, concerned with initiation of lysosome tubulation, autophagic lysosome reformation, spinal motor neuron axon outgrowth, and control of secretory vesicle maturation. SPG14 subtype is autosomal recessive arising from homozygous mutations in the *SPG14* (3q27-q28), identified in a consanguineous Italian family. A gradually progressing spastic paraplegia with pes cavus, moderate mental retardation and distal motor neuropathy dominates the clinical picture of SPG14. SPG18 subtype is considered as a syndrome with motor dysfunction, intellectual disability, and joint contractures autosomal recessive. It is an early childhood complex HSP that derives from *ERLIN2* (8p11.23) homozygous mutations. Extreme manifestations of marked psychomotor retardation, pes cavus, multiple joint contractures and chronic progressive spastic paraplegia are identified in consanguineous Arab families; the remaining subtypes were reviewed by (de Souza et al., 2017).

1.6.2 *PTPN23* Gene Structure, Expression, and Function

PTPN23 (also referred to as HD-PTP) gene encodes an enzyme called protein-tyrosine phosphatase, non-receptor type 23 (https://omim.org/). *PTPN23* gene located on the short arm of chromosome 3 (3p21.3) between WI-11814 and WI-18411 markers, particularly in the putative tumor suppressing region, and contain at minimum 1,636 amino acids (Toyooka et al. 2000). *PTPN23* consists of 25 exons and 24 introns (Xianghu et al. 2001).The domain structure of this gene composed of a tyrosine phosphatase domain, His domain, two SH3-binding motifs, C-terminal PEST motif, two proline-rich region, and N-terminal domain resemble the BRO1 domain

of the yeast which implicated in the mitogen-activated protein kinase signaling pathway **(Figure 1)** (Xianghu et al. 2001).

Figure (1): Schematic representation of human *PTPN23* protein. Human *PTPN23* protein consist of 1636 amino acids, with conserved domains indicated as blocks, including the BRO1_like domain at the N-terminus, ALIX binding domain, His domain and inactive Protein Tyrosine Phosphatase (PTP) C-terminal domain.

PTPN23 is ubiquitously expressed in all cell types including neurons. Studying of its primary structure elucidated that *PTPN23* doesn't exhibit activity of dephosphorylation on phospho-tyrosine residues. Instead, it maintains the phosphorylation patterns of tyrosine of the substrate proteins (Smigiel et al. 2018). *PTPN23* has been involved in several cellular functions. Through its interaction with the endosomal sorting complex required for transport I (ESCRT-I complex), *PTPN23* plays an important role in sorting of the endosomal cargo into multivesicular bodies (MVBs) (Doyotte et al. 2008). *PTPN23* also serves as a tumor suppressor gene, where studies have proven that loss of *PTPN23* resulted in boosted activation of pro-oncogenic signaling pathways (van der Lely et al. 2019). Of the other cellular functions, *PTPN23* have confirmed roles in ciliogenesis, negative regulation of Ras-mediated mitogenic activity, and function regulation of the survival motor neuron complex (SMN) in the U-rich small nuclear ribonucleoproteins (UsnRNPs) assembly. UsnRNPs complicate with other proteins to form spliceosomes which mediate pre-mRNA splicing (Smigiel et al. 2018) .

1.6.3 Dysfunctional *PTPN23* in Human Neurodevelopmental Diseases

Many studies reported recessive variants of *PTPN23* as a molecular cause of human neurodevelopmental diseases, particularly, in the consanguineous families. However, most of these studies are only screening and not functional studies (Smigiel et al. 2018). In a cohort study carried out by Alazami et al. in 2015 to discover novel candidate genes in neurogenetic disorders, homozygosity of the missense variant p.(Arg1332Leu) in the *PTPN23* was described in a child of consanguineous parents and suffering of epilepsy, global developmental delay, as well as brain atrophy. In 2017, Trujillano et al. also reported another bialleic missense variant p.(Met302Val) in a proband diagnosed with global developmental delay with regression and spasticity, seizures, microcephaly, and brain atrophy. Later, in 2018 Smigiel et al. described a proband of non-consanguineous parents suffering from severe developmental delay, cortical blindness, epilepsy, hypomyelination and brain atrophy. The proband was found to have two novel compound heterozygous *PTPN23* variants; p.(Asn634Lys) and p. (Leu992Tyr fs*168). In this study, *PTPN23* dysfunction was confirmed by quantification of SMN accumulations in Cajal bodies and proband's fibroblasts change in the SMN activity.

1.7 Cystic Fibrosis and *CFTR* Gene

Cystic fibrosis (CF) is a life-threatening autosomal recessive genetic disease and one of the world's most common inherited fatal disorders (Elborn 2016, Fernandez Fernandez et al. 2018) . This complex multiorgan disease is induced by a mutation in the chloride-conducting transmembrane channel gene called the cystic fibrosis transmembrane conductance regulator (*CFTR*) (Elborn 2016). *CFTR* gene was mapped in 1985 and found to be on the long arm of chromosome 7 (7q31.2). *CFTR* gene with its 27 exons encodes a single polypeptide chain of 1480 amino acids

length and of 168 kilo Dalton (KDa) molecular weight (Fernandez Fernandez et al. 2018). The structure of the CFTR protein revealed that it belongs to the ATP-binding cassette (ABC) transporter family proteins (Meng et al. 2017). The *CFTR* protein is implanted in the epithelial cells apical membrane around the body and consists of the same ABC transporter domains: two membrane-spanning domains and two nucleotide binding domains. What makes the *CFTR* different from the other ABC transporters is the unique regulatory domain ('R' region) (Meng et al. 2017,Welsh and Smith 1993). *CFTR* was found to be expressed in many exocrine tissues such as the airways, liver, lung, pancreas, intestines and sweat gland duct (Fernandez Fernandez et al. 2018). The *CFTR* protein mainly functions as an ion channel that secrets chloride to the mucus producing cells and hydrate the airway mucus to be thin (Rafeeq and Murad 2017). *CFTR* also regulate the activity of other membrane proteins like the epithelial sodium channel and it was reported that there is a relation between *CFTR* mutations and its ability to restrict salt absorption by the epithelial sodium channel (Stutts et al. 1995). *CFTR* dysfunction leads to mucus retention and chronic infection followed by local inflammation of the airway that is lungs damaging (Elborn 2016). Failure of CFTR function also affect the gastrointestinal tract (GIT) where the mucous plugs obstruct the pancreatic canaliculi and gallbladder duct; thus preventing enzyme and bile flow into the duodenum. This indeed causes malabsorption and digestive disturbances. Loss of extra salts in sweat causes an imbalance of minerals in blood; what contributes to malnutrition, arrhythmias, fatigue, exhaustion, heatstroke, and death in some cases (Rafeeq and Murad 2017). Of the CF diagnosis regard, the individual is diagnosed if has a clinical presentation of cystic fibrosis and has evidence of *CFTR* function failure. The hierarchy of *CFTR* function tests is often as follows: First, sweat chloride test, then the genetic analysis of *CFTR* and then *CFTR* physiologic tests (Farrell et al. 2017). To date, over 2000 mutations in the *CFTR* gene are estimated to cause

cystic fibrosis as listed in the cystic fibrosis mutations database (http://www.genet.sickkids.on.ca/) (Nov, 2020). The most common mutation is the deletion of phenylalanine at codon 508 (F508del) and it affects approximately 85% of CF patients worldwide (Fernandez Fernandez et al. 2018). Most of the *CFTR* mutations are missense substitutions, but other mutations like nonsense mutations, splicing, frameshifts, inframe insertions and deletions were also identified (Elborn 2016). The incidence and distribution of *CFTR* mutations vary in different populations (Shahin, Mehaney, and El-Falaki 2016).

1.7.1 *CFTR* Mutations in Palestine

In Palestine, there is still lack of adequate studies to set reliable data of *CFTR* mutations frequency and nature which in turn harper the molecular diagnosis (Essawi et al. 2015). Determining of the frequency and nature of mutations in a specific population is necessary to find sufficient and cost-effective molecular diagnosis test. Of the reported mutations among Palestinian patients in different countries, F508del, W1282X, 3120+1Kbdel8.6Kb, N1303K, and G85E were the most prevalent (Siryani et al. 2015). In a cohort study conducted by Siryani et al. on CF patients from different families residing in southern/ central Palestine to study the distribution of *CFTR* mutations, c.1393-1G>A was the most common one to be occurring in Hebron and Bethlehem, followed by F508del in Hebron. Moreover, in one of the studied families, c.1393-1G>A mutation was homozygous in one family member and compound heterozygous [c.(1393-1G>A): p.(W1282X)] in another member of the same family. In the same study, N1303K mutation was detected as a homozygous mutation in a member of a family from Hebron. All of the patients in this study clinically presented with pancreatic insufficiency and low body mass index (BMI). In addition, most of these patients at young age experienced significant lung disease with decreased lung function and severe bronchiectasis (Siryani et al. 2015). In another study carried out by

Essawi et al. on unrelated Palestinian CF patients residing in Gaza and the West Bank, the most frequent *CFTR* mutations were: F508del, deletion of exons 19–21 c.(2988+1Kbdel8.6Kb), c.(1393-1G>A), p.(Gly85Glu), and p.(Lys684Serfs∗38), respectively. There were notable differences in the *CFTR* Mutations rates between Gaza and West Bank populations. For instance, p.(Lys684Serfs∗38) and c.(1585-1G>A) mutations only existed among Palestinian patients living in Gaza, while p.(Asn1303Lys), c.(1393-1G>A), and p.(Gly85Glu) were found only in the West Bank Palestinian patients.

Precisely, [c.(2183 AA>G): p. (Lys684Serfs∗38)] is a frameshift mutation (from A to G at 2183 and deletion of A at 2184), previously reported among cystic fibrosis patients in different countries including: Palestine, Syria, Algeria, Egypt, Southern Europe, Latin America, and Iran (Alibakhshi et al. 2008, Castaldo et al. 1996). It is primarily homozygous and compound heterozygous with other variants. In the homozygous cases, it results in failure to thrive, respiratory distress with atelectasis of the right upper pulmonary lobe, *P. aeruginosa* in cultures of pharyngeal aspirate, steatorrhoea, and pancreatic insufficiency. Most of the c.(2183AA>G) carrier patients were diagnosed in the first year of life which indicates the severity of this mutation that coincides with its frameshift nature(Castaldo et al. 1996). The p.N1303K substitution mutation [c.(3909C>G, p.(Asn1303Lys)], is one of the most cystic fibrosis rare mutations and found mainly in the Mediterranean population (Noel, Sermet-Gaudelus, and Sheppard 2018). It disrupts the *CFTR* stability/delivery to the plasma membrane. It is linked with severe disease phenotype and classified as a severe pancreatic mutation, as well as early and severe lung symptoms (Noel, Sermet-Gaudelus, and Sheppard 2018, Rapino et al. 2015).

1.8 Microcephaly Primordial Dwarfism Definition, Types, and Related Genes

Growth is a highly complex and continuously regulated process under genetic, metabolic, and environmental controls (Khetarpal et al. 2016). Patterning and growth are controlled and coordinated by high-fidelity regulatory networks. The Hippo signaling cascade, mitogen-activated protein kinase (MAPK), and insulin growth factor signaling pathways are among the cellular signaling pathways that play a significant role in molding organ and organism size in addition to potent morphogen regulators such as the transforming growth factor beta (TGF-β), Wnt, and Notch (Klingseisen and Jackson 2011). Overgrowth and undergrowth disorders could result when the growth regulation pathway has failed (Khetarpal et al. 2016). One of the severe genetically growth retardation disorders is Primordial Dwarfism (PD). PD is a heterogeneous group of rare single-gene disorders (Klingseisen and Jackson 2011) where growth restriction is obvious in utero and remains post-natally (Dauber et al. 2012), with patients being exceptionally undersized for their age even as a fetus (Khetarpal et al. 2016). The main distinguishing feature of PD in most PD patients is a reduction in head size in proportion to their body size (microcephaly), which differentiates it from other types of dwarfism. (Khetarpal et al. 2016). Microcephaly primordial dwarfism (MPD) is divided into five subtypes; Seckel syndrome, Microcephalic osteodysplastic primordial dwarfism (MOPD) types I ,II and III, and Meier–Gorlin syndrome, and all are inherited in the autosomal recessive manner (Dauber et al. 2012). In addition to these subtypes, there is also a subset of PD patients presented with normal head size and classified as having Silver–Russell (SR) syndrome which could be inherited in different manners; autosomal recessive, autosomal dominant, and genomic imprinting (Khetarpal et al. 2016). Seckel syndrome was described for the first time in 1960 as a group of heterogeneous symptoms including microcephaly, severe short stature, moderate to severe mental retardation, and facial dysmorphic features including beaky

nose, narrow face, large eyes, malformed ears (Saeidi and Shahbandari 2020). MOPD types are autosomal recessive disorders mainly distinguished by bony anomalies (osteodyplastic), mental retardation, less severe growth retardation, and more severe microcephaly (Khetarpal et al. 2016). Clinical, radiological, and genetic criteria have been used to classify the different types of MOPD (Abdel-Salam et al. 2011). Many authors mentioned similarities between MOPD types I and III, implying that they demonstrate the same syndrome at different ages and are therefore now considered as a single entity (Nadjari et al. 2000). The most prevalent clinical features of these types include: Severe intrauterine growth retardation, postnatal growth retardation, micrognathia, small anterior fontanelles, prominent nose (Nadjari et al. 2000), anomalies of pelvis and clavicle, dry skin, sparsity of hairs and eyebrows (Khetarpal et al. 2016). Type II MOPD is the most common type (Brancati et al. 2005) and the most widely studied type of MPD (Monteiro et al. 2021). MOPD II patients are characterized by very severe intrauterine growth, adult height under 110 cm, typical bone dysplasia, absent or mild mental retardation, skin pigmentation, abnormal dentition (Brancati et al. 2005), a high squeaky voice, prominent eyes and nose. (Khetarpal et al. 2016). The most prevalent clinical characteristic of Meier–Gorlin syndrome (MGS) are short stature, microtia, and patellar aplasia/hypoplasia (Ting et al. 2020), where 97% of the MGS patients are present with at least two of these clinical characteristics (de Munnik et al. 2015). Depending on the number of reported cases in the literature, the prevalence of MGS is predicted to be less than 1-9/1,000,000 (X. Li et al. 2021).

Recent studies have proposed that the considerable phenotypic heterogeneity among PD patients could be attributed to exonic mutations of different functional genes associated with the related phenotypes (Khetarpal et al. 2016). Many genes have been identified for the different subtypes of MPD, encoding proteins play key roles in critical cellular processes including *ORC1*, *ORC4*,

ORC6, *CDC6*, and *CDT1* in genome replication, *PCNT*, *CEP152*, and *CPAP* in centrosome function and maintaining, *ATR* in DNA damage response, and *U4atac* in mRNA splicing (Klingseisen and Jackson 2011).

1.8.1 *TOP3A* Gene Structure, Expression, and Function

TOP3A gene is a single copy gene located on the long arm of chromosome 17 (17pll.2-12) and encodes DNA topoisomerase 3-alpha (*TopIIIα*) enzyme which was found to be homologous to the *E.coli* DNA topoisomerase I subfamily and not to the eukaryotic DNA topoisomerase I **(Figure 2)** (Hanai, Caron, and Wang 1996).

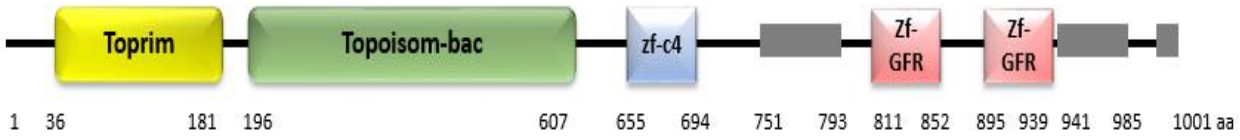


Figure (2): Schematic representation of human *TOP3A* protein. Human *TOP3A* protein domains and families are shown as blocks including: Toprim domain (Yellow), DNA topoisomerase (Green), Topoisomerase DNA binding C4 zinc finger (Blue), GRF zinc gingers (Red), and Disorder regions (Gray).

TOP3A is found to be highly expressed in many somatic tissues particularly testis, skeletal muscle, pancreas, and heart (www.uniprot.org/). Human *TopIIIα* physically binds to *BLM* (encodes a RecQ family DNA helicase) and plays an anti-recombination role with it. *TopIIIα* and *BLM*, as part of BTRR (*BLM*, *TopIIIα*, *RMI1* and *RMI2*) complex, catalyze the process of dissolution of double Holliday junctions (dHJs) that arise as an intermediate during homologous recombination (Yang et al. 2010), in order to generate non-crossover homologous recombination (HR) products (Shorrocks et al. 2021). The dissolution process of dHJs occurs via a mechanism called strand passage mechanism in two steps (Yang et al. 2010). First, creation of a hemicatenane intermediate

(two duplex DNAs interlinked via catenated single strands) through the convergent branch migration of the dHJ, promoted by the BTRR complex. Then, the hemicatenane is decatenated by *TopIIIα* in collaboration with *RMI1* and *RMI2* (Martin et al. 2018b). This mechanism can be accomplished without any exchange between genetic markers flanking the HR original site (Martin et al. 2018).

