UNIVERSITÀ
DEGLI STUDI
DI BRESCIA

UNIVERSITÀ DEGLI STUDI DI BRESCIA

DOTTORATO DI RICERCA IN
Genetica Molecolare, Biotecnologie e Medicina Sperimentale

Settore Scientifico Disciplinare:
Area 06/Scienze mediche
MED/03 Genetica Medica
MED/46 Scienze Tecniche di Medicina di laboratorio

XXXV Ciclo

Identification and interpretation of pathogenic variants following Next Generation Sequencing (NGS) analysis in human Mendelian disorders

DOTTORANDO
Mohammad Sina

RELATORE
Prof.ssa. Silvia Clara Giliani

September-2023

1. Abstract (Ita/Eng)

Durante il programma di dottorato, l'attenzione è stata rivolta al supporto del laboratorio di diagnostica nell'implementazione della convalida o nella scoperta di varianti insolite. Questo è di massima importanza per comprendere i meccanismi eziopatogenetici molecolari, ma anche per offrire la migliore consulenza alle famiglie.

Di conseguenza, sono stati portati a termine diversi progetti come segue:

I) un caso enigmatico di una femmina con un disturbo granulomatoso cronico legato all'X (CGD) con una presunta variante di splicing: (NM_000397:ex9:c.1151+2T>C) nel gene CYBB.

II) una nuova presunta variante di splicing emizigote nel gene MAGT1 (NM_032121:c.627+2T>C) situato sul cromosoma X.

III) Analisi delle variazioni del numero di copie (CNV) per aumentare il tasso di diagnosi di un pannello NGS per gli errori congeniti dell'immunità, poiché è ben noto che le CNV (inserzioni o eliminazioni di dimensioni comprese tra 2 e 50 megabasi) rappresentano circa il 12% delle anomalie genetiche. Identificare questa ampia variazione è ancora problematico, specialmente con le piattaforme Ion Torrent che utilizziamo per la diagnostica, pertanto abbiamo eseguito un'approfondita analisi in silico utilizzando diversi nuovi software.

IV) Otto famiglie con una storia personale o familiare di cancro sono state testate per un pannello di geni multipli osono state sottoposte al sequenziamento completo dell’esoma. Sono state trovate otto varianti patogeniche e verificate tramite sequenziamento di Sanger o MLPA e PCR in tempo reale. L'uso di NGS e la rilevazione di CNV hanno migliorato la diagnosi nei pazienti affetti da cancro. Alcune delle famiglie iraniane che soddisfacevano i criteri di Amsterdam sono state incluse in programmi di sorveglianza indipendentemente dal loro stato di portatori di mutazioni prima dei test genetici, mentre dopo la rivelazione del portatore solo i portatori sono stati inclusi, migliorando la conformità e riducendo i costi di gestione.

During the PhD program the focus was to support diagnostic lab implementing validation or discover of unusual variants. This is of utmost importance to understand molecular etiopathogenic mechanisms, but also in order to offer the best counselling to families.

Thus, different projects were accomplished as follows:

I) a puzzling patient of a female with X-linked chronic granulomatous disorder (CGD) with a putative splicing variant: (NM_000397:ex9:c.1151+2T>C) in the *CYBB* gene.

II) a novel hemizygous putative splicing mutation in the *MAGT1* gene (NM_032121:c.627+2T>C) located on the X-chromosome.

III) Analysis of Copy Number Variations (CNVs) to increase the diagnostic rate of a NGS panel for Inborn errors of immunity, as is well known that CNVs (indels between 2 and 50 megabases), account for roughly 12% of genetic abnormalities. Identifying this large variation is still problematic, especially with the Ion Torrent platforms we use for diagnostic, thus we performed an extensive *in silico* analysis using multiple new softwares.

IV) Eight families possessing a familial or personal history of cancer underwent multigene panel testing or whole exome sequencing. Eight pathogenic variants were found and verified through Sanger sequencing or MLPA and real-time PCR. The use of NGS and CNV detection improved the diagnostic yields in cancer patients. Some of Iranian families who met Amsterdam criteria were enrolled in surveillance programs irrespective of their mutation carrier status before genetic testing, while after carrier detection disclosures only carriers were enrolled improving compliance and decreasing the managing costs.

2. Introduction and literature review

Technological advancements in gene sequencing have profoundly impacted on genes discovery, new variants annotation, and patients' surveillance and counselling by providing better knowledge of genes, mechanisms and finally disorders.[1] Our main interest was in a better interpretation of what is now a widespread employment of next-generation sequencing (NGS) in the fields of Mendelian inherited genetic conditions, particularly primary immunodeficiencies (PIDs), and hereditary cancers [1-3].

2.1 Genetics in primary immunodeficiencies

Pediatric disorders encompass a variety of genetically diverse conditions with significant penetrance, making them suitable for genomewide diagnostic techniques. Achieving a molecular diagnosis presents difficulties; however, the lifelong advantages can be substantial. [4] Over the past decade, the median survival rate of hematopoietic stem cell transplantation (HSCT) has significantly increased to more than 84-90%, making it a highly effective curative treatment for PIDs. [5] PIDs referred to as inborn errors of immunity (IEI) arise from a genetic anomaly in one of many genes (485 at present), and cause predisposition to autoimmunity, infectious diseases, auto-inflammatory diseases, allergy, and/or cancer. [6] These syndromes arise from monogenic germline pathogenic variations that lead to loss or reduced expression, loss or gain of function of genes. [7,8] Early onset of pediatric disease symptoms in a patient facilitates genetic diagnosis by virtue of the limited occurrence of causal variants in control datasets. [9] IEI encompass a vast array of disorders that arise from harmful genetic mutations that impair both innate and adaptive immunity. The IEI syndromes can be heritable in X-linked, recessive or dominant conditions, autosomal, or with reduced penetrance. [10] Recent advances in genomic sequencing has made remarkable progress in the identification of new molecular origins for rare single-gene conditions. Consequently, this technology is gradually accessible in diagnostic clinics worldwide. [11,12] High-throughput NGS technologies have proved to be particularly advantageous for the field of pediatrics. This is primarily due to the high clinical demand for accurate diagnosis and treatment, and the significant load of pathogenic de novo variants. [4,13]

Applying targeted NGS provides a deeper insight into the functional outcome of patient's variant. [14,15] This targeted sequencing method is designed to amplify regions of interest and it can directly detect genetic alterations including single nucleotide variants (SNVs) and insertions or deletions up to 50 base pairs (InDels). [16] Genetic mutations in non-coding areas of the genome could be responsible for causing diseases by affecting RNA splicing complex, secondary structure regulation, and the degree of gene expression. [17-19]

The investigation of changes in splicing patterns can be conducted through various methods, including gene expression, computational analyses tools that predict the consequences of genetic mutations on

on splice processes. The employment of splicing reporter minigene assays that allow for the study of the mutation on splicing complex [20]. Therefore, the task of interpreting them remains arduous using current available genomic information, expecially when at the heterozygous state. [17] Abnormalities in pre-mRNA splicing often result in a Mendelian genetic disorder and contribute to 9–30% of disease-causing variants. [14,21] Functional analysis of variants with possible influence on RNA splicing would lead to higher genetic diagnostic yields.

The analysis of RNA via reverse transcription-polymerase chain reaction (RT-PCR), and then Sanger sequencing on RT-PCR products may validate possible mis-splicing variants identified by next-generation sequencing. [22-24]

The unique attributes inherent to X chromosomes have served an important function in shaping the landscape of X-linked monogenic disorders. These distinct biological characteristics have contributed to the prevalence and manifestation of these disorders and represent a vital area of study in genetics research. [25] The X-linked recessive condition results from a loss of function of a genes on the X chromosome. The condition is expressed either in males in hemizygosity or in females who carry the mutation in homozygosity. [26] The unique mode of segregation observed in pedigrees analysis for X-linked recessive syndromes can be ascribed to the presence of a single X chromosome in males (hemizygosity). This has significantly aided the identification and recognition of these types of disorders, as well as the discovery of the associated genes. [27] Male with mutated gene on X chromosome will express the condition, as he does not have another copy of the X chromosome to compensate for the loss of function. Females who inherit one mutated X chromosome will be carriers of the condition, but typically do not express it as they still have another functional copy of the X chromosome. However, if a female inherits two mutated X chromosomes, she will express the condition. [28] Among the X-linked IEI we focused particularly on two disorders in order to validate peculiar variations: chronic granulomatous disease (CGD), and XMEN syndrome. [29,30]

CGD arises from a genetic abnormality in one or more of the genes responsible for the respiratory burst process (*CYBA, CYBB, CYBC1*, *NCF1*, *NCF2, NCF4*) as shown in *Figure 1*.

CGD Patients are highly vulnerable to severe and repeated bacterial and fungal infections due to their inability to generate reactive oxygen species (ROS). [31]

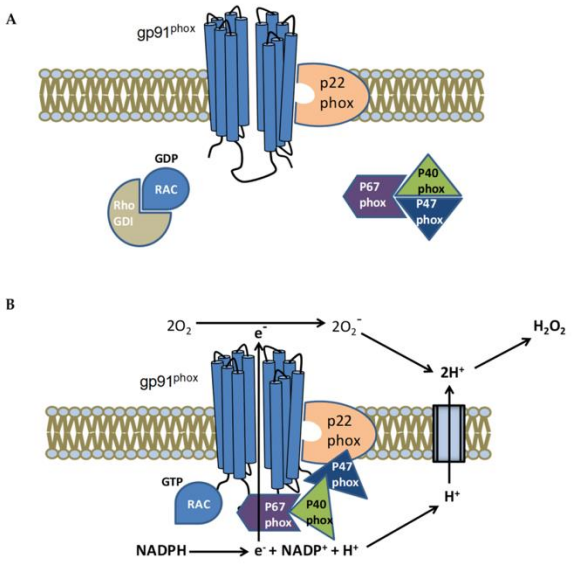


Figure 1 Schematic illustration of activation of the NADPH oxidase (NOX) complex in the cytoplasmic membrane of a phagocyte (Garay et al., Biomedicines 2022, 10, 2570)

Clinically, CGD result in an increased vulnerability to severe infections due to bacteria and fungi that are categorized as catalase-positive, as well as inflammatory complications like CGD-related colitis, which can be life-threatening. [32] The inflammatory complications can be caused by excessive stimulation of NF-κB and inflammasome in CGD phagocytes can result in the persistent generation of pro-inflammatory cytokines and the emergence of inflammatory symptoms. [5] While CGD can manifest at any point in one's life, the majority of patients are usually identified before reaching two years of age, and some may not be diagnosed until later in adulthood. [33,34] The condition inherited either as autosomal recessive or X-linked, depending on the specific genetic abnormality present in the underlying gene, impacting around 1 in every 250,000 people. [30,35] X-linked recessive CGD is the most common type of CGD (80% of the cases) and is due to germline alterations in the *CYBB* gene on the X-chromosome That affect the Cytochrome b-245 heavy protein. This is a constituent of nicotinamide adenine dinucleotide phosphate (NADPH) oxidase complex. [30,36,37] Females have two distinct populations of neutrophils, one expressing the X chromosome received from the paternal side and the other expressing the one inherited from maternal side. During early development in mammals, epigenetic mechanisms silence most genes on one randomly selected X chromosome to enable the expression of only one X chromosome equivalent, a process known as lyonization. [38] Because of lyonization, the gene dosage of the X chromosome in females becomes largely comparable to that of males. [39] Normally due to the X-chromosome location of the gene, *CYBB* deficient patients are males, but the rare occasion of affected females have been reported, usually due to skewed X-chromosome inactivation. [26] This phenomenon occurs when there is an imbalance between normal and mutated active X chromosomes, and less normal allele expression results in a heterozygous affected female. [40,41] Highly skewed X inactivation is commonly delineated by the identification of a discernible pattern

in which one X chromosome undergoes preferential inactivation in 80% of the cells, thereby denoting an uneven distribution of X-linked gene expression across both X chromosomes present in females. [42] Skewed X-chromosome inactivation can contribute to trisomy risk, neoplastic diseases, and X-linked disease incidences. [43] *HUMARA* gene in humans possesses a remarkably polymorphic site that is methylation sensitive and can correlate with X-chromosome inactivation. [44] Over the years, a technique that can measure the extent of X-chromosome inactivation based on the HUMARA polymorphism has been set up. [44]The diverse range of -CAG- repeats of the polymorphic regions was used to determine the paternal and maternal alleles in the offspring, while the digestion with methylation sensitive enzyme can discriminate the status of inactivation. [44]

While CGD is known since years and many studies have deepen the knowledge on the clinical history and the proteins involved, a new entity in PID is the "X-linked immunodeficiency with magnesium defect, Epstein-Barr virus (EBV) infection, and neoplasia" (XMEN). [45] From the clinical point of view, the few male patients known till now present with recurring ear and sinopulmonary infections, chronic EBV infection, and autoimmunity. [46] The clinical suspicion is further strengthened when there is a a record of lymphoma or immunodeficiency in male families on the mother's side of the family. [45] EBV, a gamma herpesvirus with oncogenic properties, is prevalent worldwide and infects a portion of children as well as the majority of adults. [47,48] Over 90% of people globally have been exposed to and harbor the Epstein-Barr virus. This virus marked the initial discovery of a cancer-causing viral agent and has been linked to seven different types of cancers so far. [49] EBV is a virus that solely infects humans and stays dormant within B lymphocytes. Although primary EBV infection usually happens during childhood as a mild or asymptomatic infection. When individuals are infected with EBV during adolescence or adulthood, they usually experience symptomatic infections. [50] Uncontrolled EBV infection is a key feature of X-linked immunodeficiency with X-MEN disease, which results in an increased predisposition to Lymphomas that have an association with EBV, even during childhood. [51] The disorder is an rare genetic immune anomaly resulting from a detrimental change in the *MAGT1* gene, resulting in impaired T cell functioning. [46] The *MAGT1* gene encodes the transmembrane protein responsible for transporting magnesium, which is crucial for T cell activation. This transport is essential in coordinating the signals in T cells, as it triggers calcium flux, leading to intracellular signaling. [52] (*Figure 2*)

.

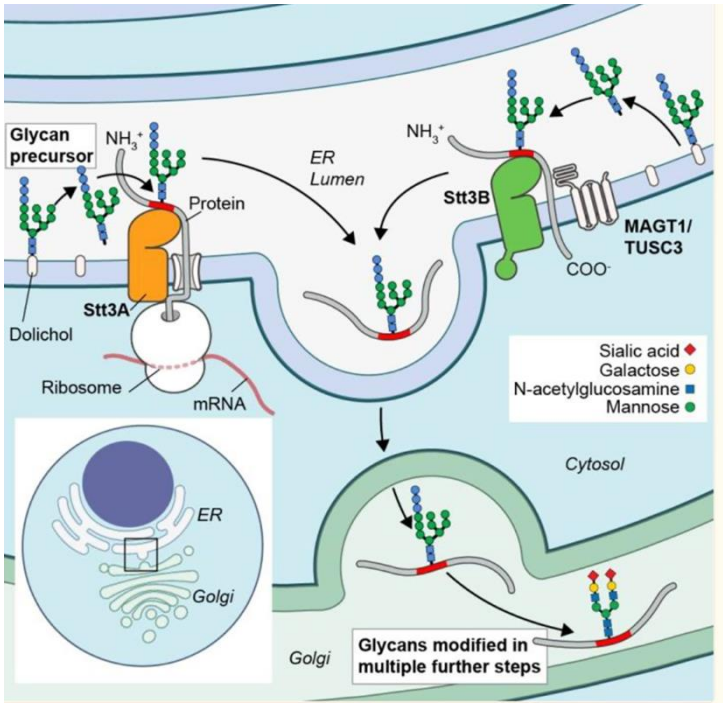


Figure 2 MAGT1 assists in the process of N-linked glycosylation. Mutations in this gene lead to reduced calcium flow and diminished activation of phospholipase C-gamma 1 (PLCγ1). Additionally, they hinder the cell's ability to undergo re-stimulation induced death when the TCR is crosslinked, Potentially because of disrupted communication via the under-glycosylated TCR and CD28 (Ravell et al, J Clin Immunol. 2020 Jul; 40(5): 671–681).

X-MEN patients commonly exhibit reduced serum levels of both IgA and IgG, along with a compromised ability to respond to polysaccharide antigens. [53] They also frequently experience ear and sinopulmonary infections, which can sometimes be linked to bronchiectasis. [54] Individuals affected by XMEN disease experience a decline in the activation of T cells and a decline in the cytolytic function of CD8+ T cells and natural killer (NK) cells. [46] Those affected by this condition typically display continuous EBV DNA concentration, followed by EBV-driven lymphoid neoplasia, Differing levels of B and T cell participation. XMEN disease diagnosis typically does not involve the consideration of ionized or total serum magnesium concentrations, as both usually remain within normal ranges. [45]At present, patients with XMEN disease have only been observed to have missense or null mutations. The *MAGT1* protein plays a crucial role in maintaining magnesium balance in eukaryotic organisms and is widely preserved across various evolutionary lineages. [22]

2.2 Detection of germline copy number variations as the underlying mutation in primary immunodeficiency disorders

PIDs are genetically heterogeneous disorders, with clinical features frequently overlapping even in the presence of different genetic defects. [6,55] Copy number variants (CNVs) are caused by indels pathogenic variants and a significant proportion of genetic diseases in humans can be attributed to them. [56] There is growing body of evidence linking CNVs to various diseases and their involvement in the evolution of human genes and genomes. [57] CNVs are defined as duplication or deletion in the genome larger than ~50 base pairs to megabases, which are responsible for up to 12% of genome DNA alterations. While heterozygous pathogenic variants may cause autosomal dominant condition through dominant negative or haploinsufficiency, hemizygous and homozygous (HMZ) pathogenic variations cause autosomal recessive condition by a loss of function or the lack of protein. [58,59] HMZ deletions of genes, whether complete or partial, frequently produce null alleles and lead to a total loss of gene activity. [60] Deletion CNVs that occur on the on the male X-chromosome, which is present in single copy (hemizygous), result in the complete loss of that genomic segment from the male genome. [61] Homozygous (biallelic) deletions are the loss of both alleles, [62] while HMZ deletions represent merely a fraction of all medically relevant CNVs, they can have a significant impact on identifying new Mendelian genes. [63-65] Recessive disease genes show clinical phenotype in homozygous deletions/insertions, which are significant for determining an individual's carrier status for recessive traits. [66,67]

WES and targeted panel sequencing focus on the exonic coding regions which constitute nearly 1% the human genome or on a small number of genes based on the targeted panel application. Identification of large CNVs through targeted sequencing can be challenging expecially with some kind of chemistry and analysis pipelines. Frequently the analysis is accomplished through split read, paired-end mapping, depth of coverage analysis, and de novo assembly. [68,69] Abundant CNV analysis software have been developed to detect them in targeted next-generation sequencing. [69] These tools mainly utilize a read-depth-based method and are devised to identify CNVs from targeted sequencing like panels and WES data and rely on measuring the read counts of target regions, assuming that it reflects the true copy number at a given locus in a linear manner. [70] Depth-based CNV algorithms calculate the copy number of a locus compared to mean read depth of the corresponding locus in reference samples. [71] The HMZDelFinder is a tool that utilizes data from WES or targeted sequencing to detect HMZ deletions. The absence of commonly-used and strongly validated rare CNV identification tools has created an obstacle for the routine evaluation of this crucial type of genetic variation across the entire exome. [69]

The GATK-gCNV algorithm is an open-source package within GATK tool that utilizes genome sequencing read-depth data to identify infrequent CNVs. [72]

The depth-based approach for detecting CNV has limitations because of biases due to library preparation and variations in sequencing performance. A high degree of variability of the read depth in targeted sequencing can be affected by various factors, including batch effect, targeted depth, GC content, sequencing efficiency, PCR duplication bias, and mappability. [73,74] As a result, the depth of reads may be underestimated or overestimated, potentially leading to misleading conclusions about the occurrence of CNVs in the targeted genomic regions [43,49].Therefore, distinguishing between genuine copy number variations and technical artifacts can be challenging results in high false-positive calling. [61,75,76] Although computational analysis of some tools for targeted-based CNV was evaluated for sensitivity and specificity, there are few reports on their false discovery rate and reproducibility in diagnostic clinics. [75]

In a clinical environment, the utmost priority is to achieve the highest level of sensitivity and the lowest rate of false discovery on a given sequencing platform. [75] Moreover is important to underline that almost all the softwares are normalized on the Illumina platforms.

The percentage of reads mapped in a genome is mainly influenced by two factors: the size of the reads generated during the amplification and sequencing, and quantity of mismatches. [77,78] It is feasible to calculate the mappability prior sequencing by utilizing the technical parameters of the sequencing experiment. Areas with high mappability are likely to generate distinct mappings, whereas regions with low mappability may result in unclear or indefinite mappings. [78]

It has been recommended to assess the effect of the mappability of regions for targeted sequencing, and to exclude from further analysis the regions with mappability lower than 75%. [54] Implementing a pre-CNV calling step to remove regions with low mappability in a particular algorithm resulted in a reduction of ‘potential’ CNVs by 33%. However, this improvement came at the expense of a 3% reduction in sensitivity. [71]

In addition to CNV analysis, standard NGS pipelines may fail to detect specific genetic mutations, such as significant insertions or deletions. A new genetic variation has been discovered in the Male Germ Cell-Associated Kinase (MAK) gene, involving an Alu insertion that cannot be detected by conventional variant detection methods based on NGS. [79] In addition, we applied a grep algorithm to detect pathogenic MAK-Alu insertion in FASTQ files of PID patients. *MAK-Alu* insertion were not detected in any cases. [79]

2.3 Molecular characterization and implications on surveillance in cases of genetic determined inherited cancer: Insight on *PMS1* gene *MSH6* genes

NGS and CNV calling are also landscape-altering approaches to identify genes accountable for hereditary familial cancers. [80] Approximately 10% of breast and colorectal cancer (CRCs) cases are caused by a genetic aberration. [81-83] Lynch syndrome (LS) is contributor to 3-5% of CRC cases and is the main reason of heritable colorectal syndrome and inherited as an autosomal dominant disorder. [84] LS individuals carry heterozygous disease-causing mutations in one of the mismatch repair (MMR) genes, namely *MLH1*, *MSH2*, *MSH6*, and *PMS2*, or *EPCAM*. [85] People with LS have a higher susceptibility to develop hereditary colorectal and endometrial tumors, as well as a diverse range of other cancers, particularly at young ages. [86] Approximately 2% of patients with LS were found to have a causative *EPCAM* deletion variant. [87,88] For individuals with *EPCAM* mutations, surveillance programs should be prioritized for the screening of CRC over endometrial cancers. [89] The development of multiple fundic gland growths in the stomach is termed Fundic Gland Polyposis (FGP), which are the most frequent type of polyp in this organ. [90,91] Syndromic FGP is linked to a genetic anomaly in the APC gene, whereas distinguishing between hereditary and sporadic FGPs is not possible through histological and histochemical analysis. [92]

Some researches indicated that variations in the *PMS1* gene associated with The development of malignancy, [93-95] while mouse model studies did not verify this hypothesis. [96]

Research has indicated that when evaluating microsatellite instability (MSI) status, the CAT25 mononucleotide marker exhibits greater sensitivity and specificity compared to other markers. [97,98] LS tumors typically exhibit a phenotype characterized by MSI and/or negative IHC staining in MRR proteins in approximately 90% of cases. [99] Some cases of LS exhibit similar symptoms to other hereditary cancer syndromes, and up to 50% of cases met the Amsterdam criteria are not LS people. [100,101] The application of NGS technology in identifying genetic abnormalities in cancer patients who fulfill clinical criteria for hereditary cancer syndromes has resulted in an increase in the diagnosis rate. This is achieved by evaluating both the gene sequences and CNVs of multiple genes simultaneously. [101-103] The analysis of CNVs using NGS data or Multiplex ligation-dependent probe amplification (MLPA) contribute in detecting up to 20% of pathogenic variants by detecting deletions or duplication spanning more than 50 base pairs, both at the heterozygous and homozygous state. [104,105] Detecting CNVs through WES data for genetic diagnosis is not yet a standard procedure because of discrepancies between different algorithms.[106-108] Validation of in silico putative identified CNVs, requires confirmation with an alternative technique such as MLPA or real-time PCR. [109,110]

In mouse models, the heterozygous recessive mutation in an MMR gene occurred before the second hit incidence. [111] This means that the insufficient dosage of MMR proteins in certain tissues caused by the loss of one allele can result in haploinsufficiency, leading to the inability to perform the corresponding protein function. This condition can promote tumorigenesis. [111]

The detection of LS mutation carriers enhances adherence to endoscopy and colonoscopy monitoring plans recommended to be implemented every 1-2 years.[85,112,113]. While Middle and North African nations have implemented clinical standards for managing LS, a significant number of LS mutation carriers in these regions remain unidentified. [114] Regular colonoscopies are commonly used as a form of surveillance for LS families in many of these nations, even though the carrier status of individuals is unknown.[85,114] It is crucial to adopt a cost-efficient strategy, for example the Amsterdam criteria, Bethesda guidelines, and personal/family cancer history, to address the significant number of undetected LS patients and limited financial resources in Iran. Additional research is necessary to gather information on insurance coverage for genetic testing and surveillance programs for Iranians suspected of having hereditary cancer.

3. Aim

The main aim of this work was to deal from different perspectives with limitations arising in a diagnostic laboratory encountering challenges in the detection and correct interpretation of variants found during the diagnostic processes and the lack of results in patients with a precise clinical diagnosis.

Moreover, we faced the crucial importance of a correct diagnosis while dealing with counselling and surveillance, expecially in cases of genetically determined tumour.

4. Material and method

4.1.1 Patient diagnosed with X-linked primary immunodeficiency

A seven-year-old girl with recurrent infection and possibly inherited immunodeficiency was admitted to Spedali Civili Hospital in Brescia, blood specimens were gathered from the proband and her parents in accordance with diagnostic protocols. The proband and her parents gave their informed approval before the collection. The patient DNA sample was subjected to the targeted gene panel analysis routinely used in the diagnostic lab including 351 PID genes, to identify the specific pathogenic variant responsible for the immunodeficiency. The targeted panel approach was applied including genes known to be involved in cases of chronic granulomatous disease, namely(refseq): *CYBB* (NM_000397), *CYBA* (NM_000101), NCF1 (NM_000265), *NCF2* (NM_000433), *NCF4* (NM_013416), *CYBC1* (NM_001033046), G6PD (NM_001360016).

After preparation of Ampliseq on demand amplicons, the NGS panel was run using IonChef automatic libraries preparation on an Ion GeneStudio™ S5 Prime (ThermoFisher Scientific). With vertical coverage under 30 reads, samples have been Sanger sequenced for the specific exons. Every variation found by NGS was confirmed by Sanger sequencing.

To visualize the depth of coverage, Integrative Genome Viewer (IGV) software was employed. [115,116] Filtering and analysis was performed using IonReporter and Wannovar.

4.1.2 Separation of PBMC AND PMN by Ficoll:

Polymorphonuclear neutrophils (PMN) and peripheral Blood Mononuclear Cells (PBMC) were isolated from peripheral blood by employing Ficoll-Hypaque density-gradient separation technique according to manufacturer's instructions (Lympholyte-H from Cedarlane Laboratories) as follows:

PBS and patient’s blood were combined, maintaining a 1:1 ratio and transferred to a separate Falcon tube containing the same volume of Ficoll. The tube was centrifuged at 2000 RPM at room temperature for 20 minutes. This process separates the various cell components as shown in *Figure 3*. After centrifugation, the interface containing PBMCs was recovered, Phosphate-buffered saline (PBS) was added and PBMCs were centrifugated and recovered as pellet. The resuspended pellet was then stored at -80°C in DMSO as vital frozen cells or as dry pellet for RNA preparation. Polymorphonuclear leukocytes (PMNs) were stored at -20°C as a source of DNA.

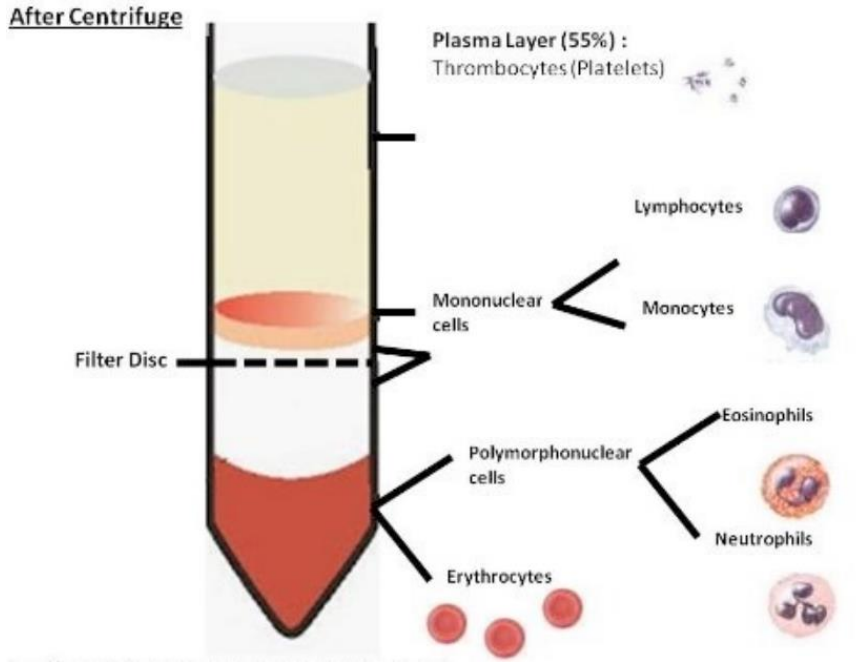


Figure 3: The separation of PBMC and PMN by Ficoll

4.1.2. RNA extraction from pellet of PBMC and analysis of exon skipping in mRNA

RNA extraction from PBMCs was performed following the guidelines of the manufacturer's kit (Macherey-Nagel, Germany). The cDNA was synthetized starting from RNA using the ImProm-IITM Reverse Transcription system kit (Promega, USA) through reverse-transcription polymerase chain reaction (RT-PCR). The putative splicing mutation due to the location of the variant in one of the invariant sites was then checked. To verify the effect of this variant on the splice-donor junction site of the mRNA of the *CYBB* gene, a 979 bp mRNA fragment was applied to the mRNA segment. The fragment covered the overlapping regions from exons 7-8 to the 3'UTR. It was amplified by means of the forward primer 5'-AACCCTCCTATGACTTGGAAATGGATAG-3'(Positioned at the junction of exons 7-8) and the reverse primer 5'-GCATTATTTGAGCATTTGGCAGCAC-3' (located in the 3'UTR). The amplification of the amplicon was conducted with AmpliTaq Gold™ DNA polymerase, utilizing the following parameters:

an initial denaturation step at 94°C for 12 minutes, followed by 40 cycles of denaturation at 94°C for 30 seconds, annealing at 69°C for 30 seconds, extension at 72°C for 30 seconds, and a final extension step at 72°C for 7 minutes. The patient's mRNA was amplified through RT-PCR, and subsequently, the upper and lower *CYBB* fragments were isolated from a 1.5% agarose gel using the QIAquick Gel

Extraction Kit from QIAGEN (Germany). with small modification, the DNA was dissolved in nuclease-free water instead of Tris-HCl buffer before sequencing the two purified fragments using Sanger sequencing. The experiment also utilized beta-actin as the housekeeping gene to assess the quality of RNA.

4.1.3. X-chromosome inactivation analysis

4.1.3.1 Enzymatic digestion of DNA and PCR amplification

The assessment of X-chromosome methylation status was conducted by amplifying HUMARA loci, utilizing a previously described method with some modifications. [44,117-119] *Figure 4*

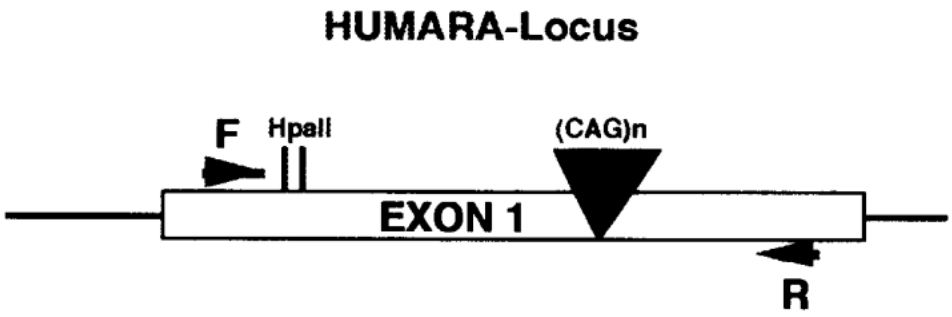


Figure 4 Exon 1 of the HUMARA site shows the placement of both the Forward (F) and Reverse (R) primers employed during PCR. These are situated in relation to the CAG repeat sequence and a pair of HpaII sites.

