ACIBADEM MEHMET ALI AYDINLAR UNIVERSITY
INSTITUTE OF HEALTH SCIENCES

INVESTIGATION OF GENETIC PREDISPOSITION IN CHILDHOOD TUMORS BY NEXT-GENERATION SEQUENCING-BASED TECHNOLOGIES

UFUK AMANVERMEZ

M.Sc. THESIS

DEPARTMENT OF GENOME STUDIES

SUPERVISOR
Prof. Özden Hatırnaz Ng

SECONDARY SUPERVISOR
Assoc. Prof. Nihat Buğra Ağaoğlu

ISTANBUL-2024

ACIBADEM MEHMET ALI AYDINLAR UNIVERSITY
INSTITUTE OF HEALTH SCIENCES

INVESTIGATION OF GENETIC PREDISPOSITION IN CHILDHOOD TUMORS BY NEXT-GENERATION SEQUENCING -BASED TECHNOLOGIES

UFUK AMANVERMEZ

M.Sc. THESIS

DEPARTMENT OF GENOME STUDIES

SUPERVISOR
Prof. Özden Hatırnaz Ng

SECONDARY SUPERVISOR
Assoc. Prof. Nihat Buğra Ağaoğlu

ISTANBUL-2024

DECLARATION

I declare that this thesis work is my own work, I had no unethical behavior at any stages from the planning to the writing of the thesis, I obtained all the information in this thesis in accordance with academic and ethical rules, I cited all the information and comments that were not obtained with this thesis work, and I provided resources in the list of references. I also declare that there was no violation of any patents and copyrights during the study and writing of this thesis.

24/06/2024

Ufuk Amanvermez

PREFACE AND ACKNOWLEDGEMENT

I would like to thank my dearest thesis advisor Prof. Özden Hatırnaz Ng who is patient, positive and energized all the time. Her positive attitude provided me a healthy and joyful learning environment. Even though I tested her limits time to time, she was always so kind and rational. She always treated me as a friend, not like a student. I feel so lucky for having the chance to work someone like you. Thanks to her knowledge and advice, I was able to move forward in my master's journey. I will always be grateful to you!

Dr. Nihat Buğra Ağaoğlu, my dear secondary advisor, has always been supportive to me. I always learned something new whenever I talked to him. He was always informative and supportive for this thesis, especially providing insights as a clinician. I feel so lucky for finding multiple chances to work with you in various studies.

I thank Prof. Dr. Uğur Özbek, Assist. Prof. Özkan Özdemir, Prof. Dr. Cengiz Canpolat and Yeşim Kızılkaya Ören who helped me throughout this thesis by their extraordinary knowledge

I also would like to thank my friends; Gizem Önder, İlayda Şahin, Alper Akkuş, Eylül Aydın, Aybike Bulut and Beyza Gürler who made my lab days informative and joyful day by day. They have been always positive to me even if I test their limits. I feel so lucky to have your friendship.

I also want to thank a special person in my life: Talat Taygun Turan has supported me throughout this thesis along with various other issues in my life. You will always have a special place for me. I feel so lucky to have you in my life, I hope it will last!

Lastly, I want to thank my parents and sisters who brought me to this point. They provided me lots of opportunities for my self-improvement.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

ABSTRACT

Investigation of Genetic Predisposition in Childhood Tumors by Next-Generation Sequencing-Based Technologies

Hereditary cancers are accounted for 5-10% of the overall cancer cancers. The ratio of the hereditary cancers increases to 10-15% in pediatric cancer cases. Due to the higher hereditary cancer risk in childhood cancer cases, in this thesis, it was aimed to identify cancer predisposition variants in pediatric cancer cases with positive family history of cancer based on Jongman's criteria. Here we examined ten different families with pediatric index cases: Wilms' tumor (n=2), B-cell acute lymphoblastic leukemia (n=2), T-cell acute lymphoblastic leukemia (n=1), lymphoma (n=1), glioneuronal tumor (n=1), medulloblastoma (n=1), glioblastoma (n=1), and pilocytic astrocytoma (n=1) with next-generation sequencing technologies. Clinical exome sequencing was performed for each index case, and after the data processing and filtering procedures, clinically relevant 23 candidate variants in ten patients were identified. These variants were validated via Sanger sequencing. Furthermore, Sanger sequencing was also performed for both affected and unaffected family members for segregation analysis. Based on the results, we detected 15 out of 23 candidate variants that are clinically relevant. Eight of these variants were classified as likely pathogenic or pathogenic, whereas seven of them were variant of unknown significance based on American College of Medical Genetics guidelines. The outcomes of this thesis include cancer predisposition awareness in physicians, uncovering the novel variants in most of the families, providing genetic counselling to the families, highlighting the importance and power of the clinical exome sequencing.

Keywords: Familial cancer, Childhood cancers, Cancer predisposition, Next-generation sequencing, Clinical exome sequencing.

ÖZET

Çocukluk Çağı Tümörlerinde Genetik Yatkınlığın Yeni Nesil Dizilemeye Dayalı Teknolojilerle Araştırılması

Kalıtsal kanserler, tüm kanser kanserlerinin %5-10'unu oluşturur. Çocukluk çağı kanserlerindeyse bu oran %10-15'e çıkmaktadır. Bu tezde, çocuk olgularda kalıtsal kanser riskinin daha yüksek olması nedeniyle, Jongman kriterlerine göre ailesinde pozitif kanser öyküsü olan çocukluk çağı kanser olgularında kansere yatkınlık varyantlarının belirlenmesi amaçlanmıştır. Bu amaca yönelik olarak, Wilms tümörü (n=2), B hücreli akut lenfoblastik lösemi (n=2), T hücreli akut lenfoblastik lösemi (n=1), lenfoma (n=1), glionöronal tümör (n=1), medulloblastoma (n=1), glioblastoma (n=1) ve pilositik astrositom (n=1) tanısı almış toplam 10 çocukluk çağı kanser vakası, yeni nesil dizileme teknolojileriyle değerlendirilmiştir. Her indeks vaka için klinik ekzom dizilimi yapılmış, veri işleme ve filtreleme prosedürlerinden sonra on hastada 23 aday varyant tanımlanmıştır. Bu varyantlar Sanger dizilimi yoluyla değerlendirilmiş ve hem etkilenen hem de etkilenmeyen aile üyelerinin segregasyon analizleri yapılmıştır. Sonuçlara dayanarak klinik açıdan anlamlı 23 aday varyanttan 13'ü klinik ile ilişkilendirilmiş olup bu varyantlardan sekizi muhtemel patojenik veya patojenik olarak sınıflandırılırken, yedisi American College of Medical Genetics kriterlerine göre önemi bilinmeyen varyant olarak sınıflandırılmıştır. Bu tez sayesinde doktorlarda kanser yatkınlığı konusunda farkındalık yaratılması, ailelerin çoğunda literatürde tanımlanmamış varyantların ortaya çıkarılması, ailelere genetik danışmanlık sağlanması, klinik ekzom dizilemenin kanser tanısında, özellikle tüm ekzom/tüm genom seçeneklerinin mümkün olmadığı bölgelerde kullanılabilirliği gösterilmiştir.

Anahtar Sözcükler: Ailesel kanser, Çocukluk çağı kanserleri, Kanser yatkınlığı, Yeni nesil dizileme, Klinik ekzom dizilimi

1 INTRODUCTION

Syndromes known as hereditary cancer syndromes are caused by the inheritance of germline variations in one or more genes, which predisposes a person to particular types of cancer. As of right now, over 50 hereditary cancer disorders have been found. When compared to healthy persons, those with variations that induce cancer predisposition have an increased risk of acquiring cancer. These cancer syndromes account for 5–10% of all cancer cases and are mostly characterized by autosomal dominant inheritance patterns(1). Furthermore, in cases of childhood hereditary malignancies, this rate rises to 10-15% (2, 3).