1.8.2 Dysfunctional *TOP3A* in Microcephaly Primordial Dwarfism

Biallelic loss-of function mutations of *BLM* cause Bloom syndrome (BS) in humans (Martin et al. 2018b). BS is a rare disorder characterized by pre and postnatal growth retardation, sensitivity to sunlight, immunodeficiency, diabetes susceptibility and high predisposition of cancer (Liu and West 2008) and marked by chromosomal instability (Shorrocks et al. 2021). BS patients' cells exhibit various markers of genomic instability, most notably a considerable increase in sister chromatid exchanges (SCEs), which serves as hallmark feature and a confirmatory diagnostic test for BS (Shorrocks et al. 2021). Increasing of SCEs in *BLM* deficient cells occurs due to the dHJs dissolution through an alternative pathway; by Holliday junction resolvases (*SLX-MUS81* and *GEN1* nucleases) which can generate crossover events (Martin et al. 2018b). Likewise, mutations in *TOP3A*, *RMI1*, and *RMI2* have been discovered in patients with Bloom syndrome-like disorders (Shorrocks et al. 2021).

In 2018, Martin et al. studied ten patients in seven unrelated families suffering from growth restriction, microcephaly, and increased SCEs. Whole-exome sequencing showed that these patients have biallelic mutations in the *TOP3A* gene, all except one of which were anticipated to cause a frameshift and premature termination consistent with loss of function. The excepted patient was compound heterozygous for both a frameshift and a missense mutation. Patients cells in

comparison with controls showed a significant decrease in *TOP3A* amounts and significant increase in SCEs; implying a significant crossover recombination.

1.9 Rett Syndrome and *MECP2* Gene

Rett syndrome (RTT) is a progressive neurological disorder described for the first time in 1966 by Andreas Rett (Austrian pediatric neurologist) among 22 of female children diagnosed with progressive cerebral atrophy and different clinical features including gait apraxia, mental retardation and stereotyped hand movements (Brunetti and Lumsden 2020). All of these girls had a history of normal early development, followed by regression and loss of deliberate hand movements (Kyle, Vashi, and Justice 2018). A hyperkinetic movement disorder (HMD), specifically hand stereotypies, is the key clinical feature in RTT diagnosis. Tremor, myoclonus, chorea, and dystonia are among the other HMDs that have been observed in RTT patients (Brunetti and Lumsden 2020). RTT is most prevalent in young females, with an incidence of 1:10,000–20,000 live births (Liyanage and Rastegar 2014) and is considered to be the second most prevalent cause of intellectual disability in females (Banerjee et al. 2019). In typical RTT clinical diagnosis, four major stages were evolved, beginning with stage I (stagnation stage), in which developmental standstill occurs between 6 and 18 months after a normal development period (Sandweiss, Brandt, and Zoghbi 2020). Microcephaly, loss of language and behavioral skills, seizures, and slowed growth, are the most notable phenotypes start to appear at this stage (Liyanage and Rastegar 2014). The second stage (rapid regression stage) occurs at age one to four years, and is mainly marked by regression (Sandweiss, Brandt, and Zoghbi 2020). In this stage, the affected child starts to lose the acquired motor skills and communication (Liyanage and Rastegar 2014), microcephaly worsens, respiratory dysrhythmias and sleep disturbances are common (Kyle, Vashi, and Justice 2018). The third stage (pseudo-stationary stage), starts at ages between 2 and 10 years (Sandweiss, Brandt,

and Zoghbi 2020). Certain functions, such as communication, can be regained by some patients so this stage is referred as wake up period but these patients would still experience disrupted sleeping patterns, anxiety, respiratory problems, and hand tremors (Liyanage and Rastegar 2014). The last stage, is also called Late motor deterioration stage, lasts from several years to decades. Patients at this stage suffer from severe physical disability, dystonia, bradykinesia and some would be completely wheelchair dependent (Kyle, Vashi, and Justice 2018). A child can be diagnosed with atypical RTT when exhibiting RTT-like symptoms without meeting all of the RTT diagnostic criteria like deviation in the onset age, severity, etc (Kyle, Vashi, and Justice 2018).

From the etiological viewpoint, RTT is considered a monogenic neurological disorder, as about 90% of typical RTT cases were found to have loss-of-function mutations in x-linked *MECP2* (methyl-CpG binding protein 2) gene **(Figure 3)** (Liyanage and Rastegar 2014) and commonly occur de *novo* (Kyle, Vashi, and Justice 2018). *MECP2* gene extends about 76 kb in the long arm of the X-chromosome (Xq28) between the *IRAK1* and the *RCP* genes and consists of four exons and three introns (Liyanage and Rastegar 2014). This gene is considered as a key epigenetic modulator in the brain (Liyanage and Rastegar 2014) that is essentially required for neuron maturations and the normal function of nerves cells since it encodes a chromatin-associated protein with a methyl-CpG binding domain that can work as transcriptional activator or repressor (Vidal et al. 2019). It has been over twenty years since the discovery that RTT is caused by mutations within the *MECP2* gene (Sandweiss, Brandt, and Zoghbi 2020). Over 500 pathogenic or likely-pathogenic *MECP2* mutations, especially those affecting exons 3 and 4 and the C-terminal region, have been reported to be associated with RTT (Banerjee et al. 2019). Most of the reported RTT cases caused by C-terminal deletions; exonic deletions; four missense, four nonsense (R270X, R168X, R255X, R106W, R133C, T158M, R294X, and R306C); and eight point mutations. Studies

on RTT confirmed the genotype-phenotype correlation between RTT and *MECP2*. For instance, truncation mutations in the *MECP2* gene have been linked to more severe RTT symptoms (Liyanage and Rastegar 2014).

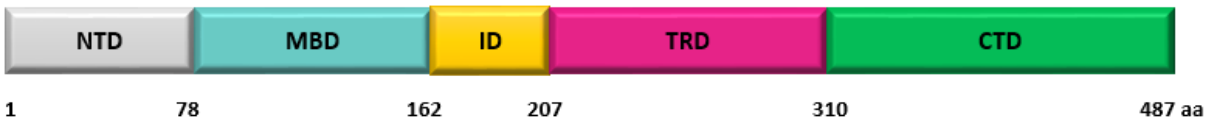


Figure (3): Schematic representation of human *MECP2* protein. Human *MECP2* domains are shown as blocks including: N-terminal domain (NTD) in gray, Methyl-CpG binding domain (MBD) in blue, Intervening domain (ID) in yellow, Transcriptional repression domain (TRD) in purple, and C-terminal domain (CTD) in green.

1.10 Thesis Statement

This study seeks to use WES and WGS approaches in order to decipher the molecular causes of HSP, MPD, and RTT disorders among Palestinian families living in Palestinian villages, and discover the molecular variants of these rare diseases. Our results shall help develop appropriate cost-effective screening tests in the future, in addition, to expand the Palestinians genomic data of rare diseases.

1.11 Study Objectives

Our study focuses on a group of large Palestinian families clinically diagnosed with different rare diseases. The first family constitutes a whole village, and has a population of about 2500 people. This family is known to have high consanguineous marriage levels and highly complex pedigree resulting in subfamilies with many individuals having rare disorders including; CF, cHSP, and developmental delay. CF patients included in our study were previously reported to have two CFTR variants: c.(2183AA>G) and p. (N1303K). Some individuals are compound heterozygotes

for these two mutations, while others are homozygous for one of them. The second family is also a multiplex consanguineous family. Affected members of this family were diagnosed with MPD, together with their first and second-degree relatives, were found to have a strong history of cancer. In the third family, a single case was clinically diagnosed with neurodevelopmental regression with appearance of behavioral problems at 2 years, cessation of speech, hyperactivity, sleep disturbances, trichotillomania, and aggression.

The study aims to:

(i) Identify of the molecular causes of cHSP, MPD, and RTT disorders in the affected individuals.

(ii) Run a segregation analysis of the disease causative variants families' members and confirm the inheritance pattern of each disorder.

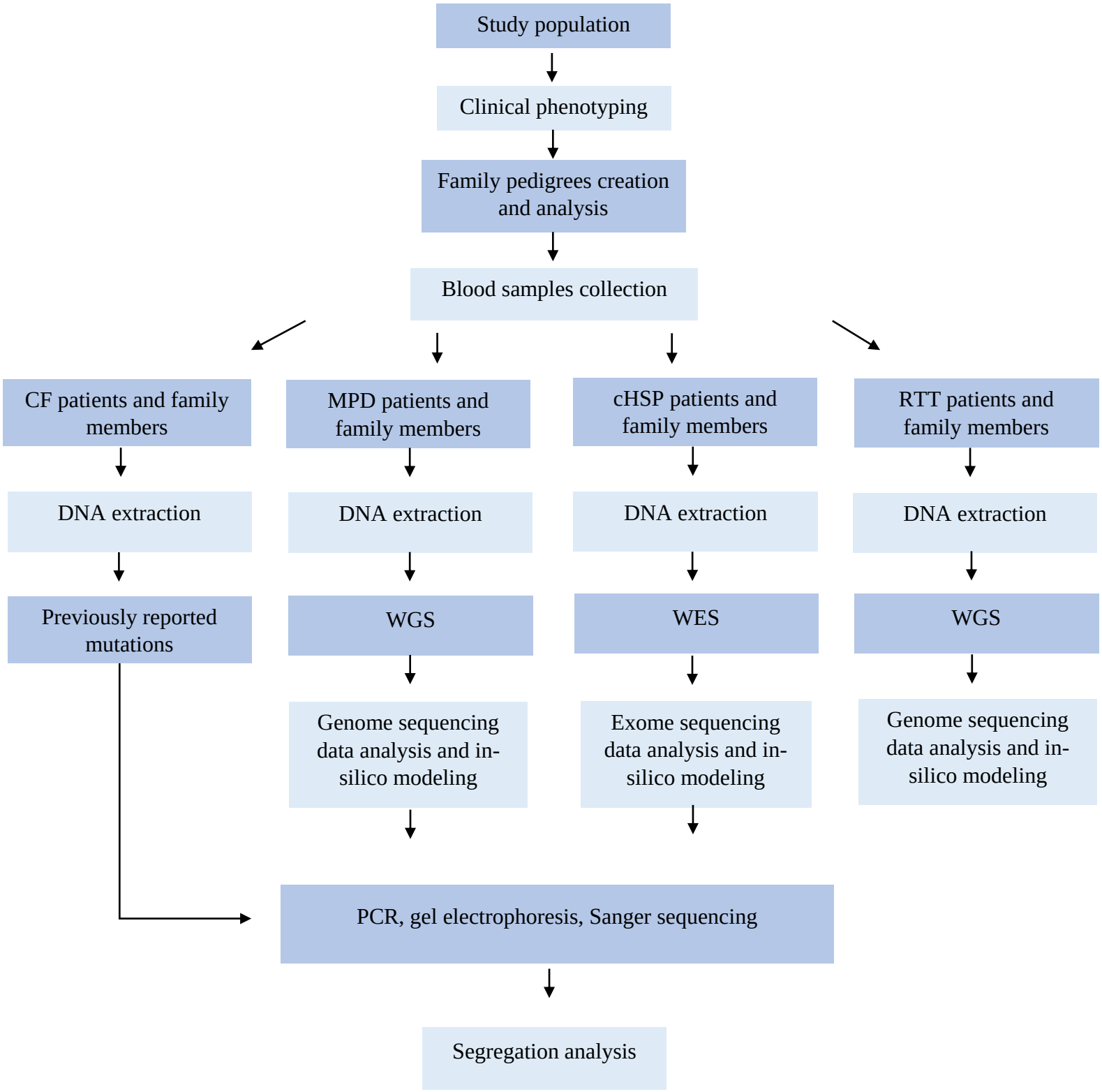
1.12 Study Significance

This project aims at deciphering the molecular causes of rare diseases among Palestinians, particularly villages and isolated communities. Preliminary data indicates that each village and community could be considered as an isolated genetic entity from other Palestinians and villages, where pathogenic variants are detected in these villages but are not detected in other Palestinian genomic data or international controls. It is therefore essential to study the genetic makeup of separate villages and communities to build a map for the founder variants in a national screening theme.

CHAPTER 2: MATERIALS and METHODS

2.1 Study Design

This study was designed as shown by the schematic below:



Our study was designed to target a population of extended Palestinian families where consanguineous marriage is common. The families to be investigated were identified. Clinical diagnosis was established for the individuals concerned, and each family pedigree was created and assessed. Blood samples were then collected from participants in the study after their consent was obtained.

This study focused on the four distinct phenotypes cHSP, CF, MPD, and RTT. WES and WGS were performed and the obtained data was curated and analyzed. Candidate genes and variants were assessed using in-silico modeling. Regions of interest were amplified using PCR, and products were detected by electrophoresis. Sanger sequencing was performed to identify the presence of the variants in family members.

2.2 Site and Population

Family #1: A total of fifty-one members of a large Palestinian family known to have high consanguineous marriage level and highly complex pedigree were recruited. Seven of them are CF patients of four subfamilies. Of the remaining members, six are cHSP patients, four of them are siblings where the other two patients are from different subfamilies. All of the recruited members belong to the same large family and reside in the same village, in the northern West Bank.

Family #2: A total of 19 members of a large multiplex consanguineous Palestinian family were recruited. Seven of them are MPD patients belonging to different four subfamilies. These family members also reside in another Palestinian village, in the northern West Bank.

Family #3: A total of 4 members of a consanguineous Palestinian family were included in this study. One of them is a RTT patient.

2.3 Clinical Phenotyping and Family Pedigrees

Clinical phenotyping was done to all affected members by specialized clinical geneticists and human genetic scientists. Separated family pedigrees were taken for each phenotype (CF and cHSP) in family #1, (MPD) in family #2, and (RTT) in family #3 and were designed using Haplopainter drawing tool (version 1.043).

2.4 Samples collection

Several field visits were made to the recruited families to accomplish samples collection. EDTA Whole-Blood samples (around 2.5 mL) were collected from all CF, cHSP, MPD, and RTT patients and their family members following standard health protocols.

2.5 Ethical Considerations

Patients and families' members recruited in this study signed consent forms prior to participation in the study voluntarily agreeing to participate in research. Patients were assigned ID numbers to ensure confidentiality. Approval was obtained by the Helsinki Committee for Ethical Approval, Palestinian Health Research Council, Gaza, Palestine. [Ethical approval number PHRC/HC/518/1].

2.6 DNA Extraction

Genomic DNA extraction was performed using the Wizard® Genomic DNA Purification Kit (Catalog No. A1120, Promega, USA). Briefly, 900µl of Cell Lysis Solution were added to a sterile 1.5ml microcentrifuge tube and then 300µl of EDTA whole blood were added and inverted several times to be mixed. The mixture was incubated at RT for 10 minutes and inverted meanwhile. The samples then centrifuged at 16,000 × g for 20 seconds at RT. After that, the supernatant was

discarded without disturbing the visible white pellet, and the pellet was vortexed to be resuspended. 300µl of Nuclei Lysis Solution were added to the resuspended cells and pipetted several times to lyse the WBCs. 100µl of the Protein Precipitation Solution were then added to nuclear lysate, vortexed vigorously, and centrifuged at 16,000 × g for 3 minutes at room temperature (RT). The supernatant was transferred to a clean microcentrifuge tube containing 300µl of RT isopropanol, gently mixed to show the white thread-like strands of DNA, and centrifuged at 16,000 × g for 1 minute at RT. After centrifugation, the supernatant was decanted, and 300 µl of RT 70% ethanol were added to wash the DNA, and then centrifuged at 16,000 × g at RT for 20 seconds. Next, the ethanol was aspirated and the pellet was air-dried for 15 minutes. Finally, the DNA was rehydrated using 100 µl of DNA Rehydration Solution and incubated at 65°C for 1 hour. DNA purity and concentration were measured using NanoDrop spectrophotometer (Thermo Scientific™ NanoDrop 2000 and 2000c, Thermo Fisher Scientific, USA).

2.7 WES and WGS of cHSP, MPD and RTT patients, respectively

WES was performed for two of the recruited cHSP patients and WGS was carried out for one MPD and one RTT recruited patients at the Exeter clinical laboratory. The WES was performed using the Twist Human Core Exome capture on an Illumina NextSeq500 sequencer. Agilent Sureselect Whole Exome v6 targeting with reads aligned (BWA-MEM,v0.7.17), mate-pairs fixed and duplicates removed (Picard v2.15.0). InDel realignment and base quality recalibration was performed using GATK (v3.7.0), SNVs/InDels were detected with GATK (HaplotypeCaller), annotation using Alamut Batch (v1.10) and CNV detection with ExomeDepth and Savvy CNV. Variants were then filtered by call quality, frequency in control datasets (gnomAD and 1000 Genomes Project), and predicted functional consequence.

WGS was performed using PCR free Illumina sequencing libraries run on a HiSeq2500 sequencing (to 45X coverage), Bcl files were aligned to hg19/GRCH build 37 and nucleotide variants called with DRAGEN (Edico Genome). Structural variants (SVs) were called using the union of four algorithms (BreakSeq, CNVnator, Delly, and Manta) and annotated to determine affected genes and regulatory regions and were compared to public and internal databases to determine minor allele frequency and filter common, or false positive SVs. InDel realignment and base quality recalibration were performed using GATK (v3.7.0). SNVs and InDels were detected using GATK HaplotypeCaller and annotated using Alamut batch (v1.8). Read depth was determined for the WE through our in-house pipeline. Copy number variants were detected using both Exome Depth (https://cran.rproject.org/web/packages/ExomeDepth/vignettes/ExomeDepthvignette.pdf and https://github.com/vplagnol/ExomeDepth) and SavvyCNV(https://www.biorxiv.org/content/10.1101/617605v1 and https://github.com/rdemolgen/SavvySuite). Variants with <5 reads, a frequency of >1% in gnomAD (V2.1.1) and/or in-house databases were excluded. De novo, homozygous or compound heterozygous variants were evaluated.