DNA was extracted from PMN of the CYBB patient and her parents (Qiagen, Germany). The patient's DNA (200 ng) was digested with 25 units of HpaII restriction enzyme (Promega, USA) at 37 °C for 16 hours, along with controls. The amplification of the HUMARA gene was conducted utilizing 1 µL (10 mM) of FAM fluorescently labeled forward 5'-TGCGCGAAGTGATCCAGAACC-3', and reverse 5'-TGGGCTTGGGGAGAACCATCC-3' primers in a total volume of 25 µL.

In general, two discrete sets of optimizations were devised, the first involving the utilization of AmpliTaq Gold™ DNA polymerases for amplification, while the second utilized GoTaq® DNA polymerases. In the former series of optimization experiments named PCR-A, the amplification was conducted utilizing 2 mM of Mg2+ and AmpliTaq Gold™ DNA polymerases enzyme. To perform PCR-A, thermal cycling was initiated with a denaturation step at 95°C for 12 min, followed by 38 amplification cycles of 95°C for 30 seconds, 69°C for 30 seconds, and 72°C for 30 seconds, with a final extension step at 72°C for 7 min.

To improve the efficiency of the PCR reaction, a longer final extension step of 1 hour was implemented in PCR set -B. This modification aims to enhance the yield and specificity of the PCR

product. The efficacy of the PCR-B protocol was evaluated by introducing two additional final concentrations 1 mM and 2.5 mM of Mg2+, into the experimental design. PCR amplification was conducted using AmpliTaq Gold™ DNA polymerase enzyme, with a final concentration of 1 mM Mg2+ and 10% DMSO under different conditions. The first condition involved PCR-B protocol and 50 ng DNA template, while the second condition involved using 100 ng of DNA template and a reduced cycle number from 40 to 28 of PCR-B protocol, referred to as PCR-C.

PCR amplification was conducted using AmpliTaq Gold™ DNA polymerase enzyme, with a final concentration of 1 mM Mg2+ and 10% DMSO under different conditions. The first condition involved PCR-B protocol and 50 ng DNA template, while the second condition involved using 100 ng of DNA template and a reduced cycle number from 40 to 28 of PCR-B protocol, referred to as PCR-C.

For the latter series of amplifications, the GoTaq® DNA polymerases from Promega (USA) were utilized along with 1 mM of Mg2+ and 50 ng of DNA template, and a final concentration of 0.5 M betaine. The amplification protocol consisted of denaturation at 95 °C for 12 minutes, followed by 48 cycles of 95 °C for 30 seconds, 69 °C for 30 seconds, and 72 °C for 30 seconds. A final extension was carried out at 72 °C for 10 minutes, resulting in PCR-D.

For the subsequent trial, the PCR-D cycles were reduced to 28 cycles, resulting in PCR-E.

4.1.3.2 Fragment Analysis

To measure the length of the fragments produced by PCR, one microliter of each sample was mixed with 10 microliters of deionized formamide containing an internal size standard. For reference, the GeneScan 500 ROX size standard (Applied Biosystems, USA) was utilized as a marker for size comparison [120]. Capillary gel electrophoresis was carried out using an ABI 3130 Genetic Analyzer, and the resulting data was analyzed using Genemapper 5 (Applied Biosystems, USA). The X-chromosome inactivation ratio was ascertained by calculating the area of the height peaks for the PCR products as determined by the following formula: X1 = (d1/u1)/[(d1/u1) + (d2/u2)]. The formula X1 calculates the fraction of peak area in digested DNA attributed to the first allele as a function of its peak area in both digested and undigested DNA, and the peak areas of the second allele in both digested and undigested DNA. Specifically, u1 and u2 are the peak areas of the first and second allele in undigested DNA, and d1 and d2 are their corresponding peak areas in digested DNA [121].

4.2. Copy number variant calling in a cohort of primary immunodeficiency patients using targeted panel NGS sequencing in a diagnostic setting

4.2.1. Patients enrolled

The 400 patients enrolled in this chapter of the work met the clinical criteria for PID analysis and underwent molecular diagnostic at the Genetic Section, Spedali Civili Brescia, Italy in a period between 2020 and 2022. Out of the total of 400 patients, 200 individuals who did not receive a genetic diagnosis were then included in the Copy Number Variation (CNV) study. *Figure 5.*

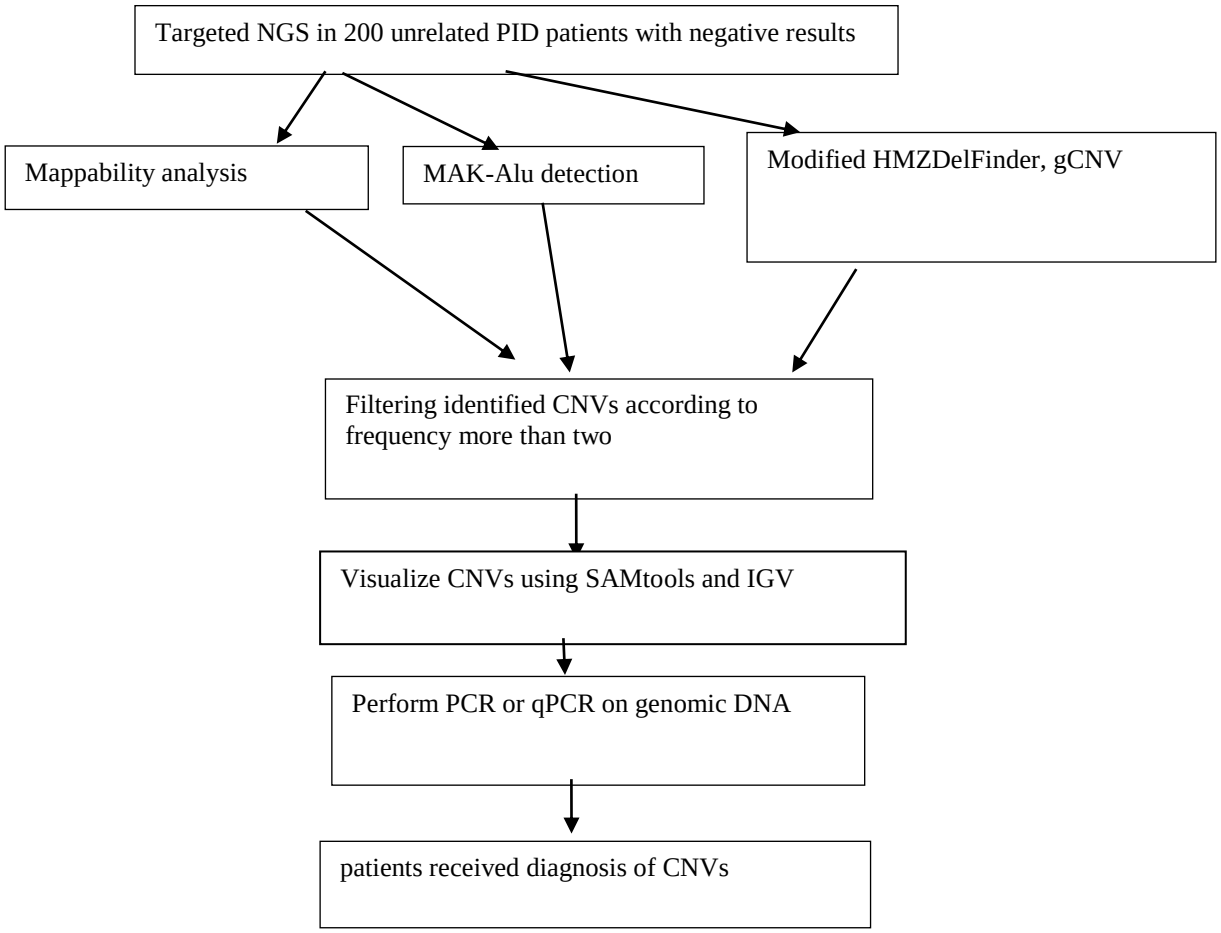


Figure 5: Genetic testing strategy for identification of patient CNV genetic variants in (PID) cases

4.2.2. Genomic DNA extraction and multigene panel testing, and data preprocessing

DNA was extracted from whole blood utilizing the Maxwell RSC (Promega). In this study 351 PID genes chosen on the IUIS classification (Tangye et al, 2019) were sequenced using tg-NGS.
The panel was built using Ampliseq designer and 299 genes were available as ready to order, while 52 genes were available as custom design. The two panels are detailed in *Table 1 and Table 2*

Table 1: Ready to order gene in PID panel

GENE NAME	GENE NAME	GENE NAME	GENE NAME	GENE NAME	GENE NAME
AICDA	ELANE	IL2RB	NOD2	PSENEN	ERCC6L2
AIRE	EPG5	IL2RG	NRAS	WIPF1	MAP3K14
AK2	FADD	IL36RN	MVK	XIAP	NFAT5
ACP5	DNASE1L3	IL6R	MYD88	FERMT1	NLRP1
ACTB	DNASE2	IL6ST	PLCG2	STAT1	OTULIN
ADA	DNMT3B	IL7R	PEPD	STAT2	POLA1
BCL10	DOCK2	IRAK1	PIK3CA	STAT3	POLE2
APOL1	DOCK8	IRAK4	PIK3CD	STAT4	PSMA3
ATM	GINS1	IRF4	PIK3CG	STAT5B	SMARCD2
B2M	FAS	IRF7	PIK3R1	STIM1	TMEM173
BACH2	FASLG	IRF8	POLR3A	STX11	TRAF3
BLM	FAT4	ISG15	RAB27A	STXBP2	ZNF341
BLNK	FCGR3A	ITCH	RAC2	ZAP70	TRAF3IP2
BTK	FCN3	KDM6A	PRF1	TAZ	HAX1
CARD11	FOXN1	LYST	PRKCD	TBK1	PSTPIP1
CARD14	FOXP3	KMT2A	PRKDC	TCF3	STK4
CARD9	G6PD	KMT2D	PSEN1	TCN2	TTC37
CASP10	GATA2	KRAS	PSMB8	FAAP24	RNF168
CASP8	HELLS	LAT	PTEN	CSF2RA	TREX1
CCBE1	HMOX1	LCK	PTPRC	VPS45	PSMB9
CDC42	ITGB2	LIG1	PMS2	TFRC	MOGS
COPA	ITK	LIG4	PNP	TGFBR1	SLC35C1
CORO1A	JAGN1	LPIN2	POLD1	TGFBR2	TYK2
CEBPE	JAK1	LRBA	POLE	RNASEH2B	AP3D1
CECR1	JAK3	LRRC8A	SAMD9	ADAR	CIB1
CFTR	ICOS	MSH6	SAMD9L	SBDS	CTPS1
CHD7	IFIH1	MSN	SAMHD1	SRP72	FPR1
CIITA	IFNAR1	MTHFD1	RAG1	TICAM1	G6PC3
CLEC7A	IFNAR2	MALT1	RAG2	TIRAP	INO80
CD19	IFNG	MBL2	RASGRP1	TLR2	IRF3
CD247	IFNGR1	MEFV	RBCK1	TLR3	LAMTOR2
CD27	IFNGR2	MKL1	REL	TLR4	MCM4
CD3D	IGLL1	MPO	RELA	TLR9	PGM3
CD3E	IKBKB	MRE11A	RELB	MAGT1	RFX5
CD3G	IKZF1	MS4A1	RORC	SMARCAL1	RFXANK
CD40	IL10	MYSM1	RPSA	TMC6	RFXAP
CD40LG	IL10RA	NBAS	SEMA3E	TMC8	RHOH

CD79A	IL10RB	NBN	SLC7A7	RNASEH2A	RLTPR
CD79B	IL12B	OAS1	SH2D1A	RNASEH2C	RNF31
CD81	IL12RB1	NCF2	SH3BP2	TNFAIP3	TPP2
DKC1	IL12RB2	NCF4	SKIV2L	TNFRSF13B	TRNT1
CR2	IL17A	NCSTN	SLC37A4	TNFRSF13C	TTC7A
CSF3R	IL17F	NFE2L2	SLC46A1	TNFRSF1A	ZBTB24
CTLA4	IL17RA	NFKB1	SP110	TNFRSF4	UNC13D
CTSC	IL17RC	NFKB2	SPINK5	AP3B1	UNC93B1
CXCR4	IL1RN	NFKBIA	SPPL2A	CSF2RB	UNG
CYBA	IL21	NHEJ1	VPS13B	ARPC1B	EXTL3
CYBB	IL21R	NLRC4	WAS	CD70	USB1
DCLRE1C	IL23R	NLRP12	WDR1	CDCA7	USP18
EIF2AK3	IL2RA	NLRP3	POMP	CLPB	

Table 2 Custom designed gene in PID panel

GENE NAME	GENE NAME
ALPI	NSMCE3
AP1S3	ORAI1
ARHGEF1	POLD2
ARPC1A	POLR3C
ATP6AP1	POLR3F
BCL11B	PSMB10
C17orf62	PSMB4
CD8A	PSMG2
CTPS2	RANBP2
DBR1	RIPK1
DEF6	SEC61A1
DNAJC21	SH3KBP1
EFL1	SLC39A7
ERBB2IP	SRP54
FCHO1	TAP1
GFI1	TAPBP
HAVCR2	TAP2
HYOU1	TBX1
ICOSLG	TNFRSF9
IKBKG	TNFSF12
IL18BP	TOP2B
IRF2BP2	IGHM
IRF9	IGKC
LACC1	RMRP

MAN2B2	RNU4ATAC-MOPD1
NCF1	TRAC

After preparation of Ampliseq on demand amplicons, the NGS panel was run using IonChef automatic libraries preparation on a Ion GeneStudio™ S5 Prime (ThermoFisher Scientific).
With vertical coverage under 30 reads, samples have been Sanger sequenced for the specific exons. Every variation found by NGS was confirmed by Sanger sequencing

4.2.3. Base calling and data analysis

The sequence information was evaluated using the Ion Torrent Suite™ Software v5.0 hosted on the Torrent Server. This data was then aligned to the human genome version GRCh37/hg19 using an alignment tool tailored for Ion Torrent outputs.
Following the genomic mapping, we considered unique variants with a minor allele frequency (MAF) below 0.0001 as per GenomAD and EXAC data. These were then annotated utilizing ANNOVAR. [116]
The samples that did not receive any diagnosis were selected for further CNV analysis.

4.2.4. Preprocess of BAM file for CNV calling

The reads obtained from Ion AmpliSeq™ were derived from multiplex PCR products, resulting in consistent position and length for each read originating from the same amplicon. However, the read depth varied significantly between amplicons and also differed among the various samples. SAMtools v0.1.19 [122] was used to sort and index Bam files. The BAM files underwent no further modifications or adjustments.

4.2.5. Modification of BED files to identify low mappability regions

The calculation of the mappability of the genes was carried out by employing the 35-mer mappability score [123], which was accessed through the UCSC genome browser[124]. The mean mappability across each exon was subsequently determined, and exons with a mean mappability score of 0.75 or below were excluded. This threshold was deliberately established to ensure the retention of exclusively the unique regions within the exome [108]. bigWigAverageOverBed were installed in Linux, using wgEncodeDukeMapabilityUniqueness35bp.bigWig. [125]

4.2.6. CNV Calling using HMZDelFinder

4.2.6.1 Read depth extraction from BAM files and CNV calling

HMZDelFinder was employed to identify hemizygous CNV deletions. In order to ensure the analysis was standardized for variations in sequencing depth and gene length, RPKM files (reads per thousand base pairs per million reads) were generated from each BAM file. To calculate RPKM (Reads Per Kilobase Million) values, the following steps were performed:

1) Determine the total number of reads in a sample and divide it by 1,000,000. This will serve as the "per million" scaling factor.
2) Divide the read counts by the "per million" scaling factor. This step normalizes the data for sequencing depth, resulting in reads per million (RPM).
3) Divide the RPM values by the length of the gene, measured in kilobases. This calculation yields RPKM values.

HMZDelFinder assesses the normalized per-interval coverage of all samples within the dataset to identify rare HMZ CNVs accurately. This approach effectively reduces false-positive calls resulting from regions with insufficient coverage. By default, the tools used to encompass all bed regions involved the utilization of an interval file containing all coding sequences provided by the manufacturer. The conditions for identifying a deletion within a given interval are as follows: (i) An RPKM value less than 0.65, and (ii) A deletion frequency within the dataset of no more than 0.5%.

Table3: PCR primers for HMZdel finder using go-Taq.

Gene-exon	Forward primer (5'-3')	Reverse primer (5'-3')
ERL-4	TTTGATTCTGAATTTGTTTGGGAGG	GCTTCTTTTACTTCTTTTTTCTTCACAC
CCBE1-26	GCGTTTCCAGTGCTGTCC	CTTTCACCATGTTCTCAGACTTCTC

To assess the accuracy of each identified variant, both the number of variants in a given sample and the plot of the CNVs made by HMZDelFinder were considered. Those deletion calls that had a frequency of more than two times were omitted from further analysis. The coverage of the remaining six regions with a homozygous mutation in a single exon was checked in the same-sex suspected samples using SAMtools and IGV. The amplification for identified variants were carried out with GoTaq® DNA Polymerase (Promega-USA) $MgCl_2$ for a final concentration of 1.5 mM, 50 ng DNA, and 0.5 M betaine under the following conditions: 94°C for 12 min, followed by 40 cycles of 94 °C for 30 sec, annealing at 69 °C for 30 sec, 72°C for 30 sec, and a final extension of 72 °C for 7 min.

4.2.6. CNV Calling using GATK-gCNV

In this study, the GATK-gCNV algorithm was utilized, which is a versatile tool for the identification of rare CNVs from read-depth information achieved through genome sequencing. The algorithm, packaged within GATK, was made available as an open-source tool. [126] The use of a probabilistic model and inference framework in GATK-gCNV allows for the incorporation of technical biases and the simultaneous prediction of CNVs. This enables self-consistency between technical depth adjustment and the identification of variants. NGS read-depth analysis was employed to infer copy-number variation. The GATK version 4. were run in docker container[127]. A total of 189 samples with gender-based separation from various Ion Torrent runs were processed collectively. To enhance accuracy, the captured areas extended by 250 bp on either end and the analysis was conducted in COHORT mode. CNVs identified in greater than two were eliminated as they were deemed potential capture artifacts or common CNVs unrelated to rare Mendelian disorders. The procedures GATK-gCNV were applied on BAM file as standard procedures of the algorithm. [128]

4.2.2.7 Real-time SYBR green PCR and PCR

Quantitative real-time PCR (qPCR) was implemented on genomic DNA (gDNA) using specific primers designed for sequences within the putative CNV regions. The *DOCK11* reference gene located on the X.chromosome was used for normalization. Each qPCR reaction contained 12 ng of gDNA, 200 nM of primer pairs, and 10 µl of Fast SYBR Green master mix (Life Technologies, Grand Island, NY). The amplification was conducted on a Stratagene Mx3000P® qPCR system (Agilent Technologies) using the standard thermocycling program: 95 °C for 3 minutes, 40 cycles of 95 °C for 20 seconds, 60 °C for 1 minute followed by a melting curve. Triplicates were performed for each sample. The 2−ddCT method was used to calculate relative changes in genomic sequence copy numbers, and standard propagation of errors was utilized to calculate error.

PCR was amplified using GoTaq® DNA Polymerase (Promega-USA) and conditions that mentioned before.

Table 4: Primers for CNV analysis identified by GATK-gCNV

Gene-exon	Forward primer (5'-3')	Reverse primer (5'-3')	Experiments
ZAP70-4	CCGCAGGTGGAGAAGCTC	CACATACAGGAACTTGCCGTC	q-PCR
MPO-7	CTGAGCGGTTGGTGAGGAG	GTGTACGGCAGCGAGGAG	q-PCR
ITK-17	ATACAGTCAACATGGAAGCACATAC	CCGATCTTCTGGTCTCTGGAG	q-PCR
PRKDC-34	CTGTTATTTTCTCTCTCAGATGGGTC	GCCCGCAAGCATAGGTTTC	q-PCR
CLPB-3	CCTGCCTCTGGAAAAGACAGC	CATCAACCGAAACAACAGGTGAG	q-PCR
ATM-31	ATTTTATCATTTATTACAGTAAGTTTTGTTGG	CGTGATATAGAGGTTTTCATTATCCTTG	q-PCR
ERBB-21	GCAGCCCGTAATGTCTTAGTG	ATTGCTCTCAAAAAGATACCCACC	q-PCR
POLE-13	GCCGCACACACAGTAAGGAGAC	GTGCCTGTTAGGAACTTGCATCTG	PCR

4.2.7 MAK-Alu detection

The Linux utility called "grep" was utilized to conduct a search for a specific 23 base pair "probe" sequence that contained the established junction sequence of the insert. [79] The analysis of sequence reads is conducted to identify the presence of the well-known pathogenic MAK-Alu insertion[129]. The NGS reads from the samples were scanned using the Linux tool "grep" to identify a 23-base pair "probe" sequence. This sequence was specifically sought after because it contained the recognized junction sequence of the insert [79].

4.3. Molecular characterization and surveillance of eight Iranian families suspected of hereditary cancer

4.3.1 Molecular characterization and clinical history of six Iranian families with targeted panel sequencing

4.3.1.1 Patients and samples collection

In this study, we evaluated eight Iranian patients with a personal or familial history of cancers. The study gained approval from the ethics committee of Motamed Cancer Institute in Tehran, Iran. Prior to genetic testing, all eligible family members or their parents received genetic counselling and provided informed consent. Through interviews, we gathered demographic information, gender, ancestral background, clinical history, personal and familial history of cancers/polyps, age at diagnosis, current surveillance for LS, insurance coverage, and detailed family trees. Pathology reports were used to confirm cancer cases mentioned in the tumor spectrum for mutation carriers. The subjects were unrelated to each other, and their histories of cancers were not connected.

4.3.1.2 Multigene panel testing, CNV analysis, and MLPA

After the isolation of genomic DNA from the peripheral blood samples of all subjects, a commercial multigene panel sequencing was employed to analyze a set of 70 genes (*Table 5*).

Table 5 genes included in the diagnostic targeted panel

GENE NAME	GENE NAME	GENE NAME	GENE NAME
AIP	FANCC	NTHL1	SDHA
ALK	FH	PALB2	SDHAF2
APC	FLCN	PHOX2B	SDHB
ATM	GALNT12	PMS1	SDHC
AXIN2	GREM1	PMS2	SDHD
BAP1	HOXB13	POLD1	SMAD4
BARD1	MAX	POLE	SMARCA4
BLM	MEN1	POT1	SMARCB1
BMPR1A	MET	PRKAR1A	STK11
BRCA1	MITF	PTEN	SUFU
BRCA2	MLH1	PTCH1	TMEM127
BRIP1	MRE11A	RAD50	TP53
CDH1	MSH2	RAD51C	TSC1
CDK4	MSH6	RAD51D	TSC2
CDKN2A	MUTYH	RB1	VHL
CHEK2	NBN	RECQL	WT1
DICER1	NF1	RET	
EPCAM	NF2	SCG5/GREM1	

The FASTQ format sequencing reads were aligned to to the hg19 reference genome (NCBI Build 37) by the HISAT2 aligner. [130] The detected indel and single nucleotide variations (SNV) variants, identified by GATK HaplotypeCaller - v4.1, were annotated using the ANNOVAR stand-alone tool. [131] Variants having a minor allele frequency greater than 1% (MAF < 1%) in the 1000 Genome Project, gnomAD, and Exome Aggregation Consortium (ExAC) were eliminated. The VarSome database, Franklin by Genoox (https://franklin.genoox.com), ClinVar, MutationTaster, dbSNP, and ACMG standard guidelines were employed to assess the pathogenicity of the detected variants.[132,133] The ExomeDepth and CONTRA software packages were utilized to perform CNV analysis for the purpose of detecting large deletions. [134,135] The control samples were used to compare the coverage of susceptible CNVs, which was examined using both IGV and SAMtools. [136] The AnnotSV software, accessible at http://lbgi.fr/AnnotSV/, was employed to annotate and interpret the pathogenicity of the CNV variants detected. [137] The verification of CNVs was conducted using multiplex ligation-dependent probe amplification (MLPA) technique with SALSA MLPA Probemix P003-D1 MLH1/MSH2, following the guidelines provided by the manufacturer.

4.3.1.3 Microsatellite instability and immunohistochemistry analysis

The CAT25 mononucleotide marker was employed to conduct MSI analysis on both CRC and blood samples obtained from the second-cousin of the proband belonging to family C. Additionally, the analysis was also performed on samples obtained from two other healthy individuals, with three replicates conducted for each individual, as previously described. [138,139] The application of GeneMapper 5 was employed for the purpose of data analysis. The examination of MMR proteins using antibodies was carried out in accordance with established methodologies, as previously described. Plagiarism was avoided in the process of rewriting the content. [140]

4.3.1.4 The examination of cascading effects and segregation analysis

To validate the existence of pathogenic mutations in both the probands and their at-risk relatives, Sanger sequencing was employed. The method involved the extraction of DNA from peripheral blood, as directed by the manufacturer's guidelines (Promega's Wizard® Genomic DNA Purification Kit, USA). The relevant exons were amplified via PCR, followed by sequencing using ABI 3130 genetic Analyzer (Applied Biosystems, USA).

4.2.2 Two families with hereditary CRC tested by WES

4.2.2.1 Subjects enrolled in the study

Two families with high burden CRCs history resided in the same village in the Mazandaran province of Northern Iran, were referred to the Motamed Cancer Institute in Tehran, Iran for genetic testing. Before undergoing testing, genetic counseling was provided to all family members, and informed consent forms were signed. Interviews were conducted to collect data on demographic information, clinical history, ancestry, history of malignancies/polyps, age of onset of cancer, and current surveillance programs. Detailed family trees were also obtained. Pathology reports were analyzed to confirm cancer incidences in individuals who tested positive for mutations.

4.2.2.2 Clinical Phenotype:

Both Family H and I met the criteria for Amsterdam I. [141,142] In the case of Family A, the patient (identified by number) was diagnosed with fundic gland in the stomach, which was later confirmed through pathology report. Subsequently, NGS was implemented on the DNA isolated from peripheral blood of this patient. The analysis of four MMR genes using IHC indicated that both MSH2 and MSH6 proteins were not present in the FFPE CRC tumor of Family A's proband.

4.2.2.3 PCR and Sanger sequencing of *MSH2* gene

The genomic DNA obtained from peripheral blood was utilized for genetic mutation testing, following the IHC report of the proband's CRC tumour, which indicated a lack of *MSH2* and *MSH6* protein expression. A blood DNA extraction kit (Promega, USA) was utilized to extract the DNA. *MSH2* primers gene were designed with flanking regions of 100 bp for all 16 exons of the *MSH2* gene (NM_000251.2) using the primers listed in *Table 6*. The amplification was carried out with GoTaq® DNA Polymerase (Promega-USA) $MgCl_2$ for a final concentration of 1.5 mM, 50 ng DNA, and 0.5 M betaine under the following conditions: 94°C for 12 min, followed by 40 cycles of 94 °C for 30 sec, annealing at 69 °C for 30 sec, 72°C for 30 sec, and a final extension of 72 °C for 7 min. Sanger sequencing on all 16 exons of the *MSH2* gene was performed in both forward and reverse directions using the 1µL BigDye Terminator and 1 reaction of ExoSAP, BigDye XTerminator Purification Kit (Applied Biosystems, USA). The ABI3130 Genetic Analyzer was used according to the manufacturer's protocol.

Table 6 – Primers for MSH2 gene amplification

Gene-exon	Forward primer (5'-3')	Reverse primer (5'-3')
MSH2-1	TTCCTTCTGATGTTACTCCCATGC	TCCGCACAAGCACCAACG
MSH2-2	CTTTCCCATGCTGTTGTGATAGTG	CACATGCTAATTGCTATTAAAGTGTCTC
MSH2-3	AATAAGGTTCATAGAGTTTGGATTTTTCC	GTAAAATAACAGGCTAAAGTTCCACTC
MSH2-4	TATCAGTACATCATATCAGTGTCTTGC	ATCATTGATACACAGTTTAGGTTTTGAG
MSH2-5	CCAAGGAAAATGAGGGACTTCAG	GTGGAGGAGGGGAGAGAAAAATAC
MSH2-6	AGTTGAACATACGGATTAAGAGGTTG	TGAGTAATTTACCCACGATTACACAC
MSH2-7	AGTTGAGACTTACGTGCTTAGTTG	GAAATCTGAATGTGTCCTAAGAGTGAG
MSH2-8	TGGGTTTACAGACGAGGTAGTG	TGTCTTTCACCAGGACAGTTATG
MSH2-9	TCCTTGGTTTGGGCAACATATAAC	TAGAAGAATTGGGCTTGGTGATCC
MSH2-10	AAGTCAGAAACTAACATTCATAAGGGAG	AATGGAGAACAGACGGACTATTTAAC
MSH2-11	GACACAGATCACAGTTTTACTCAGG	CATAGATGACCCAAGACATTAGTACG
MSH2-12	CTATGTTGAGTTTTAGGTGGGTTCC	TTTTGCGAGTTTCACATGAATATAGTG
MSH2-13	CAGGTATATTTGTCATGGCTTCTCTTG	CAAAGTATATAAAGTCCACAGGAAAACAAC
MSH2-14	GCCCATTTTTCTATTGAAGTTTTAGTGC	ATGAGTGGTCCTACTATGAGATACAG
MSH2-15	GAAATGAGAAAGCCTCAGAGATAGTG	TAACACAGAGATAGATTCTTTGCCATC
MSH2-16	GGAAATGAAACAATTTGTCACTGTCTAAC	TATTTGAAGTCACACTGCGAAGAAC

4.2.2.4 Whole exome sequencing

Two patients from family A, a female affected with CRC and her maternal aunt with a history of developing at least six polyps in her stomach every four years, underwent WES. Germline DNA was extracted from their peripheral blood and Agilent SureSelect V6 60Mb PE150 was used to prepare the

DNA libraries. The samples were then sequenced on an Illumina NovaSeq PE150, resulting in an average coverage of 150X for all cases and approximately 12 gigabases of sequenced data per sample.

4.2.2.5 Bioinformatic analysis

The sequencing reads were mapped to the reference genome using hg38 and BWA v0.7.17, and the variants were annotated using ANNOVAR 2015Dec14.

4.2.2.6 CNV analysis and validation of identified CNV and Segregation analysis

CNV calling was carried out using CNVkit and ExomeDepth algorithms and was confirmed using IGV tools. [134,143,144] Multiplex ligation-dependent probe amplification (MLPA) using SALSA MLPA Probemix P003-D *MLH1/MSH2* (MRC Holland), which contains probes for 16 exons of the *MSH2* gene as well as 3' deletions in *EPCAM*, was performed on the probands and available at-risk relatives as previously described. The amplified products were separated using an ABI 3130 Genetic Analyzer (Applied Biosystems, USA). MLPA analyses were conducted using the coffalyse.net software (MRC-Holland).

4.2.2.7 Real-time SYBR green PCR

Genomic DNA was subjected to qPCR analysis, employing custom-designed primers targeting sequences within the suspected CNV regions identified through GTAK-gcnv analysis. The following primers were used for the experiment.

Table 7 Primers used for real-time PCR in the analysis of predicted CNVs identified in EPCAM/MSH2 gene.