NGS is considered a second-generation sequencing technology that allows to sequence multiple samples simultaneously with a high number depth. In this way, disease-related conditions can be identified. In addition to whole genome/exome sequencing, NGS also enables targeted DNA sequencing. Targeted DNA sequencing is essential for both research and diagnosis. Clinical Exome Sequencing (CES) is an NGS panel that can analyze more than 3000 clinically relevant genes. (4) Clinical exome sequencing makes it possible to focus and select known genes of clinical importance, rather than using much larger datasets such as genomes or exomes. Therefore, CES can be used to detect variants in genes of interest and to identify variants in other clinically important genes. These advances in NGS have accelerated, as has the application of this technology to cancer. Simultaneous sequencing of cancer-related genes can identify specific cancer-causing variants. (5, 6)

The field of medical genetics has made considerable use of the five major variant categories—pathogenic, likely pathogenic, unknown importance, likely benign, and benign—that were published by the American College of Medical Genetics and Genomics (ACMG). (7) These groups are separated according to 28 criteria and the clinical relevance of the variation. There are sixteen pathogenic and twelve benign criteria for these categories. The benign criteria are split into stand-alone (BA1), strong (BS1–4), and supporting (BP1–6), whereas the pathogenic criteria are classified into veritably strong (PVS1), strong (PS1–4), moderate (PM1–6), and supporting (PP1–5)

depending on importance. The stated criteria are mostly derived from a variety of sources, including functional, population-based, computational, and predictive, segregation, and allelic data, in addition to other data from reliable sources. Scientists can categorize the variations according to their clinical significance using the parameters that ACMG published.

Here in this thesis, it was aimed to identify genes and variants that cause cancer predisposition in the families with positive cancer history. Based on the aim of the thesis, the index patients were studied with CES and later filtered based on the phenotype and pedigree. Later, candidate variants were confirmed by Sanger sequencing. After identification of the candidate variants, it was aimed to supply the genetic counselling needed.

2 BACKGROUND

2.1 Genome

Genome is the total data that is encoded into the genetic material in all the organisms. It is consisting of codes that are created by nucleotides. In most of the organisms, genetic code, or genome, is created as a double-stranded structure called deoxyribonucleic acid (DNA). DNA is consisting of nucleotides; and nucleotides are consisting of one of four different bases (adenine, thymine, guanine, cytosine), a phosphate group, and deoxyribose sugar. Later, nucleotides are bonded together with phosphate groups which end up in single strand of DNA. However, DNA is a double-stranded structure. Each nucleotide matches up the interested nucleotide (Adenines match with thymine, guanines match with cytosine). These matching nucleotides are paired with hydrogen bonds. However, DNA is not represented in cells as double-stranded form; DNA is wrapped around the proteins called histones. Histones are octameric structures that allow DNA to fit into the nucleus. After DNA is wrapped around histones; two different structures occur which are euchromatin and heterochromatin. Euchromatin is a region that is not dense and allow transcription. On the other hand, heterochromatin is a dense region that does not allow transcription (Figure 2.1).

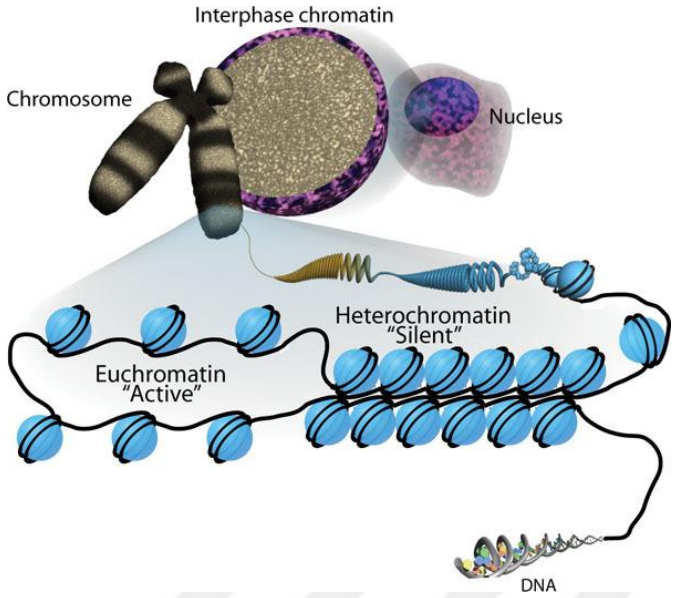

Figure 2.1. Structures of euchromatin and heterochromatin (8)

Lastly, chromosomes are being formed after the entire genome is wrapped around the histones. Genomic structure of organisms is represented in Figure 2.2.

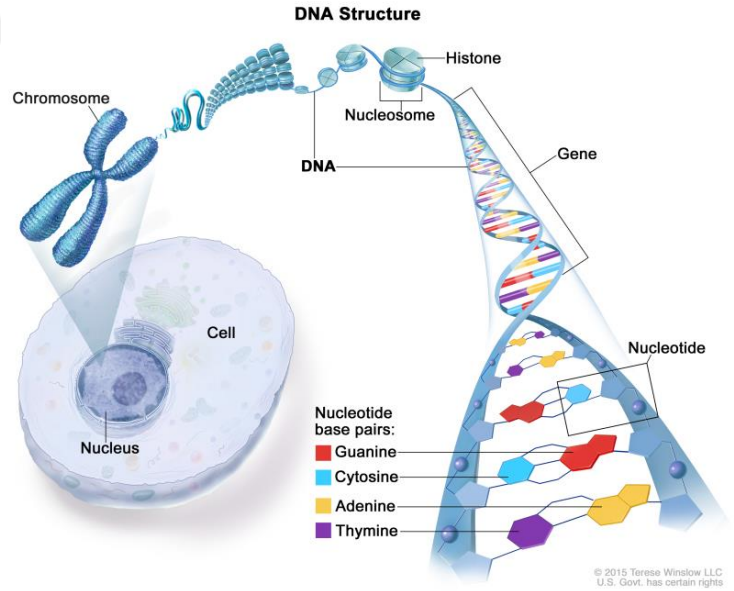

Figure 2.2. Structure of genome (9)

2.1.1 Human genome

The human genome is made up by 3 billion nucleotides in which encode the essential data for the humans. This crucial code is encoded into a macromolecule, DNA. This double-helix macromolecule is consisting of two group of nucleotides called pyrimidines (Cytosine, thymine, uracil) and purines (Guanine, adenine). Within 3 billion nucleotides, there are approximately 20,000 protein-coding genes that is required for the functioning of the organism. (10) These genes are basically located on chromosomes which are made up by dense DNA wrapped around the histones and located in the nucleus of the cells. The human chromosome set is consisting of 46 chromosomes: 22 pair of autosomes and XX (female) or XY (male). Each pair of chromosomes are called homologue chromosomes, and one comes from the mother, one comes from the father (Figure 2.3).

Figure 2.3. Complete set of human chromosomes, karyotype. (11)

2.1.2 Human Genome Project

Human Genome Project (HGP) is a revolutionary scientific initiation that is aimed to sequence and map the entire human genome. The idea was proposed firstly in 1980s and initiated in 1990 by the National Institute of Health (NIH) and US Department of Energy. By the collaboration of the scientists and institutes from different countries, it was primarily aimed to decode the complete DNA sequences and map all the genes. The human genome's initial draft was firstly released in Nature and Science Journals in February 2001 with a completeness of 90%. Two years later, in 2003, it was announced that HGP was completed. Thanks to HGP, understanding the human genome became easier which allow scientists and physicians to perform more precise experiments and diagnoses. (12-15)

2.1.3 Reference genome

A reference genome is a standardized and representative version of a genome that acts as a template or baseline for a specific and interested species. It is produced by combining the DNA sequences obtained from many members of that population. Reference genome enables scientific experiments like searching genetic diversity, identifying genes, and performing comparative genomics analysis and more. However, since reference genomes are created by sequencing the DNA of multiple individual contributors, they are unable to accurately represent the collection of genes of a single individual creature (16). Scientific community advances are causing some modifications to the reference genome sequences. Depending on the magnitude of these developments, either patches for the existing genome or a new version may be created.

2.1.4 Genes

Genes are portions of the DNA that are responsible for providing the template for transcription processes. Genes are regulated by single or multiple promotors along with other regulatory elements such as noncoding RNAs etc. When it comes to genes,

there are two terms: coding regions and non-coding regions. Coding regions are significantly smaller portion of the genome, whereas non-coding regions are the majority of the genome and do not transcribed for the protein synthesis.