2.8 Variants Analysis Using Rare Diseases Pipeline

The variants called by the Exeter laboratory were further filtered in a stepwise manner to find candidate variants of interest. In family #1, given that the mode of inheritance is autosomal recessive, as the parents do not exhibit the trait, the first criterion was that the variants must be homozygous and the variants must be shared between the two cHSP selected patients. Based on the clinical suspicion that these cHSP, MPD, and RTT clinical phenotypes resulted from deleterious variants, the previously reported variants in ClinVar or dbSNP databases as benign or likely benign, and not rare variants in gnomAD database were excluded. Then, variants were

narrowed down based on probability functional effect on the protein. To be precise, synonymous variants were excluded. PolyPhen, SIFT, and Provean algorithms were used to predict the effect of the variants, and benign or likely benign variants were excluded. Variants in genes lacking relevant phenotype based on the reported phenotypes in the OMIM database were also excluded. In addition, variants with low mapping quality or variants with allele depth (AD) lower than 0.20 were excluded. To further study the remaining variants, in-silico tools were used to predict the effect of candidates on the protein and their association to the diseases. The gene-associated phenotypes were investigated using OMIM, ClinGen, PanelApp and Decipher databases. After that, ClinVar, gnomAD, dsSNP databases and the literature were used to investigate any reported variants. Then, candidate gene features and domains were studied closely to locate these variants.

Thereafter, candidate variants and previously reported variants were plotted using ClinVar. In addition, multiple amino acid sequence alignment of Homo sapiens and other species was performed using the Clustal Omega alignment tool of the EMBL-EBI (https://www.ebi.ac.uk/Tools/msa/clustalo/) to check conservation of the site of variants. PolyPhen, Provean, and SIFT algorithms were also used to detect the effect of the missense mutations. PolyPhen uses score ranges from 0.0 (Benign) to 1.0 (Deleterious), and SIFT uses the same range but with 0.0 considered deleterious and 1.0 considered tolerated. The cutoff score for both is 0.05. The deleterious effect of the missense variant was also predicted using Align GVGD tool (http://agvgd.hci.utah.edu/agvgd_input.php). On the other hand, Mutation taster (http://www.mutationtaster.org/) was used to predict the disease-causing potential of the insertion/deletion variants. Finally, selected variants were interpreted and classified based on the ACMG guidelines.

2.9 Primers Design

Primers used to detect the presence of the variants of interest among CF, cHSP, MPD, and RTT patients and family members were designed using Primer-Blast tool of the NCBI (https://www.ncbi.nlm.nih.gov/tools/primer-blast/). First, UCSC genome browser (https://genome.ucsc.edu/)- Human GRch37/ hg19 assembly was used to search for the variant location in the genome in order to obtain the genomic DNA sequence that flanked the variant by 250 bases upstream (5') and 250 downstream (3'). Next, the obtained UCSC-PCR template sequence was copied and pasted in the PCR template box of the blast tool. Sequences that are 100% identical with the PCR template sequence were selected. After getting the specificity notion that confirms that primer pairs are specific to the input template, the best pair of the primers were picked based on the best melting temperature (Tm), GC content%, product size, runs, repeats, GC clamp, and other parameters. After choosing the best pair of primers, UCSC *in-silico* PCR tool (https://genome.ucsc.edu/cgi-bin/hgPcr) was used to further test for the primers specificity.

Table 2. Primers pairs designed to detect *CFTR, PTPN23, TOP3A* variants

Gene	Mutation	Primer sequence (5' to 3')	Product size (bp)
CFTR	c.2183AA>G	F: GGGATGTGATTCTTTCGACCA R:CAGAATCCTCTTCGATGCCA	251
	p.N1303K	F: TGGTAAGTACATGGGTGTTTCT R: TCAGTTAGCAGCCTTACCTC	237
PTPN23	g.47454479C>A	F: GTCTACCCCAATCCCATCTT R:ACCTGTTCAGGTCTTGTTGAG	310
	g.47454483del	F: CCAGAGTCTACCCCAATCC R:CTTGTTGAGTGTCCAGAGTG	303
TOP3A	g.18178274delA	F: AAGAAGGCCCGACTCCAAAG R:TGCCACTTCTGCATGTCAGT	371

2.10 PCR- Amplification Protocol

PCR amplifications were carried out for targeted fragments on the *CFTR*, *PTPN23*, *TOP3A*, and *MECP2* genes. The PCRs were performed, using GoTaq® Green Master Mix (Promega, USA), following the manufacturer's recommendations. Briefly, a 25 μL reaction mixture was prepared by adding 2 μL of the extracted DNA, 1 μLof the forward primer,1 μL of the reverse primer,15 μL of Master Mix, and 6 μL of dH2O. The primers used for the amplification of each targeted DNA fragment are listed in **Table 2**. To ensure that the PCR mixture was not contaminated, a PCR negative control was used (by preparing an extra reaction mix for each amplification and substituting the DNA with dH2O). The PCR amplifications were done with a thermocycler machine (FlexCycler2 PCR Thermal Cycler, Analytik Jena, Germany), using the following conditions for amplification of *PTPN23*, *CFTR*, *TOP3A*, and *MECP2* genes: DNA denaturation, at 95.0°C for 5 minutes; 36 cycles of: DNA denaturation, at 95.0°C for 15 seconds; primer-annealing, at 58.0°C for 45 seconds; and primer-extension, at 72.0°C for 30 seconds; the last cycle of primer-extension, at 72.0°C for 5 minutes).

2.11 Detection of PCR Products

PCR-amplification products of *CFTR*, *PTPN23*, *TOP3A*, and *MECP2* genes were analyzed by electrophoresis in agarose gels of 1.5 % (weight/volume). The gel was prepared with 1.5 g of Agarose (SigmaAldrich, St.Louis, USA) in 100 mL 1X TAE buffer (Thermo Scientific™, Lithuania). Regarding electrophoresis, gel wells were loaded with 5 μL of the PCR products and the electrophoresis was performed at 90V for 45 minutes. The DNA bands were visualized using UV-light with the ChemiDoc imaging system (ChemiDoc™, BioRad, USA).

2.12 PCR Products Purification and Sequencing

2.12.1 Cleanup of PCR products

To purify the PCR products for sequencing, 1μL of EPPiC fast mixture (A&A biotechnology, Poland) (containing exonuclease I enzyme to get rid of excess primers from previous PCR mixtures, and alkaline phosphatase enzyme to dephosphorylate dNTPs) was added to new PCR tubes, then 5 μL of each PCR product were added. After that, the mixtures were spinned down and incubated for 15 minutes in the thermal cycler (Biometra TONE thermal cycler, Analytik Jena, Germany) as the following: 10 minutes at 37.0°C (to activate the enzymes mixture), and 1 minute at 80.0°C (to completely inactivate the enzymes).

2.12.2 Loading of samples into multi-well plate

After dilution of the PCR products, 1 μL of BigDye™ terminator (Thermo Fisher Scientific, USA), 3.5 μL of sequencing buffer, 11.5 μL of water, 2 μL of forward primer (with concentration of 5 pmol), and 2 μL of PCR products were added to a multi-well plate. Then, the plate was spinned down and incubated in the thermal cycler for 25 cycles of: 1 minute and 10 seconds at 96.0°C, 5 seconds at 50.0°C, and 3 minutes at 60.0°C.

2.12.3 DNA Precipitation

To precipitate the DNA, 5 μL of 0.5 M EDTA solution and 60 μL of 100% Ethanol were added to the wells. Then, the multi-well plate was vortexed for a minute and left for about 12 minutes at 4°C. After that, the plate was centrifuged for 30 minutes at 2200 RCF at 4°C. 80 μL of 100% ethanol was then added to the wells and centrifuged again for 15 minutes at 1600 RCF. Next, the multi-well plate containing samples left to air dry for 15 minutes, and then 10 μL of formamide were

added to the wells. Then, the plate was incubated in a dry heat bath for 3 minutes at 95.0°C and on ice for another 3 minutes. Finally, the plate was placed into the Sanger sequencing machine (Hitachi3500 Genetic Analyzer, Thermo Fisher Scientific, USA) to be sequenced.

2.13 Sequences Analysis and Segregation

The sequences of the PCR products were viewed and analyzed against reference sequences obtained from GenBank (https://www.ncbi.nlm.nih.gov/genbank/) using the Sequence Scanner™ Software V 2.0. Mutations harboring status were determined for each member of the family and assigned on the family pedigrees.

CHAPTER 3: RESULTS

Our study investigated three Arab Palestinian families, each with a different clinical phenotype: cHSP, CF, MPD, and RTT. The analysis of the enrolled patients' WES and WGS data resulted in the identification of the molecular etiology of these hereditary disorders as well as the segregation of the causal variants among family members and confirming the predicted inheritance pattern using Sanger sequencing.

3.1 Clinical Findings

3.1.1 Cases of Complex Hereditary Spastic Paraplegia in Family #1:

Here, we describe an extended Palestinian family, affected with multiple autosomal recessive rare diseases. Due to the high prevalence of cases in this large family, a strong founder effect is suggested. All of the included cHSP patients are of consanguineous parents except (HSP1-V11) patient **(Figure 4) (Table 3)**. Proband 1 (HSP1-V11) is a 25 years old female. Symptoms and signs began to appear at the age of 7 years. She has been reported to have microcephaly and no dysmorphic features. She has a mild lower limb spasticity and exaggerated deep tendon reflexes in the lower limb. She is also displayed horizontal nystagmus, up-going Babinski reflexes and paresthesia. No hand weakness, or arm stiffness have been reported. She doesn’t have developmental delay or intellectual disability. She has walking on toes mode and didn’t undergo any surgery.

Patient (HSP2-V2) is a 23 years old male. The first signs and symptoms appeared between the ages of 4 and 5 years. Clinical diagnosis showed that he has microcephaly and no dysmorphic features. He has lower limb spasticity and in the lower limb. He displayed up-going Babinski

reflex. He also suffers from episodic paresthesia. Regarding the upper part, he doesn't have hand weakness, arm weakness or stiffness. The horizontal optokinetic nystagmus was clear. He had tendon release surgery at both ankles. Before the surgery, walk on toes was his preferred walking mode. He doesn't have developmental delay or intellectual disability.

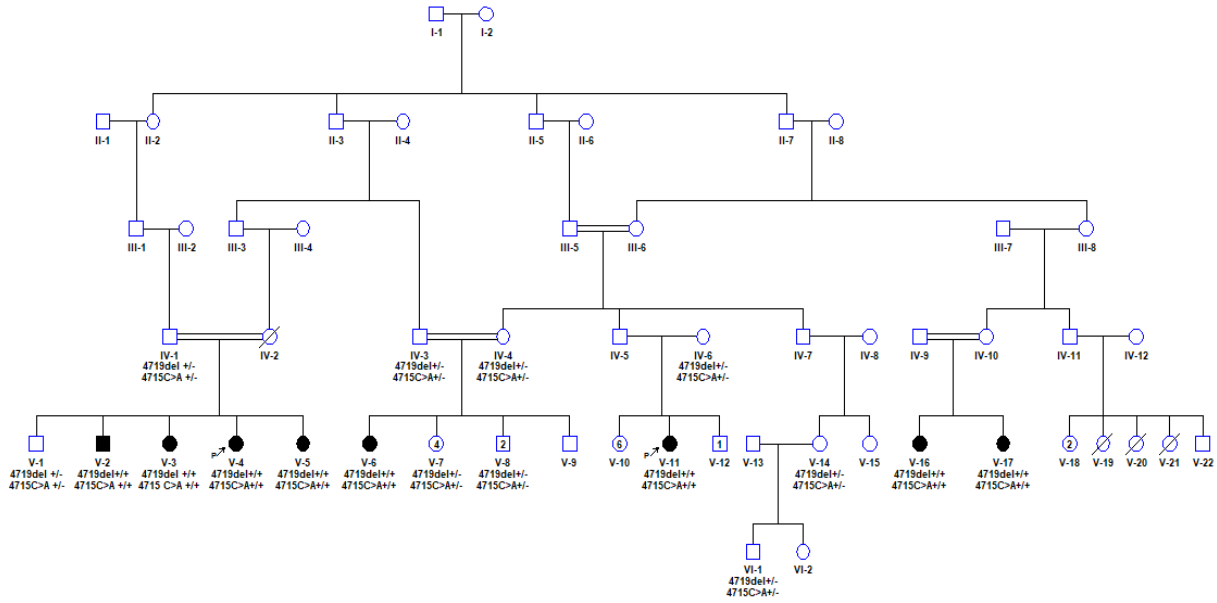


Figure (4): Family pedigree of cHSP. Pedigree structure of the Arab Palestinian family affected with cHSP, showing segregation of the *PTPN23* variants [c.(4719del) and c.(4715C>A)]. The genotype of members is indicated ['+': mutant allele, '-': wild type]. Arrows indicate proband.

Proband 2 (HSP3-V4) is a 14 years old female, and she is a sister of patient (HSP2-V2). Signs and symptoms started at the age of 6 years. She was reported to have mild developmental impairment, lower limb spasticity, and exaggerated deep tendon reflexes in the lower limb. She displayed toe walking, episodic paresthesia, and up-going Babinski reflex. Microcephaly and nystagmus were also reported in this patient. No problems in the arms or hands were reported.

Patient (HSP4-V3) is a 16 years old female. She was born with congenital hip dysplasia and had had surgery at the age of 7 months. Symptoms and signs started to appear at the age of 7 years.

She has mild developmental impairment, lower limb spasticity, and exaggerated deep tendon reflexes in the lower limb. She also presented microcephaly, toe walking and up-going Babinski reflex.

Patients (HSP5-V5) and (HSP6-V6) are 15 and 11 years old females, respectively, and they are sisters of patients (HSP2-V2) and (HSP3-V4). Signs and symptoms started in both of them at the age of 6 years and were clinically reported to have microcephaly, mild developmental impairment, horizontal optokinetic nystagmus, lower limb spasticity, and exaggerated deep tendon reflexes in the lower limb. They also manifested episodic paresthesia and up-going Babinski reflex. No problems were reported in the arms, hands, and feet lifting. Of the other medical problems, the patient (HSP6-V6) was reported to have attention deficit disorder.

Table 3. Clinical characteristics of cHSP patients homozygous for final exon frameshift *PTPN23* gene variant p. (P1572T fs*12).

	V2	**V3**	**V4**	**V5**	**V6**	**V11**
Genotype	Hom	Hom	Hom	Hom	Hom	Hom
Age	23 y	16 y	14 y	15y	11 y	25 y
Sex/Age of onset	M, 4-5 y	F, 7 y	F, 6 y	F, 6y	F, 6y	F, 7 y
Consanguinity	Yes	Yes	Yes	Yes	Yes	No
Microcephaly	Yes	Yes	Yes	Yes	Yes	Yes
Dev impairment	normal	mild	mild	mild	mild	normal
Toe walking	Yes	Yes	Yes	No	Yes	Yes
Lower limb spasticity	Yes	Yes	Yes	Yes	yes	No
Lowe limb DTRs	+++	+++	+++	+++	+++	+++
Babinski reflex	UP	UP	UP	UP	UP	UP
Upper limb neurology	normal	normal	normal	normal	normal	normal
Horizontal nystagmus	Yes	No	Yes	Yes	Yes	Yes
Hypo/Paraestehisa	episodic	normal	episodic	normal	episodic	normal

DTRs: Deep Tendon Reflexs , Hom: Homozygous, M: male, F: Female, Y: year, (+++) indicates exaggerated reflexes.

3.1.2 Cases of Cystic Fibrosis in Family #1:

All of the CF patients are of consanguineous parents (first cousin marriages) **(Figure 5)**. These patients presented with pulmonary and gastrointestinal symptoms manifested by chronic pulmonary infections, chronic diarrhea, and failure to thrive.