Gene-exon	Forward primer (5'-3')	Reverse primer (5'-3')
REAL-*MSH2*-3	ATTGGTGTTGTGGGTGTTAAAATGTC	GAGAGCCTCAAGATTGGAGAACTG
REAL-*MSH2*-10	GAATTACATTGAAAAATGGTAGTAGGTATTTATGG	CCTTACAGGTTACACGAAAGTAATATCC
REAL-EPCAM-1	GTTCGGGCTTCTGCTTGC	CTCTTGGTCCCCTCCCTATTATG

The length of the CNV in the proband family A and her father, as well as the proband of family B, was determined through the performance of SYBR green real-time PCR (iTaq Universal SYBR Green Supermix, BIORAD) on the exon at the beginning (*EPCAM* exon 1) and at the end (*MSH2* exon 8) of the variant. The carrier status of all eligible available members, including the proband from family A and her father, as well as the proband of family B, was reconfirmed by performing SYBR Green real-time PCR on exon 3 of the *MSH2* gene. The regions of interest were targeted using primers designed through Primer3 and NCBI primer blast. To assess efficiency of each primer pair, calculation of amplification efficiency was based on the creation of standard curves from a series of diluted control

genomic DNA (gDNA). [145] In order to confirm the precision of the qPCR procedure utilized to assess the copy number status of the *DOCK11* gene, located on the X chromosome, it was amplified in triplicate from a total of three males and four females simultaneously. The results from the real-time PCR analysis showed that the gene amplification had one cycle threshold (Ct) less in female samples compared to male samples. To identify carrier status in patients, the reference control gene *DOCK11* was utilized. The study conducted gene amplification in triplicate, alongside control positive DNA samples. Copy number variations were evaluated by quantifying relative to thresholds. Values below 0.7 were classified as heterozygous deletions, whereas values between 0.7 and 1.2 were considered normal for the target locus. Custom R scripts were used to visualize the data.

5. Results

5.1 A pathogenic variants in *CYBB* patient causing exon skipping

The patient was a 9-year-old female born of non-consanguineous parents of Italian descent. Family history was negative for PID, and her past medical history was uneventful. At the age of 6 years, she was admitted at the Pediatric clinic, Spedali Civili, Brescia due to right lobar pneumonia unresponsive to conventional antibiotic treatment (amoxicillin-clavulanate and clarithromycin). Then a subcutaneous abscess developed on the left leg and rapidly spontaneously fistulized; this was tested positive for *Burkholderia gladioli* which is an opportunistic Gram-negative bacillus that mainly infects immunocompromised patients; few pediatric patients with *B. gladioli* infection are reported and are predominantly affected by CGD, then the complete PID NGS panel was applied and DHR123 test upon stimulation with Phorbol 12-myristate 13-acetate (PMA) revealed partial reduction, identifying two different subset of non-reducing (87.9%) and reducing (12.1%) granulocytes (*Figure 6*) . A clinical diagnosis of CGD was done and trimethoprim-cotrimoxazole plus voriconazole prophylaxis was started. A novel germline heterozygous variant in *CYBB* (NM_000397:ex9:c.1151+2T>C) was identified. The mutation was confirmed through Sanger sequencing analysis of the patient's genomic DNA, while the parents exhibited a wild-type genotype (*Figure 7*).

To study the puzzling phenotype of an affected female due to heterozygous X-defect we perform RT-PCR analysis to check whether the point mutation was pathogenic, possibly a splicing defect.

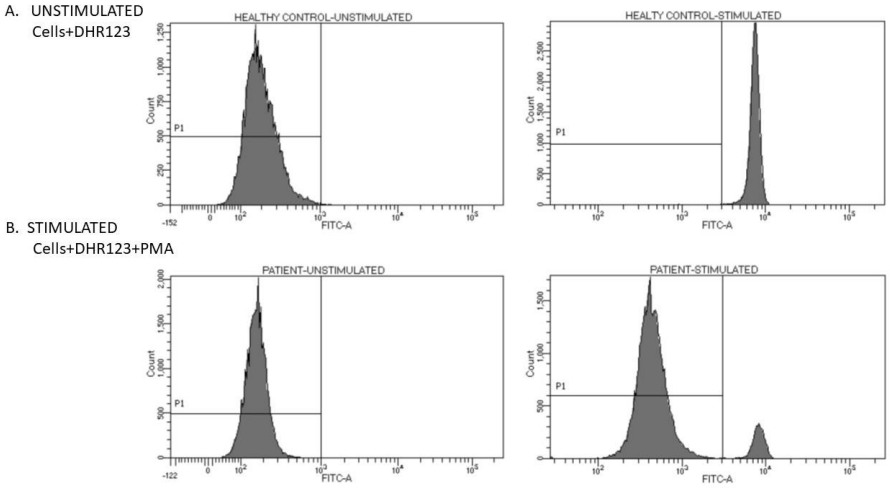


Figure 6 DHR123 test upon stimulation with Phorbol 12-myristate 13-acetate (PMA) revealed two different subset of non-reducing (87.9%) and reducing (12.1%) granulocytes

The gel electrophoresis analysis of *CYBB* transcripts from PBMCs of the patient, her parents, and two controls around exon 9, showed the presence of two distinct bands in the patient's sample, while the other samples exhibited only the upper wild-type fragment (*Figure 8*). This observation reinforces the

hypothesis that the variant could have influence on the correct splicing in the patient. The lower fragment was the extracted from the gel and Sanger sequenced confirming the complete exclusion of exon 9. This event led to the deletion of 254 nucleotides in the abnormal transcript, resulting in the direct fusion of exon 8 and 10 of the *CYBB* gene. The skipping of exon 9 probably generate a unfunctional protein, as the fusion is frameshift. Segregation analysis performed by Sanger sequencing in the patient's parents indicated that they are not carriers, thus the mutation occurred *de novo*.

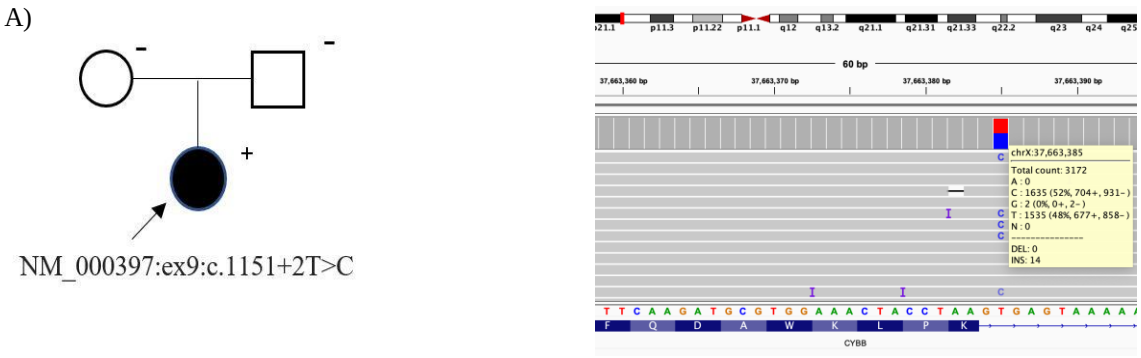


Figure 7 – A) Pedigree and NGS analysis result on genomic DNA of the proband. B) IGV representation of the heterozygous variant detected.

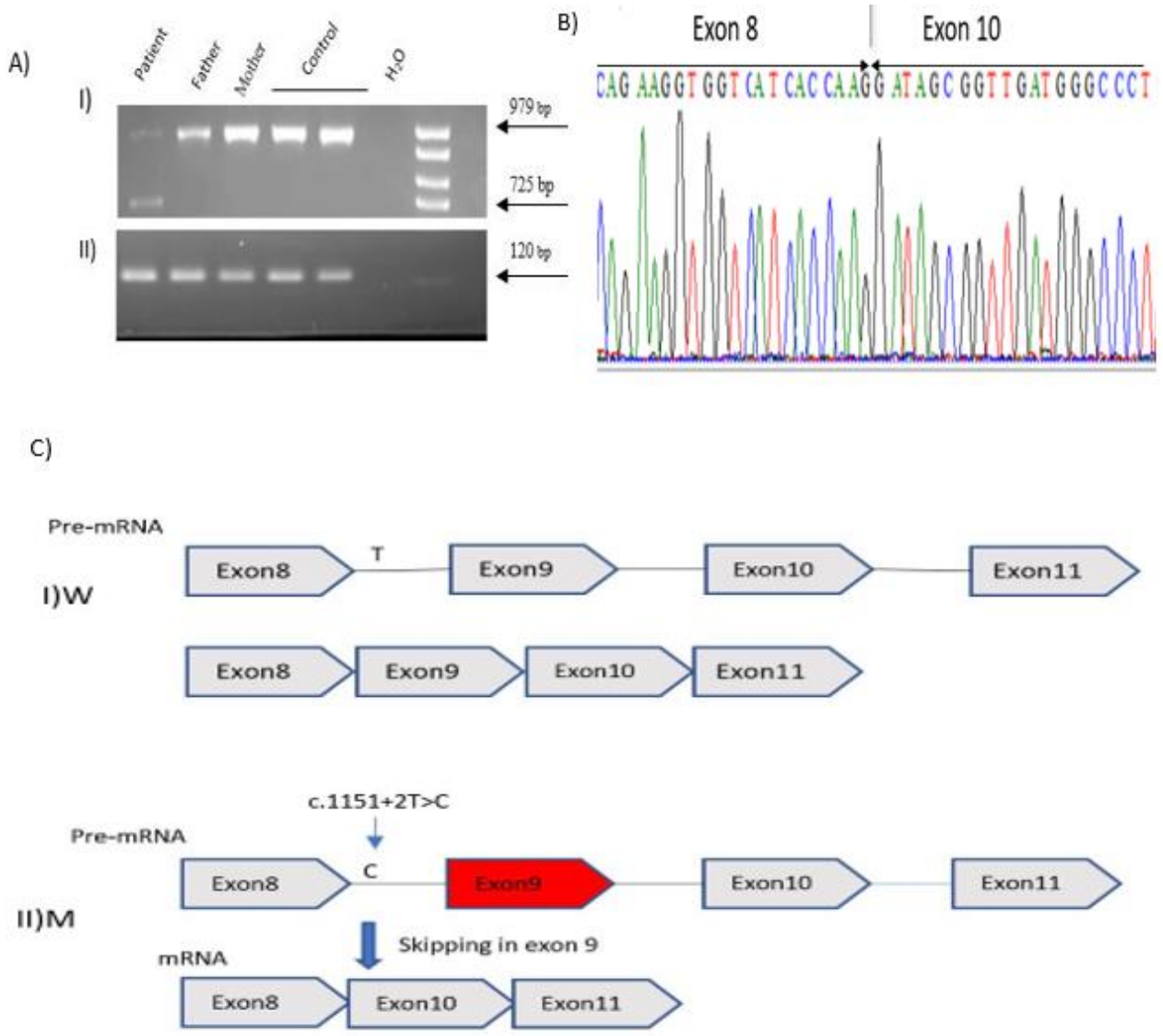


Figure 8 Exon 9 skipping in the CYBB gene in the proband mRNA. A) Agarose gel electrophoresis was used to analyze cDNA obtained from patient, father, mother, and two healthy controls for size of CYBB transcript by visualizing the RT-PCR product. I) Patient showed a smaller size band compatible with exon skipping due to a splicing defect, and a normal size band, II) Beta-actin as quality control, B) Sanger sequencing of the patient's lower CYBB fragment revealed total skipping of exon 9. C) A diagrammatic representation of the process of pre-mRNA splicing around exon 9 of the CYBB is depicted. Splicing near exon 9 in I) wild type (W), II) the mutated type (M) displays alternative splicing, which involves the skipping of exon 9, as a result of a pathogenic variant in the splicing site. The picture illustrates the exclusion of exon 9, which is highlighted in red.

5.1.2 X-chromosome inactivation analysis in *CYBB* heterozygous female patient.

The Reverse-transcription polymerase chain reaction (RT-PCR) on the *CYBB* mRNA of the patient showed two transcripts demonstrating that the putative mutation is indeed a splicing variant, thus next question was to understand why a female heterozygous for the mutation is affected.

HUMARA inactivation analysis is a test allowing the discrimination of maternal/paternal alleles and activated/inactivated alleles.

The test relies on the amplification of a CAG repeat and the digestion with a methylation sensitive enzyme. In our case we tried to set up the test as fragment analysis on an automatic sequencer, instead as an electrophoresis technique. This led to the need of optimize Plus-A artifacts and stutter fragments.

In general, two discrete sets of optimizations were chosen, the first involving the utilization of AmpliTaq Gold™ DNA polymerases for amplification, while the second utilized GoTaq® DNA polymerases.

The series of optimization experiments named PCR-A are based on AmpliTaqGod, the amplification was conducted utilizing 2 mM of Mg^{2+}. To perform PCR-A, thermal cycling was initiated with a denaturation step at 95°C for 12 min, followed by 38 amplification cycles of 95°C for 30 seconds, 69°C for 30 seconds, and 72°C for 30 seconds, with a final extension step at 72°C for 7 min.

To improve the efficiency of the PCR reaction, a longer final extension step of 1 hour was implemented in PCR set -B. This modification aims to enhance the yield and specificity of the PCR product. The efficacy of the PCR-B protocol was evaluated by introducing two additional final concentrations 1 mM and 2.5 mM of Mg^{2+}, into the experimental design. PCR amplification was conducted using AmpliTaq Gold™ DNA polymerase enzyme, with a final concentration of 1 mM Mg2+ and 10% DMSO under different conditions. The first condition involved PCR-B protocol and 50 ng DNA template, while the second condition involved using 100 ng of DNA template and a reduced cycle number from 40 to 28 of PCR-B protocol, referred to as PCR-C.

PCR amplification was conducted using AmpliTaq Gold™ DNA polymerase enzyme, with a final concentration of 1 mM Mg^{2+} and 10% DMSO under different conditions. The first condition involved PCR-B protocol and 50 ng DNA template, while the second condition involved using 100 ng of DNA template and a reduced cycle number from 40 to 28 of PCR-B protocol, referred to as PCR-C.

For the latter series of amplifications, the GoTaq® DNA polymerases from Promega (USA) were utilized along with 1 mM of Mg^{2+} and 50 ng of DNA template, and a final concentration of 0.5 M betaine. The amplification protocol consisted of denaturation at 95 °C for 12 minutes, followed by 48 cycles of 95 °C for 30 seconds, 69 °C for 30 seconds, and 72 °C for 30 seconds. A final extension was carried out at 72 °C for 10 minutes, resulting in PCR-D.

For the subsequent trial, the PCR-D cycles were reduced to 28 cycles, resulting in PCR-E.

The analysis of DNA amplicons using 50 ng of DNA, AmpliTaq Gold™ DNA Polymerase, and PCR-A protocol revealed the occurrence of Plus-A artifacts and stutter fragments, which required optimization as illustrated in *Figure 9A*. The utilization of the PCR-B protocol did not produce a noteworthy effect in eradicating plus-A artifacts or stutter fragments. A reduction in the production of smaller stutter fragments or plus-A artifact was observed as a result of amplification using AmpliTaq Gold™ DNA Polymerase, PCR-B protocol and the use of two final concentrations of 1, and 2.5 mM Mg^{2+}, as shown in *Figure 9B*. Two different strategies were applied, along with the addition of 10% DMSO (v/v) and amplification using the AmpliTaq Gold™ DNA Polymerase enzyme, a final concentration of 1 mM of Mg^{2+}: utilizing one PCR-B protocol with 50 ng of DNA template and another PCR-C protocol with 100 ng of DNA template, and both conditions resulted in to halt the PCR reactions, as demonstrated in *Figure 9C*.

Methylation analysis of X-chromosome using HUMARA assay showed skewed-inactivation (16%) in the patient. Different approaches to omit plus-A artifact and stutter products in fragments analysis of HUMARA assay were assessed [35]. Amplification using GoTaq® DNA polymerase and fewer PCR cycles eliminated plus-A artifact and reduced stutter fragments, respectively. Using AmpliTaq Gold™ DNA polymerases enzyme or the different concentrations of Mg^{2+} cannot decrease plus-A artifact and stutter fragments. While adding betaine with the final concentration of 0.5 M increased the PCR yield without any decrease in plus-A artifacts or stutter fragments, the addition of 10% DMSO stops PCR amplification. To eliminate plus-A and stutter fragments in the fragment analysis. Amplification using GoTaq® DNA polymerase and decreasing the PCR.

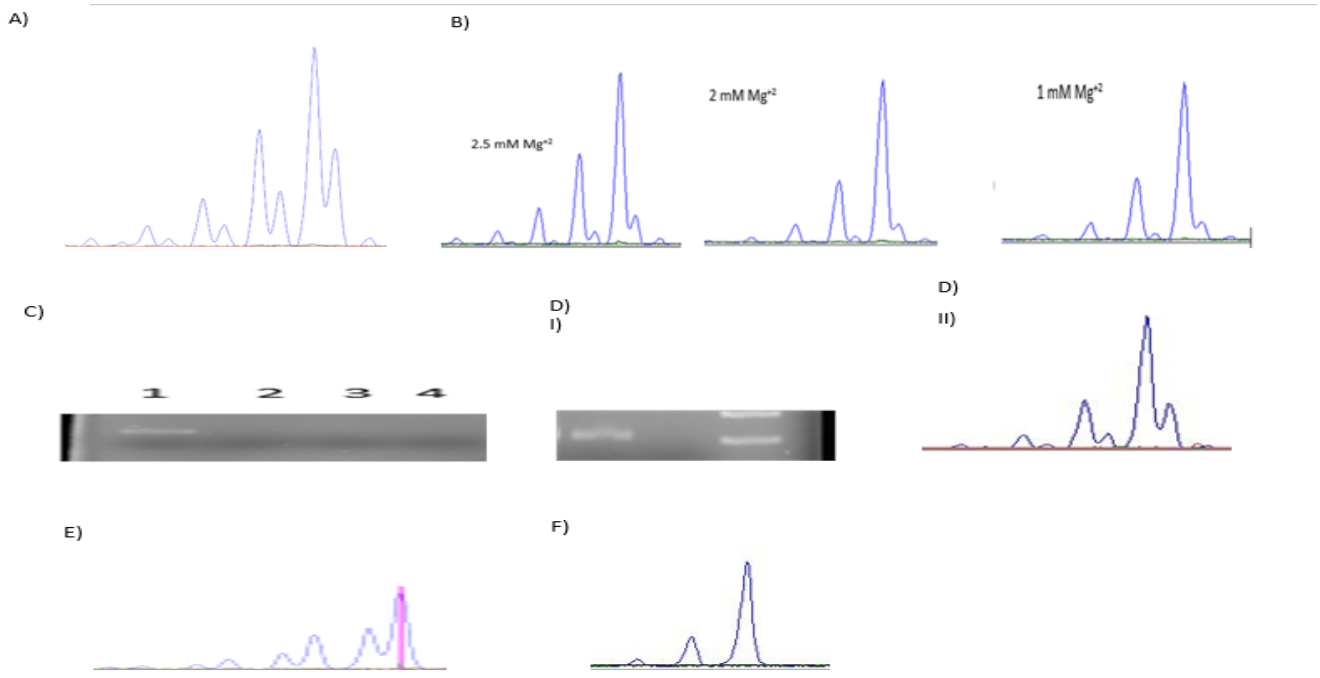


Figure 9. Analysis of the HUMARA locus through fragment assays. A) The fragment analysis of the PCR-A setting using the AmpliTaq Gold™ DNA Polymerase enzyme and 50 ng of DNA. B) The fragment analysis utilizing AmpliTaq Gold™ DNA Polymerase enzyme utilizing of 1, 2, and 2.5 mM of Mg2+ in final concentrations and the PCR-A procedure C) Investigating the influence of DMSO on AmpliTaq Gold™ DNA Polymerase enzyme amplification on agarose gel. 1: a positive control, the approach would be to utilize 50 nanograms of DNA and conduct 40 cycles of PCR, without the presence of DMSO. 2: The conditions for this step are similar to step 1, with the addition of 10% (V/V) DMSO.
3: The conditions for this step are similar to step 2, with the addition of 100 ng DNA template. Additionally, the amplification is performed for 28 cycles. 4: This step involves the use of water as template under similar conditions as step 1. D) The effect of introducing betaine at a final concentration of 0.5 M using AmpliTaq Gold™ DNA Polymerase enzyme and PCR Protocol-B. I) gel agarose II) fragment analysis E) The use of GoTaq enzyme, 1 mM Mg, 0.5M betaine, and 48 PCR cycles has an impact on the reaction. F) Condition E by reducing the number of cycles to 28.

The utilization of AmpliTaq Gold™ DNA Polymerase-tag enzyme and PCR Protocol-B, in combination with a final volume of 0.5M betaine, resulted in an amplified PCR product yield enhancement. Nevertheless, despite the presence of 0.5M betaine, fragment analysis indicated that neither plus-A artifacts nor stutter fragments were eliminated, as depicted in *Figure 9D.II* In the second round of optimizing amplification, GoTaq® DNA Polymerase (Promega, USA) was utilized. The protocol involved using a final concentration of 1 mM Mg2+, 50 ng of DNA template, a final concentration of 0.5M betaine, and PCR-D. As a result, there was an improvement in the PCR yield and the omission of the plus-A fragment. Nevertheless, there were noticeable large stutter fragments, as shown in *Figure 9E*. By reducing the PCR cycles to 28 (PCR-E protocol), the PCR conditions were optimized, and this led to a significant decrease in stutter fragments (*Figure 9F*). When the HUMARA gene was amplified using the optimized PCR conditions on undigested DNA samples from the patient and her parents, it was found that the patient and parents had two (heterozygous) and one allele, respectively (as illustrated in *Figure 10*). The X-chromosome inactivation of the patient showed an imbalanced distribution, with 16% coming from the father and 84% from the mother.

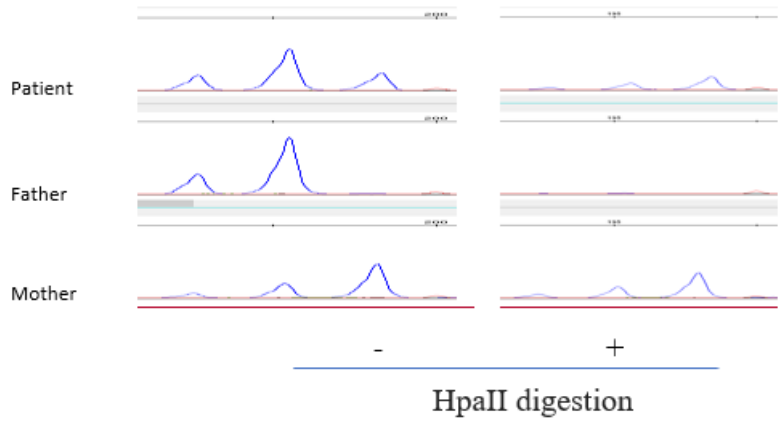


Figure 10. The analysis of X-chromosome inactivation in PBMC cells of the patient, father, and mother. PCR products of the patient and her parents were subjected to HpaII treatment (+) and a "mock" digestion (–), and fragment analysis was performed.

5.2. A patient with a pathogenic splicing variant in *MAGT1* gene

The patient was a 12-year-old male born from non-consanguineous Italian parents, referred to the diagnostic lab for a common variable immunodeficiency (CVID). Surprisingly the PID panel NGS analysis identify a newly identified hemizygous mutation in the *MAGT1* gene (NM_032121:c.627+2T>C) supposed to be a splicing variant.

The mutation analysis of aberrant mRNA splicing in the patient was done through RT-PCR and amplified a smaller fragment as compared to the normal control. Additionally, a double band was observed in the gel electrophoresis of the mother's sample, showing normal and mutated transcripts. (*Figure 11A*). The transcript of the proband was subjected to sequencing, revealing two key findings: exon 4 skipping and alternative splicing involving a heterozygous insertion of 63 base pairs from the beginning of exon 6 to the beginning of exon 5. This insertion was accompanied by some mismatched bases (*Figure 11 B, C and D*). These findings strongly indicate a causal effect of the variant.

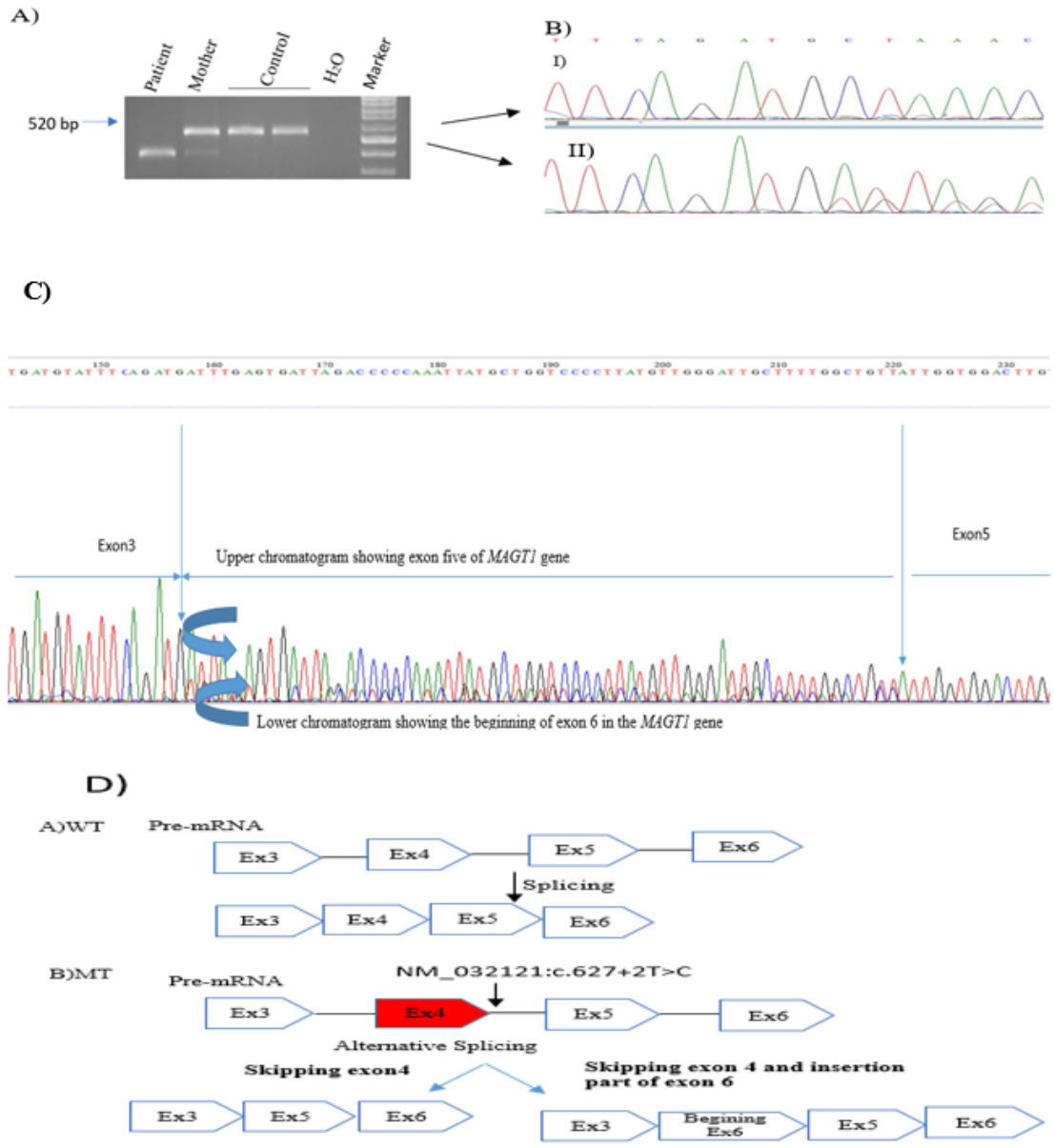


Figure 11 Exon skipping of MAGT1 in exon 4 of the proband's mRNA. A) Agarose gel electrophoresis was utilized to analyze MAGT1 transcript in blood samples from the patient, mother, and two healthy controls visualizing by the RT-PCR product overlapping regions of exons 2-6 to the 3'UTR. B) The chromatogram of MAGT1 in exon 3 and the following exons analyzed using Sanger sequencing I) Wild type and normal control of MAGT1, II) patient's fragment showing the exon 3 followed by exon5 and insertion of the beginning of exon 6 C) The chromatogram of the Sanger sequencing analysis of the MAGT1 transcript in patient showing exon 4 skipping and a heterozygous insertion of exon 6 at the beginning of exon 5. after exon 3. The upper chromatogram displays the exon 5 transcript, while the lower chromatogram specifically highlights exon 6. D) The diagram illustrates the process of pre-mRNA splicing, focusing on the adjacent exons 4 of the MAGT1 gene. It compares the normal control and the Proband in both the wild type (WT) and mutated type (MT). In the mutated type, alternative splicing occurs due to a pathogenic variant in the splicing site, leading to the skipping of exon 4. The exclusion of exon 4 is highlighted in the diagram in red.

5.3 Detection of rare CNVs on a panel of PID genes

The identification of germline CNVs across a targeted panel comprising 351 genes was accomplished through the analysis of read-depth information utilizing two advanced tools, namely GATK-gCNV and HMZDelFinder.

5.3.1 Compute mappability score in PID targeted panel genes

An adaptation was made to the standard HMZDelFinder algorithm workflow to minimize the occurrence of false positives associated with repetitive genomic regions. The mean mappability across each exon with a mean mappability score of 0.75 or below were excluded.[108] This threshold roughly corresponded to the unique regions in the target genes and caused the omission of 80 (1.2%) out of the total 6478 targets mostly in *PMS2* gene. Then different non-commercia tools (GATK-gCNV, HMZDelFinder) to detect CNVs applied.

While The HMZDelFinder algorithm using R programming in Linux environment detected after removing the regions with low mappability the GATK-gCNV was performed using standard protocol.

5.3.2 CNV Calling using HMZDelFinder

5.3.2.1 Assessment of the algorithm's performance conducted by analyzing five targeted samples obtained from other panels.

The HMZDelFinder algorithm was employed to detect possible homozygous (HMZ) deletions in the targeted NGS data by calling these variants collectively throughout the entire set of sample data. This process aimed to identify any potential HMZ deletions present in the data in classic heterogeneous laboratory patient collections, which had been produced In the long run and in diverse experimental parameters (e.g., capture kit). A R script were used to convert every BAM file into per-exon read depth, measured in reads per thousand base pairs per million reads (RPKM). To ensure the accuracy of CNV calling, the first series of tests included all the BAM files along with an additional five samples that had smaller panel genes in their panel sequencing. This was done to validate the reliability of the CNV calling method on data from genes with fewer panel genes. The inclusion of these samples provided a means of testing the accuracy and sensitivity of the CNV calling algorithm in less challenging conditions, where a subset of genes was absent. The HMZDelFinder algorithm effectively detected homozygous deletions in all five samples. Also, six samples were identified as low-quality due to the presence of multiple homozygous deletions in numerous locus regions. These samples were deemed unreliable and were subsequently excluded from further analysis to maintain the integrity and accuracy of the data.