2.1.4.1 Coding regions

About 20,000 coding genes make up the human genome, which is consisting of every nucleated cell in the body. As mentioned before, genes that create the hereditary functions and structures within the species are subunits of DNA. Since genes have such important roles in the organisms, any problem in them may cause diseases. As we are talking about the coding genes, genes are being transcribed, later translated into proteins (Figure 2.4).

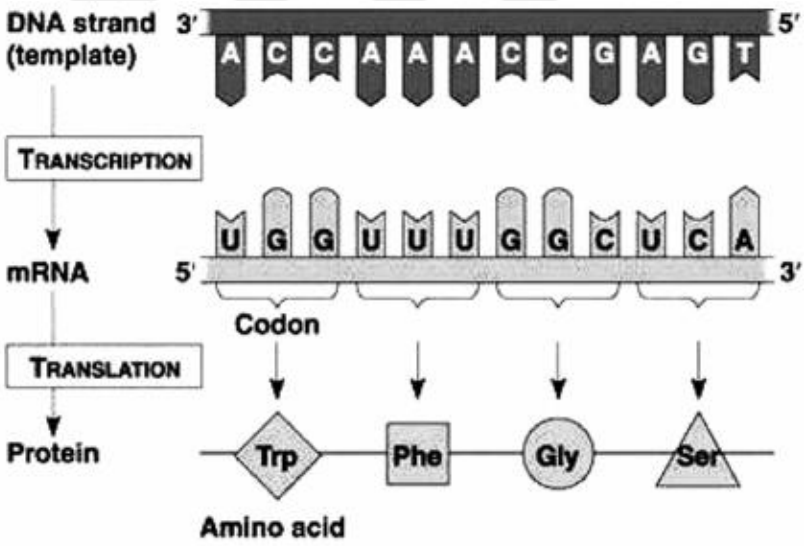

Figure 2.4. Protein synthesis steps

However, via the process called alternative splicing, same gene may create different proteins with different properties (Figure 2.5). Also mentioned before, genes have both coding and non-coding regions called exons and introns, respectively. Coding part of the DNA are being transcribed into mRNA. By the help of the splicing, different mRNAs, therefore, different proteins can be synthesized. (17, 18)

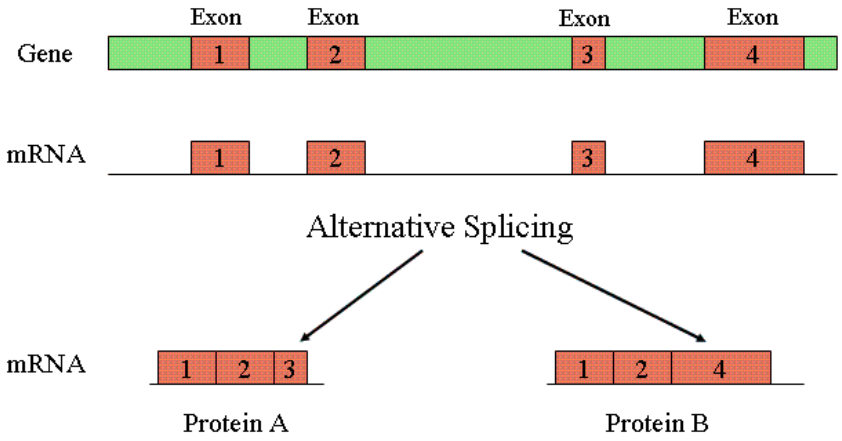

Figure 2.5. Schematic representation of alternative splicing

2.1.4.2 Non-coding regions

Non-coding regions or DNA compromise almost 98.5% of the entire human genome. Even though non-coding regions are not being used for protein synthesis, it has crucial roles regarding the cellular processes and gene regulations. For example, repeated transposable elements induce the genomic rearrangement, therefore, important from the evolutionary perspective. (19) Furthermore, human genome is full of tandemly repeated sequences. Approximately 45% of the tandemly repeated sequences are long terminal repeats (LTRs), long interspersed nuclear elements (LINEs), short interspersed nuclear elements (SINEs). Also, approximately 25% of the tandemly repeated sequences are shorted repeats called satellites, microsatellites, and minisatellites. (20) (Figure 2.6)

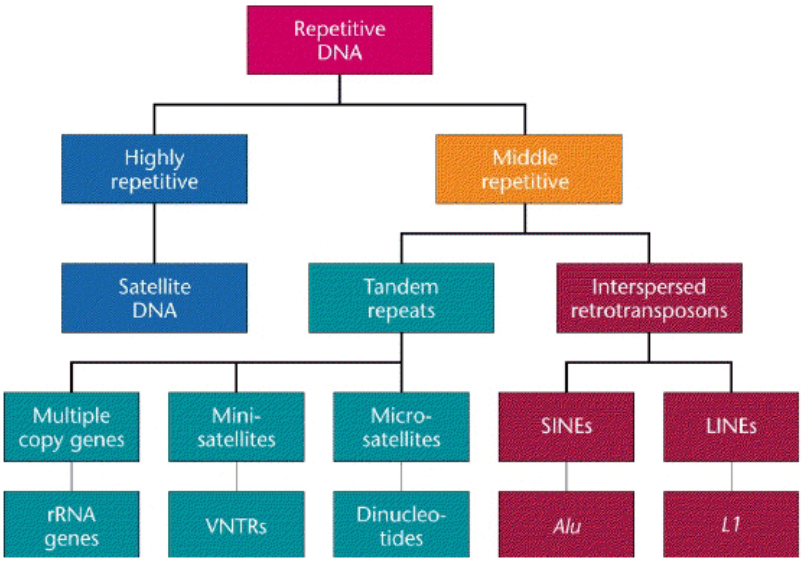

Figure 2.6. Diagram that shows repetitive DNA types. (20)

Besides the repetitive sequences, there are different sequences that are not also coded. Firstly, introns are creating genes along with exons by localizing between exons. During transcription of the gene, both introns and exons are being copied to the oligomer called pre-messenger ribonucleic acid (pre-mRNA). Later, introns are being spliced out from the oligomer. Therefore, introns are not being coded, however, it has crucial roles such as regulation. (21) Furthermore, intergenic regions are part of the non-coding DNA and have regulatory roles due to topologically associated domains. Lastly, some of the RNAs are not coding, yet, have regulatory properties such as gene expression regulation. (22, 23)

2.2 Cancer

Cancer is a type of disease caused by a group of cells that develop, spread and escape cell death due to genetic disorders occurring at any stage of normal cell development (Figure 2.7) (24).

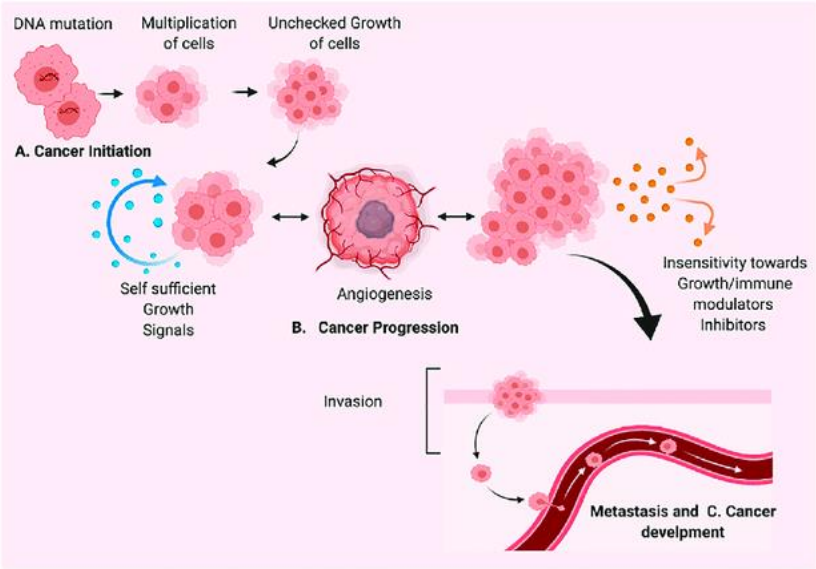

Figure 2.7. Formation and metastasis mechanism of cancer (25)

These genetic disorders can be seen in the early stages of life as well as any stage of life. It is a multifactorial disease that may be induced by combination of lifestyle, environmental exposures, and genetic background. People with cancer are mainly developing the disease due to environmental factors such as chemical exposures, sunlight etc, or habits such as smoking, diet and etc. (26, 27) In the United States, cancer is the second most common cause of death overall and the most common cause for those under the age of 85. Furthermore, it is anticipated that there will be 611,720 cancer deaths and 2,001,140 new cancer cases in the US in 2024 (28). Cancer is a serious disease that may cause the death of the individual. Therefore, special interest is needed to be given. Based on the literature, it is stated that only 5-10% of the cancers are hereditary cancers while the remaining are non-hereditary, also known as sporadic cancers (1).