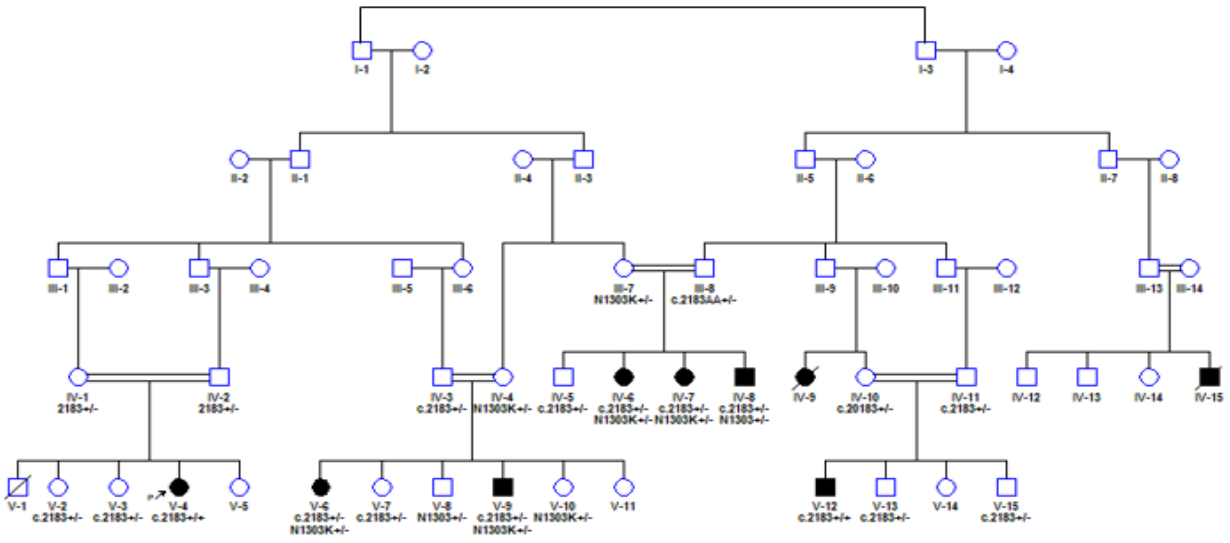


Figure (5): Family pedigree of CF. Pedigree structure of the Arab Palestinian family affected with CF, showing segregation of the *CFTR* variants [c. (2183 AA>G) and p. (N1303K]. The genotype of members is indicated ['+': mutant allele, '-': wild type]. Arrows indicate the proband.

3.1.3 Cases of Microcephaly Primordial Dwarfism in Family #2:

All of the MPD patients recruited in this study are females and have consanguineous parents **(Figure 6).** The index case is an 18-years-old female born to consanguineous parents, the parents are half first cousins, and 2nd cousins. Before birth, she was known to have growth retardation. She weighed 1900 gm at birth. Her head circumference was 30 cm, and height of 43 cm. She was admitted to neonatal ICU for 5 days due to respiratory distress. During admission, she was only on nasal oxygen treatment. Her growth parameters remained at -3 SDS. She had normal skeletal survey, and normal IGF-1 testing. Treatment by growth hormone was not advised, due to its failure

in a cousin. Her siblings are all unaffected. The major clinical features of the included MPD patients are summarized in **table 4.**

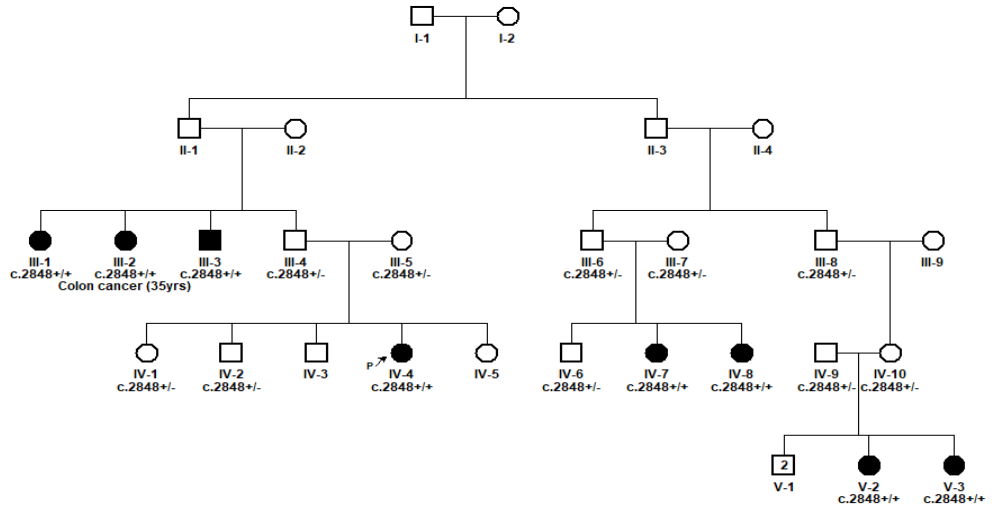


Figure (6): Family pedigree of MPD. Pedigree structure of the Arab Palestinian family affected with MPD, showing segregation of the *TOP3A* variant (c.2848del). The genotype of members is indicated ['+': mutant allele, '-': wild type]. Arrows indicate the proband.

Table4. Clinical characteristics of MPD patients homozygous for final exon frameshift *TOP3A* gene variant c.(2848del).

	Patient 1	**Patient 2**	**Patient 3**	**Patient 4**	**Patient 5**	**Patient 6**	**Patient 7**
Genotype	Hom	Hom	Hom	Hom	Hom	Hom	Hom
Sex	F	F	F	F	F	F	F
Consanguinity	Yes	Yes	Yes	Yes	Yes	Yes	Yes
Prenatal delayed growth	+	+	+	+	+	+	+
Microcephaly	+	+	+	+	+	+	+
Short stature	+	+	+	+	+	+	+
Height/Age	79cm/37y	131cm/11y	142.5cm/18y	136cm/43y	138cm/47y	104cm/7y	63cm/1y
Poor weight gain	+	+	+	+	+	+	+
Intellectual disability	-	-	-	-	-	-	-
Small facial features	+	+	+	+	+	+	+
Dysmorphic features	Triangular face Prominent nose	Triangular face Prominent nose	Triangular face Prominent nose	Triangular face Prominent nose	Triangular face Prominent nose	Triangular face Prominent nose	Triangular face Prominent nose

(+): indicates presence of a characteristic in a patient, (-): indicates absence of a characteristic in a patient, F: Female, Hom: Homozygous, m: month, y: year, cm: centimeter.

3.1.4 Cases of Rett Syndrome in Family #3:

The proband and the only affected individual in the third family is a 6 years old female of first cousins' parents **(Figure 7)**. She had a normal early postnatal life until she was clinically diagnosed

with neurodevelopmental regression with appearance of behavioral problems at 2 years, and later she exhibited different clinical signs including hyperactivity, cessation of speech, sleep disturbances, trichotillomania, and aggression.

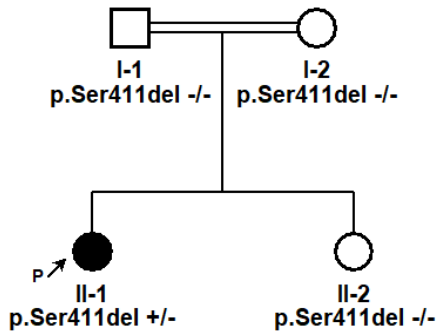


Figure (7): Family pedigree of RTT. Simplified pedigree structure of the Arab Palestinian family affected with RTT, showing segregation of the *MECP2* variant (p. Ser411del). The genotype of members is indicated ['+': mutant allele, '-': wild type]. Arrows indicate the proband.

3.2 Genetics Findings

3.2.1 Molecular-causes of cHSP

A total of 23 homozygous gene variants present in the WES data of the first cHSP patient and 28 in the second patient. Of these variants, 11 are shared between the two patients. After excluding the variants that are synonymous, benign, variants with low mapping quality, and variants in a gene with non-fitting clinical phenotype, a complex deletion/insertion variant in the *PTPN23* gene, was concluded as a candidate variant for further analysis. *PTPN23* was of particular interest because it's known to be expressed in the neurons and its associated phenotype as found in OMIM is neurodevelopmental disorder and structural brain anomalies with or without seizures and spasticity with the autosomal recessive inheritance mode. The *PTPN23* variant is a c.(4719delins)

frameshift deletion/insertion in exon 25 resulting in a p. (Pro1572Thrfs*12) mutation. This variant occurs in the C-terminal domain (CTD) next to the Protein Tyrosine Phosphatase (PTP) domain **(Figure 1).** It is predicted to escape nonsense-mediated decay and results in a partially truncated PTPN23 polypeptide by 64 amino acids (1572/1636). This variant is rare (allele frequency of 0.0004%) and not found in gnomAD v3.1, ClinVar, or HGMDPro databases. The affected residues are highly conserved across species as indicated by Clustal Omega alignment, and the shifted sequence due to this variant is also conserved **(Figure 8)**. The mutation prediction tool, mutation taster (Build of NCBI 37 / Ensembl 69) predicts that this variant is a disease causing.

This variant [c.(4719delins): p. (Pro1572Thrfs*12)] is classified as a likely-pathogenic variant based on the ACMG guidelines as it is predicted to result in a loss or disruption of normal protein function through NMD or protein truncation. In addition, it is absent from controls, cosegregated with disease in multiple affected family members in a gene definitively known to cause the disease, and multiple lines of computational evidence support its deleterious effect on the gene or gene product.

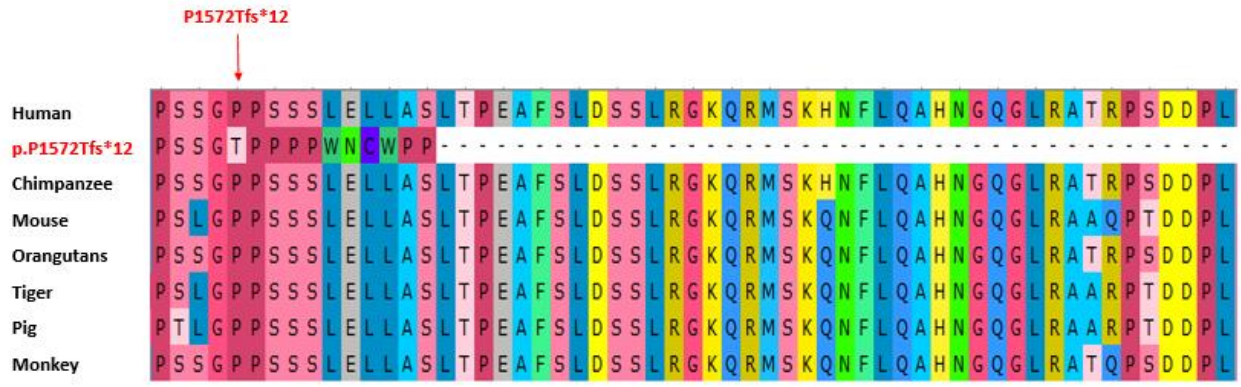


Figure (8): Comparison of conserved amino acids at positions from 1572 to 1584 of human wild type *PTPN23* with six orthologues alongside the predicted outcome of the p.(Pro1572Thr*12) frameshift variant. Multiple amino acid sequence alignment of Homo sapiens and other species (shown in bold to the left) was performed using the Clustal Omega alignment tool of EMBL-EBI (https://www.ebi.ac.uk/Tools/msa/clustalo/).The red arrow indicates the start point of the identified variant at the position 1572.

3.2.2 Molecular-causes of MPD

WGS of the proband identified 83 homozygous gene variants, 155 multiple heterozygous gene variants, and 75 single heterozygous gene variants. Because the disease inheritance pattern is AR, the analysis was confined to homozygous variants. Using the previously mentioned narrowing down strategy of the variants, *TOP3A* gene was predicted to be damaging and concluded as a candidate for further analysis. Microcephaly, growth restriction, and increased SCEs are described linked phenotypes of TOP3A in OMIM, which matches the clinical phenotype of the enrolled individuals. *TOP3A* is likewise diagnostically graded green in its relationship with severe microcephaly, bloom-syndrome, and development failure, according to PanelApp. In addition, searching ClinVar for *TOP3A* variants showed that there are previously reported pathogenic variants in patients with clinical phenotypes similar to our patients. The *TOP3A* identified variant in the analyzed genome is a (c.2848del) frameshift-deletion variant in exon 19 that causes a frameshift in 58 amino acids (p. Trp950Glyfs*58) in the CTD **(Figure 2)**, all of them are conserved according to the Clustal Omega alignment **(Figure 9)**, starting with codon Tryptophan 950, changes this amino acid to Glycine residue, and results in the loss of the original stop codon and prolongation of the protein.

This variant is predicted by MutationTaster to change the amino acids sequence, prolonged protein (+5 AA) **(Figure 9)**, and splice site changes. The (c.2848del) variant is not observed in gnomAD, ExAC, ClinVar, Ensemble, and Decipher databases (Rarity 0%). (c.2848 del) variant can be classified as a likely-pathogenic variant based on the ACMG guidelines as it is a frameshift in the *TOP3A* gene where LOF is a known mechanism of disease, absent from controls (with allele frequency of 0%), and multiple lines of computational evidence support its deleterious effect on the *TOP3A* gene and protein.

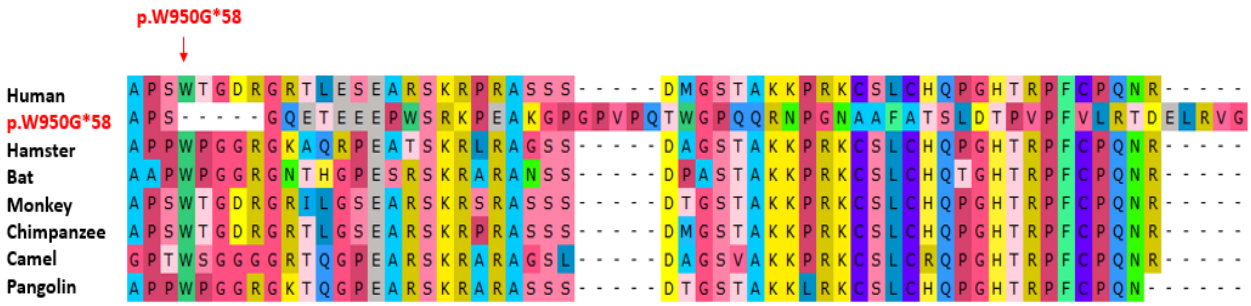


Figure (9): Comparison of conserved sequence of human wild type *TOP3A*, from tryptophan amino acid at the position 950 to the end of the protein, with six orthologues alongside the predicted outcome of the p.(W950G*58) variant. Multiple amino acid sequence alignment of Homo sapiens and other species (shown in bold to the left) was performed using the Clustal Omega alignment tool of EMBL-EBI (https://www.ebi.ac.uk/Tools/msa/clustalo/). The red arrow indicates the conserved tryptophan residue at the position 950.

3.2.3 Molecular-causes of Rett Syndrome

WGS sequencing data of the 6 years old female of first cousins' parents, clinically diagnosed with neurodevelopmental regression with appearance of behavioral problems at 2 years with cessation of speech, hyperactivity, sleep disturbances, trichotillomania, and aggression, identified 45 homozygous gene variants, 271 multiple heterozygous gene variants and 1063 single heterozygous gene variants. Homozygous gene variants weren't found to be related to her diagnosis, therefore the analysis shifted to the heterozygous variants. Firstly, hot heterozygous variants were picked based on the criteria mentioned in the methods section and each one was analyzed based on genes and variants data reported in different genes and variants databases. The most related genes were picked and then analyzed; including *CACNA1E, LNPK, MADD, ABCA2, KIF1A, and MECP2.*

Gene phenotype, phenotype inheritance mode, gene constraint scores, GeneCards reported disorders, genereviwes, and ClinVar reports were checked for each of the related genes. Variant of the *MECP2*, [ChrX:g.(153296046_153296048del): c. (1231_1233 del): p.(Ser411del)] **(Figure 10)** was suggested to be the molecular cause of this condition for many reasons, first, one of the

reported phenotype in OMIM, and ClinGen of *MECP2* gene is Rett syndrome and the patient has many common symptoms with Rett syndrome including slowed growth, sleep disturbances, abnormal behaviors, loss of communication abilities, seizures and her symptoms started at age of 2 years. Second, *MECP2* is constrained gene with pLI score of 0.89. Third, the variant is an inframe deletion causes a deletion of serine amino acid at position 411 of the *MECP2* protein in the C-terminal domain **(Figure 3)**.

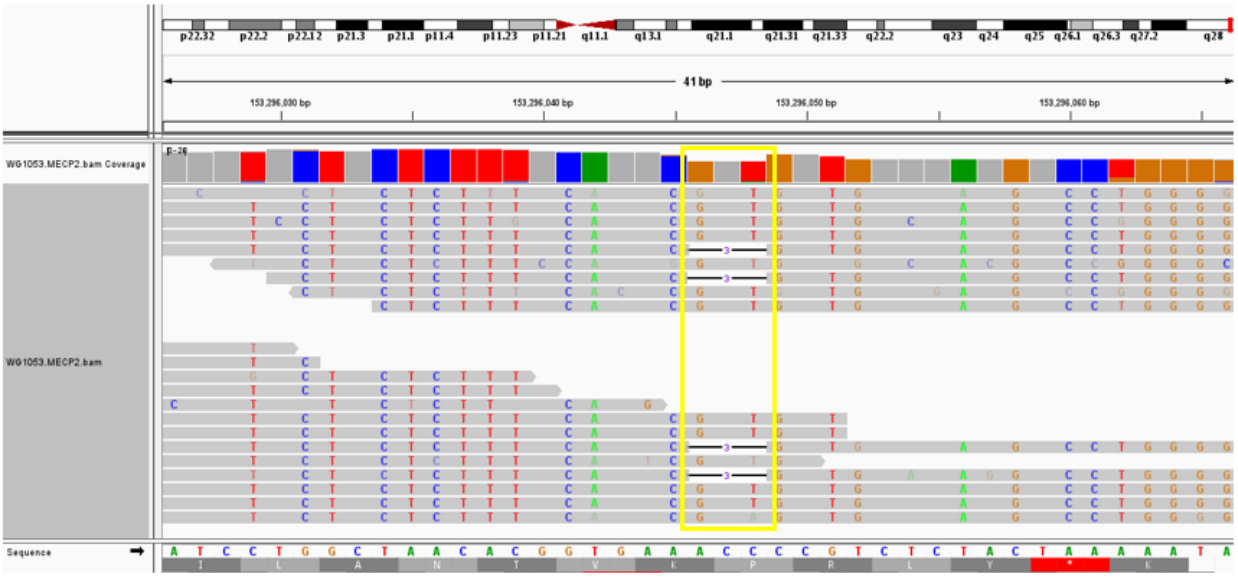


Figure (10): Integrated Genomics Viewer (IGV) screenshot of the *MECP2* [ChrX:g.(153296046_153296048del] variant. The inframe deletion variant [ChrX:g.(153296046_153296048del] of the *MECP2* gene is indicated by the yellow rectangle.