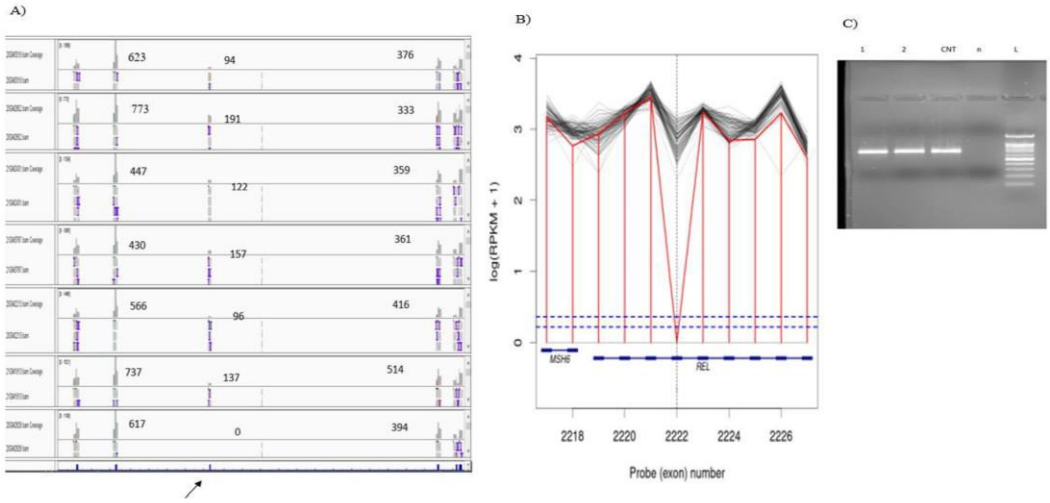
5.3.2.2 CNV calling using HMZDelFinder in the cohort of PID patients

The remaining 189 sample data sets were selected from our cohort. The analysis of panel targets was performed using the BED file, which contained the genomic coordinates of the targets utilized for targeted sequencing. Durring the process of calling HMZ CNVs, the male and female samples were separated, and individual calls were made for each of them. A total of 54 deletion calls were identified, consisting of 53 autosomal homozygous and one hemizygous (on the X-chromosome in males), without excluding low mappability regions in bed files. The codes are provided in github https://github.com/MohmSina/HMZDelFinder/blob/main/HMZpublish.R

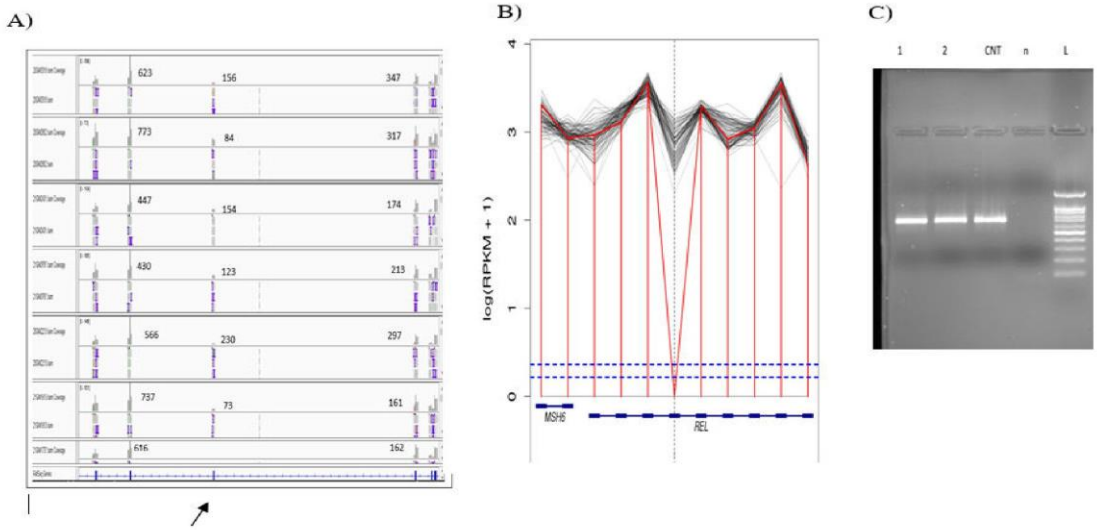
5.3.2.3 Confirmation of identified CNV calls using HMZDelFinder in silico and wetlab from PID cohort

To assess the accuracy of each identified variant, two factors were taken into account: the quantity of variants present in a given sample, and the CNV plots generated by HMZDelFinder. The plots were manually reviewed to remove potential false positive calls, also deletion calls that occurred more than twice across the identified variants were excluded from further analysis. Therefore, six samples exhibited mutations in one of the following genes: *POLA1*, *EPG5*, *CCBE1*, *IKBKB*, *CSF3R*, and *IFNGR2*. Additionally, two samples were found to have homozygous deletions in the *REL* gene. SAMtools and IGV were used to evaluate the coverage of the regions, including the upstream, the suspected predicted exons, and downstream exons of the suspected homozygous regions, in suspected same-sex samples. Those samples that suffering from low coverage in other samples too including, *POLA1*, *IKBKB*, *CSF3R*, *EPG5* and *IFNGR2* genes were excluded from further analysis. Notably, no other samples were found to have low read counts in the *CCBE1* and *REL* genes . The remaining three samples were subjected to breakpoint PCR amplification for the identified copy number variations (CNVs) to enhance the accuracy of mapping the predicted CNVs. The primers specifically designed based on the genomic coordinates associated with each predicted deletion. All of the samples were amplified during PCR reaction, this means that none of the CNVs detected by HMZDelFinder was verified as shown in *Figure 12*.

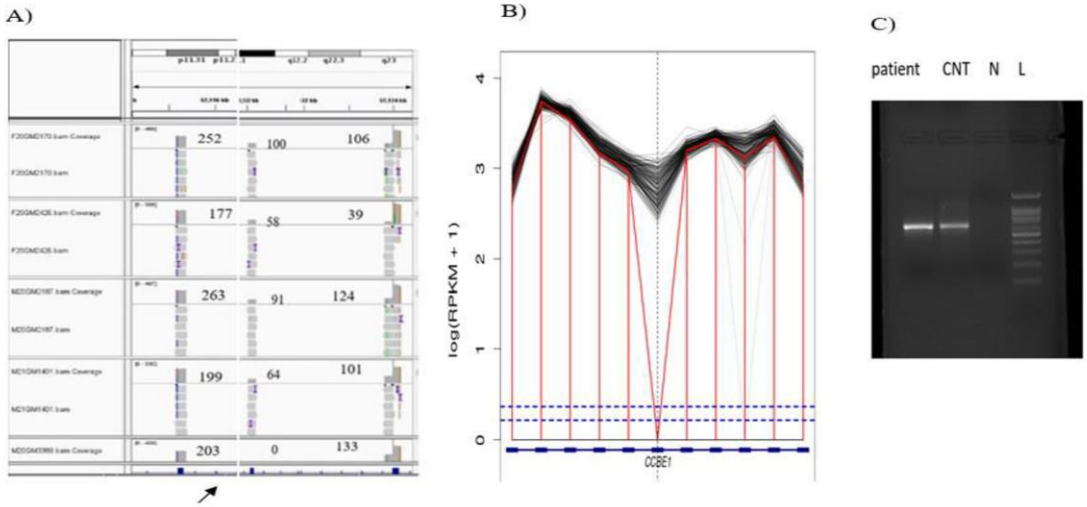
I)


A)
623
94
376
773
191
333
447
122
359
430
157
361
566
96
416
737
137
514
617
0
394
B)
log(RPKM + 1)
MSH6
REL
2218
2220
2222
2224
2226
Probe (exon) number
C)
1
2
CNT
n
L

II)


A)
623
156
347
773
84
317
447
154
174
430
123
213
566
230
297
737
73
161
616
162
B)
log(RPKM + 1)
MSH6
REL
C)
1
2
CNT
n
L

III)


A)
252
100
106
177
58
39
263
91
124
199
64
101
203
0
133
B)
log(RPKM + 1)
CCBE1
C)
patient
CNT
N
L

Supplementary Figure 1: Validation of the predicted deletions identified using HMZdelfinder. I) patient number one suspected of REL mutation, II) patient number 2 suspected of REL mutation, III) patient number 3 suspected of mutation CCBE1 A) Integrative genomics viewer (IGV) visualization of the predicted CNV regions in samples. The regions were shown by arrows in the samples both upstream and downstream sequence realignment. B) The HMZ deletion plots generated by HMZdelfinder algorithm exhibit loci containing deleted exons and neighboring exons, with RPKM values displayed on the Y-axis using a log scale. A vertical black dashed line shows the exon deletion, and a red vertical line joins RPKM quantity at the deleted exon and neighboring exons for the sample. RPKM values for the remaining samples are shown with black lines. The threshold RPKM value used in the study is indicated by a lower blue dashed line. Each plot includes details of the call, such as location, number of removed exons, and z-score at the top. C) Applying PCR and agarose gel electrophoresis to analyze the homozygosity of deletions in the patients, alongside a healthy control, I) In patient number 1 suspected of REL mutation) II) patient number 2 suspected of REL mutation. III) In patient suspected of CCBE1 mutation.

5.3.3 CNV Calling using GATK gCNV

Out of the 186 samples, seven had to be omitted from the analysis because of a high number of predicted CNVs, indicating their low sample quality or different amplicon efficiencies. Upon combining adjacent identified CNVs, a comprehensive total of 500 variants were ascertained, spanning across a dataset of 181 individual samples. CNV selection was based on the prioritization of their occurrence within our study population. Only Rare CNVs, which were found only once were deemed eligible for subsequent analysis. The extraction of exon depth preceding and following the predicted CNVs was performed using Integrative Genomics Viewer (IGV) and SAMtools. The read depth for each genomic region was obtained by analyzing all the samples using the SAMtools software tool. Then samples with low frequency were inspected. In the analysis process, certain samples were identified as having low coverage and were subsequently excluded from further analysis. Among the remaining 181 samples, 65 samples have at least one CNV with frequency of one, totally 106 CNVs. These findings encompass a total of 88 heterozygous deletions and 12 regions with increase in copy number, along with 6 homozygous deletions. Finally After checking varant with IGV and SAMtools, a total of seven regions were selected for real-time PCR analysis to validate the presence of heterozygous or homozygous deletions (*Figure 13*). Real-time PCR confirmed the mutation in the *CLBP* gene (*Figure 14 A*). Also PCR amplification for a patient suspected of having a homozygous deletion in POLE gene identified by GATK-gCNV tools, and reamplification with smaller regions of suspected genes identified by HMZdelfinder did not verfity the presence of predicted homozygous deletions (*Figure 14 B*).

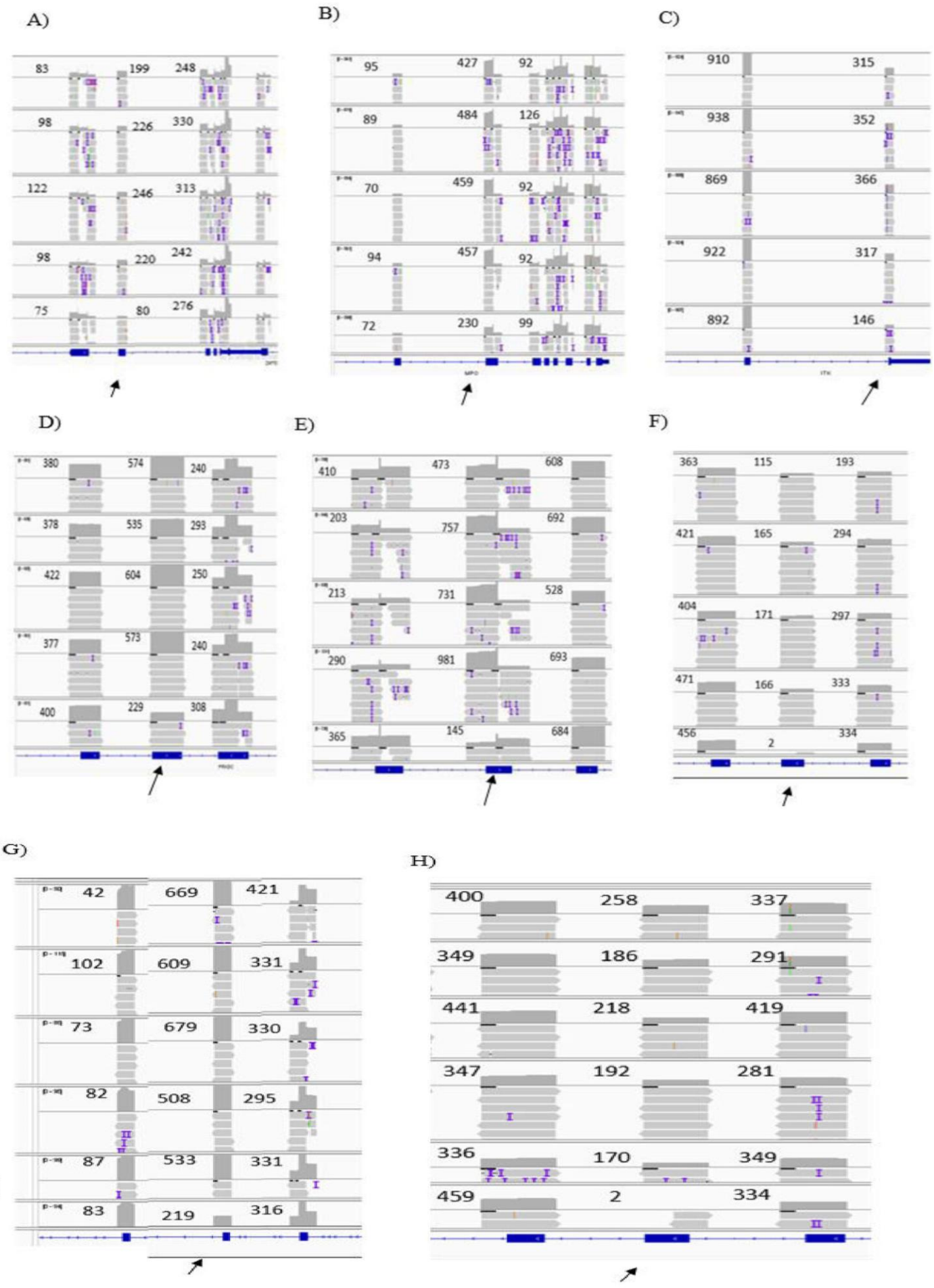

A)
83 199 248
98 226 330
122 246 313
98 220 242
75 80 276
B)
95 427 92
89 484 126
70 459 92
94 457 92
72 230 99
C)
910 315
938 352
869 366
922 317
892 146
D)
380 574 240
378 535 293
422 604 250
377 573 240
400 229 308
E)
410 473 608
203 757 692
213 731 528
290 981 693
365 145 684
F)
363 115 193
421 165 294
404 171 297
471 166 333
456 2 334
G)
42 669 421
102 609 331
73 679 330
82 508 295
87 533 331
83 219 316
H)
400 258 337
349 186 291
441 218 419
347 192 281
336 170 349
459 2 334

Figure 13: Integrative genomics viewer (IGV) visualization of the predicted CNV regions in samples predicted deletions identified using GATK-gCNV. The deletion regions were shown by arrows. The numbers showing depth of coverage in that area in the samples both upstream and downstream sequence realignment. A) patients with mutation in ZAP70, B) patients with mutation in MPO, C) patient with mutation in ITK, D) patient with mutation in PRKDC, E) patient with mutation in ATM, F) patient with mutation in POLE G) patient with mutation in CLPB, H) patient with mutation in POLE.

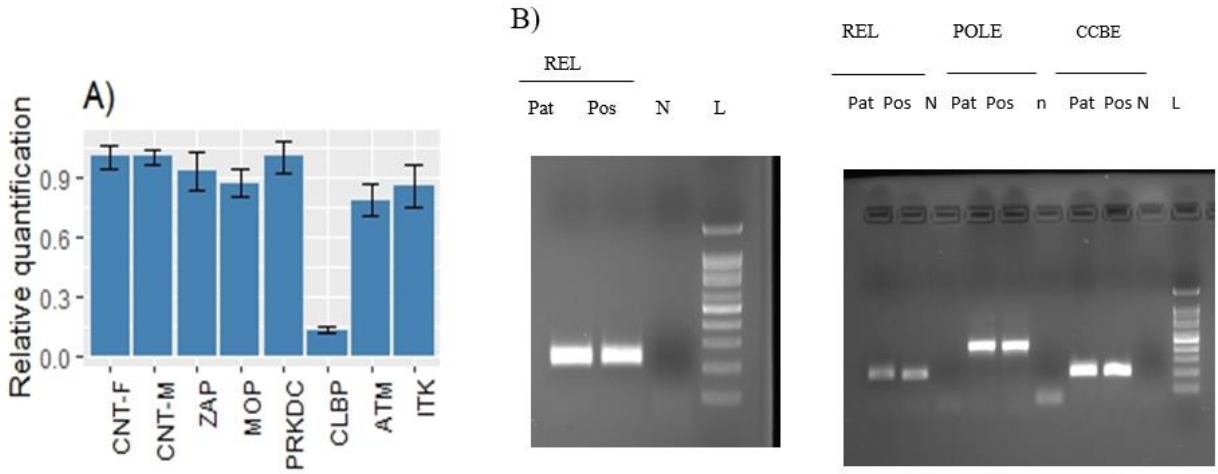


Figure 14: Validation of CNVs identified using GATK-gCNV tools and reconfirmation of previous PCR result of predicted genes identified by HMZdelfinder. A) Validation of heterozygous deletion using SYBR-Green Real-time PCR normalized against the DOCK11 gene. The control group consisted of individuals of the same gender as the subjects being investigated, with CNT-F and CNT-M representing female and male controls, respectively. The legend below the image represents the identified genes in the algorithm. B) PCR amplification for a patient suspected of having a homozygous deletion in POLE gene identified by GATK-gCNV tools, and reamplification with smaller regions of suspected genes identified by HMZdelfinder. The legend below the image represents the amplified genes. Pat: patient, pos: control positive, N: control negative.

5.3.4 MAK-Alu insertion

The investigation focused on the identification of Alu transposable element insertion within the MAK gene, a prevalent factor associated with inherited retinal diseases (IRD) in individuals of Ashkenazi Jewish heritage. *MAK*-Alu insertion were not detected in any cases.

5. 4 Molecular characterization and surveillance of nine Iranian families suspected of hereditary cancer

5.4.1. Molecular characterization and surveillance of seven Iranian families with targeted panel

In our study on hereditary cancers, we conducted an analysis on seven unrelated Iranian families and identified the presence of seven distinct pathogenic variants, which are outlined in *Table 8*. Remarkably, three of these variants were discovered to be novel, shedding light on previously unknown genetic mutations. Among the affected families, six were diagnosed with LS. In contrast, the remaining one family exhibited a defect in the *PMS1* gene. The chromatogram of SNV variations were shown in *Figure 15*.

In Family A, the proband with ovarian cancer was identified as a carrier of a frameshift pathogenic variant, specifically c.2294_2295 del in the *PMS1* gene (*Figure 15 A*). CRC was developed in the maternal aunt of the proband (*Figure 16 A*), who was found not to be a carrier of the variant. Intact expression of MMR proteins was observed through IHC analysis of her CRC tumor. In family B, a heterozygous exon 3 deletion in the *MSH2* gene was identified in the proband, a 29-year-old female CRC patient. Four cases of LS-associated cancers had occurred in her maternal family (*Figure 16 B*). The proband of family C harbored a novel frameshift pathogenic variant, c.705dupA in the *MSH2* gene (*Figure 15 B*), leading to the premature truncation of the *MSH2* protein. He developed gastric cancer and fulfilled the criteria for both LS and hereditary diffuse gastric cancer (HDGC) (*Figure 16 C*). In this family, a total of seven cases were excluded from the family analysis due to their refusal to undergo genetic testing, either personally or through their parents. The analysis of the familial mutation incorporated the other 30 relatives considered to be at risk. It was found that cascade testing was able to detect the pathogenic variant in 12 out of the 30 alive individuals who were tested. The genetic variation was identified in proband's mother, and LS-associated cancers were diagnosed in eight other affected relatives. It was observed that the maternal uncle and maternal cousin of the proband were non-carriers and had not experienced any history of cancers or polyps at 72 and 50 years old respectively. The co-segregation of the variant with cancer occurrence was confirmed by these results, as evidenced by the absence of cancers in carriers of the wild-type variant, and the presence of the pathogenic variant in members affected by cancer (*Figure 16 C*). The loss of MSH2/MSH6 protein expression was observed in the CRC sample of the second-cousin of the proband, who was diagnosed at 23 years old, as determined by the IHC analysis that was performed within this family.

Families	Gene (NM_)	Clinical criteria	Nucleotide change	MSI status using CAT25	type of IHC analysis	Location of tumor for IHC analysis	IHC analysis results of MMR proteins	MLPA	Effect in protein must be near the nucleotide change	Novel or reported
Fam. A	*PMS1*(NM_000534.5)	Revised Bethesda Guidelines	c.2295_2296del	N.A.	N.A	OC	N.A	N.A	p.His765GlufsX19	Novel[146]
Fam. B	*MSH2 (NM_000251.2)*	Revised Bethesda Guidelines	c.(366+1_367-1)_(645+1_646-1)del(EX3del)	N.A.	All MMR genes	CRC	Loss of *MSH2/MSH6*	Yes	N.A.	Reported©[146]
Fam. C	*MSH2* (NM_000251.2)	Amsterdam I	c.705dupA	MSI-Instable	All MMR genes	CRC	Loss of *MSH2/MSH6*	N.A	p.Asp236Argfs*20	Novel
Fam. D	*MSH2* (NM_000251.2)	Amsterdam criteria I	c.842C>G	N.A.	All MMR genes	OC	Loss of *MSH2*	N.A	p.Ser281Ter	Reported [147]
Fam. E	*MSH6* (NM_000179.3)	Family history of Breast cancer and CRC	c.3226C>T	N.A.	All MMR genes	CRC	N.A	N.A	p.Arg1076Cys	Reported [148,149]
Fam. F	*MSH6 (NM_000179.3)*	Breast cancer < 40 yrs	c.(3801+1_3802-1)_(4001+1_4002-1)del (EX9del)	N.A.	All MMR genes	Breast	A weak staining in *MSH2/MSH6* in both tumor and normal	Yes	N.A.	Novel
Fam. G	*PMS2 NM_000535.6)*	Amsterdam I	c.943C>T	MSI-Instable	All MMR genes	CRC	Loss of *PMS2*	N.A	(p.Arg315Ter)*	Reported[150,151]

Table 8: A comprehensive summary of the outcomes obtained from genetic testing, immunohistochemistry (IHC), and microsatellite instability (MSI) analysis performed in seven distinct families. The "NA" abbreviation denotes instances where the result was not applicable. In situations where the precise location of a deletion was unidentifiable, we relied on the MLPA outcomes and marked the deletion with an ©.

In family D, the proband had been affected by ovarian cancer (OC) and was found to be a carrier of c.842C>G in the *MSH2* gene (*Figure 15 C*). Confirmation of the loss of protein expression in her tumor tissue was achieved through IHC. Upon conducting segregation analysis, it was revealed that both parents carried the mutation and the mother had been diagnosed with CRC at 59. The proband's

two sisters were also examined, with one being identified as a carrier of the genetic anomaly and developing CRC at 37 years old, while the other was found to be wild-type and had no history of cancer (*Figure 16 D*). In family E, the presence of c.3226C>T in the *MSH6* gene was confirmed in the proband's father by Sanger sequencing, followed by IHC in his tumor tissue (*Figure 15 D*). Howevr, the maternal side needs additional investigations (*Figure 16 E*). Exon 9 deletion in the *MSH6* gene was demonstrated by CNV analysis in Family F (*Figure 16 F*). The affected patient's breast cancer was further confirmed by MLPA analysis. Weakly positive staining was observed in both MSH2/MSH6 proteins, as evidenced by IHC analysis on breast tumor tissue (*Figure 17*). Genetic testing was declined by the patient's relatives. The family G carries a *PMS2* variant, specifically the c943C>T mutation (*Figure. 15 E*), which has been identified in four individuals affected by cancer . One of these cases involves a person affected by both ovarian cancer and breast cancer as illustrated *Figure 16 G*.

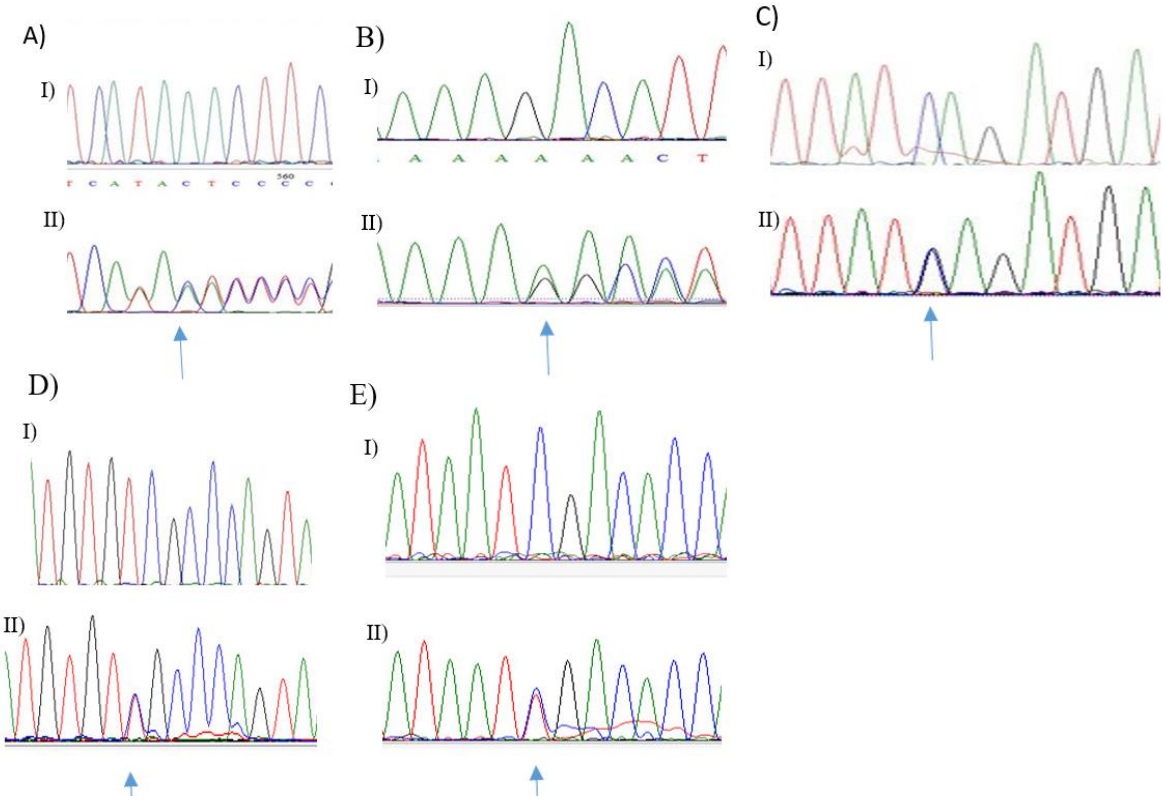
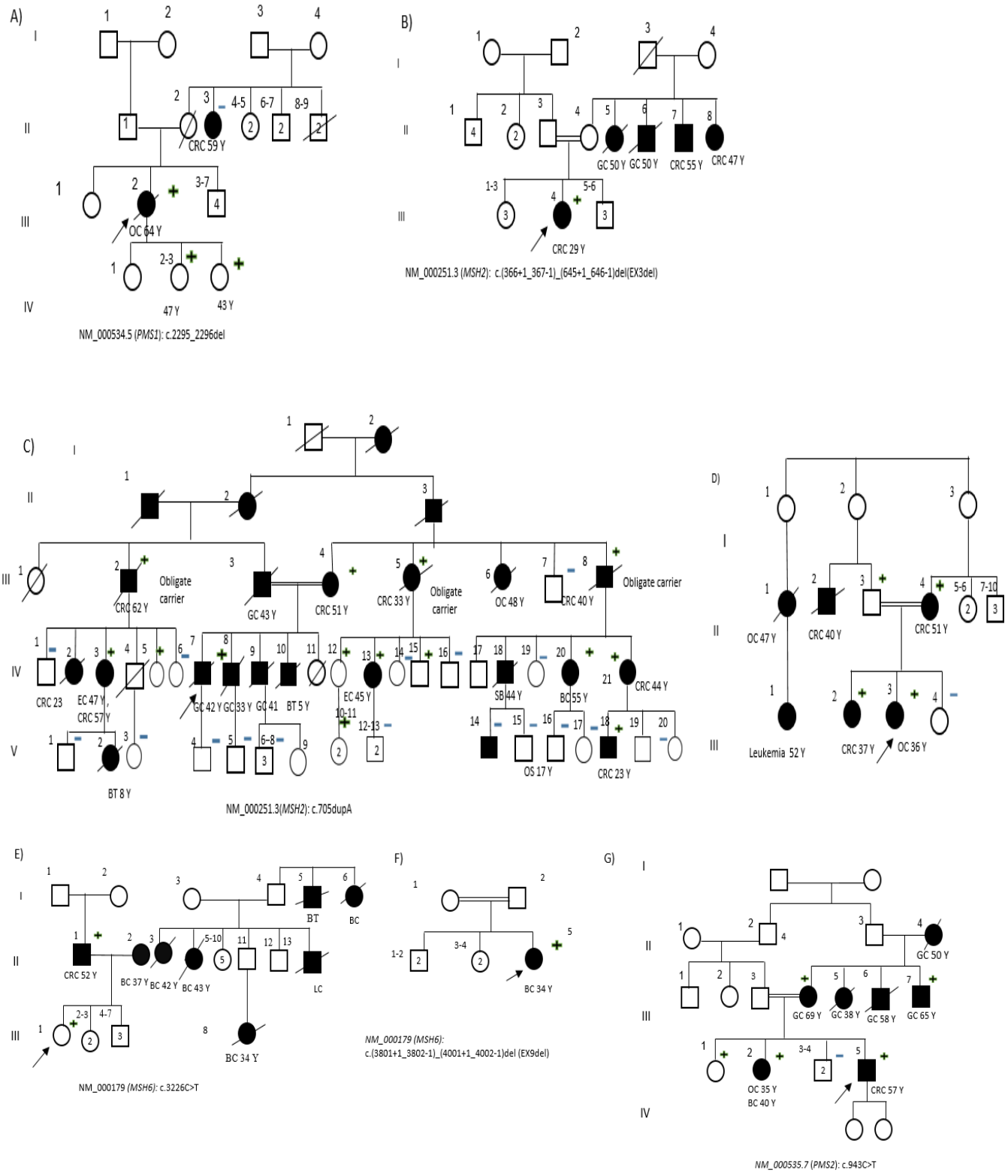


Figure 15. The chromatogram presents the DNA sequence of the pathogenic variants found in the probands. The chromatogram shows the wild-type, and the mutation carrier sequences for the following variants: A) PMS1 c.2295_2296del, B) MSH2 c.705dupA, C) MSH2 c.842C>G, D) MSH6 c.3226C>T, and E) PMS2 c.943C>T.



The *Figure 16. Pedigrees of seven families are shown: A) family A, B) family B, C) family C, D) family D, E) family E, F) family F, G) family G. The figures illustrate individuals with cancer (represented in black) and those without cancer (represented in white). Arrows indicate probands, and the age at diagnosis is denoted by Y. The cancers observed in the families include breast cancer (BC), brain tumors (BT), colorectal cancer (CRC), gastric cancer (GC), ovarian cancer (OC), osteosarcoma (OS), and small bowel cancer (SB). The individuals who underwent testing are identified by a plus sign if they carry the mutation and a minus sign if they have the wild-type. The number in each shape indicates the number of individuals in that generation.*

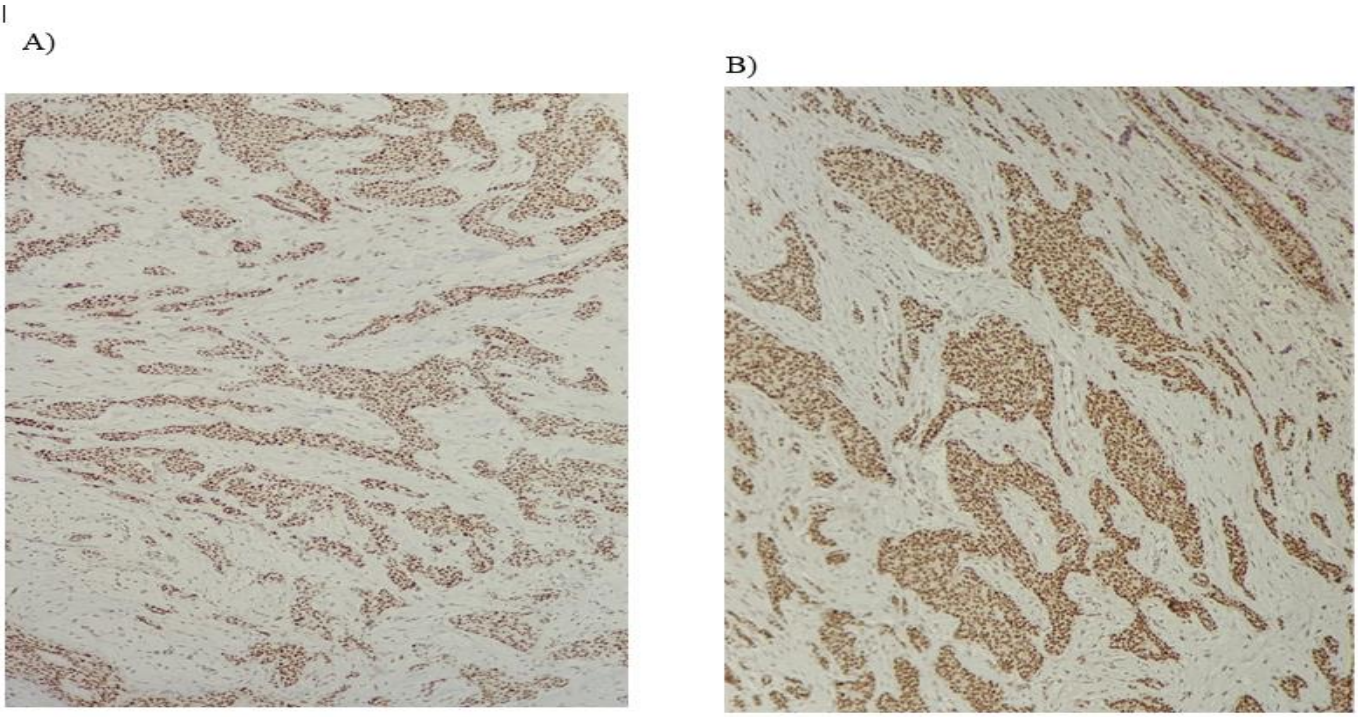


Fig.17 the IHC analysis conducted on FFPE tissue on the breast cancer tissue of the patient with EX9del showed a weak positive staining of MSH2 protein in both normal and tumor tissues (A) MSH2 staining, as well as a weak positive staining of MSH6 protein in both normal and tumor tissues (B).