2.2.1 Adult cancers

Adult cancers are variety of malignancies that may affect different systems, organs, and tissues. These malignancies occur based on the age, sex, genetic background, environmental factors, and habits. Based on where the development of the malignancy start, cancers may be classified. For example, malignancies occurring

in the milk duct cells or lobules cause breast cancer (mostly in females) (29). Moreover, lung cancer is another example of adult cancers which is mostly associated with smoking (30). Another example of adult cancers is colorectal cancer which is caused by the abnormal cell proliferation in the lining of colon or rectum, also called as polyps. As it was stated before, many adult cancers are sporadic. Also, in hereditary cancers, age onset of the malignancies is lower than the sporadic cancers due to the positive genetic background that allow cells to evade cell death. (31)

2.2.2 Childhood cancers

Childhood cancers are malignancies that mainly affect children and adolescents. Childhood cancers are seen more rare than adult cancers due to cellular differences, immune system, less exposure to carcinogenic chemicals, and early detection & intervention. As children and adolescents have less mutation load compared to adults (32), furthermore, immune system of the children and adolescents works more efficiently than the adult immune systems to detect and eradicate the cancerous cells (33). Moreover, children and adolescents are usually exposed to carcinogens less than adults which causes less incidence of cancer in children and adolescents. Lastly, childhood cancers are usually more aggressive than the adult cancers which allow physicians to diagnose and treat the disease earlier (34). When it comes to genetic background of the childhood cancers, it is stated that 10-15% of the childhood cancers are hereditary, whereas 5-10% in adults (2, 3). Some examples of hereditary childhood cancers include brain and central nervous system (CNS) tumors such as glioblastoma, medulloblastoma, glioma and more; hematologic malignancies such as T cell acute lymphoblastic leukemia (T-ALL), B cell acute lymphoblastic leukemia (B-ALL), lymphoma; Wilms' tumor, retinoblastoma and more. (35)

2.3 Cancer Predisposition

Cancer predisposition is the term that indicates the increased susceptibility to develop cancer for an individual. In 1890s, Dr. Aldred Warthin who is a pathologist, created the pedigree of a family with a significant amount of gynecological and colon

cancers seen in four generations. This was the first described family with cancer predisposition (36). Later, Dr. Lynch discovered a similar family with a significant amount of digestive system malignancies. After the discovery of the Dr. Lynch, he and his colleagues identified the first hereditary cancer syndrome called Lynch Syndrome (OMIM #120435) and breast and ovarian cancer syndrome (OMIM #604370). (37) It was clearly thought that inheritance is the main key point that causes familial cancers. In 1971, another physician called Dr. Alfred G. Knudson published an article based on his studies on retinoblastoma. This study suggests "two-hit theory": as the "first hit", the patient inherits one allele from one parent, thus, increases the risk of cancer development in hereditary cancer cases. However, the other or "healthy" allele of the gene mutates and leads to "second hit". The study clarifies why cancer in hereditary/sporadic cases differ in terms of onset and prevalence. Cancer typically strikes earlier in life and may run in families more frequently in hereditary cases, when one mutation is already present from birth. In sporadic examples, which might take longer and be less frequent, both hits must happen separately in a cell within a person's lifespan. (38) This study was mainly on retinoblastoma patients; however, it is attributable for neuroblastoma and Wilms tumor. Knudson's two-hit theory has been crucial to cancer research and genetic counseling since it reveals the fundamentals of how genetic alterations contribute to the development of cancer. (39) After the advancements in the molecular genetics and sequencing technologies, we identified more genes that are responsible for familial malignancies, and they are named as predisposition genes. (40) Cancer predisposition may arise from the certain types of variants in different cancer predisposition genes. For example, as a well-known cancer predisposition genes *BRCA1* and *BRCA2*, their deleterious variants are known to cause Hereditary Breast and Ovarian Cancer Syndrome (HBOC). People with HBOC have greater risk of developing breast and ovarian cancers compared to people without HBOC (41, 42). Furthermore, *MLH1*, *MSH2*, *MSH6* and *PMS2* genes are responsible for DNA mismatch repair, and deleterious variants are associated with the Lynch Syndrome and colorectal malignancies. Another example is the *TP53* gene which is associated with the Li-Fraumeni syndrome. All of the indicated syndromes are associated with the cancer predisposition which are only a few examples of cancer predisposition syndromes. It is for sure that there are other genes and syndromes to be

identified in the future. (40) Another example of cancer predisposition syndromes is Familial Adenomatous Polyposis which is caused by deleterious *APC* variants and frequently associated with colorectal cancer due to development of significant number of polyps (43). Cowden syndrome is another type of cancer predisposition syndromes and caused by deleterious *PTEN* variants. People with Cowden syndrome have greater risk of developing renal cell malignancies along with CNS tumors (44). Furthermore, *NF1* and *NF2* are two genes, and they are known to cause neurofibromatosis which is another example of cancer predisposition syndrome (45).

In 1991, Narod and his colleagues concluded the predisposition rate as 4% and probable disposition as 10% in more than 16,000 cases. Furthermore, in 2005, Merks and his colleagues detected the childhood cancer predisposition ratio as 4% whereas probable predisposition as 3.3%. Similarly, other researchers also indicated the hereditary fraction approximately as 10%. However, based on the advancements in the sequencing technologies, our knowledge of hereditary fractions regarding the childhood cancers are being expanded. (142-144)

2.4 Sequencing Technologies

2.4.1 First-generation sequencing

Development of the DNA sequencing technology that uses dideoxy method by Frederic Sanger has changed the path of the genomics in 1977. (46) Working mechanism of Sanger sequencing is based on chain-termination by using dideoxynucleotides. To perform Sanger sequencing, deoxytriphosphates (dNTPs) and di-deoxytriphosphates (ddNTPs) are required. ddNTPs are individually labeled with fluorescents based on their bases. After the reaction is completed, all products are being run on the gel, later, the sequence of the DNA segment is being derived based on their sizes by using a detector that measures the fluorescent wavelength (Figure 2.8). (47)

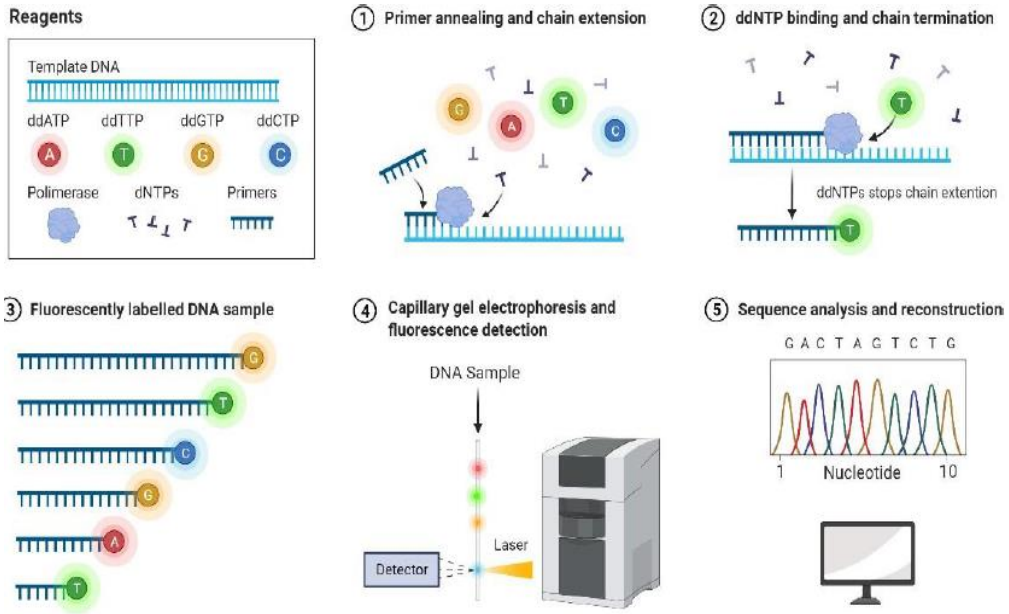

Figure 2.8. Workflow of the Sanger sequencing (47)