This variant is rare with gnomAD frequency of 0% and occurs on a conserved site in the *MECP2* protein **(Figure 11)**. In addition, it is predicted to be disease-causing mutation by MutationTaster and phosphoserine residue is predicted to be lost. This variant is needed to be confirmed by techniques other than Sanger sequencing.

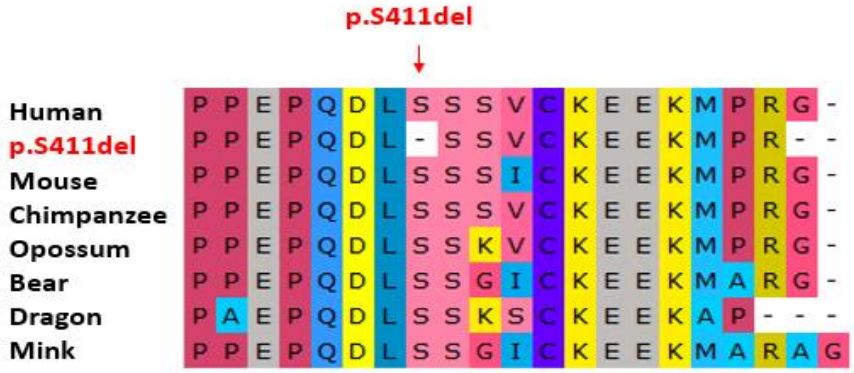


Figure (11): Comparison of conserved Serin-411 of human *MECP2* with six orthologues, alongside the predicted outcome of the p.(Ser411del). Multiple amino acid sequence alignment of Homo sapiens and other species (shown in bold to the left) was performed using the Clustal Omega alignment tool of EMBL-EBI (https://www.ebi.ac.uk/Tools/msa/clustalo/). The red arrow indicates the conserved Serine residues at the position 411.

3.3 Segregation Analysis

3.3.1 Segregation analysis of *PTPN23* Variants

WES analyses identified a homozygous frameshift deletion/insertion variant [c.(4719delins): p.(Pro1572Thrfs*12)] in the *PTPN23* gene of two cHSP affected individuals. This variant was confirmed using dideoxy sequencing (Sanger sequencing). Sanger sequencing also showed the segregation of this variant among all of family members as expected for an autosomal recessive cause of the cHSP **(Figure 4)**. The segregation analysis showed that all of the cHSP affected subjects including in this study are homozygous for the c.(4719delins) variant **(Figure 12)**.

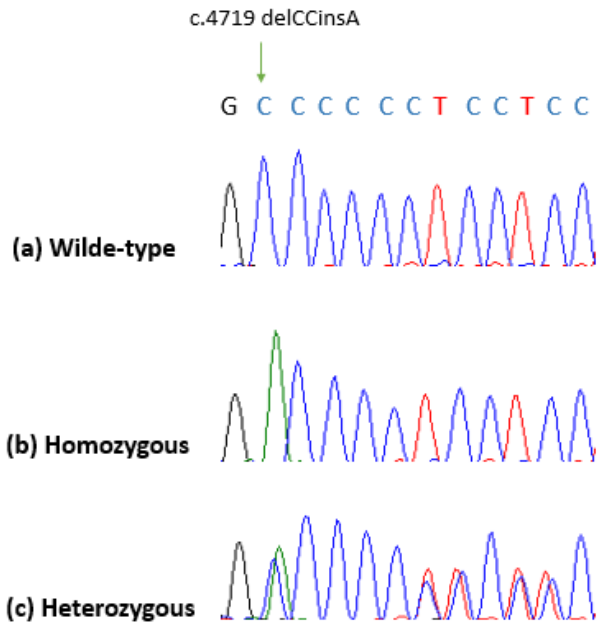


Figure (12): Electropherogram showing the DNA sequence of the *PTPN23* (c.4719delins) frameshift variant (green arrow). (a) A wild-type state in an unaffected subject (PL0240), **(b)** homozygous state in a cHSP affected subject (PL0233), and **(c)** heterozygous state in a carrier unaffected individual (PL0235) are compared.

3.3.2 Segregation Analysis of *CFTR* Variants

As mentioned earlier, CF patients were reported to have two *CFTR* variants: c.(2183AA>G) and p. N1303K. Sanger confirmation and segregation showed that of the seven studied patients, five are compound heterozygous for both of the variants, at which one of the parent is heterozygous for the first variant and the other is heterozygous for the second mutation **(Figure 13)**, and two of the patients are c.(2183AA>G) homozygous for two parents' carrier for the variant **(Figure 5)**.

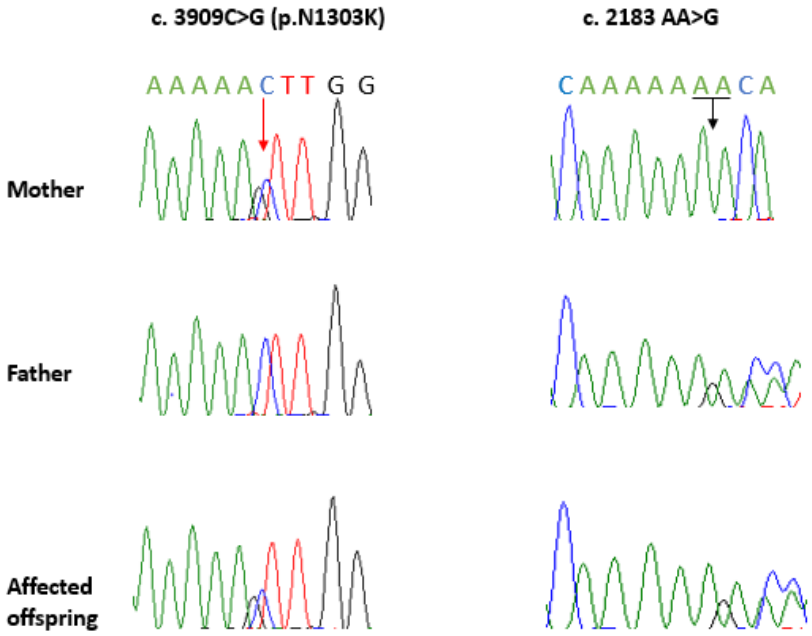


Figure (13): Segregation of *CFTR* mutations in one of the CF studied patients. The Electropherograms to the right and left showing the DNA sequence of the of *CFTR* (p.N1303K) (red arrow) and (c.2183AA>G) substitution missense variants (green arrow) variants, respectively. The mother (CF23) is heterozygous carrier for the p.N1303K variants, and the father (CF22) is heterozygous carrier for the c.2183 AA>G variants. The affected offspring (CF18) is compound heterozygous carrier for both variants.

3.3.3 Segregation analysis of *TOP3A* Variants

WGS identified a frameshift-deletion variant [c.(2848del): p.(Trp950Glyfs*58)] in *TOP3A* gene of the MPD proband patient. Sanger segregation revealed that all of the enrolled MPD affected individuals are homozygous carrier for this variant. Parents of the affected individuals have been verified to be heterozygous carriers which confirm that this variant is inherited in an autosomal recessive manner to cause MPD among family members **(Figure 14)**.

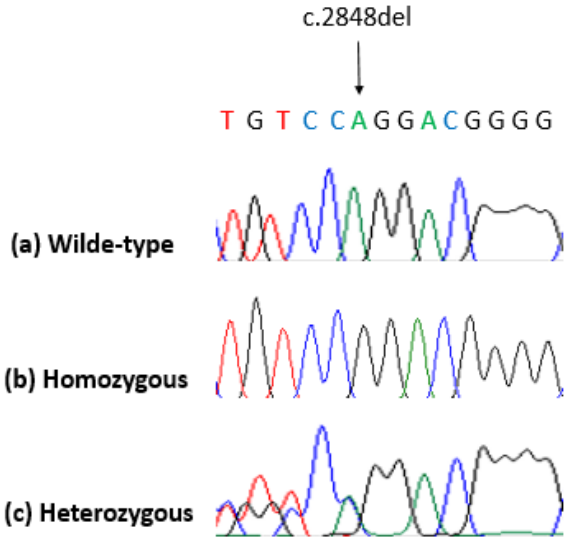


Figure (14): Electropherogram showing the DNA sequence of the *TOP3A* (c.2848del) frameshift deletion variant (black arrow). (a) A wild-type state in an unaffected subject (PL0496), **(b)** homozygous state in an MPD affected subject (PL0504), and **(c)** heterozygous state in a carrier unaffected individual (PL0509).

CHAPTER 4: DISSCUSSION

We investigated clinical manifestations and genetic components of a complex form of hereditary spastic paraplegia (cHSP) in six of eight affected individuals. The affected individuals belong to an extended Palestinian family that constitutes a whole village located in the southwest of Jenin. Bioinformatics analysis of WES data of two of the affected individuals revealed that cHSP is associated with a biallelic *PTPN23* variant. The identified variant, c. (4719 delCCinsA), is located in the last exon of *PTPN23* gene (exon 25) and is predicted to escape nonsense-mediated decay (NMD). This frameshift deletion in exon 25 is predicted to produce *PTPN23* polypeptide that modestly truncated with an altered C-terminus (1572/1636 amino acids). This variant occurs in the CTD region which is rich in proline and is located subsequent to the tyrosine–protein phosphatase (TPP) domain **(Figure 1)**. This region does not yet have a specific function. In that regard, the HIS domain of *PTPN23*, which is likewise a proline-rich domain, was found to be important for different signaling pathways through its interaction with the Grb2 family of protein (Tanase 2010). However, in order to confirm the effect of the p. (P1572T fs*12) on *PTPN23* functionality, molecular studies should be conducted.

In previous studies, *PTPN23* was suggested as a candidate gene for severe neurological impairment and seizures. In a study conducted by *Sowada et al.* on a single case diagnosed with a rare form of autosomal-recessive developmental and epileptic encephalopathy (DEE), two heterozygous pathogenic *PTPN23* variants [c.(1595C>T): p.(Pro532Leu)] and [c.(3586C>T): p.(Arg1196*)] were reported to be the causative variants (Sowada et al. 2017). *Smigiel et al.* also reported two *PTPN23* pathogenic variants [c.(1902C>G): p.(Asn634Lys)] and [c.(2974delC): p.(Leu992Tyrfs*168)] in compound heterozygous state in a single pediatric case affected by developmental epileptic encephalopathy with hypomyelination and brain atrophy (Smigiel et al.

2018). Another three individual studies identified three *PTPN23* pathogenic variants in homozygous state in three unrelated single cases diagnosed with seizures, spasticity, global developmental delay, and motor delay **(Table S1)** (Alazami et al. 2015, Bend et al. 2020,Trujillano et al. 2017). The pathogenicity of two of the previously identified biallelic candidate *PTPN23* gene variants by *Bend et al.* in individuals diagnosed with varying degrees of neurodevelopmental impairment and structural brain abnormalities are recently reported to be not supported by the accessible gnomAD allele frequency data (Khalaf-Nazzal et al. 2021). In addition, clinical manifestations described in another three patients by *Bend et al.* are also found to be non-specific and to significantly differ from the mostly neurological consequences and outcomes related to the originally identified pathogenic *PTPN23* variants (Khalaf-Nazzal et al. 2021).

Regarding clinical manifestations of the cHSP patients investigated in our study, all of them displayed microcephaly, lower limb weakness and stiffness, hyperreflexia, muscle wasting, and extensor plantar response. The steppage gait of the patients shows characteristics of spasticity and foot drop. The sensory examination was normal in all of them, and mild intellectual impairment was noted in the majority. Heterogeneity of the clinical manifestations that presented in those patients suggests the involvement of upper and lower motor neuropathy. This noted complicated form of hereditary spastic paraplegia is similar to that described in patients affected by Silver syndrome/SPG17 (Musacchio et al. 2017). In the study conducted by *Bend et al*., a European eleven years old female was reported to have a homozygous pathogenic *PTPN23* variant [c.(4651_4652dup): p.(Leu1552Hisfs*33)]. The patient shared similar clinical signs to those manifested by our patients, in particular microcephaly, developmental regression at 2 years, mild intellectual disability, optic atrophy, spastic diplegia, hyperreflexia, hypotonia, slow gait, delayed myelination, and language delay. The study reported a frameshift variant located in the last exon

which was predicted to escape non-mediated decay (NMD), and thus resulting in a truncated polypeptide product. This variant, similar to our identified variant, is found to be associated with mild severity manifestations compared to the previously identified variants in the major conserved *PTPN23* domains; BRO1-like domain, ALIX domain, and tyrosine-protein phosphatase domain (Alazami et al. 2015, Smigiel et al. 2018,Trujillano et al. 2017, Sowada et al. 2017). Variants located in these domains result in more severe neurological clinical signs including seizures, progressive spasticity, brain atrophy, and severe white matter volume loss. This suggests that the variant we identified in the CTD could result in a partially functioning polypeptide.

As mentioned previously, one of the important *PTPN23* roles is sorting of the endosomal cargo into multivesicular bodies through its interaction with the ESCRT-I complex. ALIX/ESCRT interaction was reported to play critical roles in the survival of embryonic neural progenitors, as well as the growth and maintenance of axons and dendrites, all of which are required for the development of a functional brain (Sadoul et al. 2018). In previous studies, several truncating variants in *UBAP1* (Ubiquitin Associated Protein 1) gene, encodes a subunit of the ESCRT-I, were found to be associated with childhood-onset progressive SPG (Gu et al. 2020, Farazi Fard et al. 2019). Furthermore, *Connell et al*. also documented the role interaction between spastin (the most common cause of HSP4) and the ESCRT-III complex-associated endosomal protein CHMP1B in regulating polarized membrane traffic (Connell et al.2020). These together may suggest the critical function of *PTPN23* in motor neurons integrity.

In regards to the cystic fibrosis (CF) patients belonging to the same extended Palestinian family, we segregated two of the CFTR variants in all of the patients and family members. All of the CF patients in our study presented with pulmonary and gastrointestinal symptoms manifested chronic pulmonary infections, chronic diarrhea, and failure to thrive. The first variant, [c. (2183 AA>G):

p.(Lys684Serfs∗38)], is a deletion/insertion frameshift variant which occurs in exon 14 of the *CFTR* gene. This variant is rare and is not found in gnomAD or ExAC. It is considered to be a target for NMD, and results in a half-truncated *CFTR* polypeptide product (721/1481).

The second variant [(c. (3909C>G): p. (N1303K)] is a missense substitution which occurs in exon 21 of the CFTR gene and resulting in changing of the conserved asparagine residue at position 1303 to lysine. It's a pathogenic variant according to the ACMG classification guidelines (https://varsome.com/). p. (N1303K) is a common variant, with the highest recorded frequency in Lebanon, and is found to be associated with severe pancreatic and pulmonary phenotypes in approximately all of the Lebanese CF patients (Farhat et al. 2015). However, the presence of this variant in compound heterozygous with other *CFTR* variants may contribute to the clinical phenotype severity (Farhat et al. 2015).

The segregation analysis showed that two of our CF patients carry the c. (2183 AA>G) variant in a homozygous state and both parents are carriers for the variant. The remaining five patients carry the same variant in a compound heterozygous state with the second variant p. (N1303K), while one parent is heterozygous for the first variant and the other is heterozygous for the second variant **(Figure 13)**. Altogether, twelve individuals are heterozygous carrier for the c. (2183 AA>G) variant, and two are homozygous, four are heterozygous carrier for the p. (N1303K) variant, five are compound heterozygous for both of variants **(Figure 5)**.

In a previous study conducted on Palestinian CF patients residing in Gaza and the West Bank, the c. (2183 AA>G) variant was found to exist among Palestinian patients living in Gaza in a homozygous state (Essawi et al. 2015). However, here, we report here the presence of this variant among Palestinians living in the West Bank. In another cohort study of *CFTR* variants among CF

patients in 21 unrelated Palestinian families residing in the central and southern of Palestine, c.1393-1G>A variant was reported to be the most prevalent one (Siryani et al. 2015). In this study, the p. (N1303K) variant was identified in homozygous state in one patient residing in Hebron (Siryani et al. 2015). Narratives state that the extended Palestinian family we are studying traces back to south of Palestine, particularly the city of Hebron. This suggests that the p. (N1303K) variant is a founder mutation previously identified in Hebron.

Despite the fact that CF is less common in Palestine, many cases exist, and the causative variants must be discovered. The Palestinian population has a narrow range of CF mutations due to its homozygous ethnic heritage. Identifying the most prevalent *CFTR* mutations in the Palestinian population and establishing a particular mutations screening test would allow to diagnose patients earlier and more quickly. It will also qualify the population-screening program aimed at detecting carriers of the *CFTR* variants. Further studies are needed to elucidate the distribution and frequency of *CFTR* gene mutations in the different districts of Palestine.