5.4.2 Tumours spectrum in LS-*MSH2* gene

A total of 21 mutation carriers of *MSH2* pathogenic variants were identified, including 18 alive mutation carriers and three deceased obligate carriers. Among these carriers, 14 cases were affected by cancer, with one case presenting with both endometrial cancer and CRC (EC and CRC). In addition, pathogenic variants in *MSH6* were confirmed in CRC and breast cancer patients. This information is presented in *Table 9*.

Table 9. Tumor spectrum and average ages at diagnosis in individuals with confirmed pathogenic variants in the MSH2 gene, or obligate carriers.

Tumors types	Number of affected cases	Average age at diagnosis
Colorectal cancer	10	42
Gastric cancer	1	42
Endometrial cancer	2	46
Ovarian Cancer	1	36
Breast cancer	1	55

5.4.3 Surveillance programs

In the surveillance programs for LS before genetic testing, only 14 members of family C, including six cancer-affected individuals and four non-mutation carriers, had participated among all seven families. within this extended family, it was observed that six individuals, aged between 18-25 years, had not participated in any surveillance program. Through genetic analysis, the family members' carrier status was ascertained, resulting in a significant optimization of LS surveillance in family C. It was found that 28% (4/14) of the patients who were undergoing LS surveillance based solely on medical suspicion were determined to be non-mutation carriers. Furthermore, LS surveillance was not deemed necessary for six non-mutation carriers between the ages of 18 and 25 years who were planning to enroll in a surveillance program after reaching 25 years of age.

5.4.2 Molecular characterization and surveillance of two Iranian families identified by WES and CNV

Family H displayed a notable pattern of CRC incidences across successive generations, as depicted in *Figure 18H*. Additionally, the family met the diagnostic criteria known as Amsterdam criteria I, which are used to identify cases of hereditary non-polyposis colorectal cancer (HNPCC) [152]. In the medical report of proband family H, the IHC analysis of four MMR genes exhibited the lack of *MSH2* and MSH6 proteins in the FFPE CRC tumor. To further explore the matter, the *MSH2* gene was analyzed using Sanger sequencing. However, no pathogenic variant was identified in the *MSH2* gene. In order to identify potential pathogenic variants, both WES and CNV analysis were conducted. The WES analysis did not identify any pathogenic SNVs within the MMR or *EPCAM* genes. However, the CNV analysis identified a significant heterozygous deletion of approximately 76.7 kb (NC_000002.12:g.(?-47368993_47445665 -?)del in GRCh38). This deletion encompassed exons 1-9 of the *EPCAM* gene, followed by exons 1-8 of the *MSH2* gene. To verify the size of the CNV found in the proband, a real-time PCR analysis was conducted on specific exons. The target exons included *EPCAM* exon 1, which corresponds to the beginning of the identified variant, and *MSH2* exon 8, which represents the end of the variant. These details are illustrated in *Figure 19A*. Furthermore, the validation of previous findings was carried out using MLPA Probemix P003-D1. This particular Probemix includes probes designed for *EPCAM* exon 9 and the entire *MSH2* gene. The MLPA analysis revealed heterozygous deletions in *EPCAM* exon 9 and exons 1-8 of the *MSH2* gene, as illustrated in *Figure 20A*. The carrier status of all 23 patients in family A was assessed through MLPA and SYBR green real-time PCR targeting exon 3 of the *MSH2* gene (*Figure 19C*). Fifteen individuals were found to have novel heterozygous CNV deletions in the *EPCAM* and *MSH2* genes. Out of these individuals, six had developed CRCs, while the other nine individuals were asymptomatic carriers (*Figure 18H*). The medical endoscopy report of a patient in family H, specifically patient number eight in the second

generation, indicated the presence of fundic gland polyps in the stomach. This diagnosis was further confirmed by the patient's pathology report. Notably, there was no prior record of fundic gland polyposis within the patient's family medical history. Despite conducting extensive genetic analyses, including whole-exome sequencing (WES), CNV analysis, multiplex ligation-dependent probe amplification (MLPA) of *MLH1/MSH2* genes, and real-time PCR, no pathogenic variants associated with LS or fundic gland phenotypes were detected in the patient. The specific details and results of these analyses were not provided.

In *Figure 18I*, it is observed that the family I also fulfilled Amsterdam criteria I [152]. The proband of family I developed CRC at 25 years old, and the maternal aunt of the proband in family H,. Subsequently, the individual underwent surgery and long-term LS surveillance for over 30 years. Based on the higher occurrence of CRCs within Family I (*Figure 18I*) and their shared residence in the same village as Family H, there is a suspicion that Family I might have the same genetic variant as found in Family H. Real-time PCR analysis of the probands from family I was performed on exon 1 of the *EPCAM* gene and exons 3 and 8 of the *MSH2* gene. The results revealed a heterozygous deletion in these regions in probands and those unaffted people with a wild-type allele (*Figure 19*). Furthermore, MLPA analysis conducted on proband I demonstrated a heterozygous deletion in exon 9 of the *EPCAM* gene and exons 1-8 of the *MSH2* gene (*Figure 20*). After obtaining conclusive evidence, it has been determined that the genetic deletion identified in Family H is also present in Family I. The results are shown in *Table 10*.

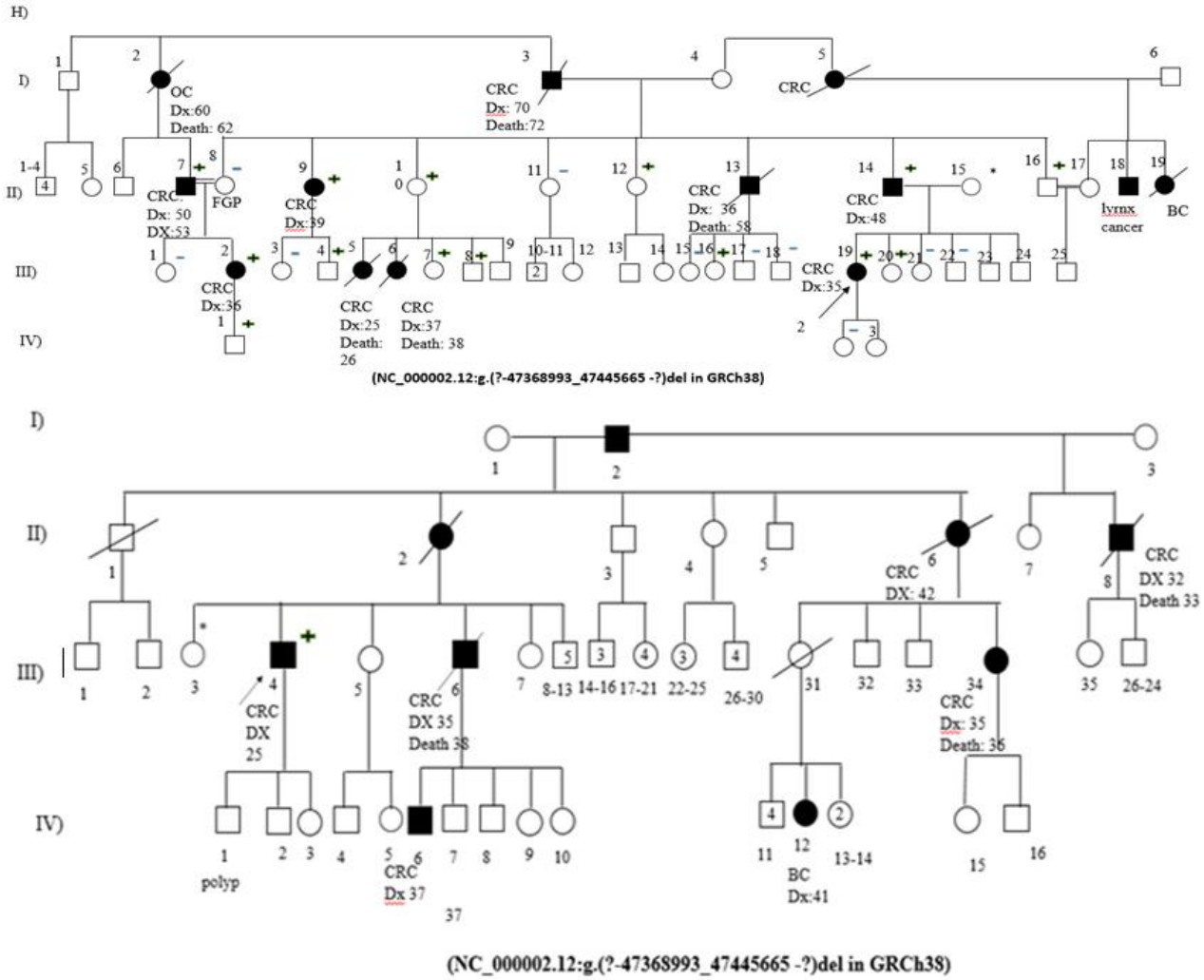


Figure 18. The pedigree of families H and I diagnosed with LS, depicting their cancer family history and the carrier status of family members. Families H and I exhibit a heterozygous deletion (NC_000002.12:g.(?-47368993_47445665-?)del in GRCh38) spanning exons 1-9 of the EPCAM gene and exons 1-8 of the MSH2 gene. Each individual is assigned a generation using Roman numerals (I to IV) and a unique number within each generation for identification purposes. In family H, all individuals in generation IV were under the age of 40. Individuals affected by cancer are represented by black shaded shape symbols, and their specific cancer diagnoses (BC: breast cancer, CRC: colorectal cancer, OC: ovarian cancer, FGP: fundic gland polyps) are provided next to or below the symbol. "+" indicates a carrier of the mutation, while "-" indicates a wild-type carrier based on genetic testing. A diagonal line through a shape signifies deceased individuals, and probands are denoted by arrows. Additionally, the asterisk (*) indicates the same person appearing in both Family H and Family I.

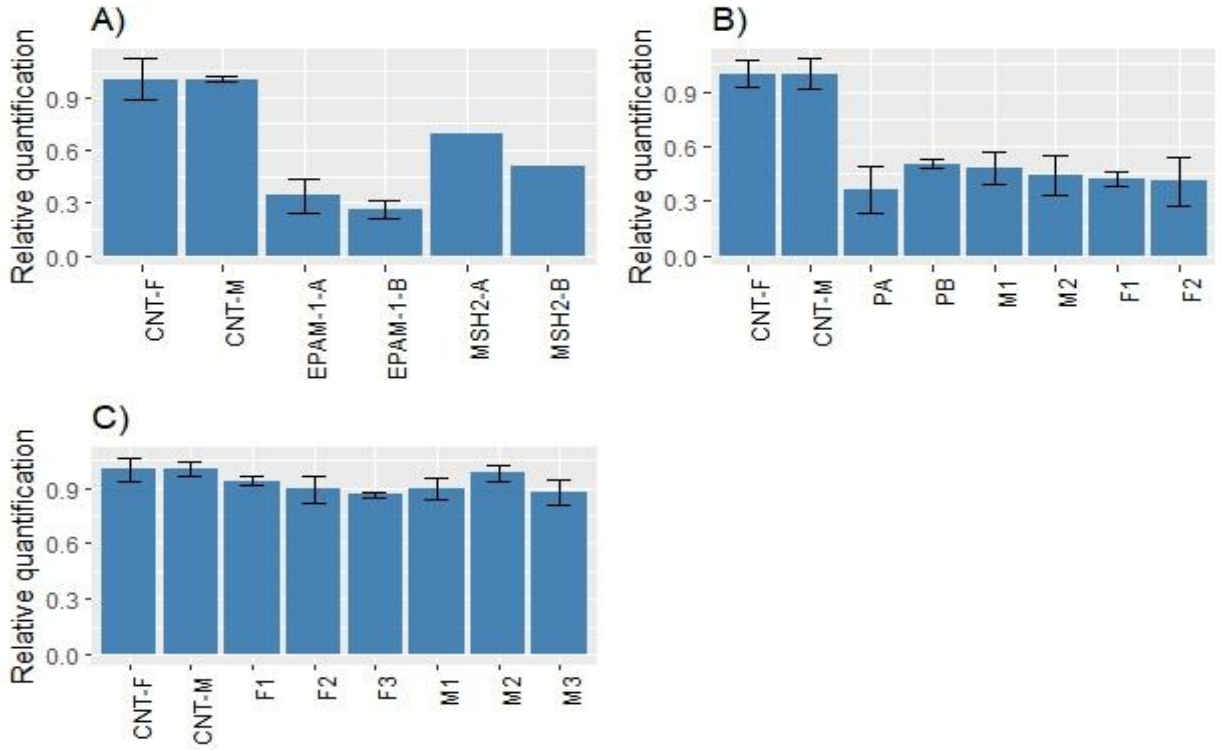


Figure 19: Histograms presenting the validation of identified CNVs using SYBR-Green Real-time PCR, normalized against the DOCK11 gene. The following abbreviations are used: CNT - normal positive control from subjects of the same sex as the tested individuals; CNT-F - female normal positive control; CNT-M - male normal positive control. A) The histograms show the validation of CNVs at the gross boundaries of the identified regions in families A and B, specifically for the EPCAM-1 exon 1 (EP) and MSH2 exon 8 (MS) genes B) The carrier status of the CNV was examined in the probands of families A (PA) and B (PB), as well as in two additional affected males (M1, M2) and females (F1, F2) from family A in exon 3 of MSH2 gene. The analysis focused on the MSH2 gene in exon 3. C) Three males and females from family A exhibited a wild-type allele in exon 3 of the MSH2 gene.

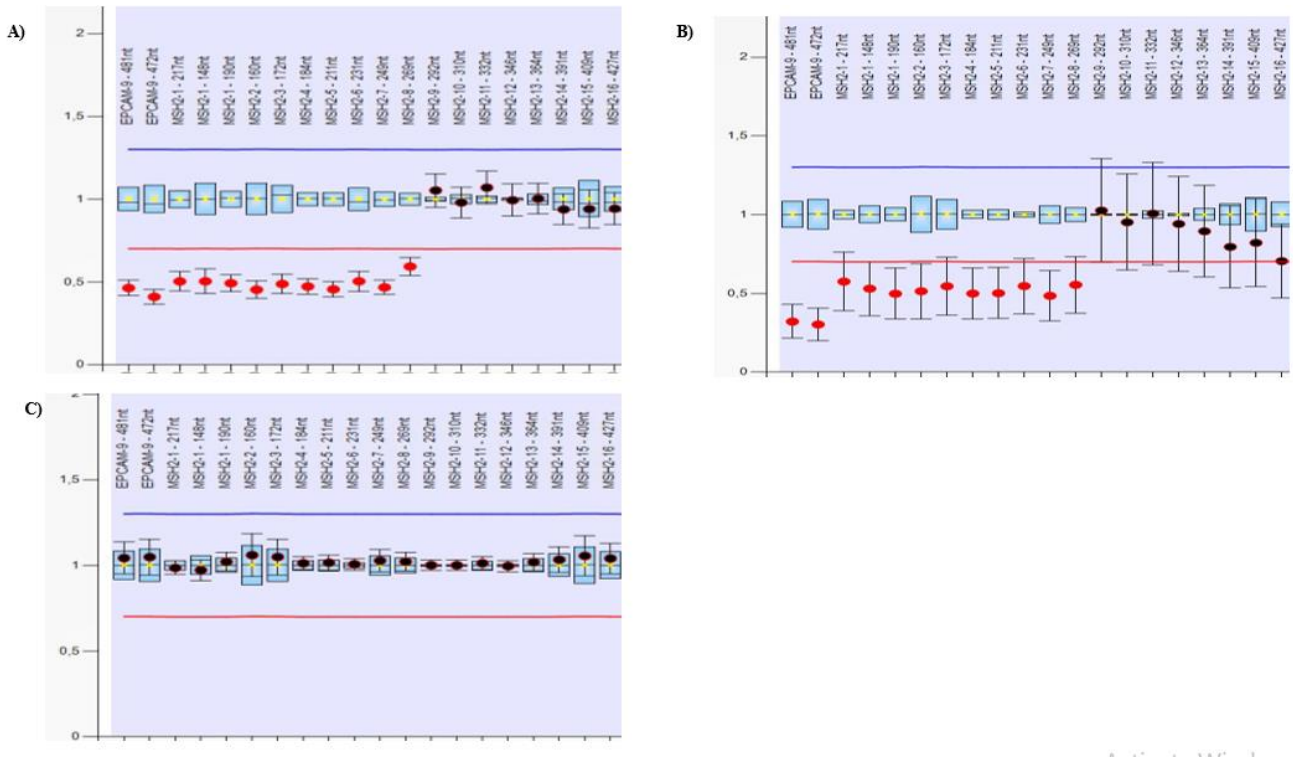


Figure 20. The ratio chart of MLPA analysis utilizing SALSA MLPA Probemix P003-D1 MLH1/MSH2. Deletions are identified reducing by 0.5 in peak height relative to a normal control, indicating heterozygous deletion. Panels A) and B) exhibit heterozygous deletions observed in families A and B, specifically in exon 9 of the EPCAM gene and exons 1-8 of the MSH2 gene. Panel C) showcases the normal control sample.

families	Gene (NM_)	Clinical criteria	Nucleotide change	type of IHC analysis	Location of tumor for IHC analysis	IHC analysis results of MMR proteins	MLPA	Effect in protein must be near the nucleotide change	Novel or reported
Fam. H		Amsterdam I	(NC_000002.12:g.(?-47368993_47445665 -?)del in GRCh38),	All MMR genes	CRC	Loss of MSH2/ MSH6	yes	absence or reduced expression of MSH2 and MSH6	Novel©
Fam. I		Amsterdam I	(NC_000002.12:g.(?-47368993_47445665 -?)del in GRCh38),	N.A.	N.A.	N.A.	yes		Novel ©

Table 10: A comprehensive summary of the outcomes obtained from genetic testing, immunohistochemistry (IHC). The "NA" abbreviation denotes instances where the result was not applicable. In situations where the precise location of a deletion was unidentifiable, we relied on the MLPA outcomes and marked the deletion with an ©

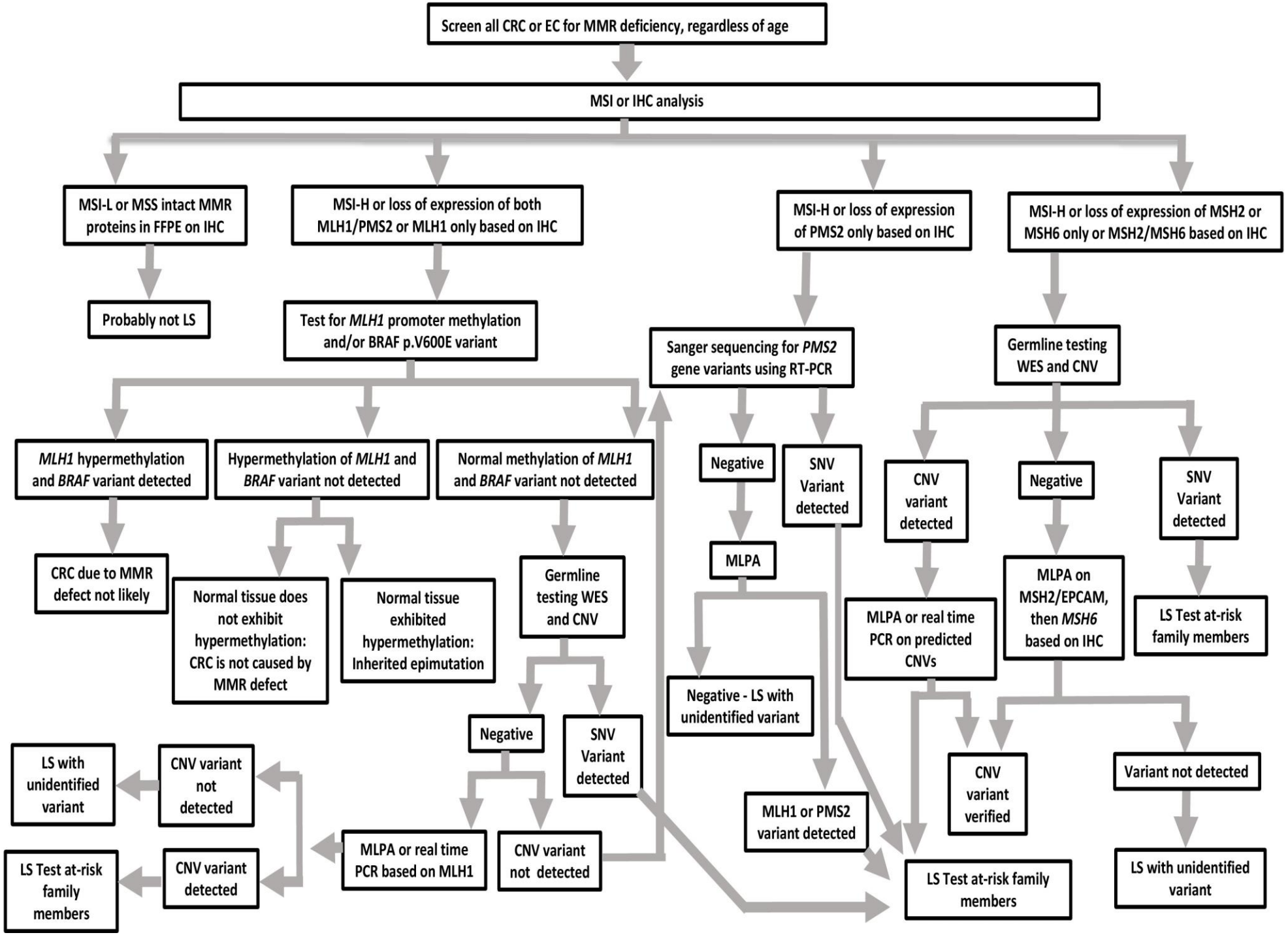
5.3.2 Tumours spectrum in LS- *EPCAM/MSH2* gene

In families H and I, all six individuals affected by cancer and carrying a confirmed mutation in the EPCAM/MSH2 gene were CRC affected before the age of 55. In families H and I, a total of 24 individuals underwent genetic testing. Among these individuals, it was determined that 19 of them were aged 25 years or older. The age upon diagnosis was between 35 and 53 years. Among the group of 19 individuals, 15 individuals had taken part in the LS surveillance program, while two individuals did not carry the mutation associated with the syndrome. Furthermore, two individuals who carried the mutation were not enrolled in any surveillance program.

5.3.2 Mappability analysis

Mappability analysis of MMR genes showed that a total of 30% of *MLH1*, 10% of *MSH2*, 4% of *MSH6*, and 100% of *PMS2* gene amplified regions are not unique exons across these genes. Regarding mappability and CNV analysis, we have developed a novel strategy for LS testing, outlined in *Figure 21*.

Figure 21. Flowchart of a suggested genetic testing strategy for Lynch syndrome. CNV: Copy NumberVariation, CRC: Colorectal Cancer, IHC: Immunohistochemistry on FFPE Tumor, MMR: Mismatch Repair Gene, MSI-H: Microsatellite Instability High, MSI-L: Microsatellite Instability Low, MSS: Microsatellite Stable


Screen all CRC or EC for MMR deficiency, regardless of age
MSI or IHC analysis
MSI-L or MSS intact MMR proteins in FFPE on IHC
MSI-H or loss of expression of both MLH1/PMS2 or MLH1 only based on IHC
MSI-H or loss of expression of PMS2 only based on IHC
MSI-H or loss of expression of MSH2 or MSH6 only or MSH2/MSH6 based on IHC
Probably not LS
Test for MLH1 promoter methylation and/or BRAF p.V600E variant
Sanger sequencing for PMS2 gene variants using RT-PCR
Germline testing WES and CNV
MLH1 hypermethylation and BRAF variant detected
Hypermethylation of MLH1 and BRAF variant not detected
Normal methylation of MLH1 and BRAF variant not detected
Negative
SNV Variant detected
CNV variant detected
Negative
SNV Variant detected
CRC due to MMR defect not likely
Normal tissue does not exhibit hypermethylation: CRC is not caused by MMR defect
Normal tissue exhibited hypermethylation: Inherited epimutation
Germline testing WES and CNV
MLPA
MLPA or real time PCR on predicted CNVs
MLPA on MSH2/EPCAM, then MSH6 based on IHC
LS Test at-risk family members
Negative - LS with unidentified variant
Negative
SNV Variant detected
LS with unidentified variant
CNV variant not detected
MLPA or real time PCR based on MLH1
CNV variant not detected
MLH1 or PMS2 variant detected
CNV variant verified
Variant not detected
LS Test at-risk family members
CNV variant detected
LS Test at-risk family members
LS Test at-risk family members
LS with unidentified variant

6. Discussion

As a whole, this PhD thesis deals with the problems arising in a diagnostic laboratory facing difficulties in the correct interpretation of variants found during the different diagnostic processes and the lack of results in patients with a precise clinical diagnosis.
The different chapters deal with specific crucial aspects of interpretation and technical limits of the different approaches used in genetic diagnosis.

6.1. Pathogenic variant in an X-linked gene in a female patient diagnosed with *CGD.*

From a biological standpoint, females typically enjoy a significant advantage, as they tend to display either no illness or encounter less severe symptoms when compared to males harbouring the same genetic mutation on X chromosome. [153] Males are the primary individuals impacted by X-linked pathogenic variations. [154] The only exception is when a female is homozygous for the harmful allele, or if the mutation exert a dominant effect. [153] The variations in the impact of X-linked pathological variations on different sexes arise from the mechanism of for equalizing X chromosome gene expression between genders, known as X-inactivation. [155,156] Most placental mammals and Humans employ a mechanism to counterbalance the inequality in the count of X chromosomes between the sexes (XY in males and XX in females) by selectively activating only one of the paired X chromosomes in females. This biological mechanism, known as X-inactivation, effectively silences one X chromosomes, ensuring that both genders and possess a only one functional X chromosome.[157,158] The phenomenon occurs in the initial stages of embryonic development in female organisms[159], and contributes to the distinct outcomes experienced by males and females in X-linked genetic conditions. Most females do not display X-linked disorders due to two main reasons. Firstly, they are typically not homozygous for the harmful genetic mutation. Secondly, their variant cells, which contain the detrimental allele, receive an adequate amount of gene product from the cells that transcribe the normal allele, allowing them to perform the necessary metabolic functions.[153] X-linked CGD primarily is caused by pathogenic variations in the *CYBB* gene[160]. Advances in diagnostic techniques have significantly expanded our insight of the genetic underpinnings of CGD in the past decades. Consequently, our comprehension of the clinicopathological characteristics associated with this condition has greatly improved. [161] The NGS analysis on a seven-year-old girl with CGD, using a targeted panel for all the genes involved in X-linked and autosomal recessive CGD, showed a heterozygous possibly pathogenic variant NM_000397:ex9:c.1151+2T>C in the *CYBB* gene. Her parents were demonstrated wild type for the gene; thus, the variant seems to be a *de novo* one. The genetic anomaly was a potential splice donor site variant, predicted to possibly impact on the correct splicing of the gene.

The variant was was designated as probably pathogenic (class 4) in line with the AMCG classification guidelines as the mutation resides within the +/−1 of 2 splice sites (PVS1) and has no frequency in the general population.
A functional analysis using RT-PCR and sequencing around the exon's boundaries, revealed exon 9 total skipping, confirming the variant pathogenicity.

The patient anyhow carried also a wild-type *CYBB* allele and exhibited symptoms indicative of reduced expression of the wild-type variant. The presentation of X-linked recessive disorders, like X-linked-CGD, in women might be influenced by skewed X-inactivation. [162] Skewed X-inactivation is frequently attributed to negative selection acting upon one of the alleles. [163] This arising from a non-random X-chromosome inactivation (XCI) led to a diminished wild type allele's normal expression. [164] X-chromosome inactivation serves as a remarkable instance of an epigenetic silencing that extends throughout the whole human X-chromosome, encompassing approximately 160 Mbp. [165] The non-active allele of the X-chromosome undergoes significant methylation, displaying a high concentration of inactive histone modifications while lacking active ones. In particular, in the study of primary immunodeficiencies, the evaluation of the X-chromosome inactivation status has been exploited for years as a tool to discriminate carrier females. The X chromosome inactivation technique relies on the distinctive methylation patterns found in a gene on the X chromosome known as HUMARA (Human Androgen Receptor). [166] When XCI occurs within a cell, it consequently leads to the inactivation of the identical X chromosome in total subsequent cells derived from the initial cell. [167] The HUMARA gene possesses three key attributes that render it exceptionally advantageous for the intended objective. I) In normal embryogenesis of a female infant, the gene is placed on the X chromosome and undergoes inactivation through methylation. While most genes on the X chromosome experience this process, there are exceptions II) The Human Androgen Receptor gene exhibit different numbers of CAG repeats alleles. [168]PCR amplification of a specific region of the gene from a normal female generates two distinct DNA bands on the gel, clearly separated from each other. [169] III) The foundation of this analysis relies on the presence of methylation-sensitive restriction sites on HUMARA locus in PCR products, rendering them susceptible to digestion by the HpaII (or HhaI) enzyme when unmethylated. [170] These characteristics enable the differentiation between a methylated allele and an unmethylated one. In this study, the HUMARA assay was conducted to explore the extent of imbalanced XCI. The HUMARA assay in our patient and her parents detected a 195 bp fragment in the proband's father, while the mother exhibited two fragments, 195 bp and 198 bp. The amplification showed that the patient possessed 195 bp and 198 bp fragments. The treatment of probands' amplicon by HpaII enzyme showed a great proportion of the X chromosome with a length of 195 bp subjected to digestion. This led to a shorter peak of this amplicon during fragment analysis in comparison with the untreated

amplicon, indicating that the fragment with 195 bp on the X chromosome carries the mutated *CYBB* gene, so it was unmethylated and transcriptionally active. The calculation revealed that the allele with a length of 195 bp has an inactivation of 16%, while it was 84% for the other allele with a length of 198 bp. Given that the patient inherited a locus of 198 bp from the mother and 195 bp from the father, and the more digestion of the 195 bp locus, thus, it is evident that the X chromosome, which underwent transcriptional inactivation, is located on the paternal side.

Conversely, it can be inferred that the transcriptionally silenced X chromosome is located on the maternal side, even though the mother carries both wild-type alleles in the *CYBB* gene. The HUMARA assay identified a non-uniform distribution of X-inactivation within the DNA samples extracted from the patient's polymorphonuclear leukocytes (PMN), highlighting skewed inactivation. The confirmation of a genuine reduction in wild-type *CYBB* expression within the patient's PBMC cells was established through the RT-PCR, with a dominant expression of the altered variant while demonstrating a decreased expression of the wild-type variant. The overall findings from the RT-PCR and HUMURA assays support the conclusion that the pathogenic allele is the dominant transcript, which is active and unmethylated. Therefore, the symptom of the disease in the patient could be attributed to the predominant expression of the mutated *CYBB* gene, which is influenced by skewed XCI due to unknown reasons.