2.4.2 Second-generation sequencing

Second-generation sequencing, also known as next-generation sequencing (NGS) or short-read sequencing, was a revolutionary development in genomics that allowed the parallel sequencing of multiple samples at the same time in a cheaper manner (48). NGS includes several steps, which start with the isolation of the genomic sample. Later, the genomic sample is fragmented by approximately 50-600 base pairs (bp) depending on the sequencer, and by the addition of the adaptors, a library is prepared. Afterward, a process called target enrichment takes place to increase the number of copies of the target region (or the complete genome in whole genome sequencing). After target enrichment, copies of the genome regions are being sequenced by using different fluorophores attached to each nucleotide. After the sequencing process of the genomic targets, the FASTQ file is generated for further bioinformatics processes and analysis. The workflow of the NGS is shown in Figure 2.9.

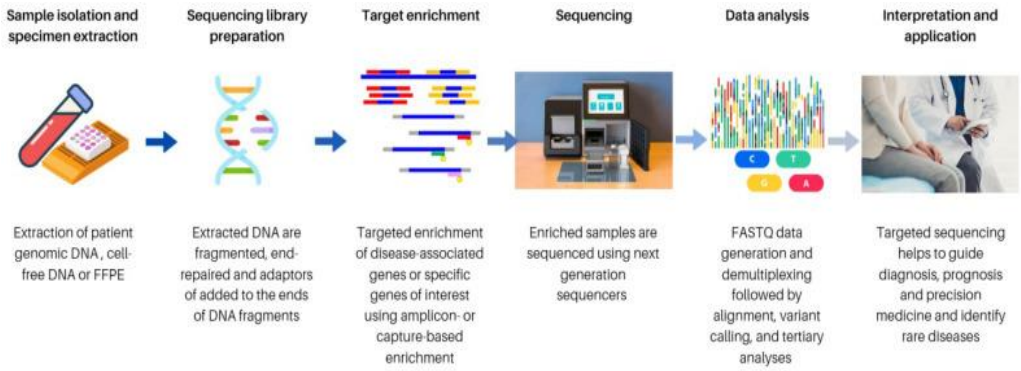

Figure 2.9. Workflow of NGS (49)

2.4.3 Third-generation sequencing

Third-generation sequencing, also known single molecule sequencing, is a relatively new technique that overcomes the limitations of the NGS. Even though NGS is still a widely used method in the world, NGS has limitations as it uses short fragments of DNA (short-reading sequencing) and has challenges sequencing the repetitive sequences. Since third-generation sequencing reads approximately 10 kilobase (kb) bp, it is possible to detect structural variants and/or repetitive regions that are challenging for NGS. (50, 51)

2.5 NGS Analysis

NGS analysis is the bioinformatics and computational procedures that are used for clarification of the vast amount of data produced by NGS technology and finding relevant information. Several different data types and processes are involved in NGS analysis. These data types and processes are essential for analyzing the data derived by NGS technologies.

2.5.1 Data types

2.5.1.1 FASTQ

FASTQ is a text-based file that is widely used in NGS. It consists of the sequence of the sample along with the quality score that corresponds to each base (FASTQ is the combination of FASTA and Quality, which indicates the storage of both quality scores and the sequence). An example of a FASTQ file is given in Figure 2.10.

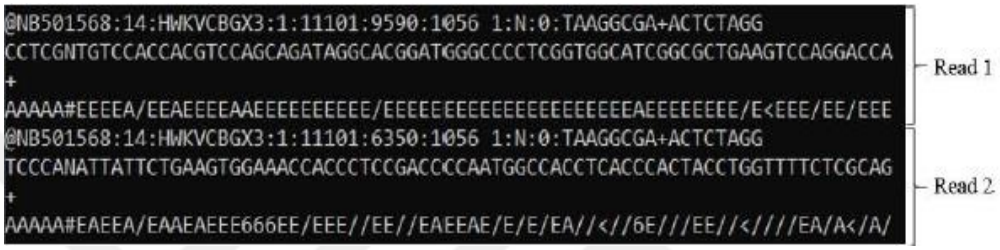

Figure 2.10. An example of FASTQ file. FASTQ files are consisting of four lines for each read that indicate the title and the description (the first line), the sequence (the second line), a plus sign to separate between the second and the fourth lines (the third line), and lastly the quality scores (the fourth line). (52)

2.5.1.2 SAM and BAM

In NGS, sequence alignment data is typically represented by using the SAM (Sequence Alignment/Map) and BAM (Binary Alignment/Map) file formats. These formats are crucial for many bioinformatics applications. These include a variety of processes like variant calling. Since they hold data about the alignment of sequence reads to a reference genome, it is crucial for the analysis of the NGS derived data. (53) For genomic data, SAM is a text-based format that stores quality measurements and aligned sequence information. Aligned raw data is presented in the binary alignment map (BAM) format. Furthermore, BAM is the compressed version of the Sequence Alignment Map (SAM) in a binary manner. Multiple sequencing plate outputs and short and long reads are supported by these formats. There is a section and a header for every alignment. Headers begin with the symbol "@." These symbols help in the identification of alignments one by one. SAM files retain excessive amounts of space

for archiving, increasing the requirement for storage. It is possible to compress SAM files to BAM to increase performance and reduce storage requirements. (54)

2.5.1.3 VCF

Variant calling format (VCF) is a standardized file type in genomics that is used for representing SNPs, indels, and structural variants, and the information belongs to these genomic alterations. As per the standardized format of the VCFs, they have a header section that includes the information about the VCF file and a body section divided into eight different columns: “CHROM”, “POS”, “ID”, “REF”, “ALT”, “QUAL”, “FILTER”, and “INFO”, respectively (Figure 2.11). The “CHROM” column indicates the chromosome that the interested variant is located. The “POS” column indicates the chromosomal location where the interested variant is located. Also, “ID” column represents the unique identifier of the interested variant. The “REF” column indicates the nucleotide base in the reference genome at that location, whereas the “ALT” column indicates the nucleotide base that is different than the reference genome at that location. “QUAL” column indicates the quality scores for the interested variant. The “FILTER” column indicates the variant acceptability for analysis. However, requirements and parameters for evaluating the variant acceptability should be given during the variant calling process. Lastly, the “INFO” column includes information like allele count, allele frequency, ancestral information, and more. (55)

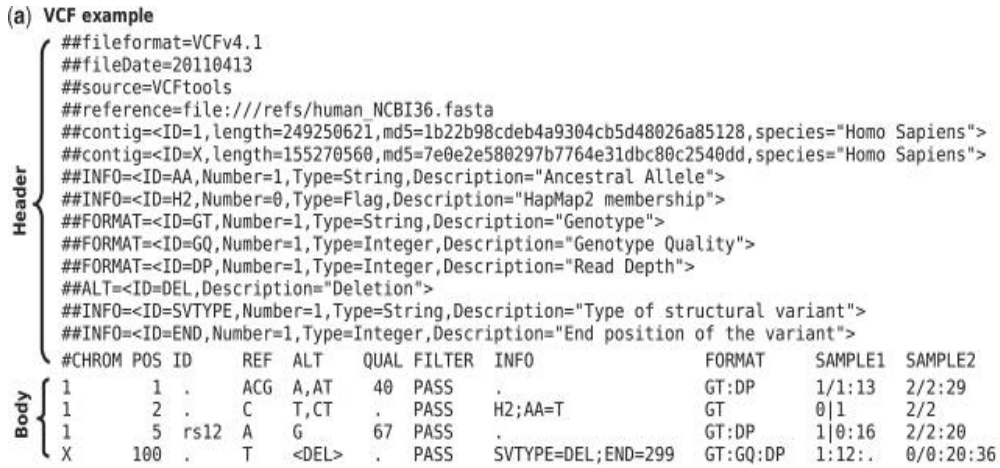

Figure 2.11. An example of a VCF file. Adapted from Danecek *et. al*, 2011 (55)