Our study investigates clinical and genetic findings of microcephaly primordial dwarfism (MPD) associated with biallelic *TOP3A* variant in eight patients belonging to an extended Palestinian family residing in a village located east of Jenin. The discovered variant in our study, [c.(2848del): p.(Trp950Glyfs*58)], is a frameshift-deletion occurring in the last exon (exon 19) of *TOP3A* gene. This variant causes a frameshift in conserved 58 amino acids among different species, thus, it is predicted to change the amino acid sequence. In addition, it results in changing in the C-terminal region due to loss of the original stop codon which in turn results in prolongation of the produced polypeptide (+5 aa) to a length close to its original length. This change in the amino acid sequence may impact protein structure and function, which can be validated by further protein modeling and molecular studies. The identified variant was validated using Sanger sequencing to be homozygous

in all affected individuals and heterozygous in all parents. This variant is considered rare and is not reported in gnomAD, ExAC, ClinVar, Ensemble or Decipher.

All of affected individuals recruited in our study presented prenatal delayed growth, microcephaly, short statute, poor weight gain, and small facial features with large nose **(Table 3)**. There were no signs of intellectual disability or dysmorphic features in any of them. These symptoms overlap with symptoms of Bloom-syndrome (BS) patients except for the no presence of photosensitive malar rash (J. German and Sanz 2013). Generally, Bloom-like syndrome that associated with *TOP3A* deficiency is a recent reported human disease (Jiang et al. 2021). So far, ten different *TOP3A* pathogenic variants of this disease have been identified in thirteen affected individuals in two recent separate studies **(Table S2-A)** (Jiang et al. 2021, Martin et al. 2018). Consistent with our clinical findings, all of studied affected individuals by *Martin et al.* and *Jiang et al.* were found to have restriction of prenatal-onset growth, microcephaly, short stature, and poor weight gain. However, some of patients presented with other clinical phenotypes including mild developmental delay, gastroesophageal reflux, recurrent infections, cardiomyopathy, hepatomegaly, fatty liver, Café-au-lait macules, and reduced subcutaneous fat **(Table S2A-C)**. Mild signs in our patients may indicate that the p.(Trp950Glyfs*58) variant gives a polypeptide product that retains partial functionality of *TOP3A*. Contrary to what was found in our study and the study by Martin et al. about failure of growth-hormone (GH) treatment in certain patients, *Jiang et al.* reported good response to treatment in one patient. However, due to the given overlap with BS, and the possibility of cancer predisposition, it is important to call against growth hormone treatment for these patients to avoid the development of GH treatment associated malignancies.

In regards to cytogenetic examination, SCEs were found to be elevated in all of the available patients in the two previous studies, implying that the *TOP3A* variants impair the BTR complex

function. As mentioned previously, elevated SCEs is used as a confirmatory diagnostic test for BS. However, the presence of this clinical feature in patients with *TOP3A*, *RMI1* and *RMI2* mutations in Bloom-syndrome-like disorders suggests that individuals who were previously identified as BS patients, based on the cytogenetic examination, should be reevaluated.

Genomic instability arise from BS makes it to be considered a cancer-prone disorder (James German 1997). Malignancies are typically developing at early adulthood in BS patients (Arora et al. 2014). In our study, one affected individual developed colon cancer at the age of 35 years. This is considered the first case where TOP3A mutations are associated with cancer. Four of the remaining patients are still children and two are in their forties and have not been diagnosed with cancer. Whether cancer occurring in our patients is due to *TOP3A* mutations, is to be further delineated by identifying more cases, and following up with our patients.

It is well established that three subfamilies of DNA topoisomerases (IA, IIA, IB) exist in human (Wang, Lyu, and Wang 2002). Topoisomerase IA is found in two isoforms, topo IIIα (*TOP3A*) and IIIβ (*TOP3B*), with different roles in cellular development (Bizard and Hickson 2020). Knocking out *TOP3A* in mice was reported to be lethal, indicating its essential role in the cell viability (W. Li and Wang 1998). *TOP3A* enzyme is found to be localized in the nucleus and the mitochondria (Wang, Lyu, and Wang 2002). In the nucleus, *TOP3A* acts as a dissolution enzyme (dissolvasome) with the *BLM* helicase, *RMI1*, and *RMI2* proteins to form the BTR complex which catalyze dHJs dissolution to yield non-crossover products (Sarbajna and West 2014). This reaction is critical in somatic cell for avoiding crossover which if uncontrolled can result in loss of heterozygosity and cancer susceptibility (Sarbajna and West 2014). *TOP3A* has also been recently considered as an essential component of the mitochondrial DNA (mtDNA) replication and mtDNA separation machinery (Nicholls et al. 2018). In agreement with this function, Martin et al. reported

one patient with *TOP3A* mutation to have chronic progressive external ophthalmoplegia (PEO) and cerebellar ataxia which are considered as clinical signs associated with mitochondrial dysfunction (Martin et al. 2018). However, this patient has no signs of short stature or elevated SCEs. Cellular and molecular studies of *Jiang et al*. in two patients who carried *TOP3A* variants and displayed hepatomegaly showed that these variants resulted in impaired mitochondrial metabolism (reduced mitochondrial ATP generation and oxygen consumption rates) **(Table S3)**. The authors reported that the variants didn't affect genome stability (Jiang et al. 2021).

In agreement with the previously mentioned essential roles of *TOP3A* in the cell **(Table S3)**, the clinical characteristics of our patients indicate that the discovered variant is a severe hypomorph. The notable growth restriction and absence of clinical features of mitochondrial dysfunction suggest that the pathogenesis mechanism is related to the nuclear function of *TOP3A* as part of the BTR complex. In summary, mutations in *TOP3A* reduced the functional performance of the BTR complex in mitotic cells as in the *BLM* mutations in BS. Thus, resulting in microcephaly and growth restriction due to genome instability and impaired viability during development that emerge from micronuclei, chromosomal breaks, and chromosomal aberrations (Payne and Hickson 2009). Further molecular and cellular studies are required to confirm the pathogenesis mechanism of p. (Trp950Glyfs*58) variant in *TOP3A* gene.

In the present study, we also investigated the clinical characteristics and genetic cause of Rett syndrome (RTT) in a single case in a Palestinian family residing in another village in Jenin. Bioinformatics analysis of WGS data of the affected individual revealed that the RTT is associated with one heterozygous variant in x-linked methyl-CpG-binding protein 2 (*MECP2)* gene. The identified variant, [c. (1231_1233 del): p.(Ser411del)], is an inframe deletion located in the last exon of *MECP2* gene (exon 4). This variant is considered rare with frequency of 0% in gnomAD.

It results in deletion of the conserved serine amino acid at position 411 of the C-terminal domain (CTD) of the MeCP2 protein. It is predicted by MutationTaster to change the amino acid sequence and some of the protein features including loss of phosphorylation sites in the disordered region of the protein. The functional effect of this variant on MeCP2 can be further investigated by molecular studies.

MECP2 de novo mutations are considered to be responsible for 95% of typical RTT cases and more than 50% of atypical RTT cases (Pejhan and Rastegar 2021). In our study, Sanger sequencing verified that both parents, as well as the second daughter, are healthy and do not carry the identified variant, p.(Ser411del). Therefore, this variant is predicted to be a *de novo* mutation.

The majority of RTT patients are female heterozygotes with somatic mosaics for both wild-type and mutant *MECP2* (Katz et al. 2016). The affected individual in our study is a 6 years-old female, of first cousins' parents, clinically diagnosed with neurodevelopmental regression with exhibition of behavioral problems at 2 years, and in later ages, cessation of speech, hyperactivity, sleep disturbances, trichotillomania, and aggression. These progressive clinical manifestations are consistent with the established RTT symptoms and stages of the disease (Hagberg 2002). Severity of the RTT symptoms mainly depends on the type of *MECP2* gene variant (Neul et al. 2008). The most severe RTT symptoms were found in individuals who are carrier for mutations in the major MeCP2 domains (Pejhan and Rastegar 2021). The milder phenotype of the p.(Ser411del) variant in our patient compared to the other phenotypes previously reported in RTT patients with specific *MECP2* variants could be attributed to its occurrence in the CTD. Together with the methyl-CpG-binding domain (MBD), CTD has been shown to contribute to MeCP2 binding to the nucleosome core (Chandler et al. 1999). *Smeets et al.* found that truncation of hot spot deletions in the C-

terminal domain which occurs beyond the coding regions that doesn't involve the nuclear localization signal, explains their milder clinical symptoms in RTT patients.

As a global transcriptional repressor, MeCP2 represses the expression of the Brain-derived neurotrophic factor (*BDNF*) gene via its selectively binding to BDNF promoter III. Separation of MeCP2 from BDNF promoter III through depolarization of neurons membrane, which triggers calcium-dependent phosphorylation of *MECP2*, allows the expression of *BDNF* (Chen et al. 2003). MeCP2 regulation by calcium influx could be crucial for nervous system function as it plays important role in regulating certain programs of activity-dependent gene transcription in brain (Zhou et al. 2006). Disruption of MeCP2 phosphorylation by *MECP2* variants on serine residues in the C-terminal domain was found to affect the neuronal activity. Zhou et al reported that mutation in serine residue at position 421 of *MECP2* resulted in blocking the ability of MeCP2 in controlling of spine morphogenesis, dendritic growth restriction, and induction of BDNF transcription (Zhou et al. 2006). It is expected that this disruption may underlie the neural-specific pathogenesis that was diagnosed in the RTT patient in our study. Sanger sequencing failed to confirm our identified variant p.(Ser411del), therefore this variant is predicted to be a longer deletion and it may affect Ser421, which has been evidenced to be an important site for MeCP2 phosphorylation. Techniques other than Sanger sequencing are ongoing to confirm this variant.

In our study, we deciphered the causative variants of three rare diseases, cHSP, MPD, and RTT, that were found to have occurred among extended families in isolated communities in Palestine. Discovering these variants will help in studying the genetic makeup of these isolated communities and building a map for the founder variants in a national screening theme. In addition, it will help highlight the genes that need to be included on the international gene testing-diagnostic panels for the HSP and MPD diseases. Furthermore, studying of rare diseases can be considered as human

genetic models that may help in the identification of common molecular pathways and thus better understanding of common diseases.

CONCLUSION

By all of our findings, we call attention to the importance of WES and WGS as a promising tools for rare-disease diagnosis and management in Palestine. We define a complex form of HSP, that may include intellectual impairment or may not besides microcephaly, caused by pathogenic biallelic variant in the PTPN23 gene that was detected by WES. Along with the findings of the previous studies, we support the importance of the inclusion of PTPN23 on the international HSP gene testing-diagnostic panels. We also define a novel pathogenic biallelic variant in TOP3A gene associated with mild phenotype of MPD using WGS. Furthermore, WGS is proven as a powerful tool in the detection of *de novo* CNV variant that may disrupt the MeCP2 phosphorylation, affecting the neuronal activity and resulting in RTT syndrome. In the Palestinian isolated populations and the large extended consanguineous families, rare diseases are no longer rare. Failure to address this issue may exacerbate the transmission of these diseases from one generation to the next. Therefore, scientific collaboration with such populations is essential to investigate and characterize such diseases, find the causative pathogenic mutations, and establish future management plans.

REFERECNES

Abdel-Salam, Ghada M.H. et al. 2011. “A Homozygous Mutation in RNU4ATAC as a Cause of Microcephalic Osteodysplastic Primordial Dwarfism Type I (MOPD I) with Associated Pigmentary Disorder.” *American Journal of Medical Genetics, Part A*.

Al-Gazali, Lihadh, and Hanan Hamamy. 2014. “Consanguinity and Dysmorphology in Arabs.” *Human Heredity*.

Al-Gazali, Lihadh, Hanan Hamamy, and Shaikha Al-Arrayad. 2006. “Genetic Disorders in the Arab World.” *British Medical Journal*.

Alazami, Anas M. et al. 2015. “Accelerating Novel Candidate Gene Discovery in Neurogenetic Disorders via Whole-Exome Sequencing of Prescreened Multiplex Consanguineous Families.” *Cell Reports*.

Alibakhshi, Reza et al. 2008. “Analysis of the CFTR Gene in Iranian Cystic Fibrosis Patients: Identification of Eight Novel Mutations.” *Journal of Cystic Fibrosis*.

Arora, Harleen et al. 2014. “Bloom Syndrome.” *International Journal of Dermatology*.

Banerjee, Abhishek et al. 2019. “Towards a Better Diagnosis and Treatment of Rett Syndrome: A Model Synaptic Disorder.” *Brain*.

Bend, Renee et al. 2020. “Phenotype and Mutation Expansion of the PTPN23 Associated Disorder Characterized by Neurodevelopmental Delay and Structural Brain Abnormalities.” *European Journal of Human Genetics*.

Bick, David et al. 2019. “Case for Genome Sequencing in Infants and Children with Rare, Undiagnosed or Genetic Diseases.” *Journal of Medical Genetics*.

Bizard, Anna H., and Ian D. Hickson. 2020. “The Many Lives of Type IA Topoisomerases.” *Journal of Biological Chemistry*.

Blackstone, Craig. 2018. “Hereditary Spastic Paraplegia.” In *Handbook of Clinical Neurology*,.

Boycott, Kym M. et al. 2019. “A Diagnosis for All Rare Genetic Diseases: The Horizon and the Next Frontiers.” *Cell*.

Brancati, Francesco, Marco Castori, Rita Mingarelli, and Bruno Dallapiccola. 2005. “Majewski Osteodysplastic Primordial Dwarfism Type II (MOPD II) Complicated by Stroke: Clinical Report and Review of Cerebral Vascular Anomalies.” *American Journal of Medical Genetics*.

Brunetti, Sara, and Daniel E. Lumsden. 2020. “Rett Syndrome as a Movement and Motor Disorder – A Narrative Review.” *European Journal of Paediatric Neurology*.

Castaldo, G. et al. 1996. “Clinical Features of Cystic Fibrosis Patients with Rare Genotypes.” *Journal of Medical Genetics*.

Chandler, Simon P., Dmitry Guschin, Nicoletta Landsberger, and Alan P. Wolffe. 1999. “The Methyl-CpG Binding Transcriptional Repressor MeCP2 Stably Associates with Nucleosomal DNA.” *Biochemistry*.

Chen, Wen G. et al. 2003. “Derepression of BDNF Transcription Involves Calcium-Dependent Phosphorylation of MeCP2.” *Science*.

Cheptou, Pierre-Olivier, Anna L. Hargreaves, Dries Bonte, and Hans Jacquemyn. 2017. “Correction to ‘Adaptation to Fragmentation: Evolutionary Dynamics Driven by Human Influences.’” *Philosophical Transactions of the Royal Society B: Biological Sciences*.

Connell, James W. et al. 2020. “ESCRT-III-Associated Proteins and Spastin Inhibit Protrudin-Dependent Polarised Membrane Traffic.” *Cellular and Molecular Life Sciences*.

Dauber, Andrew et al. 2012. “Novel Microcephalic Primordial Dwarfism Disorder Associated with Variants in the Centrosomal Protein Ninein.” *Journal of Clinical Endocrinology and Metabolism*.

Dawkins, Hugh J.S. et al. 2018. “Progress in Rare Diseases Research 2010–2016: An IRDiRC Perspective.” *Clinical and Translational Science*.

Doyotte, Aurelie, Aleksandr Mironov, Edward McKenzie, and Philip Woodman. 2008. “The Bro1-Related Protein HD-PTP/PTPN23 Is Required for Endosomal Cargo Sorting and Multivesicular Body Morphogenesis.” *Proceedings of the National Academy of Sciences of the United States of America*.

Elborn, J. Stuart. 2016. “Cystic Fibrosis.” *The Lancet*.

Ellard, Sian et al. 2020. “ACGS Best Practice Guidelines for Variant Classification 2020 v4.01.” *Association for Clinical Genetic Science*.

Essawi, O. et al. 2015. “Next Generation Sequencing to Determine the Cystic Fibrosis Mutation Spectrum in Palestinian Population.” *Disease Markers*.

Faber, Ingrid et al. 2017. “Hereditary Spastic Paraplegia from 1880 to 2017: An Historical Review.” *Arquivos de Neuro-Psiquiatria*.

Farazi Fard, Mohammad Ali et al. 2019. “Truncating Mutations in UBAP1 Cause Hereditary Spastic Paraplegia.” *American Journal of Human Genetics*.

Farhat, Raëd et al. 2015. “N1303K (c.3909C>G) Mutation and Splicing: Implication of Its c.[744-33GATT(6); 869+11C>T] Complex Allele in CFTR Exon 7 Aberrant Splicing.” *BioMed Research International*.

Farnaes, Lauge et al. 2018. “Rapid Whole-Genome Sequencing Decreases Infant Morbidity and Cost of Hospitalization.” *npj Genomic Medicine*.

Farrell, Philip M. et al. 2017. “Diagnosis of Cystic Fibrosis: Consensus Guidelines from the Cystic Fibrosis Foundation.” *Journal of Pediatrics*.