Some variants of Taq polymerases add additional nucleotides, predominantly adenine, to the 3' end of fragments. This phenomenon, commonly referred to as "plus-A-artifact" results in the production of two sets of amplified fragments that differ in length by nucleotides. [117] Variation in fragment length can mislead when attempting to interpret the resulting fragments. [117] Based on our findings, replacing Taq polymerase is the most effective approach for eliminating ""plus-A-artefact". In this study, it was found that the GoTaq ® DNA Polymerase (Promega, USA) demonstrated better performance in excluding plus-A than the AmpliTaq Gold™ DNA Polymerase. Other optimization factors, such as modifying PCR protocols or adjusting the concentrations of Mg^{+2} and/or DNA template, cannot eliminate the " plus-A-artefact " fragments. The most efficient approach for reducing stutter fragments is to decrease the number of PCR cycles. To sum up, we have outlined a new *CYBB* pathogenic variation that led to an exon 9 skipping, resulting in a shortened *CYBB* protein. The integration of cDNA sequencing for splicing mutation analysis, coupled with the examination of X-chromosome inactivation, holds the potential to significantly enhance the characterization of female individuals harbouring splicing variants on the X chromosome, thereby advancing our understanding in the field of genomic research.

6.2. A new pathogenic splicing variant in the MAGT1 gene.

X-linked immunodeficiency with magnesium defect, Epstein-Barr virus (EBV) infection, and neoplasia (XMEN) disease is a rare monogenic genetic disorder. The MAGT1 gene is associated with magnesium transport, and the inheritance pattern for this trait is X-linked recessive. [46] The gene consists of 10 exons, with the primary variant encoding a protein comprising 335 amino acids. [171] *MAGT1* is present in all mammalian cells and is highly conserved throughout evolution. Its primary function is in maintaining magnesium homeostasis. [53] Diagnosing XMEN poses a significant challenge due to its variable clinical presentation, which can often overlap with other conditions like common variable immunodeficiency (CVID). [53] Consequently, accurately identifying XMEN remains a complex task. [172] Patients with XMEN syndrome are vulnerable to EBV infections, along with chronically diminished levels of intracellular magnesium (Mg^{2+}). [173] XMEN syndrome is classified by the inability to effectively eradicate the EBV, leading to the development of lymphoma. EBV is a prevalent human virus that typically remains in the body for many years without causing remarkable complications in individuals with a fully functioning immune system. [173] Overall, 36 unique male patients with the condition were reported, and our investigation identified a new hemizygous *MAGT1* putative splicing mutation in a male patient while using a targeted gene panel for the study of cases of CVID. [174] The mutation is on the single X-chromosome and thus we only need to demonstrate its pathogenicity. The novel hemizygous variant identified using NGS is located in the splicing region of the *MAGT1* gene (NM_032121:c.627+2T>C) in a 12-year-old male with clinical phenotype of CVID. Several software applications were developed to forecast splicing events that involve the formation or elimination of splice sites within exons or introns. [175] These tools include GeneSplicer [176], Human Splicing Finder [177], MaxEntScan[178], NetGene2[179], NNSplice [180], and FSPLICE[181]. These tools employ computational algorithms and techniques to enhance the accuracy and reliability of splicing site identification. By adhering to the ACMG guidelines, researchers and clinicians can utilize these tools effectively in predicting pathogenicity of a mutation in splicing sites. [182] However, these RNA splicing predictors generally exhibited high sensitivity (~90–100%) but lower specificity (~60–80%) regarding identifying abnormalities in splice sites. [133,183] Despite the wide range of software applications developed for the prediction of splicing sites, they all share a foundation in common biological concepts. Therefore, when assessing the influence of a DNA variant on pre-mRNA splicing, it is essential to consider the combined predictions from various computational tools as a single body of corroborating evidence. [133]

To determine the functional consequences of the variant, anyhow, a functional analysis using mRNA in RT-PCR and gel electrophoresis were performed. The patient's RT-PCR showed a shortened single band as compared to the normal control analyzed in parallel, while sequencing analysis of transcript

revealed two smaller aberrant fragment amplifications compared to the normal control. Characterization of the patient's transcript through sequencing provided insights into the splicing abnormalities. Two key findings were observed: exon 4 skipping and alternative splicing involving a heterozygous insertion of 63 base pairs from the beginning of exon 6 to the beginning of exon 5. The transcript supports the computational tools used in classifying the variant as pathogenic, as no normal transcript was detected. Exon 4 skipping, can cause a frameshift or loss of critical functional domains, affecting the protein's structure and function. Also, alternative splicing involving the insertion of 63 base pairs may disrupt the reading frame and introduce abnormal protein sequences. The skipping and addition of partial exons can be attributed to pre-mRNA splicing, a fundamental process in gene expression. mRNA splicing involves the eliminating of introns and connecting of adjacent exons, resulting in the alignment of all gene exons in a mature mRNA molecule. This alignment mirrors the sequential order of exons found in genomic DNA sequence[24] Specifically, the presence of GT and AG dinucleotides consistently at the 5'- and 3'-ends of introns undergoing splicing. The two ends of an intron are the 5' and 3' splicing sites or, more specifically, the splicing donor (5' SS) and acceptor (3' SS) sites, respectively. Any change in this area could affect splicing. Approximately 10% of the total mutations in the public database have been found to occur on canonical splicing sites, thereby impacting the procedure of pre-mRNA splicing and contributing to human genetic diseases. These variants, present in the pre-mRNA stage, affect the splicing process, leading to the addition of introns or the removal of exons from the final mRNA molecule[184]. As a result, this alteration affects both the resulting mRNA and its protein-coding sequence. [133]

In conclusion, the functional characterization of the hemizygous splicing variant in the *MAGT1* gene provides evidence of its impact on mRNA splicing. The identified variant results in exon skipping and alternative splicing events, disrupting the protein's structure and function. These findings can contribute to our understanding of the pathogenic mechanisms associated with *MAGT1* gene mutations and related disorders and to an appropriate counselling to the family.

6.3 Detection of rare CNVs on a panel of PID genes

The rapid advancement in NGS has rapidly increased the discovery of gene mutations responsible for diseases. [185] Moreover, the utilization of NGS in molecular genetic diagnostics has emerged as an effective approach for the identification of new mutations. [186] Most studies conducted on the application of targeted panels have solely focused on SNVs and indels, overlooking the assessment and utilization of CNVs in both clinical and research settings. [128,186] CNVs refer to structural variations of intermediate scale in the genome, encompassing changes in the quantity of copies of DNA fragments ranging from 1 kilobase (kb) to 5 megabases (Mb). [187] Due to the limitations of traditional

Sanger sequencing, which is unable to detect CNVs at the heterozygous state, the impact of CNVs on the genetic burden has been largely overlooked. [69] Detecting CNVs from NGS data poses a significant challenge despite their equal importance to SNVs. [187] The primary obstacles encountered in detecting CNVs involve experimental biases resulting from factors like repetitive DNA regions [188] and GC content. [189] Additional challenges include biases arising from technical inconsistency during library preparation, capture, and sequencing. [69] To address the issue of GC-content bias, several CNV detection tools (including ExomeDepth[134], ExomeCopy[190], CODEX[191], DECoN[192], CLAMMS[193], and CANOES[190]) incorporate GC-content correction methods. Conversely, some tools (such as CoNVaDING [194], and CONTRA[195]) mitigate this bias by utilizing sample ratios instead. [69] The clinical application of detecting CNVs from WES is problematic because of numerous previous studies indicating significant issues with high false-positive rates and limited sensitivity. [75,107,196-199]

The routine practice in clinical settings involves employing the same gene panel over several years on a large quantity of samples. [200] The CNV of a particular region in a sample is examined through algorithms that utilize the depth of coverage to make CNV predictions. [75] These algorithms assess the depth of the region by comparing it to the average depth observed in an extensive training set consisting of other samples. [69]

In this study, we evaluated the possibility of using the AmpliSeq™ sequencing PGM platform for the identification of CNVs in a diagnostic setting to improve the clinical sensitivity. The platform has a distinct approach compared to Hybridization-Based Target Enrichment for NGS systems. [201] Ion AmpliSeq™ reads are obtained as multiplex PCR products, and the sequencing depth data for each amplicon is included in the BAM file to conduct copy number detection analysis. [202] Various sections of the genome may exhibit varying levels of accessibility, resulting in differing amplification efficiencies and relative read depths. [123] It is probable that a specific region will maintain consistent accessibility across multiple runs of the sequencing.

Several algorithms are available to identify CNVs in NGS analysis most of them for Hybridization-Based Target Enrichment. [197] In previous studies that assessed the effectiveness of commonly employed CNV calling tools, a notable proportion of the generated calls were found to be false positives. [75] Furthermore, the absence of gold standard CNV tools posed limitations and hindered the comparability of the results. [75] In a clinical environment, achieving the highest sensitivity and minimizing false discovery rates are imperative when selecting a tool. [203] CNV analysis was accomplished on a cohort of Italian patients diagnosed with PID for whom the target panel used in the diagnostic lab was negative for pathogenic/probably pathogenic variants. [204] The NGS experiments were conducted over a period of two years, necessitating the use of multiple batches. [205] Our objective was to assess two CNV tools for detecting germline CNVs. This was defined as the ability to

accurately detect genuine CNVs (true positives) and reduce the risk of identifying false positives (false positive discoveries).

In this study, a group of 200 individuals were utilized to thoroughly assess and evaluate the effectiveness of the GATK-gCNV and HMZDelFinder pipeline in detecting CNVs from targeted sequencing data. It is crucial to acknowledge that all our samples were produced within the same sequencing facility, gene panel, and following an identical protocol, but is well known that ampliSeq produced data are not consistent in efficiency of amplification even among the same analytical session.

6.3.1 Computer mappability score and CNV Calling using HMZDelFinder on Targeted gene panels.

The HMZDelFinder is capable solely of identifying a reduction in the genomic copies. [206] The algorithm demonstrated exceptional efficacy in detecting rare HMZ deletions within targeted data, surpassing the performance of other widely utilized tools. [207] This particular group of CNVs has the potential to lead to null alleles, causing a total absence of gene functionality. [206] The identification of these individuals could potentially unveil previously unknown genes or variations associated with Mendelian diseases. [207] HMZDelFinder performs a comprehensive assessment of the normalized per-interval coverage for entire samples. [206] Although the frequency of variant replication across various control groups can suggest the reliability of the finding, it is essential to consider that technical artifacts may also contribute to reproducibility. The overall count of detected CNVs was significantly influenced by the process of DNA extraction, library preparation, and sequencing of samples in batches. [208] Variants with high frequency can be excluded due to either technical artifacts or the presence of benign CNVs in the general population. This approach enables the identification of rare exonic HMZ deletions while reducing the occurrence of false-positive detections caused by regions with low coverage. [207] Current approaches to identifying HMZ deletions involve analyzing the discrepancy in sequencing depth between a specific exome and the remaining exomes within the dataset. Nevertheless, the depth of coverage is greatly influenced by the sequencing conditions, which undergo constant changes in standard laboratory environments. HMZDelFinder was designed to identify HMZ deletions in a comprehensive dataset of over 500 homogeneous exome samples. [209] However, its effectiveness may be suboptimal when applied to diverse laboratory cohorts. These cohorts typically involve exome data obtained at different points in time and under varying conditions. [207] Consequently, the NGS data obtained in the long run inevitably exhibit heterogeneity, thereby adding complexity to the identification of deletions. [209] In this study, HMZDelFinder algorithm was executed to a cohort of Italian patients with PID to detect CNVs generated in Ion AmpliSeq NGS Panels for Targeted Sequencing. The tool has been meticulously designed for HMZ deletion, as it

effectively displays the identified regions, gene symbols, samples, merges the neighbouring exons, and shows them in a visually comprehensible manner. [206] This feature enables convenient data visualization and facilitates thorough examination. The utilization of IGV for visualizing identified regions can aid in distinguishing false positives homozygous regions. [116] The algorithm's performance was assessed by analyzing five targeted samples obtained from other panels, in which a significant number of exons were not included. It successfully detected samples with fewer genes panel compared to others in the cohort study. Therefore, the tool is suitable for identifying samples with extremely low quality from various batches. After removing these five samples, reanalysis of samples in each gender separately was conducted on the remaining samples, and a total of 54 deletions were identified. Collectively 35 CNVs were identified in 10 genes, including *VPS13B, PMS2, KMT2A, RNASEH2B, KMT2A, RNASEH2B, DOCK8, IL6ST, IL6ST*, and *PLCG2*.

Certain obstacles persist in CNV calling; one, such challenge involves accurately ascertaining the copy number status in regions of the genome that exhibit high variability, regardless of the employed technology. The identification of CNVs in genes that have closely related counterparts in different regions of the genome remains a challenging task, especially when utilizing short-read sequencing data. These areas often cause the occurrence of either false positive or false negative CNV identifications, [108] as we observe the 35 recurring CNVs across 10 specific regions. Employing a threshold based on the average mappability score at the exon level aids in mitigating the false positive instances. However, this approach also inadvertently excludes certain genes of significant clinical relevance due to the inherent challenges in aligning short reads within. [108] Testing these genes through short-read sequencing poses challenges due to their overlap with segmental duplications. Further research is essential to enhance the integration of reproducibility and other quality metrics, which would ultimately result in better assessment and ranking of potentially valid CNVs.

We calculated the mappability score in the identified called CNVs, and only the *PMS2* gene had a score below 0.75%. However, a significant decrease in the amplification plots of other mentioned regions suggests an association between the reduction of depth of coverage and the nature of amplified regions. These findings underscore the importance of our analysis of the plot generated by the tools and shed light on the limitations associated with short-read sequencing when examining CNVs. To address the false positive issue resulting from PCR-based NGS methods, additional investigations employing long-read sequencing may offer a potential solution. However, Long-read sequencing remains prohibitively expensive for regular diagnostic purposes. [143] Filtration was utilized to eliminate the repetitive deletions that could potentially be false positive CNVs. These methods were implemented at the variant and sample levels to ensure enhanced accuracy and reliability. The remaining HMZ deletions were examined at the 5' and 3' ends of the identified variants by IGV. "Samtools depth view" was utilized to extract the read depth for every position or region extracted

from IGV tools. This method was employed to select samples demonstrating consistent coverage depth in neighbouring exons of the identified variants. Consequently, only those samples that exhibited the same coverage in both the 3' and 5' regions of the identified variant were used as reference samples. Also, "samtools depth view" was used to extract the depth of coverage in deleted regions to be sure that only the suspected sampled had zero coverage at the locus. Also, a comprehensive manual review utilizing IGV excluded lower-quality CNVs, while retaining the possible true positive findings that warrant further investigation in the wet lab. Only five HMZ were approved after rigorous manual inspections and thorough filtration processes, which revealed a concerning false positive rate of 90%. The high rate can be attributed to the limitations of the tools, which fail to recognize the significant decrease in coverage depth in these areas. The five remaining regions consistently showed non-zero in all other samples, as confirmed by samtools and IGV. This observation confirmed that the tool is operating effectively, as none of the other samples in the cohort displayed these particular HMZ deletions. It is essential to perform a manual review to address the significant decline in amplification indicated by the documented plots. Confirmation of the five identified CNV calls was carried out using PCR to ensure that no amplification occurred in those regions. However, none of the identified variants were validated as predicted homozygous deletions were amplified. The disparity between in silico and wet lab results can be attributed to the inherent characteristics of the Ion AmpliSeq PGM.

The handling of duplicate reads differs significantly between Ion AmpliSeq and hybridization platforms, marking a crucial distinction. In contrast to hybridization platforms, flagging duplicate reads is not well-suited for Ion AmpliSeq data. This is primarily because Ion AmpliSeq data often involves multiple independent reads that may share identical 5' alignment positions and 3' adapter flows. As a result, the conventional method of identifying duplicate reads becomes problematic when applied to Ion AmpliSeq data. The procedure of identifying duplicates in an Ion AmpliSeq carries the potential of incorrectly flagging multiple reads as duplicates, even though they are distinct from one another. It is challenging to eliminate PCR artifacts from the NGS data, rendering their removal practically impossible. This suggests that employing a hybridization capture technique is advantageous, as it requires deduplication methods to efficiently handle any false positives that may arise from this source.

6.3.3 CNV Calling using GATK gCNV

We employed GATK-gCNV, a versatile algorithm available as an open-source package within GATK, for detecting rare CNVs using read-depth information obtained from targeted sequencing. [210] GATK-gCNV presents a customizable method for detecting CNVs in NGS, making it suitable for various applications, such as trait association studies and clinical screening while ensuring sensitivity and specificity. In principle, NGS data should enable the identification of a majority of CNVs that modify genes with comparable recall to genome sequencing. In practical application, the fluctuation in sequencing coverage caused by hybridization-based exome enrichment and other biases associated with NGS library preparation can alter the accurate read-depth signals. The inherent technical variability has posed considerable obstacles to achieve a balance between the recall of CNV and high precision based on NGS platforms. These alterations also depend on the specific characteristics of each sample.

During the process of CNV calling, it is crucial to consider the limitations associated with the algorithms in use. These limitations encompass factors such as the number of CNVs, their expected length, and the efficiency of the calling process. [197] The GATK-gCNV utilizes a probabilistic model that considers technical inconsistencies. This ensuring the prediction of CNVs while maintaining integrity between variant calling and read-depth normalization. The GATK-gCNV employs a Bayesian approach to discern both global and individual sample deviations in read-depth data from extensive populations, all while identifying CNVs. [128]. The steps of the GATK-gCNV pipeline as follows:

Information on coverage is gathered by aligning reads to the genome across specific genomic regions.

II) The original interval list is filtered to remove coverage outliers, unmappable genomic sequences, and regions of segmental duplications.

I) III) Batches are clustered based on read-depth profile similarity, and each batch is processed separately.

II) IV) Total chromosomal coverage of is used to infer chromosomal ploidies.

III) V) The GATK-gCNV acquires knowledge of read-depth bias and noise, it continuously refines the posterior probabilities of copy number states until achieving a self-consistent state[128]. Concerning these procedures, we did not eliminate the exons with low mappability from our BED file, and the analysis was conducted using default procedures. [128]

The unfiltered results generated by GATK-gCNV are designed to be highly sensitive, enabling thorough exploration of potential CNVs. On average, it identifies 6.3 CNV calls per sample, with a rarity level below 1% (variant site frequency). These calls consist of 2.4 deletions and 3.9 duplications, accurately pinpointed at a resolution surpassing two well-captured exons.

In this study, we conducted thorough benchmarking of GATK-gCNV on Ion Torrent Sequencing. In this evaluation, we have effectively demonstrated the consistent processing and generation of CNV through the employed approach. To enhance the specificity of the analysis, filters at both the sample and variant levels were employed, following the same approach utilized in previous instances of HMZdelFinder analysis. This approach allows for accurate detection while removing technical artifacts. In our study, qPCR was utilized to verify and validate the observed findings in a practical context. According to the findings, a prevalence rate of one in seven cases was ascertained through the real-time PCR analysis. Specifically, this investigation identified a heterozygous deletion within the *CLBP* gene. So, it is crucial to conduct further investigations to address the issue of calling high false positive rates associated with the identified CNVs. Our finding is in accordance with the previous research on GATK-gCNV results demonstrated a 95% recall rate across samples.
[128] However, the accuracy drops to 22% when utilizing unfiltered outputs, indicating a significant lack of precision. [128] Another study showed that the GATK gCNV caller is the most sensitive tool for identifying CNVs in WES and with a mean precision of less than 13%. [143] In the event of detecting small CNVs, particularly those encompassing a single exon, the GATK-gCNV demonstrated good performance in terms of accurately identifying and predicting such variations. However, the use of other softwares is also recommended if the objective is to achieve a low false positive rate by selecting CNVs that intersect. [75] However, limited consistencies in CNV detection among three distinct read-depth based programs suggest that the current utilization of WES for CNV detection is not yet mature. [203] The limited ability to identify CNVs and the associated ambiguity in their specificity when relying on WES, as opposed to chromosomal microarray (CMA) based CNV detection, implies that CMA will remain crucial in the identification of clinically significant CNVs during the NGS era, wherein WES or targeted panels are predominantly employed. [203] In line with this study, the low specificity observed in our study among two distinct read-depth-based programs suggests the detection of CNVs through targeted sequencing is still in its early stages, especially when using Ion Torrent technology. [203] The coverage of all CNVs necessitates at the moment the use of at least 10 tools. [197]

6.3.4 Searching for MAK-Alu sequence within a cohort of PID

Large insertions or deletions in heterozygosity might be undetected or under detected when employing standard NGS workflows. An instance of such a mutation is the recent identification of an Alu insertion in the Male Germ Cell-Associated Kinase (MAK) gene, which is overlooked genomic alterations. [79]
The use of the known junctional insertion sequence to previously identified insertions (or deletions), which may encompass founder mutations within the specific population, is under investigation. As part of the research, we analyzed the raw NGS data from 200 samples. Our investigation revealed that none

of the samples contained MAK-Alu insertions. The MAK-Alu mutation is not a prevalent genetic cause of PID diseases. Nevertheless, this straightforward method eliminates the need for laboratory experiments or computationally intensive algorithms. Additionally, it can be applied to identify other known disease-causing insertions and deletions. [79]

6.3.5 Conclusion of CNV Identification in PID

Incorporating CNVs analysis into a clinical workflow can enhance the identification of molecular diagnoses for patients with unresolved (SNV/indel) cases. Considering the cost-effectiveness, focused methodology, and reduced analytical workload compared to genome sequencing, targeted NGS is anticipated to remain a valuable tool in diagnostics over long run. [143] Therefore, it remains crucial to allocate further investments towards enhancing resources and refining the current tools to improve the effectiveness of targeted panels as a diagnostic test. It is crucial to follow appropriate procedures when establishing and implementing a pipeline that includes rigorous quality control measures in a clinical setting. By CNV analysis, the workload associated with CNVs detection using independent approaches can be significantly diminished.

However, to minimize the false report, validations through alternative methods are essential. Prior to conveying the results to patients, it is crucial to validate clinically significant CNVs using an alternative approach. This validation step is necessary because the detection of CNVs-based methods remains dependent on other samples employed for comparison. To ensure accuracy and reliability, employing an orthogonal method for validation becomes imperative. So, we have used a restrict CNVs calling from targeted panel data by reviewing manually and verifying the CNVs, and the clustering quality of the probe underlying the CNVs. Thereby minimizing the potential errors and providing more reliable results for patients.

There are several critical considerations regarding the identification of CNVs using NGS data with strategies to prioritize superior CNVs with clinical significance. Ensuring the samples are generated utilizing identical sequencing platforms, library preparation techniques, and target capture kits are of utmost significance. Ultimately, we strongly advise confirming the CNVs through alternative techniques such as ddPCR or quantitative PCR (qPCR), aCGH (array comparative genomic hybridization)[211,212], MLPA[213,214], and long-read genome sequencing. [199,215]

In summary, both computational CNV prediction tools exhibited a significant number of false positive predictions. Consequently, it is imperative to validate the identified using the appropriate methods. Our research highlights the need for a more reliable and reproducible approach for detecting CNVs from NGS data, specifically designed for clinical applications. The conclusions of this research should be viewed in light of its limitations. The first is the analysis was carried out over time. Since the

establishing of a CNV benchmark set is a notably difficult and costly procedure, the use of control positive samples for assessing the performance algorithms in batch is limited. So, selecting samples from the same batch and using a single control tools like CONTRA might be beneficial. [135] Increasing the sample size can help scientists thoroughly examine the attributes of the anticipated CNV and study the outcomes at varying sequencing depths. Another constraint is that consistent conditions might not always be ideal for computational methods. For instance, the default parameters may not always indicate the optimal number of samples needed for reliable identification.

6.4. Molecular characterization and surveillance of nine Iranian families suspected of hereditary cancer

5.4.1. Molecular characterization and surveillance of seven Iranian families with targeted panel sequencing

A medical necessity arises from the high prevalence of LS in Iran and the Middle East, requiring the detection of LS mutation carriers. By excluding non-carriers from the surveillance program, the compliance of carriers with cancer surveillance programs will be enhanced. This targeted approach will result in a considerable reduction in incidence of illness and death associated with LS-related malignancies and patients' financial burden. [216,217] In the current investigation, the identification of seven distinct putative pathogenic variants, encompassing three newly discovered mutations, responsible for hereditary cancers in seven unrelated families was achieved through multigene panel sequencing. Limited research studies have indicated the development of hereditary cancers is resulting from a genetic defect in the *PMS1* gene. [95,218] The role of the *PMS1* gene in CRC predisposition, however, was found unproven by most studies. [96,219,220] A pathogenic variant in the *PMS1* gene, expected to be be disease-causing, was identified through NGS analysis in an ovarian cancer-affected member in our study. Further analysis was conducted on the maternal aunt of the proband, who had developed CRC. To ensure no genetic defect in MMR genes is overlooked, as was observed in a previous study. [221] The investigation of protein expression in her tumour tissue was conducted through immunohistochemistry (IHC) analysis. Additionally, her sample was examined to determine the presence of the suspected pathogenic variant in the *PMS1* gene. The intact expression of MMR proteins revealed through IHC analysis, and she did not exhibit the *PMS1* mutation, thus suggesting the occurrence of sporadic CRC in her case. In the general population, ovarian cancer typically manifests with an average age of 63 years. [222] The absence of co-segregation between the *PMS1* pathogenic variants observed in the aunt and the occurrence of ovarian cancer in the proband at the age of 64 is a crucial factor that needs to be acknowledged. There is a possibility that the anticipated pathogenic variation in the *PMS1* gene should not be considered as having a causative link to the

inherited susceptibility of ovarian cancer, as supported by previous research findings. [96,223], Despite the assessment made by Varsome and Franklin regarding the pathogenicity of this variant, there is no available evidence suggesting that this particular pathogenic variant plays a causative role in a hereditary cancer syndrome within this family.

An important discovery in the current research is that among the seven index patients examined, two were carrier of significant genomic deletions in either the *MSH2* or *MSH6* genes. This finding highlights the critical significance of identifying large pathogenic variations through CNV analysis in diagnostic laboratories. [146] A research conducted in the northern region of Iran revealed that 10.9% of CRC patients satisfied the Amsterdam criteria. [224] Nevertheless, the examination of genetic disorders in families that met the clinical criteria from this particular area was only conducted in a single study. [225] The C family, carrying the c.705dupA mutation in the MSH2 gene, represents the second confirmed LS family identified in the Mazandaran province of Iran. According to NCCN guidelines, to commence colonoscopic surveillance between the ages of 20 and 25, or 2-5 years prior to the earliest documented occurrence of CRC in LS families is recommended. [226] In this family, even though there was a case of CRC at the age of 23, it was observed that some of the younger members were advised to commence a surveillance program after age 25. This highlights the need for genetic and clinical counselling, which is currently lacking in the Iranian healthcare system, to be included as an essential component of surveillance protocols. Prior to genetic testing, several individuals from family C were included in LS surveillance, irrespective of their carrier status, for an extended period. Ensuring broader utilization of genetic counselling and genetic testing is of utmost importance in minimizing the occurrence of hereditary cancers. The existing healthcare system in Iran lacks accuracy in LS surveillance and fails to assess the carrier status, as highlighted in our earlier research. [217] Therefore, it is imperative to prioritize various measures such as examining familial backgrounds, creating a comprehensive national LS registry, and implementing genetic testing as crucial steps to enhance the surveillance programs for LS in Iran. Based on our understanding, there are only a limited number of genetic centers in Iran dedicated to the diagnosis of the syndrome, and they maintain a local registry for LS cases. [227] Therefore, conducting genetic analysis on extensive pedigrees, like the ones we have presented, has the potential to contribute valuable insights into the understanding of the disorder. Additionally, it can enhance our understanding of the specific genes involved in the occurrence of this syndrome. Surprisingly, the participation and agreement for genetic testing among the relatives of families C and D surpassed that of the other five families. This discrepancy can be attributed to the proactive involvement of the probands and the notable prevalence of cancer within these families. Additionally, the specialists implemented a direct management approach that involved effective communication with relatives who were deemed at risk. These specialists took the initiative to organize the collection of blood samples from these individuals, ensuring that the process was

convenient for them. The reluctance to share genetic information with relatives in other families has undeniably been influenced by significant cultural and social factors. The previously documented case of the pathogenic variant c.842C>G within the *MSH2* gene originated from Tunisia. [147] Our study has provided definitive confirmation of its role in causing diseases. Based on our study, we have identified three distinct *MSH2* pathogenic variants. The range of tumours observed in individuals confirmed to carry these *MSH2* pathogenic variants closely aligns with the tumour spectrum reported in Western countries, where CRC and EC have the highest incidence rates. [228] Compound heterozygosity for the variant c.3226C>T in the *MSH6* gene has been documented in existing academic literature, with reports of its occurrence in three separate unrelated cases. Additionally, this genetic abnormality has been observed in two sisters diagnosed with constitutional mismatch repair cancer syndrome. [148,229-231] This variant is classified as a rare variant, described in the general population with an average frequency of 0.00009, as reported by the GenomAD and ExAC databases. In the present investigation, we examined an individual (the father of the probands) from family E who was affected by CRC. Our findings revealed the presence of a specific genetic alteration, namely the c.3226C>T variant in the *MSH6* gene, in this individual. Additionally, we observed a loss of protein expression in MSH2/MSH6. Furthermore, upon conducting a pedigree analysis for the proband, it was observed that five instances of breast cancer had occurred on the maternal side. This finding highlights the necessity for additional investigations aimed at uncovering the specific gene responsible for this pattern. Moreover, an examination of the proband's family lineage revealed the presence of five instances of breast cancer among maternal relatives. This observation highlights the necessity for additional investigations aimed at elucidating the specific gene responsible for this inherited condition. A mutation in the MSH6 gene has the potential to cause reduced immunohistochemical (IHC) staining of the MSH2/MSH6 protein in both tumour and normal tissues. [232] Haploinsufficiency refers to the condition where one allele of specific tumour suppressor genes (TSGs), like MMR genes, experiences a loss-of-function, which can lead to tumorigenesis. [59] In our current investigation, we identified a deficiency in MSH2/MSH6 protein expression within both the tumour and normal tissues of a breast cancer patient. This individual carries the EX9del variant in the *MSH6* gene. This suggests that the tumorigenesis process did not involve a second (somatic) inactivating mutation that would lead to a total loss of the MSH2/MSH6 protein. Our findings align with prior research indicating that a pathogenic mutation in an MMR gene may result in mildly positive staining of MMR proteins within tumour tissue. This staining pattern signifies inadequate expression of the MMR protein, resembling a state of haploinsufficiency. [232-235] Our research aligns with earlier findings that validate the high susceptibility of individuals with germline mutation c.943C>T in the *PMS2* gene to develop cancers associated with LS. This confirmation is supported by IHC analysis. [150,236,237]

6.4.1. 2 Conclusion:

The use of NGS data for identifying CNVs enhances the effectiveness of cancer genetic diagnostic clinics. Therefore, leading to improved patient surveillance and a reduction in costly and invasive procedures, as they are selectively performed solely on individuals identified as mutation carriers. All the LS mutations identified within the examined families verified for their direct association with the pathology, either through their inherent characteristics or functional demonstrations. An interesting finding deriving from our examination of an Iranian family lineage suggests that the *PMS1* gene is unlikely to be a predisposing factor for cancer. This conclusion stems from the absence of the presumed disease-causing mutation in the affected individuals. Tumorigenesis can occur even in the absence of a second hit when there is a loss-of-function mutation in one allele of the *MSH6* gene, leading to inadequate expression of the MMR MSH6 and MSH2 proteins. Hence, the *MSH6* gene can be defined as a haploinsufficient gene. The range of tumours observed in individuals carrying LS in the *MSH2* gene is comparable to that documented in Western nations, primarily consisting of CRC and endometrial cancers. Enhancing genetic analysis among individuals meeting clinical criteria for hereditary cancers could significantly improve the management of LS in Iranian patients. This advancement would involve the exclusion of LS non-carriers from surveillance programs, thereby optimizing the allocation of resources and focusing attention on those individuals at higher risk.