2.5.2 Processing of the data

To work with the NGS data, there are several steps to create more analyzable and convenient data, which are mandatory for the analysis. These steps are cumulatively called "pipeline". A pipeline in genomics is a collection of analytical or computational tools. These tools are carried out in a certain order for interpreting the genomic data. Pipelines' main purpose it to process raw genomic data and turn it into scientifically interpretable data. Pipelines are intended to manage the large datasets derived from high-throughput sequencing methods like next-generation sequencing (NGS). A standard genomics pipeline consists of numerous phases, each with multiple procedures, such as variant calling, alignment, quality control, data preprocessing, annotation, and interpretation. There are different pipelines available online based on the needs of the investigators. Furthermore, investigators can create their own pipelines or make changes on the pipelines based on their needs. An example workflow of processing the NGS data is shown in Figure 2.12. (56)

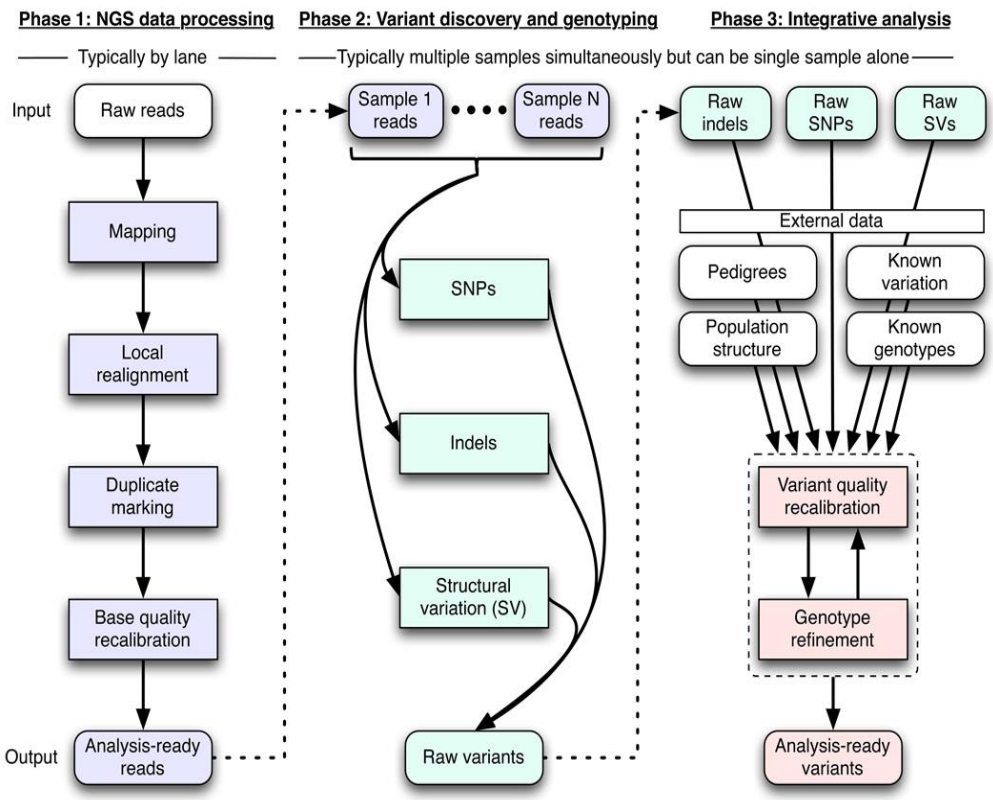

Figure 2.12. A workflow example that represent the processing of the NGS data. (56)

2.5.2.1 From FASTQ to VCF

As the first step of the NGS data processing, raw FASTQ files must be processed which will lead to another type of data called VCF files, as mentioned above. For this purpose, firstly, FASTQ files must pass through the quality check. After the quality of the FASTQ file is confirmed, adaptor sequences must be trimmed off along with removing the low-quality reads. Later, remaining reads must be mapped to the reference genome by using alignment tools. After the mapping process, SAM and BAM files must be created. After the removal of the duplicate reads, variant calling must be performed to generate VCF file which further will be used for the annotation step. (56)

2.5.2.2 Annotation

The process of adding more details to the genomic variants observed during analysis is referred as annotation. Since an individual variant may have a vast amount of information in both literature and databases. This information includes multiple parameters that researchers may use such as allele frequency data, in-silico pathogenicity predictors, disease-associated databases, ontology, and/or clinical significance-associated databases. Allele frequency is the proportion of the number of observed copies over the total allele count in a population. Also, *in-silico* pathogenicity predictors are another data to have insights about the ACMG classification of the variant by using different methods. Some of them include evolutionary conservations, whereas some of the others include structural predictions. There is no specific way to perform annotation as per it depends on the needs of the researcher. (57)

2.5.2.3 Filtering

As the variants annotated based on the researcher's needs, interested variants must be filtered out according to the phenotype and the family tree (or pedigree) of the individual. There is no certain way of filtering, as the annotation strategy, filtering strategy also depends on the needs of the researcher. The main idea of the filtering

process is to observe the interested variant(s) from the raw variant list by using the phenotypic and familial data of the individual.

2.6 Variants

A variant in genomics is a specific form of a gene or DNA sequence that differ from the reference genome for the interested species. Variants may affect physical features, diseases, or other biological aspects differently and may induced by nature as a component of evolutionary process as well as other components like chemicals.

2.6.1 Types of variants

2.6.1.1 Nucleotide substitutions

Nucleotide substitutions, also known as single nucleotide variations (SNVs), are variants that is caused by the change of a single nucleotide to a different base than the reference genome. It is known that the most frequently observed variant type is the nucleotide substitutions. (Figure 2.8) (58)

2.6.1.2 Insertion and deletion variants

Based on the reference genome, any addition and/or removal of one or more nucleotides are called insertion and deletion variants (indels), respectively. Indels may have significant consequences due to the risk of altering, therefore, obstructing the protein structures as per changing the amino acid sequence. (Figure 2.8) (59)

2.6.1.3 Copy number variations

Copy number variations (CNVs) are the structural variants that include diverse dosage of specific DNA segments. These variants include unbalanced translocations with insertions, large indels bigger than one kilobase, duplications. (Figure 2.13) (60)

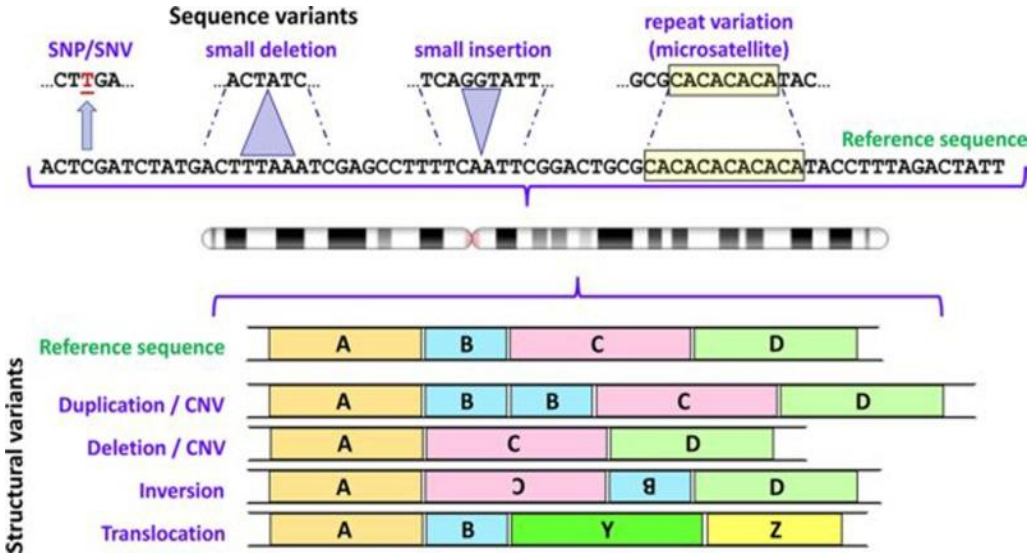

Figure 2.13. Schematic representation of variant types (61)