Fernandez Fernandez, Elena, Chiara De Santi, Virginia De Rose, and Catherine M. Greene. 2018. “CFTR Dysfunction in Cystic Fibrosis and Chronic Obstructive Pulmonary Disease.” *Expert Review of Respiratory Medicine*.

Frésard, Laure, and Stephen B. Montgomery. 2018. “Diagnosing Rare Diseases after the Exome.” *Cold Spring Harbor Molecular Case Studies*.

German, J., and M. Sanz. 2013. “Bloom’s Syndrome.” In *Brenner’s Encyclopedia of Genetics: Second Edition,*.

German, James. 1997. “Bloom’s Syndrome. XX. The First 100 Cancers.” *Cancer Genetics and Cytogenetics*.

Gu, Shen et al. 2020. “Truncating Variants in UBAP1 Associated with Childhood-Onset

Nonsyndromic Hereditary Spastic Paraplegia." *Human Mutation*.

Hagberg, Bengt. 2002. "Clinical Manifestations and Stages of Rett Syndrome." *Mental Retardation and Developmental Disabilities Research Reviews*.

Hanai, Ryo, Paul R. Caron, and James C. Wang. 1996. "Human TOP3: A Single-Copy Gene Encoding DNA Topoisomerase III." *Proceedings of the National Academy of Sciences of the United States of America*.

Harripaul, R. et al. 2017. "Mapping Autosomal Recessive Intellectual Disability: Combined Microarray and Exome Sequencing Identifies 26 Novel Candidate Genes in 192 Consanguineous Families." *Molecular Psychiatry*.

Hindorff, Lucia A. et al. 2018. "Prioritizing Diversity in Human Genomics Research." *Nature Reviews Genetics*.

Jiang, Wenjun et al. 2021. "Predominant Cellular Mitochondrial Dysfunction in the TOP3A Gene-Caused Bloom Syndrome-like Disorder." *Biochimica et Biophysica Acta - Molecular Basis of Disease*.

Jooma, Sonya, Michael J. Hahn, Lucia A. Hindorff, and Vence L. Bonham. 2019. "Defining and Achieving Health Equity in Genomic Medicine." *Ethnicity and Disease*.

Katz, David M. et al. 2016. "Rett Syndrome: Crossing the Threshold to Clinical Translation." *Trends in Neurosciences*.

Khalaf-Nazzal, Reham et al. 2021. "Final Exon Frameshift Biallelic Ptpn23 Variants Are Associated with Microcephalic Complex Hereditary Spastic Paraplegia." *Brain Sciences*.

Khetarpal, Preeti, Satrupa Das, Inusha Panigrahi, and Anjana Munshi. 2016. "Primordial Dwarfism: Overview of Clinical and Genetic Aspects." *Molecular Genetics and Genomics*.

Klebe, S., G. Stevanin, and C. Depienne. 2015. "Clinical and Genetic Heterogeneity in Hereditary Spastic Paraplegias: From SPG1 to SPG72 and Still Counting." *Revue Neurologique*.

Klingseisen, Anna, and Andrew P. Jackson. 2011. "Mechanisms and Pathways of Growth Failure in Primordial Dwarfism." *Genes and Development*.

Kristiansson, Kati, Jussi Naukkarinen, and Leena Peltonen. 2008. “Isolated Populations and Complex Disease Gene Identification.” *Genome Biology*.

Kyle, Stephanie M., Neeti Vashi, and Monica J. Justice. 2018. “Rett Syndrome: A Neurological Disorder with Metabolic Components.” *Open Biology*.

Li, Wei, and James C. Wang. 1998. “Mammalian DNA Topoisomerase IIIα Is Essential in Early Embryogenesis.” *Proceedings of the National Academy of Sciences of the United States of America*.

Li, Xia et al. 2021. “Prenatal Diagnosis of Meier-Gorlin Syndrome 7: A Case Presentation.” *BMC Pregnancy and Childbirth*.

Lionel, Anath C. et al. 2018. “Improved Diagnostic Yield Compared with Targeted Gene Sequencing Panels Suggests a Role for Whole-Genome Sequencing as a First-Tier Genetic Test.” *Genetics in Medicine*.

Liu, Yilun, and Stephen C. West. 2008. “More Complexity to the Bloom’s Syndrome Complex.” *Genes and Development*.

Liyanage, Vichithra R.B., and Mojgan Rastegar. 2014. “Rett Syndrome and MeCP2.” *NeuroMolecular Medicine*.

M., Hussain et al. 2016. “Exome Sequencing of Pakistani Consanguineous Families Identifies 30 Novel Candidate Genes for Recessive Intellectual Disability.” *Molecular Psychiatry*.

Major, Tamás et al. 2021. “Founder Effects in Hereditary Hemorrhagic Telangiectasia.” *Journal of Clinical Medicine*.

Martin, Carol Anne et al. 2018a. “Erratum: Mutations in TOP3A Cause a Bloom Syndrome-like Disorder (The American Journal of Human Genetics (2018) 103(2) (221–231), (S0002929718302283) (10.1016/j.Ajhg.2018.07.001)).” *American Journal of Human Genetics*.

Martin, C. A. 2018b. “Mutations in TOP3A Cause a Bloom Syndrome-like Disorder.” *American Journal of Human Genetics*.

Meng, Xin et al. 2017. “The Cystic Fibrosis Transmembrane Conductance Regulator (CFTR)

and Its Stability." *Cellular and Molecular Life Sciences*.

Monteiro, Andre et al. 2021. "Intracranial Aneurysms in Microcephalic Primordial Dwarfism: A Systematic Review." *Journal of NeuroInterventional Surgery*.

Mukherjee, Kusumika. 2019. "Care for Rare: Spotlight on Rare Diseases." *Trends in Pharmacological Sciences*.

de Munnik, Sonja A. et al. 2015. "Meier-Gorlin Syndrome." *Orphanet Journal of Rare Diseases*.

Musacchio, Thomas et al. 2017. "ALS and MMN Mimics in Patients with BSCL2 Mutations: The Expanding Clinical Spectrum of SPG17 Hereditary Spastic Paraplegia." *Journal of Neurology*.

Muzzey, Dale, Eric A. Evans, and Caroline Lieber. 2015. "Understanding the Basics of NGS: From Mechanism to Variant Calling." *Current Genetic Medicine Reports*.

Nadjari, Michael et al. 2000. "Ultrasonographic Prenatal Diagnosis of Microcephalic Osteodysplastic Primordial Dwarfism Types I/III." *Prenatal Diagnosis*.

Neul, J. L. et al. 2008. "Specific Mutations in Methyl-CpG-Binding Protein 2 Confer Different Severity in Rett Syndrome." *Neurology*.

Nguengang Wakap, Stéphanie et al. 2020. "Estimating Cumulative Point Prevalence of Rare Diseases: Analysis of the Orphanet Database." *European Journal of Human Genetics*.

Nicholls, Thomas J. et al. 2018. "Topoisomerase 3α Is Required for Decatenation and Segregation of Human MtDNA." *Molecular Cell*.

Noel, Sabrina, Isabelle Sermet-Gaudelus, and David N. Sheppard. 2018. "N1303K: Leaving No Stone Unturned in the Search for Transformational Therapeutics." *Journal of Cystic Fibrosis*.

Nykamp, Keith et al. 2017. "Sherloc: A Comprehensive Refinement of the ACMG-AMP Variant Classification Criteria." *Genetics in Medicine*.

Ortega, Roberta Paiva Magalhães, and Sérgio Rosemberg. 2019. "Hereditary Spastic Paraplegia: A Clinical and Epidemiological Study of a Brazilian Pediatric Population." *Arquivos de*

Neuro-Psiquiatria.

Payne, Miranda, and Ian D. Hickson. 2009. "Genomic Instability and Cancer: Lessons from Analysis of Bloom's Syndrome." In *Biochemical Society Transactions,*.

Pejhan, Shervin, and Mojgan Rastegar. 2021. "Role of Dna Methyl-Cpg-Binding Protein Mecp2 in Rett Syndrome Pathobiology and Mechanism of Disease." *Biomolecules*.

Pogue, Robert E. et al. 2018. "Rare Genetic Diseases: Update on Diagnosis, Treatment and Online Resources." *Drug Discovery Today*.

Polymeris, Alexandros A. et al. 2016. "A Series of Greek Children with Pure Hereditary Spastic Paraplegia: Clinical Features and Genetic Findings." *Journal of Neurology*.

Prokop, Jeremy W. et al. 2018. "Genome Sequencing in the Clinic: The Past, Present, and Future of Genomic Medicine." *Physiological Genomics*.

Rafeeq, Misbahuddin M., and Hussam Aly Sayed Murad. 2017. "Cystic Fibrosis: Current Therapeutic Targets and Future Approaches." *Journal of Translational Medicine*.

Richards, Sue et al. 2015. "Standards and Guidelines for the Interpretation of Sequence Variants: A Joint Consensus Recommendation of the American College of Medical Genetics and Genomics and the Association for Molecular Pathology." *Genetics in Medicine*.

Sadoul, Rémy et al. 2018. "The Role of ESCRT during Development and Functioning of the Nervous System." *Seminars in Cell and Developmental Biology*.

Saeidi, Minoo, and Morteza Shahbandari. 2020. "A Child with Seckel Syndrome and Arterial Stenosis: Case Report and Literature Review." *International Medical Case Reports Journal*.

Sandweiss, Alexander J., Vicky L. Brandt, and Huda Y. Zoghbi. 2020. "Advances in Understanding of Rett Syndrome and MECP2 Duplication Syndrome: Prospects for Future Therapies." *The Lancet Neurology*.

Sarbajna, Shriparna, and Stephen C. West. 2014. "Holliday Junction Processing Enzymes as Guardians of Genome Stability." *Trends in Biochemical Sciences*.

Shahin, Walaa Aboulkasem, Dina Ahmed Mehaney, and Mona Mostafa El-Falaki. 2016.

"Mutation Spectrum of Egyptian Children with Cystic Fibrosis." *SpringerPlus*.

Shalev, S. A. 2019. "Characteristics of Genetic Diseases in Consanguineous Populations in the Genomic Era: Lessons from Arab Communities in North Israel." *Clinical Genetics*.

Shorrocks, Ann Marie K. et al. 2021. "The Bloom Syndrome Complex Senses RPA-Coated Single-Stranded DNA to Restart Stalled Replication Forks." *Nature Communications*.

Siryani, Issa et al. 2015. "Distribution of Cystic Fibrosis Transmembrane Conductance Regulator (CFTR) Mutations in a Cohort of Patients Residing in Palestine." *PLoS ONE*.

Smeets, E. et al. 2005. "Rett Syndrome in Females with CTS Hot Spot Deletions: A Disorder Profile." *American Journal of Medical Genetics*.

Smigiel, Robert et al. 2018. "Developmental Epileptic Encephalopathy with Hypomyelination and Brain Atrophy Associated with PTPN23 Variants Affecting the Assembly of UsnRNPs." *European Journal of Human Genetics*.

de Souza, Paulo Victor Sgobbi et al. 2017. "Hereditary Spastic Paraplegia: Clinical and Genetic Hallmarks." *Cerebellum*.

Souza, Paulo Victor Sgobbi de et al. 2017. "New Genetic Causes for Complex Hereditary Spastic Paraplegia." *Journal of the Neurological Sciences*.

Sowada, Nadine et al. 2017. "Mutations of PTPN23 in Developmental and Epileptic Encephalopathy." *Human Genetics*.

Stutts, M. Jackson et al. 1995. "CFTR as a CAMP-Dependent Regulator of Sodium Channels." *Science*.

Tadmouri, G. O. 2006. "CTGA: The Database for Genetic Disorders in Arab Populations." *Nucleic Acids Research*.

Tanase, Carmen Alexandra. 2010. "Histidine Domain-Protein Tyrosine Phosphatase Interacts with Grb2 and GrpL." *PLoS ONE*.

Ting, Chun Yi et al. 2020. "Further Delineation of CDC45-Related Meier-Gorlin Syndrome with Craniosynostosis and Review of Literature." *European Journal of Medical Genetics*.

Toft, Anders et al. 2019. “Peripheral Neuropathy in Hereditary Spastic Paraplegia Caused by REEP1 Variants.” *Journal of Neurology*.

Toyooka, Shin ichi et al. 2000. “HD-PTP: A Novel Protein Tyrosine Phosphatase Gene on Human Chromosome 3p21.3.” *Biochemical and Biophysical Research Communications*.

Trujillano, Daniel et al. 2017. “Clinical Exome Sequencing: Results from 2819 Samples Reflecting 1000 Families.” *European Journal of Human Genetics*.

Ulintz, Peter J., Weisheng Wu, and Chris M. Gates. 2019. “Bioinformatics Analysis of Whole Exome Sequencing Data.” In *Methods in Molecular Biology*,.

van der Lely, Lisa et al. 2019. “Loss of PTPN23 Promotes Proliferation and Epithelial-to-Mesenchymal Transition in Human Intestinal Cancer Cells.” *Inflammatory Intestinal Diseases*.

Vidal, Silvia et al. 2019. “Genetic Landscape of Rett Syndrome Spectrum: Improvements and Challenges.” *International Journal of Molecular Sciences*.

Wang, Yong, Yi Lisa Lyu, and James C. Wang. 2002. “Dual Localization of Human DNA Topoisomerase IIIα to Mitochondria and Nucleus.” *Proceedings of the National Academy of Sciences of the United States of America*.

Welsh, Michael J., and Alan E. Smith. 1993. “Molecular Mechanisms of CFTR Chloride Channel Dysfunction in Cystic Fibrosis.” *Cell*.

Xianghu, Qu et al. 2001. “Cloning, Chromosomal Assignment and Tissue Expression of a Novel Human Protein Tyrosine Phosphatase (PTP-TD14) Gene.” *Yi chuan = Hereditas* 23(6): 503—510. http://europepmc.org/abstract/CBA/354754.

Yang, Jay et al. 2010. “Human Topoisomerase IIIα Is a Single-Stranded DNA Decatenase That Is Stimulated by BLM and RMI1.” *Journal of Biological Chemistry*.

Yohe, Sophia, and Bharat Thyagarajan. 2017. “Review of Clinical Next-Generation Sequencing.” *Archives of Pathology and Laboratory Medicine*.

Zhou, Zhaolan et al. 2006. “Brain-Specific Phosphorylation of MeCP2 Regulates Activity-Dependent Bdnf Transcription, Dendritic Growth, and Spine Maturation.” *Neuron*.

Zlotogora, Joël. 1997. “Autosomal Recessive Diseases among Palestinian Arabs.” *Journal of Medical Genetics*.

Zlotogora J. 2019. “Autosomal Recessive Diseases among the Israeli Arabs.” *Human Genetics*.

Zlotogora, Joël, Saleh Barges, Bishara Bisharat, and Stavit A. Shalev. 2006. “Genetic Disorders among Palestinian Arabs. 4: Genetic Clinics in the Community.” *American Journal of Medical Genetics, Part A.*

APPENDIX A: SUPPLEMENTARY TABLES

Table S1. Candidate PTPN23 gene variants identified in patients with severe neurological and developmental impairments and seizures.

Reference	**[Alazami]**	**[Trujillano]**	**[Bend]**
Genotype (NM_015466.4)	Homozygous c.(3995G>T): p.(Arg1332Leu)	Homozygous c.(904A>G: p.(Met302Val)	Homozygous c.(2568_2594del27): p.(Val857_Pro865del)
Family history			
Ethnicity	Saudi Arabia	NK	Syria
Sex	M	NK	F
Age	5y	NK	7y
Microcephaly	√	NK	√
Developmental impairment	√ Global developmental delay	√ Global developmental delay Developmental regression	√ Profound
Seizures	√	√	√
Neurological impairment	√ Cerebral palsy	√ Spasticity	√ Spasticity and contractures Little purposeful movement Wheelchair bound
MRI	Mild cerebellar atrophy Severe loss in the white matter volume with enlarged ventricles Thin corpus callosum Cystic encephalomalacia	Brain atrophy	Brain atrophy and lissencephaly
Other	GERD		

GERD: Gastroesophageal reflux disease, NK: not known, M: male, F: female, y: year.

Table S2-A. Clinical signs associated with Bloom-Syndrome like disorder in patients' carrier for *TOP3A* variants.

Reference	Family	Pt	Genotype [NM_004618.4]	Ethnicity	Family History	Sex	Age	Microcephaly	Short stature
[Martin]	1	1	Homozygous c.[2718del] p.[Thr907LeufsTer101]	USA		F	5m	√	√
	2	2	Homozygous c.[2271dup]:p.[Arg758GlnfsTer3]	UAE	Another brother died at 13y as a result of cardiomyopathy	F	3y	√	√
		3		UAE		F	8y	√	√
		4		UAE		M	10y	√	√
	3	5	Compound Heterozygous c.[527C>T]: p.[Ala176Val] c.[1072_1073dup]:p.[Tyr359GlyfsTer17]	Japan		F	15y	√	√
	4	6	Homozygous c.[2271dup]:p.[Arg758GlnfsTer3]	Syria		F	19m	√	√
	5	7	Homozygous c.[2428del]:p. [Ser810LeufsTer2]	Spain		F	3y,3m	√	√
	6	8		Spain		M	7y	√	√
		9	Homozygous c.[2271dup]:p.[Arg758GlnfsTer3]	Syria		M	11y	√	√
	7	10		Saudi Arabia			4y,7m	√	√
	8	11	Compound Heterozygous c.[298A>G]:p. [Met100Val] c.[403C>T]:p.[Arg135Ter]	UK		F	adult	X	X
[Jiang]	9	12	Compound Heterozygous c.[1436A > G]:p.[D479G] c.[2362C > T]:p.[Q778*]	China	Non consanguineous parents	M	6y	√	√
		13		China		F	5y	√	√

"√": indicates presence of the clinical sign, "X": indicate absence of the clinical sign Pt: patient, F: female, M: male, y: year.