6.4.2. Two LS families identified by WES and CNV analysis

WES focuses on approximately 3% of the entire genome, specifically targeting protein-coding genes. [238] WES on families with CRC along with functional computational variant prioritization, can lead to the discovery of variants that exhibited consistent inheritance patterns with the disease. [239] In this research using WES, the initial identification of *EPCAM* and *MSH2* deletions within Iranian families marked a significant milestone. These findings shed light on the genetic landscape of these regions and contribute to the broader understanding of hereditary conditions associated with LS. Given that the extended H and I families presently reside in the northern region of Iran, it is conceivable that they share a common ancestor within the first four generations. Unfortunately, no further historical information was accessible at this time. Out of the 19 individuals from families H and I who underwent genetic testing and were above the age of 25, it was found that 15 individuals had taken part in the LS surveillance program. This outcome brings positive news since early identification and treatment play a crucial role in managing cancers associated with LS. Nevertheless, it is essential to emphasize that, based on our findings, two individuals did not carry the mutation. This exposed a clear deficiency in comprehension regarding the genetic testing procedure and the ramifications of its outcomes. In addition, it is concerning to note that two individuals carrying the mutation, both above

the age of 25, were not part of any surveillance program. This discovery is deeply worrisome since regular surveillance plays a vital role in detecting and treating cancers associated with LS at an early stage. To ensure that individuals with LS-associated mutations are well-informed about the significance of surveillance and receive appropriate care, it is crucial to conduct further investigation and provide adequate education. The epigenetic silencing of the *MSH2* gene, caused by deletions in *EPCAM* led to LS. [240] In relation to *EPCAM* deletions and their association with CRC, the probability of developing the disease is similar to that of individuals who carry *MSH2* mutations. In contrast to individuals harbouring an *MSH2* mutation, the likelihood of developing endometrial cancer is diminished. [241,242] In accordance with our research findings, it has been observed that individuals carrying the *EPCAM* mutation are more prone to developing CRCs during their younger years. Out of the nine reported cases of CRC, eight occurred in individuals below the age of 40, while one involved an individual above the age of 50. Therefore, it is crucial to focus gastrointestinal prevention strategies on the colon and rectum for individuals with *EPCAM* deletion, considering the occurrence of CRC and the potential for metachronous CRCs. [87] According to the latest findings, a mere 8% of individuals who were diagnosed positive with WES opted for cascade testing. The average number of relatives tested per patient was approximately 1.5. [243] Nevertheless, our research focused solely on a single family, wherein over 20 individuals underwent cascade testing. The significant number of individuals tested within Family A can be attributed to the proactive involvement of the proband. Following the proband's awareness of the genetic anomaly and subsequent genetic consultation, she informed other members of the family, leading them to seek additional testing. This emphasizes the pivotal role of increasing awareness among probands as a crucial measure in preventive efforts.

Numerous global laboratories, corporations, and educational establishments are employing NGS to conduct tests for LS. [244] The current recommendations advocate employing IHC analysis as a preliminary step, followed by Sanger sequencing to identify any loss-of-function mutations in the respective gene. This approach ensures accurate and reliable results while investigating gene function impairment. [182] As a result, certain diagnostic laboratories still employ Sanger sequencing to analyze MMR genes in individuals displaying a deficiency of MMR protein. [87,245,246] In situations where there is a depletion of MSH2/MSH6 protein expression, it is advisable to employ a sequential testing methodology that utilizes Sanger sequencing. In terms of prioritizing genetic testing, it is recommended to begin with *MSH2* sequencing as the first step. If the specific variant responsible for the condition is not detected, the next recommended course of action would be to proceed with *EPCAM* gene testing. In instances where no variant is identified within the *EPCAM* gene, it is advisable to pursue *MSH6* gene sequencing. [247] Performing Sanger sequencing on the MMR genes is a time-consuming and expensive task, primarily because of the considerable length of these genes. The inclusion of supplementary genes for testing can further escalate the intricacy and financial burden

associated with the procedure. [247] Furthermore, the diagnostic efficacy of Sanger sequencing is hindered by its incapacity to identify CNVs, thereby limiting its utility in approximately 20% of LS cases. When considering the cost analysis of examining the MSH2 gene through Sanger sequencing versus WES, our investigation demonstrated a noteworthy 12% decrease in expenses per sample when utilizing WES in comparison to Sanger sequencing. In simple terms, sequencing approximately 20,000 genes through WES is more cost-effective compared to sequencing just the MSH2 gene, which consists of 16 exons, using Sanger sequencing. The analysis of WES data and CNV analysis pose considerable difficulties when interpreting incidental findings with variant uncertain significance (VUS). To tackle these difficulties, IHC as a functional assay can be employed in instances involving VUS to ascertain the pathogenic nature of germline variants. [248] Accurate risk assessments of cancer in patients and their families hold immense importance for clinicians and genetic counsellors. This information plays a crucial role in their practice, enabling them to provide reliable guidance and recommendations. Considerable progress has been made in contemporary studies, leading to valuable revelations regarding the functional ramifications associated with the missense variant and its role in the development of cancer. This analysis involved a thorough investigation into the characteristics of the tumour, encompassing the assessment of microsatellite instability and the lack of mismatch repair protein expression. By integrating this data into the analysis of germline missense variants, we can enhance the precision of genetic testing and optimize the care provided to individuals at higher susceptibility to cancer. [248] The accurate interpretation of massively parallel sequencing experiments was significantly influenced by the mapping and alignment capabilities of these regions. [78] An exploration was conducted to examine the properties of mappability within the MMR gene family. The investigation revealed remarkable variations in mappability among the MMR genes, with particular emphasis on the *PMS2* gene. The difficulties associated with accurately mapping *PMS2* in WES due to its low mappability, it is recommended to utilize an RNA-based sequencing for detecting mutations. This strategy will enable the differentiation between the *PMS2* gene and its pseudogene counterpart, *PMS2CL*, thereby addressing the issue. This method had made notable progress in detecting PMS2 pathogenic variations, exhibiting enhancements of up to 20%. [249] A refined approach is suggested considering the outcomes acquired, as depicted in *Figure 15*. Nonetheless, the constrained number of samples utilized in our investigation poses a considerable challenge in extrapolating these findings to other research facilities. Therefore, it is crucial to assess the testing expenses across diverse regions. The identification of significant fluctuations in LS genes using MLPA has gained extensive acceptance and constitutes approximately 5% to 20% of all MMR gene mutations. [250,251] While MLPA is known for its efficacy, it has limited throughput due to its reliance on multiplex PCR. As a result, it is not suitable for conducting high-throughput screenings of all cancer susceptibility genes. CNV analysis using WES read depth data is a valuable screening tool that complements existing methods, aiding in

the identification of clinically significant CNVs within the exome. This approach has the potential to enhance diagnostic yields significantly. Incorporating supplementary analysis of CNVs holds the potential to uncover a majority disease-causing CNVs, as well as a considerable number of smaller CNVs. [252] The use of SYBR-Green real-time PCR offers several major advantages. First, it requires only a small amount of DNA and can be completed within two hours. Second, it is relatively simple and less costly than using TaqMan probes. [145] In addition, the SYBR-Green real-time PCR technique offers a reliable means of quantifying gene deletion with precision. This assay's fundamental principle can be effectively applied to various biological systems, offering immense potential to enhance patient care through its ability to deliver faster and more precise outcomes.

In summary, our suggested revision to the current LS genetics diagnosis guideline offers a comprehensive solution to address the challenging aspects of the guideline, including its arduous nature, high costs, and labor-intensive requirements. Our primary objective is to maximize cost savings and improve diagnostic yields. We propose the adoption of WES and CNV analysis as the preferred sequencing method for individuals with CRC suspected of having LS, as opposed to using Sanger sequencing. The significance of this method becomes especially apparent in cases where the absence of MMR protein expression is observable within tumours. *EPCAM* deletions present a comparable threat to CRC as *MSH2* mutations, albeit with a reduced risk of developing other cancers associated with LS. The integration of genetic testing for individuals who fulfil the Amsterdam criteria has the potential to improve the existing LS surveillance in Iran, a country known for its high incidence of LS attributed to elevated consanguinity rates. [253] The strategy involves the omission of individuals without the mutation from the surveillance process, while actively identifying and including those who carry LS mutations for participation in the surveillance program.

7. Conclusions

The main goal of this PhD program was to identify and interpret pathogenic variants following NGS analysis in human Mendelian disorders. The following four main projects were conducted to achieve this goal:

A case of X-linked CGD in a female with pathogenic variants in the splicing site was investigated. A novel de novo pathogenic variant that caused exon skipping was identified in the CGD case and the study of X-chromosome inactivation revealed a skewed pattern impacting on the clinical phenotype. This variant was not previously reported in the literature, and it provides new insights into the clinical basis of CGD.

A novel hemizygous variant in the MAGT1 gene located on the X-chromosome was identified adding a new case of XMEN to the few described, moreover the mutation is the first splicing defect found as causing variant in the gene.

A new analysis of CNVs using many different softwares available was conducted to increase the diagnostic rate of a NGS panel for Inborn errors of immunity on the IonTorrent platform.

The HMZDelFinder algorithm was found to be the most effective method for detecting CNVs in the IonTorrent platform. This finding has important implications for the use of NGS in the diagnostic setting, as it suggests that the HMZDelFinder algorithm can be used to improve the diagnostic yield of NGS panels for inborn errors of immunity, even if the chemistry used for the amplicon based technique is not suitable for such an analysis.

Nine Iranian families with a personal or familial history of cancer were identified, and the use of NGS and CNV detection was demonstrated useful to improve the diagnostic yields.

Eight pathogenic variants were identified in nine Iranian families with a personal or familial history of cancer using a combination of multigene panel testing and whole exome sequencing. The use of NGS and CNV detection in this project significantly improved the diagnostic yield, and it provided new insights into the molecular basis of LS. our proposed modification to the existing LS genetics diagnosis guideline effectively tackles the arduous, expensive, and labor-intensive elements associated with the current guideline.

8. Acknowledgment

I would like to thank Prof. Silvia Giliani, supervisor of this doctoral thesis, for the innumerable opportunities for personal and professional growth, and for the support, as well as for the top-level scientific supervision, which she have given me offered in these years. I thank the many collaborators, including Prof Keivan Majidzadeh-A, Shiva Zarinfam, Dr Rosalba Monica Ferraro, and Dr Elena Laura Mazzoldi, who participated in the study through the referral of patients, the support in the experiments and discussion of the results and data. I thank all the individual members of the various research teams that I have had the pleasure of meeting and colleagues at the "Breast Cancer Research Center, Motamed Cancer Institute, ACECR, Tehran, Iran". Also special thanks for colleague, including Dr Chiara Romani and Dr Manuela Baronio at the laboratory at the "A. Nocivelli" Institute for Molecular Medicine Spedali Civili and University of Brescia, Brescia. I could never properly put into words how invaluable each of you has been during the different stages of my PhD journey.

I would like to express my heartfelt appreciation to my family, with a special mention of my wife and her parents, as well as my brother, mother, and grandmother. Their steadfast love, unwavering support, and invaluable guidance have been the driving force behind my academic journey. Their unwavering belief in my abilities has served as a perpetual wellspring of inspiration, and I am profoundly grateful for their instrumental role in shaping my path and contributing to my accomplishments.

9. List of abbreviations

- aCGH: array comparative genomic hybridization
- BC: breast cancer
- BT: brain tumors
- CGD: chronic granulomatous disorder
- CMA: chromosomal microarray
- CNVs: Copy Number Variations
- CRC: colorectal cancer
- CVID: common variable immunodeficiency
- EBV: Epstein-Barr virus
- ExAC: Exome Aggregation Consortium
- FGP: Fundic gland polyposis
- GC: gastric cancer
- HDGC: hereditary diffuse gastric cancer
- HMZ: hemizygous and homozygous
- HSCT: hematopoietic stem cell transplantation
- HNPCC: hereditary non-polyposis colorectal cancer
- IGV: Integrative Genome Viewer
- IHC: immunohistochemistry
- InDels: insertions or deletions
- IRD: inherited retinal diseases
- Kb: kilobase
- MAF: minor allele frequency
- MAK: Male Germ Cell-Associated Kinase
- MB: megabases
- MLPA: Multiplex ligation-dependent probe amplification
- MSI: microsatellite instability
- NADPH: nicotinamide adenine dinucleotide phosphate
- NC: natural killer
- NGS: next-generation sequencing
- OC: ovarian cancer
- OS: osteosarcoma
- PBMC: Peripheral Blood Mononuclear Cells

• PBS: Phosphate-buffered saline
• PMN: polymorphonuclear neutrophils
• PIDs: primary immunodeficiencies
• qPCR: quantitative PCR
• RT-PCR: reverse transcription-polymerase chain reaction
• RPKM: reads per thousand base pairs per million reads
• SB: small bowel cancer
• SNVs: single nucleotide variants
• TSGs: tumor suppressor genes
• VUS: variant uncertain significance
• XCI: X-chromosome inactivation

10. References

1. Vorsteveld EE, Hoischen A, van der Made CI. Next-Generation Sequencing in the Field of Primary Immunodeficiencies: Current Yield, Challenges, and Future Perspectives. Clinical reviews in allergy & immunology. 2021;61(2):212-225.
2. Kamps R, Brandão RD, Bosch BJ, et al. Next-Generation Sequencing in Oncology: Genetic Diagnosis, Risk Prediction and Cancer Classification. International journal of molecular sciences. 2017;18(2).
3. Lorenzi D, Fernández C, Bilinski M, et al. First custom next-generation sequencing infertility panel in Latin America: design and first results. JBRA assisted reproduction. 2020;24(2):104-114.
4. Wright CF, Campbell P, Eberhardt RY, et al. Genomic Diagnosis of Rare Pediatric Disease in the United Kingdom and Ireland. 2023.
5. Yu H-H, Yang Y-H, Chiang B-LJCria, immunology. Chronic granulomatous disease: a comprehensive review. 2021;61:101-113.
6. Tangye SG, Al-Herz W, Bousfiha A, et al. Human Inborn Errors of Immunity: 2019 Update on the Classification from the International Union of Immunological Societies Expert Committee. Journal of clinical immunology. 2020;40(1):24-64.
7. Picard C, Bobby Gaspar H, Al-Herz W, et al. International Union of Immunological Societies: 2017 Primary Immunodeficiency Diseases Committee Report on Inborn Errors of Immunity. Journal of clinical immunology. 2018;38(1):96-128.
8. Bousfiha A, Jeddane L, Picard C, et al. The 2017 IUIS phenotypic classification for primary immunodeficiencies. 2018;38:129-143.
9. Bennett CA, Petrovski S, Oliver KL, Berkovic SFJNG. ExACtly zero or once: A clinically helpful guide to assessing genetic variants in mild epilepsies. 2017;3(4).
10. Bousfiha A, Moundir A, Tangye SG, et al. The 2022 Update of IUIS Phenotypical Classification for Human Inborn Errors of Immunity. Journal of clinical immunology. 2022;42(7):1508-1520.
11. Smedley D, Smith KR, Martin A, et al. 100,000 Genomes Pilot on Rare-Disease Diagnosis in Health Care - Preliminary Report. The New England journal of medicine. 2021;385(20):1868-1880.
12. Adams DR, Eng CMJNEJoM. Next-generation sequencing to diagnose suspected genetic disorders. 2018;379(14):1353-1362.
13. Lalonde E, Rentas S, Lin F, Dulik MC, Skraban CM, Spinner NBJFiP. Genomic diagnosis for pediatric disorders: revolution and evolution. 2020;8:373.
14. Rowlands CF, Baralle D, Ellingford JM. Machine Learning Approaches for the Prioritization of Genomic Variants Impacting Pre-mRNA Splicing. Cells. 2019;8(12).
15. Peng F, Zhong L, Zhang B, et al. Successful application of next-generation sequencing for pre-natal diagnosis in a pedigree with chronic granulomatosis disease. Experimental and therapeutic medicine. 2019;17(4):2931-2936.
16. Similuk MN, Yan J, Ghosh R, et al. Clinical exome sequencing of 1000 families with complex immune phenotypes: Toward comprehensive genomic evaluations. The Journal of allergy and clinical immunology. 2022;150(4):947-954.
17. Schulz J, Mah N, Neuenschwander M, et al. Loss-of-function uORF mutations in human malignancies. Scientific reports. 2018;8(1):2395.
18. Gutierrez-Rodrigues F, Donaires FS, Pinto A, et al. Pathogenic TERT promoter variants in telomere diseases. 2019;21(7):1594-1602.
19. Jang YJ, LaBella AL, Feeney TP, et al. Disease-causing mutations in the promoter and enhancer of the ornithine transcarbamylase gene. 2018;39(4):527-536.
20. Putscher E, Hecker M, Fitzner B, Lorenz P, Zettl UK. Principles and Practical Considerations for the Analysis of Disease-Associated Alternative Splicing Events Using the Gateway Cloning-Based Minigene Vectors pDESTsplice and pSpliceExpress. International journal of molecular sciences. 2021;22(10).

21. Stenson PD, Mort M, Ball EV, et al. The Human Gene Mutation Database: towards a comprehensive repository of inherited mutation data for medical research, genetic diagnosis and next-generation sequencing studies. Human genetics. 2017;136(6):665-677.
22. Köker MY, Sanal O, de Boer M, et al. Skewing of X-chromosome inactivation in three generations of carriers with X-linked chronic granulomatous disease within one family. European journal of clinical investigation. 2006;36(4):257-264.
23. Leparc GG, Mitra RD. A sensitive procedure to detect alternatively spliced mRNA in pooled-tissue samples. Nucleic acids research. 2007;35(21):e146.
24. Dufner-Almeida LG, do Carmo RT, Masotti C, Haddad LA. Understanding human DNA variants affecting pre-mRNA splicing in the NGS era. Advances in genetics. 2019;103:39-90.
25. Martin HC, Gardner EJ, Samocha KE, et al. The contribution of X-linked coding variation to severe developmental disorders. 2021;12(1):627.
26. Migeon BRJGiM. X-linked diseases: susceptible females. 2020;22(7):1156-1174.
27. Tarpey PS, Smith R, Pleasance E, et al. A systematic, large-scale resequencing screen of X-chromosome coding exons in mental retardation. 2009;41(5):535-543.
28. Casanova JL, Abel L. Human genetics of infectious diseases: Unique insights into immunological redundancy. Seminars in immunology. 2018;36:1-12.
29. Matsuda-Lennikov M, Biancalana M, Zou J, et al. Magnesium transporter 1 (MAGT1) deficiency causes selective defects in N-linked glycosylation and expression of immune-response genes. The Journal of biological chemistry. 2019;294(37):13638-13656.
30. Yu HH, Yang YH, Chiang BL. Chronic Granulomatous Disease: a Comprehensive Review. Clinical reviews in allergy & immunology. 2021;61(2):101-113.
31. Rider NL, Jameson MB, Creech CB. Chronic Granulomatous Disease: Epidemiology, Pathophysiology, and Genetic Basis of Disease. Journal of the Pediatric Infectious Diseases Society. 2018;7(suppl_1):S2-s5.
32. Nunoi H, Nakamura H, Nishimura T, Matsukura M. Recent topics and advanced therapies in chronic granulomatous disease. Human cell. 2023;36(2):515-527.
33. Kumar GJBCRC. Lytic lesion of skull: a rare presentation of chronic granulomatous disease. 2020;13(9):e235423.
34. Anjani G, Vignesh P, Joshi V, et al. Recent advances in chronic granulomatous disease. 2020;7(1):84-92.
35. Roos D, Kuhns DB, Maddalena A, et al. Hematologically important mutations: X-linked chronic granulomatous disease (third update). 2010;45(3):246-265.
36. Marciano BE, Spalding C, Fitzgerald A, et al. Common severe infections in chronic granulomatous disease. Clinical infectious diseases : an official publication of the Infectious Diseases Society of America. 2015;60(8):1176-1183.
37. Roos D, Kuhns DB, Maddalena A, et al. Hematologically important mutations: the autosomal recessive forms of chronic granulomatous disease (second update). Blood cells, molecules & diseases. 2010;44(4):291-299.
38. Lyon MFJn. Gene action in the X-chromosome of the mouse (Mus musculus L.). 1961;190(4773):372-373.
39. Agrelo R, Wutz A. X inactivation and disease. Paper presented at: Seminars in cell & developmental biology2010.
40. Vowells SJ, Fleisher TA, Sekhsaria S, Alling DW, Maguire TE, Malech HLJTJop. Genotype-dependent variability in flow cytometric evaluation of reduced nicotinamide adenine dinucleotide phosphate oxidase function in patients with chronic granulomatous disease. 1996;128(1):104-107.
41. Roesler J, Hecht M, Freihorst J, Lohmann-Matthes ML, Emmendörffer A. Diagnosis of chronic granulomatous disease and of its mode of inheritance by dihydrorhodamine 123 and flow microcytofluorometry. European journal of pediatrics. 1991;150(3):161-165.
42. Talebizadeh Z, Bittel DC, Veatch OJ, Kibiryeva N, Butler MG. Brief report: non-random X chromosome inactivation in females with autism. Journal of autism and developmental disorders. 2005;35(5):675-681.

43. Hatakeyama C, Anderson CL, Beever CL, Peñaherrera MS, Brown CJ, Robinson WP. The dynamics of X-inactivation skewing as women age. Clinical genetics. 2004;66(4):327-332.
44. Allen RC, Zoghbi HY, Moseley AB, Rosenblatt HM, Belmont JW. Methylation of HpaII and HhaI sites near the polymorphic CAG repeat in the human androgen-receptor gene correlates with X chromosome inactivation. American journal of human genetics. 1992;51(6):1229-1239.
45. Ravell JC, Chauvin SD, He T, Lenardo M. An Update on XMEN Disease. Journal of clinical immunology. 2020;40(5):671-681.
46. Ravell J, Chaigne-Delalande B, Lenardo M. X-linked immunodeficiency with magnesium defect, Epstein-Barr virus infection, and neoplasia disease: a combined immune deficiency with magnesium defect. Current opinion in pediatrics. 2014;26(6):713-719.
47. Cohen JIJNEjom. Epstein–Barr virus infection. 2000;343(7):481-492.
48. Balfour Jr HH, Odumade OA, Schmeling DO, et al. Behavioral, virologic, and immunologic factors associated with acquisition and severity of primary Epstein–Barr virus infection in university students. 2013;207(1):80-88.
49. Haskologlu S, Baskin K, Aytekin C, et al. Scales of Magt1 Gene: Novel Mutations, Different Presentations. Iranian journal of allergy, asthma, and immunology. 2022;21(1):92-97.
50. Higgins CD, Swerdlow AJ, Macsween KF, et al. A study of risk factors for acquisition of Epstein-Barr virus and its subtypes. The Journal of infectious diseases. 2007;195(4):474-482.
51. Brigida I, Chiriaco M, Di Cesare S, et al. Large Deletion of MAGT1 Gene in a Patient with Classic Kaposi Sarcoma, CD4 Lymphopenia, and EBV Infection. Journal of clinical immunology. 2017;37(1):32-35.
52. Vaeth M, Feske S. Ion channelopathies of the immune system. Current opinion in immunology. 2018;52:39-50.
53. Au EYL, Tung EKK, Ip RWK, Li PHJCRiI. Novel MAGT1 Mutation Found in the First Chinese XMEN in Hong Kong. 2022;2022.
54. Ravell JC, Matsuda-Lennikov M, Chauvin SD, et al. Defective glycosylation and multisystem abnormalities characterize the primary immunodeficiency XMEN disease. 2020;130(1):507-522.
55. King JR, Hammarström LJJoci. Newborn screening for primary immunodeficiency diseases: history, current and future practice. 2018;38(1):56-66.
56. Wan R, Schieck M, Caballero-Oteyza A, et al. Copy Number Analysis in a Large Cohort Suggestive of Inborn Errors of Immunity Indicates a Wide Spectrum of Relevant Chromosomal Losses and Gains. Journal of clinical immunology. 2022;42(5):1083-1092.
57. Lupski JRJE, mutagenesis m. Structural variation mutagenesis of the human genome: Impact on disease and evolution. 2015;56(5):419-436.
58. Gambin T, Akdemir ZC, Yuan B, et al. Homozygous and hemizygous CNV detection from exome sequencing data in a Mendelian disease cohort. Nucleic acids research. 2017;45(4):1633-1648.
59. Zschocke J, Byers PH, Wilkie AOM. Mendelian inheritance revisited: dominance and recessiveness in medical genetics. Nature reviews Genetics. 2023.
60. Alkuraya FSJGM. Natural human knockouts and the era of genotype to phenotype. 2015;7:1-3.
61. Gambin T, Akdemir ZC, Yuan B, et al. Homozygous and hemizygous CNV detection from exome sequencing data in a Mendelian disease cohort. 2017;45(4):1633-1648.
62. Liu W, Xie CC, Zhu Y, et al. Homozygous deletions and recurrent amplifications implicate new genes involved in prostate cancer. Neoplasia (New York, NY). 2008;10(8):897-907.
63. Gonzaga-Jauregui C, Lupski JR, Gibbs RA. Human genome sequencing in health and disease. Annual review of medicine. 2012;63:35-61.
64. Hjeij R, Lindstrand A, Francis R, et al. ARMC4 mutations cause primary ciliary dyskinesia with randomization of left/right body asymmetry. American journal of human genetics. 2013;93(2):357-367.
65. Day-Williams AG, Sun C, Jelcic I, et al. Whole Genome Sequencing Reveals a Chromosome 9p Deletion Causing DOCK8 Deficiency in an Adult Diagnosed with Hyper IgE Syndrome Who Developed Progressive Multifocal Leukoencephalopathy. Journal of clinical immunology. 2015;35(1):92-96.

66. Boone PM, Campbell IM, Baggett BC, et al. Deletions of recessive disease genes: CNV contribution to carrier states and disease-causing alleles. 2013;23(9):1383-1394.
67. Fang M, Abolhassani H, Lim CK, Zhang J, Hammarström LJJoci. Next generation sequencing data analysis in primary immunodeficiency disorders–future directions. 2016;36:68-75.
68. Carvalho CM, Lupski JR. Mechanisms underlying structural variant formation in genomic disorders. Nature reviews Genetics. 2016;17(4):224-238.
69. Roca I, González-Castro L, Fernández H, Couce ML, Fernández-Marmiesse A. Free-access copy-number variant detection tools for targeted next-generation sequencing data. Mutation research Reviews in mutation research. 2019;779:114-125.
70. Moreno-Cabrera JM, Del Valle J, Castellanos E, et al. Evaluation of CNV detection tools for NGS panel data in genetic diagnostics. European journal of human genetics : EJHG. 2020;28(12):1645-1655.
71. Rajagopalan R, Murrell JR, Luo M, Conlin LK. A highly sensitive and specific workflow for detecting rare copy-number variants from exome sequencing data. Genome medicine. 2020;12(1):14.
72. Babadi M, Lee SK, Smirnov A, et al. Precise common and rare germline CNV calling with GATK. 2018;78(13_Supplement):2287-2287.
73. Sanghvi RV, Buhay CJ, Powell BC, et al. Characterizing reduced coverage regions through comparison of exome and genome sequencing data across 10 centers. Genetics in medicine : official journal of the American College of Medical Genetics. 2018;20(8):855-866.
74. Mandelker D, Schmidt RJ, Ankala A, et al. Navigating highly homologous genes in a molecular diagnostic setting: a resource for clinical next-generation sequencing. Genetics in medicine : official journal of the American College of Medical Genetics. 2016;18(12):1282-1289.
75. Gordeeva V, Sharova E, Babalyan K, Sultanov R, Govorun VM, Arapidi G. Benchmarking germline CNV calling tools from exome sequencing data. Scientific reports. 2021;11(1):14416.
76. de Ligt J, Boone PM, Pfundt R, et al. Detection of clinically relevant copy number variants with whole-exome sequencing. 2013;34(10):1439-1448.
77. Whiteford N, Haslam N, Weber G, et al. An analysis of the feasibility of short read sequencing. Nucleic acids research. 2005;33(19):e171.
78. Derrien T, Estellé J, Marco Sola S, et al. Fast computation and applications of genome mappability. PloS one. 2012;7(1):e30377.
79. Bujakowska KM, White J, Place E, Consugar M, Comander J. Efficient In Silico Identification of a Common Insertion in the MAK Gene which Causes Retinitis Pigmentosa. PloS one. 2015;10(11):e0142614.
80. Kayser K, Degenhardt F, Holzapfel S, et al. Copy number variation analysis and targeted NGS in 77 families with suspected Lynch syndrome reveals novel potential causative genes. 2018;143(11):2800-2813.
81. Li X, Liu G, Wu W. Recent advances in Lynch syndrome. Experimental hematology & oncology. 2021;10(1):37.
82. Yurgelun MB, Hampel H. Recent Advances in Lynch Syndrome: Diagnosis, Treatment, and Cancer Prevention. American Society of Clinical Oncology educational book American Society of Clinical Oncology Annual Meeting. 2018;38:101-109.
83. Beitsch PD, Whitworth PW, Hughes K, et al. Underdiagnosis of Hereditary Breast Cancer: Are Genetic Testing Guidelines a Tool or an Obstacle? Journal of clinical oncology : official journal of the American Society of Clinical Oncology. 2019;37(6):453-460.
84. Umar A, Boland CR, Terdiman JP, et al. Revised Bethesda Guidelines for hereditary nonpolyposis colorectal cancer (Lynch syndrome) and microsatellite instability. Journal of the National Cancer Institute. 2004;96(4):261-268.
85. Sina M, Ghorbanoghli Z, Abedrabbo A, et al. Identification and management of Lynch syndrome in the Middle East and North African countries: outcome of a survey in 12 countries. Familial cancer. 2021;20(3):215-221.

86. Møller P, Seppälä TT, Bernstein I, et al. Cancer risk and survival in path_MMR carriers by gene and gender up to 75 years of age: a report from the Prospective Lynch Syndrome Database. Gut. 2018;67(7):1306-1316.
87. Cini G, Quaia M, Canzonieri V, et al. Toward a better definition of EPCAM deletions in Lynch Syndrome: Report of new variants in Italy and the associated molecular phenotype. Molecular genetics & genomic medicine. 2019;7(5):e587.
88. da Silva JA, Castedo S, Pedroto I, Marcos-Pinto RJEJoMG. Extracolonic tumours in a pedigree with EPCAM-related Lynch Syndrome. 2022;65(5):104479.
89. Lynch HT, Riegert-Johnson DL, Snyder C, et al. Lynch syndrome-associated extracolonic tumors are rare in two extended families with the same EPCAM deletion. 2011;106(10):1829.
90. Martin FC, Chenevix-Trench G, Yeomans ND. Systematic review with meta-analysis: fundic gland polyps and proton pump inhibitors. Alimentary pharmacology & therapeutics. 2016;44(9):915-925.
91. Akanuma N, Rabinovitch PS, Mattis AN, Lauwers GY, Choi WT. Fundic Gland Polyps Lack DNA Content Abnormality Characteristic of Other Adenomatous Precursor Lesions in the Gastrointestinal Tract. Modern pathology : an official journal of the United States and Canadian Academy of Pathology, Inc. 2023;36(5):100117.
92. Hashimoto T, Ogawa R, Matsubara A, et al. Familial adenomatous polyposis-associated and sporadic pyloric gland adenomas of the upper gastrointestinal tract share common genetic features. Histopathology. 2015;67(5):689-698.
93. Nicolaides NC, Papadopoulos N, Liu B, et al. Mutations of two PMS homologues in hereditary nonpolyposis colon cancer. Nature. 1994;371(6492):75-80.
94. Betti M, Casalone E, Ferrante D, et al. Germline mutations in DNA repair genes predispose asbestos-exposed patients to malignant pleural mesothelioma. Cancer letters. 2017;405:38-45.
95. Castéra L, Krieger S, Rousselin A, et al. Next-generation sequencing for the diagnosis of hereditary breast and ovarian cancer using genomic capture targeting multiple candidate genes. European journal of human genetics : EJHG. 2014;22(11):1305-1313.
96. Prolla TA, Baker SM, Harris AC, et al. Tumour susceptibility and spontaneous mutation in mice deficient in Mlh1, Pms1 and Pms2 DNA mismatch repair. Nature genetics. 1998;18(3):276-279.
97. Schwitalle Y, Kloor M, Eiermann S, et al. Immune response against frameshift-induced neopeptides in HNPCC patients and healthy HNPCC mutation carriers. Gastroenterology. 2008;134(4):988-997.
98. Babaei H, Zeinalian M, Emami MH, Hashemzadeh M, Farahani N, Salehi R. Simplified microsatellite instability detection protocol provides equivalent sensitivity to robust detection strategies in Lynch syndrome patients. Cancer biology & medicine. 2017;14(2):142-150.
99. Lee CT, Chow NH, Chen YL, et al. Clinicopathological features of mismatch repair protein expression patterns in colorectal cancer. Pathology, research and practice. 2021;217:153288.
100. Hansen MF, Johansen J, Sylvander AE, et al. Use of multigene-panel identifies pathogenic variants in several CRC-predisposing genes in patients previously tested for Lynch Syndrome. Clinical genetics. 2017;92(4):405-414.
101. Yurgelun MB, Allen B, Kaldate RR, et al. Identification of a Variety of Mutations in Cancer Predisposition Genes in Patients With Suspected Lynch Syndrome. Gastroenterology. 2015;149(3):604-613.e620.
102. Bhai P, Levy MA, Rooney K, et al. Analysis of Sequence and Copy Number Variants in Canadian Patient Cohort With Familial Cancer Syndromes Using a Unique Next Generation Sequencing Based Approach. Frontiers in genetics. 2021;12:698595.
103. Stoffel EM, Koeppe E, Everett J, et al. Germline Genetic Features of Young Individuals With Colorectal Cancer. Gastroenterology. 2018;154(4):897-905.e891.
104. Kerkhof J, Schenkel LC, Reilly J, et al. Clinical Validation of Copy Number Variant Detection from Targeted Next-Generation Sequencing Panels. The Journal of molecular diagnostics : JMD. 2017;19(6):905-920.
105. Hampel HJSOC. Genetic testing for hereditary colorectal cancer. 2009;18(4):687-703.