2.6.2 Molecular consequences of variants

2.6.2.1 Synonymous variants

Synonymous variants are the DNA alterations that do not alter the amino acid sequence of the produced protein. Since it does not change the structural form of the protein, they are usually classified as functionally neutral variants. (62) However, it is also known that synonymous variants may impact the splicing mechanism efficiency in which leads to non-functioning proteins. (63)

2.6.2.2 Missense variants

Missense variants are the change of a single nucleotide that causes the change of the amino acid residue. Since the missense variants alter the amino acid sequence, it means that the missense variants are only seen in the coding region of the genome. Missense variants may have functional consequences on the proteins based on the amino acid change and/or where the missense variant locates on (missense variants on functional domains of the proteins may have higher impacts on the function of the protein). (64, 65)

2.6.2.3 Nonsense variants

Nonsense variants are the single nucleotide alterations in the genome that create a premature stop codon in the gene, in which result in premature protein. The produced premature proteins, also known as truncated protein, are usually evaluated as damaging and non-functional proteins. (66, 67)

2.6.2.4 Frameshift variants

Proteins are being translated based on frame, therefore, the product results in a certain amino acid sequence. Frameshift variants are usually indels which cause the disruption of the reading frame, later results in either truncated protein or non-functional protein. Frameshift variants are observed in the coding region of the DNA as one or more indel variants are present in the coding regions. However, indels with multiple of three nucleotides do not cause frameshift variants as per amino acid codons are coded as three nucleotides. (68, 69)

2.6.3 Germline and somatic variants

The variants are divided into two main groups based on their inheritance which are named germline variants and somatic variants. Germline variants, often referred to as hereditary variants, are genetic alterations which are found in the sperm or egg cells. When they fertilize, as a result, these variants are being inherited to the offspring. Therefore, each of the cell in the organism carries the variant including germ cells. On the other hand, in the somatic variant, also known as acquired variants, are genetic changes that occur in a single cell after fertilization. These changes are not inherited but can be passed on to the progeny of the mutated cell.(70, 71) (Figure 2.13)

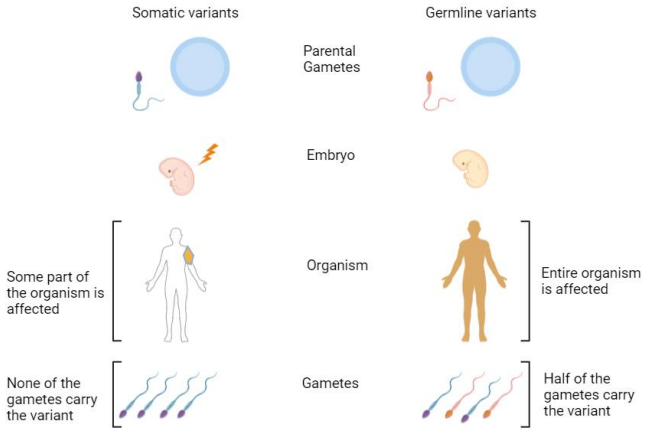

Figure 2.14. Main mechanisms of germline and somatic mutations.

2.7 ACMG Guidelines and Variant Interpretation

Guideline for the evaluating variants that are detected via sequencing technologies were firstly created in 2008 by the American College of Medical Genetics and Genomics (ACMG) (72). Both sequencing technologies and human knowledge have advanced since then; ACMG updated its parameters and guidelines for the interpretation of variations in response to achievements in the genomics. These guidelines cover all typical genetic tests performed in clinical laboratories, including genotyping, single genes, panels, exomes, and genomes. This guideline provides a classification pathway for known disease-gene associations. It recommends using standard terms like 'pathogenic', 'likely pathogenic', 'uncertain significance', 'likely benign', and 'benign' with evidence-based parameters to describe the variations. There are very strong, strong, moderate, and supportive evidence of pathogenicity, along with stand-alone, strong, and supportive evidence for benign (Figure 2.14). (7)

	Benign		Pathogenic			
	Strong	Supporting	Supporting	Moderate	Strong	Very Strong
Population data	Allele frequency is too high for disorder BA1/BS1 OR observation in controls inconsistent with disease penetrance BS2			Absent in population databases	Prevalance in affecteds statistically increased over controls PS4	
Computational data		Multiple lines of computational evidence suggest no impact on gene /gene product BP4 Missense in gene where only truncating cause disease BP1 Silent variant with non predicted splice impact BP7 In-frame indels in repeat w/out known function BP3	Multiple lines of computational evidence support a deleterious effect on the gene /gene product PP3	Novel missense change at an amino acid residue where a different pathogenic missense change has been seen before PM5 Protein lenght changing variant PM4	Same amino acid change as an established pathogenic variant PS1	Predicted null variant in a gene where LOF is a known mechanism of disease PVS1
Functional data	Well-established functional studies show no deleterious effect BS3		Missense in a gene with low rate of benign missense variants and pathogenic missenses common PP2	Mutational hotspot or well-studied functional domain without being benign variant PM1	Well-established functional studies show the deleterious effect PS3	
Segregation data	Nonsegregation with disease BS4		Cosegregation with disease in multiple affected family members PP1			
De novo data				De novo without maternal and paternal confirmation PM6	De novo with maternal and paternal confirmation PS2	
Allelic data		Observed in trans with a dominant variant BP2 Obsereved in cis with a pathogenic variant BP2		For recessive disorders, detected in trans with a pathogenic variant PM3		
Other database		Reputable source w/out shared data = benign BP6	Reputable source = pathogenic PP5			
Other database		Found in case with an alternate cause BP5	Patient's phenotype or FH highly spesific for gene PP4			

Figure 2.15. ACMG Guideline parameters to classify the variants based on their pathogenicity. (7)

3 MATERIALS AND METHODS

3.1 Materials

3.1.1 Sample collection and DNA isolation

Five c.c. blood samples of the patients and their family members who were eligible for the study were collected by EDTA tube. Following sample collection, DNA isolation was performed with Zymo Research Quick-DNA Miniprep Plus Kit. After isolating DNAs, Thermo Scientific™ NanoDrop™ One/Onec Microvolume UV-Vis Spectrophotometer was utilized to measure DNA concentrations and purity.

3.1.2 Clinical exome sequencing

For the clinical exome sequencing, Sophia Genetics Clinical Exome Solution (CES_v2) kit that covers exonic and flanking intronic regions of 4490 genes was used via Illumina NextSeq platform.

3.1.3 Data processing

For the data processing part, ACUGEN pipeline which is developed by ACURARE team utilized.

3.1.4 Variant filtering

Variant filtering part was performed by using Gennext Platform (gennext.bio) (https://github.com/GenivaInformatics/gennext-workflows/tree/main).

3.1.5 Sanger sequencing

For the confirmation and segregation analyses, Genemark PCR Dye Master Mix II (5X) was used for PCR reactions (Table 3.1). Furthermore, designed primer pairs are given in Table 3.2.