Table S2-B. Clinical signs associated with Bloom-Syndrome like disorder in patients' carrier for *TOP3A* variants.

Reference	Pt	Poor weight gain	Parental growth restriction	Elevated SC	Café-au-lait macules	Developmental delay	Decreased subcutaneous fat	Cancer	Malar rash
[Martin]	1	✓	✓	✓	✓	mild	✗	✗	✗
	2	✓	✓	✓	✓	✗	✗	✗	✗
	3	✓	✓	✓	✓	✗	✗	✗	✗
	4	✓	✓	NA	✓	✗	✗	✗	✗
	5	✓	✓	✓	✗	✗	✗	✗	✗
	6	✓	✓	NA	✓	mild	✗	✗	✗
	7	✓	✓	✓	✗	✗	✓	✗	✗
	8	✓	✓	✓	✓	mild	✓	✗	✗
	9	✓	✓	✓	✓	✗	✗	✗	✗
	10	✓	✓	NA	✓	mild	✓	✗	✗
	11	✗	✗	✗	NA	✗	NA	✗	✗
[Jiang]	12	✓	✓	✓	✗	✗	✓	✗	✗
	13	✓	✓	✓	✗	✗	✓	✗	✗

NA: not available, SC: sister chromatids.

Table S2-C. Clinical signs associated with Bloom-Syndrome like disorder in patients' carrier for *TOP3A* variants.

Reference	Patient	Gastroesophageal reflux disease	mtDNA depletion in muscles	Elevated SC	Dilated Cardiomyopathy	Other
[Martin]	1	√	NA	√	✗	CDH, gastrostomy
	2	NA	NA	√	✗	
	3	NA	NA	√	√	abnormal movements and tics
	4	NA	NA	NA	√(severe decreased)	
	5	✗	√(87%)	√	√ (after heart transplant)	hearing loss, CMAMMA
	6	√	NA	NA	normal echocardiogram	
	7	✗	NA	√	NA	
	8	NA	NA	√	√ (asymptomatic)	No response to GH treatment
	9	✗	NA	√	HCM	Microcytic anemia
	10	NA	NA	NA	✗	
	11	NA	√(>80%)	✗	NA	PEO, ataxia
[Jiang]	12	✗	NA	√	√(enlargement of left atrial and ventricular and heart failure)	Triangular face, Hepatomegaly, Fatty liver, mildly elevated AFP, good response to GH treatment
	13	✗	NA	√	√(enlargement of left atrial and ventricular and heart failure)	Triangular face, Hepatomegaly, mildly elevated AFP, good response to GH treatment

NA: not available, HCM: Hypertrophic cardiomyopathy, PEO: progressive external ophthalmoplegia , CDH: Congenital diaphragmatic hernia, AFP:alpha-fetoprotein, GH: Growth hormone

Table S3. Molecular and Cellular studies on *TOP3A* variants identified in patients with Bloom-Syndrome like disorder.

Reference	Pt	Genotype	Assessment of	Cellular/Molecular approach	Findings
[Martin]	1	Homozygous c.[2718del] p.[Thr907LeufsTer101]	Cellular amounts of mutated *TOP3A*	Immunoblotting	Marked reduction in cellular *TOP3A* enzyme amount (Depletion of *TOP3A* enzymatic activity)
	7,8	Homozygous c.[2428del]:p. [Ser810LeufsTer2]			
	1	Homozygous c.[2718del] p.[Thr907LeufsTer101]	Decatenation activity of *TOP3A*	Quantification of the dHJ dissolution	Normal activity
	2,3	Homozygous c.[2271dup]:p.[Arg758GlnfsTer3]	Impairment of dHJ dissolution	SCE Assay	Marked elevation in SCEs
	5	Compound Heterozygous c.[527C>T]: p.[Ala176Val] c.[1072_1073dup]:p.[Tyr359GlyfsTer17]			
	7,8	Homozygous c.[2428del]:p. [Ser810LeufsTer2]			
	1	Homozygous c.[2718del] p.[Thr907LeufsTer101]	Mitotic abnormalities and G1-associated defects	Immunofluorescence and microscopy	Increased numbers of ultrafine DNA bridges Elevated amounts of chromatin bridges and lagging DNA Elevated amounts of micronuclei
Cell viability and proliferation	12,13	Compound Heterozygous c.[1436A > G]:p.[D479G] c.[2362C > T]:p.[Q778*]	Mutated *Top3A* expression	RT-PCR	Low expression
	12,13		Mutated *TOP3A* protein stability	Western blotting	Stable *TOP3A*
	12,13	c.[1436A > G]:p.[D479G]	Mitochondrial function (ATP generation levels)	ATP analysis	Significant reduction in ATP generation levels
			Mitochondrial function (mitochondrial respiration)	Oxygen consumption rates analysis	Significant reduction in the oxygen consumption rates
			Impairment of dHJ dissolution	SCE Assay	Moderate elevation of SCEs
			Association among *TOP3A*, *RMI1*, and *RMI2*	CRISPR/cas9	Large reduction in the cellular protein level of *RMI1* and *RMI2* No significant effect in BLM level
			Cell proliferation		Detectable decreasing in cell growth rate
			Sun-sensitivity problems		No UV sensitivity

Pt: patient, SCEs: sister chromatid exchanges, dHJs: double Holiday Junctions, UV: Ultra-violet

APPENDIX B: CONSENT FORMS

Participation Information Leaflet

الجامعة العربية الأمريكية
ARAB AMERICAN UNIVERSITY

حكايات الأمل

مشروع الأمراض الوراثية النادرة في فلسطين

نموذج المعلومات المتعلقة بالدراسة

مجموعة من الأطباء الباحثين ورفاقهم من الجامعة العربية الأمريكية وأطباء الأطفال في فلسطين يجرون دراسة بحثية لمعرفة المزيد عن أسباب الأمراض الجينية والمتوقع انتقالها بالوراثة. لقد تم اختيار طفلك للمشاركة في الدراسة لأنه يعاني من مرض يعتقد أنه يعود الى وجود سبب وراثي (جيني). الهدف من هذه الدراسة هو التعرف و تحديد السبب الوراثي (الجيني) الكامن وراء هذه الحالة في عائلتك. إذا كنت توافق على انضمام طفلك ومشاركته في هذه الدراسة فإننا سنقوم بإتخاذ الخطوات العملية التالية :

1. سيتم سحب عينة من الدم حوالي (3-5 مل) من الوريد, و إرسالها الى المختبرات الجينية- الجزيئية التابعة للجامعة العربية الأمريكية, لتحضير الحمض النووي من هذه العينة واجراء الدراسة اللازمة.

في حال رفض الطفل لعطاء عينة الدم بسبب الخوف من وخز الابرة, سنستعيض عنها بأخذ عينة من لعاب الطفل.

2. من أجل تحديد المرض الجيني في عائلتك, قد نكون بحاجة للحصول على عينة دم لتحضير الحمض النووي من بعض أو كل فرد من أفراد أسرتك, لأنّ هناك احتمالية حملهم للجينات المعنية دون ظهور الأعراض عليهم. هذا قرار يحدد الحاجة إليه الفريق الطبي و فريق البحث التابع لنا. من الممكن ان تشمل العينات والديك, أطفال أقرباءك مثل أولاد العم/ العمة و أولاد الخال/الخالة. ننوه بأنه لن يؤخذ عينة من الحمض النووي من أفراد العائلة الذين ليست لديهم احتمالية حمل الجينات المسببة للمرض.
3. في سياق جمع المعلومات السريرية قد نحتاج الى الملفات الطبية الخاصة بالمريض للمراجعة, مثل التصويربالرنين المغناطيسي و تقارير علم الأمراض و المختبر أو أية تقارير تتعلق بالحالة الطبية للمريض. نود الإشارة الى ان هذه المعلومات جميعها سيتم الحفاظ على سريتها التامة بحسب ما هو معمول به حسب قوانين السرية في الجامعة, ولن يكون هناك وصول الى هذه المعلومات الا لأعضاء الفريق البحثي .
4. اذا أردت سيتم إعلامك بالنتائج التي تم الحصول عليها من خلال هذه الدراسة.

Participation Information Leaflet – Continued

5. سوف تتحمل ميزانية البحث كافة التكاليف المالية الناجمة عن مشاركتك في هذه الدراسة و التي تشمل المواصلات و الزيارات الميدانية و الفحوصات اللازمة لهذا البحث. بالإضافة لن يكون هناك مقابل مادي للمشاركة في هذه الدراسة.

ليس هناك عقبات أساسية أو مضاعفات للشتراك في هذه الدراسة, الاّ أننا ننوّه للنقاط التالية:

- لا يوجد أي مضاعفات أساسية أو هامة من سحب عينة الدم .
- ان نتائج هذه الدراسة سرية للغاية , سيتم التعامل مع المرضى من خلال ترميز بالأرقام وليس بأسمائهم الصريحة ولن يتم استعمالها من أي جهة خارج نطاق البحث العلمي
- لك الخيار باعلامك بنتائج البحث التي تخصك والتي تم الحصول عليها من خلال هذه الدراسة .

- من أجل المساعدة في تحسين الإستراتيجيات المستقبلية للتشخيص الجيني و الوراثي للمرضى الجدد, قد ندخل المعلومات الخاصة بالمريض المتعلقة بترتيب الاحماض النووية في المادة الوراثية وتاريخ الشجرة العائلية الوراثية لديك, دون الكشف عن إسمك أوعمرك وأي بيانات خاصة بك قد تكشف عن هويتك أو أي أحد من أفراد عائلتك, في قاعدة بيانات جينية عالمية, بحسب ما هو معمول به في المختبرات الجينية العالمية المرموقة.
- في حال الكشف أن الخلل الجيني يمكن أن ينتقل الى أبنائكم المستقبليين سيتم إبلاغكم بذلك من أجل إتخاذ كافة الإجراءات اللازمة لتقليل إحتمالية ولادة أطفال مصابين قدر الأمكان.
- سيتم تخزين العينة لمدة لا تقل عن عشر سنوات.
- في حال ظهور الحاجة الى إستخدام عينة المشارك مستقبلا فإن ذلك يعتمد على رغبة المريض في إستخدامها,ولك الحق في الرفض او الموافقة منذ الآن. ولن يتم إستخدام هذه العينات لأي هدف غير الأهداف المتعلقة بالبحث حول المرض المُعنى به.
- نود الإشارة الى انه لا يوجد أي فائدة تجارية أو مادية من هذه الدراسة (البحث) .
- للمريض الحق في الإنسحاب من الدراسة في أي وقت. و في حال إنسحابه وعدم إكمال الدراسة لك الحق في المطالبة في تدمير العينات و التخلص منها.

Dr. Reham Khalaf-Nazzal, MD, PhD
Arab American University

Pediatric Consent Forms

الجامعة العربية الأمريكية
ARAB AMERICAN UNIVERSITY

حكايات الأمل

مشروع الأمراض الوراثية النادرة في فلسطين

موافقة وتفويض

التاريخ

_____

الرقم

_____

أنا الموقع أدناه:

_____

وبعد أن شرح لي الدكتور/ة _____ أوافق بمحض إرادتي على دخول (ابني/ابنتي) و المشاركة بالدراسة المتعلقة بالاعتلالات الجينية، وأوافق على دراسة عينات الدم أو اللعاب المأخوذة مني للغايات المذكورة أعلاه. كما أفوض فريق الدراسة بإعطاء ما يلزم من المعلومات عني للاستخدامات العلمية شريطة ألا يعطى اسمي الصريح في حالة النشر.

أعلم أني أستطيع الانسحاب من الدراسة في أي وقت وأن مشاركتي في الدراسة لن تكلفني أي تكلفة مالية وعلى ذلك أوقع.

توقيع ولي أمر المريض:

_____

توقيع الطبيب:

_____

المشروع البحثي: الأمراض الوراثية النادرة المنتشرة في أوساط المجتمع الفلسطيني
Dr. Reham Khalaf-Nazzal, MD, PhD
Arab American University

Adult Consent Forms

الجامعة العربية الأمريكية
ARAB AMERICAN UNIVERSITY

حكايات الأمل

مشروع الأمراض الوراثية النادرة في فلسطين

موافقة وتفويض

التاريخ

_____

الرقم

_____

أنا الموقع أدناه:

_____ المولود في _____

وبعد أن شرحت لي الدكتورة رهام خلف نزال أوافق بمحض إرادتي على دخولي و المشاركة بالدراسة المتعلقة بالاعتلالات الجينية، وأوافق على دراسة عينات الدم أو اللعاب المأخوذة مني للغايات المذكورة أعلاه. كما أفوض فريق الدراسة بإعطاء ما يلزم من المعلومات عني للاستخدامات العلمية شريطة ألا يعطى اسمي الصريح في حالة النشر.

أعلم أني أستطيع الانسحاب من الدراسة في أي وقت وأن مشاركتي في الدراسة لن تكلفني أي تكلفة مالية وعلى ذلك أوقع.

توقيع المريض:

_____

توقيع الطبيب:

_____

المشروع البحثي: الأمراض الوراثية النادرة المنتشرة في أوساط المجتمع الفلسطيني
Dr. Reham Khalaf-Nazzal, MD, PhD
Arab American University

تحديد طفرات جينية جديدة مُسببة للأمراض في عائلاتٍ فلسطينيةٍ مُشّخصة بأمراضٍ وراثيّة نادرة باستخدام تقنيتيّ تسلسل الإكسوم الكامل وتسلسل الجينوم الكامل

إنّ انتشار الأمراض الوراثية النادرة في فلسطين يُعزى بشكلٍ أساسي إلى توارث طفراتٍ جينية محددة ناتجة عن تأثير المؤسس الوراثية في المجتمعات السُّكانية المعزولة والّتي غالباً ما يشيع فيها زواج الأقارب. تصنف غالبية هذه الطفرات الجينية على أنّها نادرة الحدوث استناداً إلى قاعدة بيانات الحمض النووي في فلسطين وقواعد بيانات الجينات العالمية. لذلك، فإنّه من الضرورة بمكان دراسة التركيب الجيني للسكان وإنشاء خارطة جينية تشمل الطفرات الجينية الناتجة عن تأثير المؤسس وتوزيعها في المجتمعات الفلسطينية والأمراض النادرة الموروثة الّتي تُعزى إلى هذه الطفرات. وعلى ذلك، قمنا في هذه الدراسة بتوظيف تقنيتيّ تحليل التسلسل الكامل للجينوم وتحليل التسلسل الكامل للاكسوم للكشف عن الطفرات الجينية الممرضة لمجموعة من الاضطرابات الوراثية النادرة في ثلاث عائلات فلسطينية تقطن قرىً فلسطينية تشكل مجتمعات معزولة. تشمل هذه الاضطرابات شكلاً معقداً من الشلل النصفي التشنجي الوراثي (cHSP)،التليف الكيسي (CF)، التقزم البدائي المصحوب بصغر الرأس (MPD)، بالإضافة إلى متلازمة ريت (RTT). فيما يخص العائلة الأولى، أظهرت نتائج تحليل تسلسل الحمض النووي (سانجر) وجود عدة أنماط جينية للاضطرابيين المتنحيين (cHSP) و (CF) لأفرادٍ مختلفين في نفس العائلة. إضافةً إلى ذلك، تبين أنّ الطفرة الجينية الممرضة المسببة للشلل النصفي التشنجي الوراثي للمرضى المدرجين في هذه الدراسة هي طفرة جينية جديدة متنحيّة (متماثلة اللواقح) تحدث في الجين المسمّى (*PTPN23*) . كما بينت نتائج تحليل تسلسل سانجر أنّ طفرةً جينية جديدة متنحيّة في الجين (*TOP3A*) هي المتسببة بحدوث اضطراب التقزم البدائي المصحوب بصغر الرأس عند الأفراد المصابين في العائلة الثانية. أمّا عن متلازمة ريت الّتي شُخصت بها حالة واحدة في العائلة الثالثة ، فإنّ تحليل التسلسل الكامل للجينوم وجد أنّها مرتبطة بطفرةٍ جينية ذاتية في الجين (*MECP2*). من خلال هذه الدراسة، نُسلط الضوء على أهمية تقنيات تحليل التسلسل الكامل للجينوم والاكسوم في الإجابة على الأسئلة البحثية الجينومية المعقدة وتشخيص الأمراض الوراثية النادرة المستعصية والمساهمة في تقديم خططٍ مستقبلية لإدارة التعامل مع مثل هذه الأمراض والتي لم يتم الكشف عن أسبابها البيولوجية الجزيئية في فلسطين لسنواتٍ طويلة. بالإضافة إلى ذلك، فإنّ دراستنا تؤكد على وجود أنماطٍ وراثية مختلفة للأمراض الوراثية النادرة في العائلات التي يشيع فيها زواج الأقارب، والّتي يُعتقد غالباً أنها موروثة من خلال النمط الوراثي المتنحي الصبغي الجسدي.