106. Yao R, Zhang C, Yu T, et al. Evaluation of three read-depth based CNV detection tools using whole-exome sequencing data. 2017;10:1-7.
107. Hong CS, Singh LN, Mullikin JC, Biesecker LGJGm. Assessing the reproducibility of exome copy number variations predictions. 2016;8(1):1-11.
108. Rajagopalan R, Murrell JR, Luo M, Conlin LKJGm. A highly sensitive and specific workflow for detecting rare copy-number variants from exome sequencing data. 2020;12(1):1-11.
109. Ma L, Chung WK. Quantitative analysis of copy number variants based on real-time LightCycler PCR. Current protocols in human genetics. 2014;80:7.21.21-27.21.28.
110. Agaoglu NB, Unal B, Akgun Dogan O, et al. Determining the accuracy of next generation sequencing based copy number variation analysis in Hereditary Breast and Ovarian Cancer. 2022;22(2):239-246.
111. Shrestha KS, Aska EM, Tuominen MM, Kauppi L. Tissue-specific reduction in MLH1 expression induces microsatellite instability in intestine of Mlh1(+/-) mice. DNA repair. 2021;106:103178.
112. Underhill ML, Germansky KA, Yurgelun MB. Advances in Hereditary Colorectal and Pancreatic Cancers. Clinical therapeutics. 2016;38(7):1600-1621.
113. Vasen HF, Blanco I, Aktan-Collan K, et al. Revised guidelines for the clinical management of Lynch syndrome (HNPCC): recommendations by a group of European experts. Gut. 2013;62(6):812-823.
114. Fakheri H, Bari Z, Merat S. Familial aspects of colorectal cancers in southern littoral of Caspian Sea. Archives of Iranian medicine. 2011;14(3):175-178.
115. Desvignes JP, Bartoli M, Delague V, et al. VarAFT: a variant annotation and filtration system for human next generation sequencing data. Nucleic acids research. 2018;46(W1):W545-w553.
116. Thorvaldsdóttir H, Robinson JT, Mesirov JP. Integrative Genomics Viewer (IGV): high-performance genomics data visualization and exploration. Briefings in bioinformatics. 2013;14(2):178-192.
117. Falk N, Maire N, Sama W, et al. Comparison of PCR-RFLP and Genescan-based genotyping for analyzing infection dynamics of Plasmodium falciparum. The American journal of tropical medicine and hygiene. 2006;74(6):944-950.
118. Jurinke C, van den Boom D, Cantor CR, Köster H. Automated genotyping using the DNA MassArray technology. Methods in molecular biology (Clifton, NJ). 2002;187:179-192.
119. Yokota K, Nozu K, Minamikawa S, et al. Female X-linked Alport syndrome with somatic mosaicism. Clinical and experimental nephrology. 2017;21(5):877-883.
120. Boudewijns M, van Dongen JJ, Langerak AW. The human androgen receptor X-chromosome inactivation assay for clonality diagnostics of natural killer cell proliferations. The Journal of molecular diagnostics : JMD. 2007;9(3):337-344.
121. Janczar S, Babol-Pokora K, Jatczak-Pawlik I, et al. Six molecular patterns leading to hemophilia A phenotype in 18 females from Poland. Thrombosis research. 2020;193:9-14.
122. Danecek P, Bonfield JK, Liddle J, et al. Twelve years of SAMtools and BCFtools. GigaScience. 2021;10(2).
123. Derrien T, Estellé J, Marco Sola S, et al. Fast computation and applications of genome mappability. 2012;7(1):e30377.
124. score. EM. Accessed 2 Dec 2018.; http://hgdownload.cse.ucsc.edu/goldenPath/hg19/encodeDCC/wgEncodeMapability/wgEncodeDukeMapabilityUniqueness35bp.bigWig.
125. Sina M. How to apply a mappability filter to exclude exons that have difficulty in mapping short-read. [https://www.linkedin.com/posts/mohammad-sina-02b8b588_mappability-ngs-cnv-activity-7012899533726605312-DPe-/?utm_source=share&utm_medium=member_desktop].
126. Caetano-Anolles D. (How to) Call common and rare germline copy number variants. March 21, 2023; https://gatk.broadinstitute.org/hc/en-us/articles/360035531152--How-to-Call-common-and-rare-germline-copy-number-variants.

127. Sina M. How to Call common and rare germline copy number variants using GATK gCNV in cohort mode. 2023.
128. Babadi M, Fu JM, Lee SK, et al. GATK-gCNV: A Rare Copy Number Variant Discovery Algorithm and Its Application to Exome Sequencing in the UK Biobank. 2022.
129. Sina M. Identification of samples with MAK-Alu insertions. 2023; https://www.youtube.com/watch?v=Qsl5cluptFI.
130. Kim D, Paggi JM, Park C, Bennett C, Salzberg SL. Graph-based genome alignment and genotyping with HISAT2 and HISAT-genotype. Nature biotechnology. 2019;37(8):907-915.
131. Wang K, Li M, Hakonarson H. ANNOVAR: functional annotation of genetic variants from high-throughput sequencing data. Nucleic acids research. 2010;38(16):e164.
132. Kopanos C, Tsiolkas V, Kouris A, et al. VarSome: the human genomic variant search engine. Bioinformatics (Oxford, England). 2019;35(11):1978-1980.
133. Richards S, Aziz N, Bale S, et al. 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 : official journal of the American College of Medical Genetics. 2015;17(5):405-424.
134. Plagnol V, Curtis J, Epstein M, et al. A robust model for read count data in exome sequencing experiments and implications for copy number variant calling. Bioinformatics (Oxford, England). 2012;28(21):2747-2754.
135. Li J, Lupat R, Amarasinghe KC, et al. CONTRA: copy number analysis for targeted resequencing. Bioinformatics (Oxford, England). 2012;28(10):1307-1313.
136. Robinson JT, Thorvaldsdóttir H, Winckler W, et al. Integrative genomics viewer. Nature biotechnology. 2011;29(1):24-26.
137. Geoffroy V, Herenger Y, Kress A, et al. AnnotSV: an integrated tool for structural variations annotation. Bioinformatics (Oxford, England). 2018;34(20):3572-3574.
138. Bianchi F, Galizia E, Catalani R, et al. CAT25 is a mononucleotide marker to identify HNPCC patients. The Journal of molecular diagnostics : JMD. 2009;11(3):248-252.
139. Deschoolmeester V, Baay M, Wuyts W, et al. Detection of microsatellite instability in colorectal cancer using an alternative multiplex assay of quasi-monomorphic mononucleotide markers. The Journal of molecular diagnostics : JMD. 2008;10(2):154-159.
140. Goshayeshi L, Khooiee A, Ghaffarzadegan K, et al. Screening for Lynch Syndrome in Cases with Colorectal Carcinoma from Mashhad. Archives of Iranian medicine. 2017;20(6):332-337.
141. Vasen HF, Mecklin JP, Khan PM, Lynch HT. The International Collaborative Group on Hereditary Non-Polyposis Colorectal Cancer (ICG-HNPCC). Diseases of the colon and rectum. 1991;34(5):424-425.
142. Vasen HF, Watson P, Mecklin JP, Lynch HT. New clinical criteria for hereditary nonpolyposis colorectal cancer (HNPCC, Lynch syndrome) proposed by the International Collaborative group on HNPCC. Gastroenterology. 1999;116(6):1453-1456.
143. Gabrielaite M, Torp MH, Rasmussen MS, et al. A Comparison of Tools for Copy-Number Variation Detection in Germline Whole Exome and Whole Genome Sequencing Data. Cancers. 2021;13(24).
144. Talevich E, Shain AH, Botton T, Bastian BC. CNVkit: Genome-Wide Copy Number Detection and Visualization from Targeted DNA Sequencing. PLoS computational biology. 2016;12(4):e1004873.
145. Ponchel F, Toomes C, Bransfield K, et al. Real-time PCR based on SYBR-Green I fluorescence: an alternative to the TaqMan assay for a relative quantification of gene rearrangements, gene amplifications and micro gene deletions. BMC biotechnology. 2003;3:18.
146. Mangold E, Pagenstecher C, Friedl W, et al. Spectrum and frequencies of mutations in MSH2 and MLH1 identified in 1,721 German families suspected of hereditary nonpolyposis colorectal cancer. Int J Cancer. 2005;116(5):692-702.
147. Moussa SA, Moussa A, Kourda N, et al. Lynch syndrome in Tunisia: first description of clinical features and germline mutations. Int J Colorectal Dis. 2011;26(4):455-467.

148. Rahner N, Höefler G, Högenauer C, et al. Compound heterozygosity for two MSH6 mutations in a patient with early onset colorectal cancer, vitiligo and systemic lupus erythematosus. Am J Med Genet A. 2008;146a(10):1314-1319.
149. Plaschke J, Engel C, Krüger S, et al. Lower incidence of colorectal cancer and later age of disease onset in 27 families with pathogenic MSH6 germline mutations compared with families with MLH1 or MSH2 mutations: the German Hereditary Nonpolyposis Colorectal Cancer Consortium. Journal of clinical oncology : official journal of the American Society of Clinical Oncology. 2004;22(22):4486-4494.
150. Guo X, Wu W, Gao H, et al. PMS2 germline mutation c.943C>T (p.Arg315*)-induced Lynch syndrome-associated ovarian cancer. Mol Genet Genomic Med. 2019;7(6):e721.
151. Sugano K, Nakajima T, Sekine S, et al. Germline PMS2 mutation screened by mismatch repair protein immunohistochemistry of colorectal cancer in Japan. Cancer science. 2016;107(11):1677-1686.
152. Syngal S, Fox EA, Eng C, Kolodner RD, Garber JE. Sensitivity and specificity of clinical criteria for hereditary non-polyposis colorectal cancer associated mutations in MSH2 and MLH1. Journal of medical genetics. 2000;37(9):641-645.
153. Migeon BR. X-linked diseases: susceptible females. Genetics in medicine : official journal of the American College of Medical Genetics. 2020;22(7):1156-1174.
154. Migeon BRJJ. The role of X inactivation and cellular mosaicism in women's health and sex-specific diseases. 2006;295(12):1428-1433.
155. Ercan S. Mechanisms of x chromosome dosage compensation. Journal of genomics. 2015;3:1-19.
156. Lyon MFJAjohg. Sex chromatin and gene action in the mammalian X-chromosome. 1962;14(2):135.
157. Migeon BRJTiG. Choosing the active X: the human version of X inactivation. 2017;33(12):899-909.
158. Migeon BR, Beer MA, Bjornsson HTJPO. Embryonic loss of human females with partial trisomy 19 identifies region critical for the single active X. 2017;12(4):e0170403.
159. Monk M, Boubelik M, Lehnert SJD. Temporal and regional changes in DNA methylation in the embryonic, extraembryonic and germ cell lineages during mouse embryo development. 1987;99(3):371-382.
160. Leiding JW, Holland SM. Chronic Granulomatous Disease. In: Adam MP, Mirzaa GM, Pagon RA, et al., eds. GeneReviews(®). Seattle (WA): University of Washington, Seattle
Copyright © 1993-2023, University of Washington, Seattle. GeneReviews is a registered trademark of the University of Washington, Seattle. All rights reserved.; 1993.
161. Beghin A, Comini M, Soresina A, et al. Chronic Granulomatous Disease in children: a single center experience. Clinical immunology (Orlando, Fla). 2018;188:12-19.
162. Eguchi M, Yagi C, Tauchi H, Kobayashi M, Ishii E, Eguchi-Ishimae M. Exon skipping in CYBB mRNA and skewed inactivation of X chromosome cause late-onset chronic granulomatous disease. Pediatric hematology and oncology. 2018;35(5-6):341-349.
163. Shvetsova E, Sofronova A, Monajemi R, et al. Skewed X-inactivation is common in the general female population. European journal of human genetics : EJHG. 2019;27(3):455-465.
164. Wu CY, Chen YC, Lee WI, et al. Clinical Features of Female Taiwanese Carriers with X-linked Chronic Granulomatous Disease from 2004 to 2019. Journal of clinical immunology. 2021;41(6):1303-1314.
165. Wutz AJNRG. Gene silencing in X-chromosome inactivation: advances in understanding facultative heterochromatin formation. 2011;12(8):542-553.
166. Juchniewicz P, Kloska A, Portalska K, et al. X-chromosome inactivation patterns depend on age and tissue but not conception method in humans. Chromosome research : an international journal on the molecular, supramolecular and evolutionary aspects of chromosome biology. 2023;31(1):4.
167. Brown C, Robinson WJCg. The causes and consequences of random and non-random X chromosome inactivation in humans. 2000;58(5):353-363.

168. Butler MG, Manzardo AM. Androgen receptor (AR) gene CAG trinucleotide repeat length associated with body composition measures in non-syndromic obese, non-obese and Prader-Willi syndrome individuals. Journal of assisted reproduction and genetics. 2015;32(6):909-915.
169. Mochizuki H, Goto-Koshino Y, Takahashi M, Fujino Y, Ohno K, Tsujimoto H. Demonstration of the cell clonality in canine hematopoietic tumors by X-chromosome inactivation pattern analysis. Veterinary pathology. 2015;52(1):61-69.
170. Comertpay S, Pastorino S, Tanji M, et al. Evaluation of clonal origin of malignant mesothelioma. Journal of translational medicine. 2014;12:301.
171. Li F-Y, Chaigne-Delalande B, Su H, Uzel G, Matthews H, Lenardo MJJB, The Journal of the American Society of Hematology. XMEN disease: a new primary immunodeficiency affecting Mg2+ regulation of immunity against Epstein-Barr virus. 2014;123(14):2148-2152.
172. Taylor GS, Long HM, Brooks JM, Rickinson AB, Hislop ADJAroi. The immunology of Epstein-Barr virus–induced disease. 2015;33:787-821.
173. Watson CM, Nadat F, Ahmed S, et al. Identification of a novel MAGT1 mutation supports a diagnosis of XMEN disease. Genes and immunity. 2022;23(2):66-72.
174. Ravell JC, Chauvin SD, He T, Lenardo MJJoci. An update on XMEN disease. 2020;40:671-681.
175. Jian X, Boerwinkle E, Liu XJGiM. In silico tools for splicing defect prediction: a survey from the viewpoint of end users. 2014;16(7):497-503.
176. Pertea M, Lin X, Salzberg SLJNar. GeneSplicer: a new computational method for splice site prediction. 2001;29(5):1185-1190.
177. Desmet F-O, Hamroun D, Lalande M, Collod-Béroud G, Claustres M, Béroud CJNar. Human Splicing Finder: an online bioinformatics tool to predict splicing signals. 2009;37(9):e67-e67.
178. Yeo G, Burge CB. Maximum entropy modeling of short sequence motifs with applications to RNA splicing signals. Paper presented at: Proceedings of the seventh annual international conference on Research in computational molecular biology2003.
179. Brunak S, Engelbrecht J, Knudsen SJJomb. Prediction of human mRNA donor and acceptor sites from the DNA sequence. 1991;220(1):49-65.
180. Reese MG, Eeckman FH, Kulp D, Haussler D. Improved splice site detection in Genie. Paper presented at: Proceedings of the first annual international conference on Computational molecular biology1997.
181. Xiong F, Gao J, Li J, et al. Noncanonical and canonical splice sites: a novel mutation at the rare noncanonical splice-donor cut site (IVS4+ 1A> G) of SEDL causes variable splicing isoforms in X-linked spondyloepiphyseal dysplasia tarda. 2009;17(4):510-516.
182. Medicine ALQACJGi. Erratum to: ACMG technical standards and guidelines for genetic testing for inherited colorectal cancer (Lynch syndrome, familial adenomatous polyposis, and MYH-associated polyposis)(Genetics in Medicine,(2014), 16, 1,(101-116), 10.1038/gim. 2013.166). 2021;23(10):1807-1817.
183. Houdayer C, Caux-Moncoutier V, Krieger S, et al. Guidelines for splicing analysis in molecular diagnosis derived from a set of 327 combined in silico/in vitro studies on BRCA1 and BRCA2 variants. 2012;33(8):1228-1238.
184. Cariola F, Disciglio V, Valentini AM, et al. Characterization of a rare variant (c. 2635-2A> G) of the MSH2 gene in a family with Lynch syndrome. 2018;33(4):534-539.
185. Chiu FP, Doolan BJ, McGrath JA, Onoufriadis A. A decade of next-generation sequencing in genodermatoses: the impact on gene discovery and clinical diagnostics. The British journal of dermatology. 2021;184(4):606-616.
186. Cifaldi C, Brigida I, Barzaghi F, et al. Targeted NGS Platforms for Genetic Screening and Gene Discovery in Primary Immunodeficiencies. Frontiers in immunology. 2019;10:316.
187. Gordeeva V, Sharova E, Babalyan K, Sultanov R, Govorun VM, Arapidi GJSR. Benchmarking germline CNV calling tools from exome sequencing data. 2021;11(1):1-11.
188. Benjamini Y, Speed TP. Summarizing and correcting the GC content bias in high-throughput sequencing. Nucleic acids research. 2012;40(10):e72.

189. Treangen TJ, Salzberg SLJNRG. Repetitive DNA and next-generation sequencing: computational challenges and solutions. 2012;13(1):36-46.
190. Backenroth D, Homsy J, Murillo LR, et al. CANOES: detecting rare copy number variants from whole exome sequencing data. 2014;42(12):e97-e97.
191. Jiang Y, Oldridge DA, Diskin SJ, Zhang NR. CODEX: a normalization and copy number variation detection method for whole exome sequencing. Nucleic acids research. 2015;43(6):e39.
192. Fowler A, Mahamdallie S, Ruark E, et al. Accurate clinical detection of exon copy number variants in a targeted NGS panel using DECoN. 2016;1.
193. Packer JS, Maxwell EK, O'Dushlaine C, et al. CLAMMS: a scalable algorithm for calling common and rare copy number variants from exome sequencing data. 2016;32(1):133-135.
194. Johansson LF, van Dijk F, de Boer EN, et al. CoNVaDING: single exon variation detection in targeted NGS data. 2016;37(5):457-464.
195. Li J, Lupat R, Amarasinghe KC, et al. CONTRA: copy number analysis for targeted resequencing. 2012;28(10):1307-1313.
196. Ruderfer DM, Hamamsy T, Lek M, et al. Patterns of genic intolerance of rare copy number variation in 59,898 human exomes. 2016;48(10):1107-1111.
197. Roca I, González-Castro L, Fernández H, Couce ML, Fernández-Marmiesse AJMRRiMR. Free-access copy-number variant detection tools for targeted next-generation sequencing data. 2019;779:114-125.
198. Moreno-Cabrera JM, Del Valle J, Castellanos E, et al. Evaluation of CNV detection tools for NGS panel data in genetic diagnostics. 2020;28(12):1645-1655.
199. Mason-Suares H, Landry L, S Lebo MJCGMR. Detecting copy number variation via next generation technology. 2016;4:74-85.
200. Miller N, Bouma M, Sabatini L, Gulukota K. SILO: A Computational Method for Detecting Copy Number Gain in Clinical Specimens Analyzed on a Next-Generation Sequencing Platform. The Journal of molecular diagnostics : JMD. 2021;23(10):1241-1248.
201. Millat G, Chanavat V, Rousson RJCca. Evaluation of a new NGS method based on a custom AmpliSeq library and Ion Torrent PGM sequencing for the fast detection of genetic variations in cardiomyopathies. 2014;433:266-271.
202. Nishio Sy, Moteki H, Usami SiJMg, medicine g. Simple and efficient germline copy number variant visualization method for the Ion AmpliSeq™ custom panel. 2018;6(4):678-686.
203. Yao R, Zhang C, Yu T, et al. Evaluation of three read-depth based CNV detection tools using whole-exome sequencing data. Molecular cytogenetics. 2017;10:30.
204. Nishio SY, Moteki H, Usami SI. Simple and efficient germline copy number variant visualization method for the Ion AmpliSeq™ custom panel. Molecular genetics & genomic medicine. 2018;6(4):678-686.
205. Fromer M, Moran JL, Chambert K, et al. Discovery and statistical genotyping of copy-number variation from whole-exome sequencing depth. American journal of human genetics. 2012;91(4):597-607.
206. Stray-Pedersen A, Sorte HS, Samarakoon P, et al. Primary immunodeficiency diseases: genomic approaches delineate heterogeneous Mendelian disorders. 2017;139(1):232-245.
207. Bigio B, Seeleuthner Y, Kerner G, et al. Detection of homozygous and hemizygous partial exon deletions by whole-exome sequencing. 2020:2020.2007. 2023.217976.
208. Zhang Z, Cheng H, Hong X, et al. EnsembleCNV: an ensemble machine learning algorithm to identify and genotype copy number variation using SNP array data. Nucleic acids research. 2019;47(7):e39.
209. Bigio B, Seeleuthner Y, Kerner G, et al. Detection of homozygous and hemizygous complete or partial exon deletions by whole-exome sequencing. NAR genomics and bioinformatics. 2021;3(2):lqab037.
210. Babadi M, Fu JM, Lee SK, et al. GATK-gCNV: A Rare Copy Number Variant Discovery Algorithm and Its Application to Exome Sequencing in the UK Biobank. 2022:2022.2008. 2025.504851.

211. Hackmann K, Kuhlee F, Betcheva-Krajcir E, et al. Ready to clone: CNV detection and breakpoint fine-mapping in breast and ovarian cancer susceptibility genes by high-resolution array CGH. 2016;159:585-590.
212. Oostlander AE, Meijer G, Ylstra BJCg. Microarray-based comparative genomic hybridization and its applications in human genetics. 2004;66(6):488-495.
213. Schouten JP, McElgunn CJ, Waaijer R, Zwijnenburg D, Diepvens F, Pals GJNar. Relative quantification of 40 nucleic acid sequences by multiplex ligation-dependent probe amplification. 2002;30(12):e57-e57.
214. Concolino P, Capoluongo EJEromd. Detection of BRCA1/2 large genomic rearrangements in breast and ovarian cancer patients: an overview of the current methods. 2019;19(9):795-802.
215. Huddleston J, Chaisson MJ, Steinberg KM, et al. Discovery and genotyping of structural variation from long-read haploid genome sequence data. 2017;27(5):677-685.
216. Tanakaya K. Current clinical topics of Lynch syndrome. Int J Clin Oncol. 2019;24(9):1013-1019.
217. Sina M, Ghorbanoghli Z, Abedrabbo A, et al. Identification and management of Lynch syndrome in the Middle East and North African countries: outcome of a survey in 12 countries. Familial cancer. 2020.
218. Li X, Wu Y, Suo P, et al. Identification of a novel germline frameshift mutation p.D300fs of PMS1 in a patient with hepatocellular carcinoma: A case report and literature review. Medicine (Baltimore). 2020;99(5):e19076-e19076.
219. Wang Q, Lasset C, Desseigne F, et al. Prevalence of germline mutations of hMLH1, hMSH2, hPMS1, hPMS2, and hMSH6 genes in 75 French kindreds with nonpolyposis colorectal cancer. Hum Genet. 1999;105(1-2):79-85.
220. Maehara Y, Egashira A, Oki E, Kakeji Y, Tsuzuki T. DNA repair dysfunction in gastrointestinal tract cancers. Cancer Sci. 2008;99(3):451-458.
221. Liu T, Yan H, Kuismanen S, et al. The role of hPMS1 and hPMS2 in predisposing to colorectal cancer. Cancer research. 2001;61(21):7798-7802.
222. Cancer. ACSCfafSSO. Special Section: Ovarian Cancer. (Accessed 30 December 2018). https://www.cancer.org/content/dam/cancer-org/research/cancer-facts-and-statistics/annual-cancer-facts-and-figures/2018/cancer-facts-and-figures-special-section-ovarian-cancer-2018.pdf.
223. Landry KK, Seward DJ, Dragon JA, et al. Investigation of discordant sibling pairs from hereditary breast cancer families and analysis of a rare PMS1 variant. Cancer genetics. 2022;260-261:30-36.
224. Hafez Fakheri M, Zohre Bari M, Shahin Merat M. Familial aspects of colorectal cancers in southern littoral of Caspian Sea. Archives of Iranian medicine. 2011;14(3):175.
225. Ghaedi H, Ramsheh SM, Omidvar ME, et al. Whole-exome sequencing identified a novel mutation of MLH1 in an extended family with lynch syndrome. Genes & Diseases. 2019.
226. Samir Gupta JMW, Lisen Axell, Lee-May Chen, Daniel C. Chung, Katherine M. Clayback. Genetic/familial high-risk assessment: colorectal version 1.2022, NCCN Clinical Practice Guidelines in Oncology (NCCN Guidelines®). In:2022:1-154.
227. Ladan Goshayeshi AE, Kambiz Akhavan Rezayat, Hooman Moosanen Mozaffari,Omid Ghanaee, Benyamin Hosein FAMILIAL COLORECTALREGISTRY IN NORTHEAST OF IRAN. Acta HealthMedica. 2017;2(2):172.
228. Dominguez-Valentin M, Sampson JR, Seppälä TT, et al. Cancer risks by gene, age, and gender in 6350 carriers of pathogenic mismatch repair variants: findings from the Prospective Lynch Syndrome Database. Genetics in medicine : official journal of the American College of Medical Genetics. 2020;22(1):15-25.
229. Plaschke J, Linnebacher M, Kloor M, et al. Compound heterozygosity for two MSH6 mutations in a patient with early onset of HNPCC-associated cancers, but without hematological malignancy and brain tumor. European journal of human genetics : EJHG. 2006;14(5):561-566.

230. Okkels H, Sunde L, Lindorff-Larsen K, et al. Polyposis and early cancer in a patient with low penetrant mutations in MSH6 and APC: hereditary colorectal cancer as a polygenic trait. Int J Colorectal Dis. 2006;21(8):847-850.
231. Jasperson KW, Samowitz WS, Burt RW. Constitutional mismatch repair-deficiency syndrome presenting as colonic adenomatous polyposis: clues from the skin. Clinical genetics. 2011;80(4):394-397.
232. Shia J, Klimstra DS, Nafa K, et al. Value of immunohistochemical detection of DNA mismatch repair proteins in predicting germline mutation in hereditary colorectal neoplasms. Am J Surg Pathol. 2005;29(1):96-104.
233. Kets CM, Hoogerbrugge N, van Krieken JH, Goossens M, Brunner HG, Ligtenberg MJ. Compound heterozygosity for two MSH2 mutations suggests mild consequences of the initiation codon variant c.1A>G of MSH2. European journal of human genetics : EJHG. 2009;17(2):159-164.
234. Curtius K, Gupta S, Boland CR. Review article: Lynch Syndrome-a mechanistic and clinical management update. Alimentary pharmacology & therapeutics. 2022;55(8):960-977.
235. Shrestha KS, Aska E-M, Tuominen MM, Kauppi LJDr. Tissue-specific reduction in MLH1 expression induces microsatellite instability in intestine of Mlh1+/− mice. 2021;106:103178.
236. Vaughn CP, Robles J, Swensen JJ, et al. Clinical analysis of PMS2: mutation detection and avoidance of pseudogenes. Human mutation. 2010;31(5):588-593.
237. Borràs E, Pineda M, Cadiñanos J, et al. Refining the role of PMS2 in Lynch syndrome: germline mutational analysis improved by comprehensive assessment of variants. Journal of medical genetics. 2013;50(8):552-563.
238. Suwinski P, Ong C, Ling MHT, Poh YM, Khan AM, Ong HS. Advancing Personalized Medicine Through the Application of Whole Exome Sequencing and Big Data Analytics. Frontiers in genetics. 2019;10:49.
239. Skopelitou D, Srivastava A, Miao B, et al. Whole exome sequencing identifies novel germline variants of SLC15A4 gene as potentially cancer predisposing in familial colorectal cancer. Molecular genetics and genomics : MGG. 2022;297(4):965-979.
240. Pathak SJ, Mueller JL, Okamoto K, et al. EPCAM mutation update: Variants associated with congenital tufting enteropathy and Lynch syndrome. Human mutation. 2019;40(2):142-161.
241. Ligtenberg MJ, Kuiper RP, Geurts van Kessel A, Hoogerbrugge NJFc. EPCAM deletion carriers constitute a unique subgroup of Lynch syndrome patients. 2013;12:169-174.
242. Sobocińska J, Kolenda T, Teresiak A, et al. Diagnostics of Mutations in MMR/EPCAM Genes and Their Role in the Treatment and Care of Patients with Lynch Syndrome. Diagnostics (Basel, Switzerland). 2020;10(10).
243. Stefka J, Streff H, Liu P, Towne M, Smith HS. Cascade testing after exome sequencing: Retrospective analysis of linked family data at 2 US laboratories. Genetics in medicine : official journal of the American College of Medical Genetics. 2023;25(5):100818.
244. Tafe LJ. Targeted Next-Generation Sequencing for Hereditary Cancer Syndromes: A Focus on Lynch Syndrome and Associated Endometrial Cancer. The Journal of molecular diagnostics : JMD. 2015;17(5):472-482.
245. Li Y, Fan L, Zheng J, et al. Lynch syndrome pre-screening and comprehensive characterization in a multi-center large cohort of Chinese patients with colorectal cancer. Cancer biology & medicine. 2022;19(8):1235-1248.
246. Kayser K, Degenhardt F, Holzapfel S, et al. Copy number variation analysis and targeted NGS in 77 families with suspected Lynch syndrome reveals novel potential causative genes. International journal of cancer. 2018;143(11):2800-2813.
247. Hegde M, Ferber M, Mao R, Samowitz W, Ganguly A. ACMG technical standards and guidelines for genetic testing for inherited colorectal cancer (Lynch syndrome, familial adenomatous polyposis, and MYH-associated polyposis). Genetics in medicine : official journal of the American College of Medical Genetics. 2014;16(1):101-116.
248. Li S, Qian D, Thompson BA, et al. Tumour characteristics provide evidence for germline mismatch repair missense variant pathogenicity. Journal of medical genetics. 2020;57(1):62-69.

249. van der Klift HM, Mensenkamp AR, Drost M, et al. Comprehensive Mutation Analysis of PMS2 in a Large Cohort of Probands Suspected of Lynch Syndrome or Constitutional Mismatch Repair Deficiency Syndrome. Human mutation. 2016;37(11):1162-1179.
250. Snowsill T, Coelho H, Huxley N, et al. Molecular testing for Lynch syndrome in people with colorectal cancer: systematic reviews and economic evaluation. 2017;21(51):1-280.
251. Perez-Cabornero L, Velasco E, Infante M, et al. A new strategy to screen MMR genes in Lynch Syndrome: HA-CAE, MLPA and RT-PCR. 2009;45(8):1485-1493.
252. Zampaglione E, Kinde B, Place EM, et al. Copy-number variation contributes 9% of pathogenicity in the inherited retinal degenerations. 2020;22(6):1079-1087.
253. Saadat M, Ansari-Lari M, Farhud DD. Consanguineous marriage in Iran. Annals of human biology. 2004;31(2):263-269.