Table 3.1. PCR reactions for each sample

Ingredient	Volume
Genemark PCR Dye Master Mix II	5 µl
Forward Primer (5 µM)	0.5 µl
Reverse Primer (5 µM)	0.5 µl
Distilled Water	17 µl
DNA sample (10ng/µl)	2 µl

Table 3.2. Designed primers pairs for each patient and gene

Patient	Gene	Forward Primer	Reverse Primer
1	*XPA*	AGCTAGGTC CTCGGAGTGG	CGGGGAGAA TCTGCACACAT
1	*AP4B1*	TCTTTTGTCT TTCACTGGTGGG	CCTTATAGAT CATGGCAGCCTG A
2	*MLH1*	ACCAGCACA GAGAAGTTGCT	GCTGATGCTA TTGTGGGTTAGT
3	*RET*	GCTGTGTCC ACCCCCTTAC	ACTGCATCCC CTGTGAGATC
5	*BRCA1*	TCCTGATGG GTTGTGTTTGGT	ACAAAATGA AGCGGCCCATC
6	*ERBB4*	ATGCATCCA ATGGTCCACCC	GGTGTCTGTA AGGTGGAGGC
7	*BRCA2*	CAATAGCTG ACGAAGAACTTG CA	CACATTCTTC CGTACTGGCCT
9	*FLT3*	ACCTCACCT AACTCTGGCTTG	TGAACAGGG GAGAAAAGGCA
10	*TGFBR2*	CCACATCCA ACTCCTTCTCTCC	CAGGGTGTCC AGCTCAATGG

PCR reactions were performed by Bio-Rad T100 Thermal Cycler with the conditions indicated in Table 3.3.

Table 3.3. PCR conditions

Step	Temperature	Time
Initial Denaturation	95 ºC	5 mins
Denaturation	95 ºC	30 secs (35 cycles)
Annealing	Depends on primer Tm	30 secs (35 cycles)
Extension	72 ºC	45 secs (35 cycles)
Final Extension	72 ºC	5 mins

After PCR reactions, samples were run on 2% agarose gel. For the preparation of agarose gel, 3 gr of agarose and 150 ml of TAE buffer was used. Lastly, Bio-Rad PowerPac™ Basic Power Supply was used for running the samples on agarose gel with GeneMark ZGen 50bp DNA Ladder. Lastly, bands in the gel were visualized by using Bio-Rad ChemiDoc™ MP Imaging System.

3.2 Methods

3.2.1 Sample collection and DNA isolation

Families with positive cancer history who applied to Pediatric Hematology Unit of Altunizade Hospital were evaluated based on Jongmans' criteria (73). A total of 10 patients were enrolled to the study based on the eligibility. There were eight girls and two boys with the mean age of 11.3 (min:3 – max: 17). In the study, the index cases were diagnosed with Wilms' tumor (n=2), B-ALL (n=2), T-ALL (n=1), lymphoma (n=1), glioneuronal tumor (n=1), medulloblastoma(n=1), glioblastoma(n=1), and pilocytic astrocytoma (n=1) (Table 3.4). After collecting the peripheral blood samples of the patients and their family members, DNA isolations were performed according to Zymo Research Quick-DNA Miniprep Plus Kit manual. After DNA isolation,

Thermo Scientific™ NanoDrop™ One/Onec Microvolume UV-Vis Spectrophotometer was utilized for measuring the DNA concentration and purity. After measurement, all samples diluted to 10 ng/µl concentration, and samples were stored at -20 ^{o}C.

Table 3.4. General information about the patient enrolled to the study.

Sex	
Male	2
Female	8
Age (year) (n=10)	
Mean	11.3
Median	14.0 (max: 17, min:3)
Sample source (n=10)	
Peripheral blood	100%
Tumor tissue	10%
Referred center	
Acıbadem Altunizade Hospital	100%

Table 3.5. Clinical characteristics of the patients

Patient Number	Age	Sex	Diagnosis	Samples from family member	Family history
1	11	F	Wilms' tumor	Index, Brother, Mother	-Index's maternal grandmother (uterine ca.) -Index's paternal grandfather (lung ca) -mother's paternal aunt (lymphoma) -mother's paternal uncle (lung ca) -mother's maternal grandmother (pancreatic ca) -mother's paternal grandparents (lung ca)
2	12	F	Lymphoma	Index, Mother, Father	-Father's mother (breast ca) -Father's maternal aunt (breast ca) -Father's maternal grandmother (breast ca) -Index's paternal grandmother's paternal cousine (neoplasm) -Index's paternal grandmother's paternal cousine (colon ca) -Mother's maternal grandmother (breast ca)
3	3	M	Glioneuronal tumor	Index, Mother, Father	-Mother's father (laryngeal ca) -Mother's paternal uncle (laryngeal ca) -Mother's paternal aunt (uterine ca) -Mother's paternal grandmother (colon ca) -Father's paternal uncle (brain cyst)
4	5	F	Wilms' tumor	Index, Mother, Father, Brother	-Father's paternal grandfather (prostate ca) -Father's paternal grandmother (pancreatic ca) -Father's paternal grandmother's brother (neoplasm) -Mother's father (prostate ca + colon ca) -Mother's paternal uncle (lung ca) -Mother's paternal uncle (prostate ca) -Mother's paternal uncle (lymphoma) -Mother's paternal aunt (neoplasm)
5	8	F	B-ALL	Index, Mother, Father, Sister	-Index's maternal grandmother (breast ca + lung ca) -Index's maternal grandfather (kidney ca) -Father's paternal grandfather (gastric ca)
6	16	F	Medulloblastoma	Index, Mother, Father, Brother, Twin sister	-Index's father (bladder ca) -Index's paternal grandfather (gastric ca)
7	17	F	T-ALL	Index, Mother, Father	-Index's paternal aunt (kidney ca) -Index's paternal aunt (breast ca) -Father's paternal uncle (prostate ca) -Mother's paternal cousin (lung ca)
8	13	F	B-ALL	Index, Mother, Father	-Index's maternal aunt (breast ca) -Index's paternal grandfather (lymphoma) -Mother's maternal uncle (prostate ca) -Mother's maternal grandmother (colon ca) -Father's paternal grandmother (breast ca)
9	15	F	Pilocytic astrocytoma	Index, Mother, Father, Sister, Twin sister	-Index's paternal aunt (breast ca) -Index's paternal grandfather (prostate ca) -Index's maternal grandmother (lung ca) -Father's paternal grandmother (breast ca)
10	13	M	Glioblastoma	Index, Mother, Father, Brother	-Index's maternal aunt (thyroid ca) -Index's paternal aunt (thyroid ca) -Father's paternal grandfather (prostate ca) -Father's paternal uncle (colon ca)

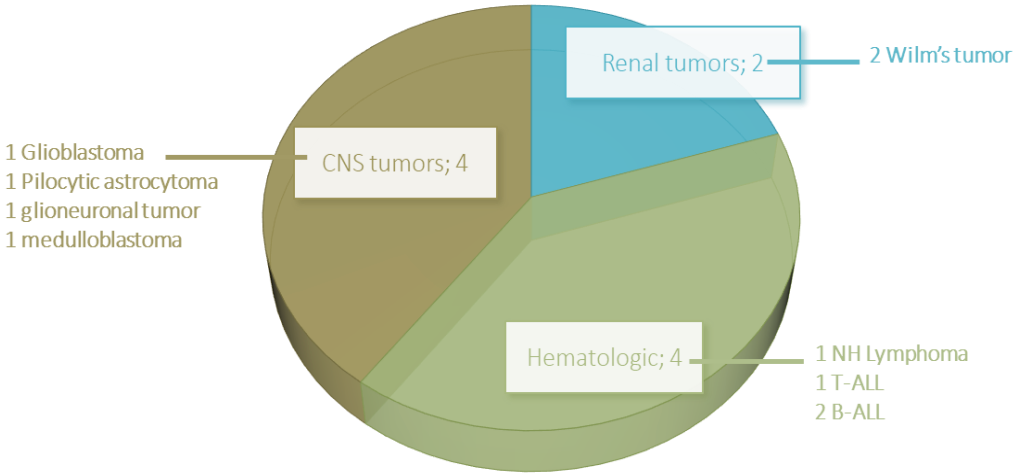

Figure 3.1. Patient distribution based on the malignancy types

3.2.2 Clinical exome sequencing

For the clinical exome sequencing, DNA samples were studied with Sophia Genetics Clinical Exome Solution (CES_v2) kit that sequences 4490 clinically related genes' exons and flanking intronic regions. After library preparation and sequencing processes, output data (fastq files) were stored in hard drive for the further data processing.

3.2.3 Data processing

Gennext Platform (www.gennext.bio) was used for the processing of the output data. At this stage of the study, paired end fastq files was checked in terms of quality by utilizing FastQC program (http://www.bioinformatics.babraham.ac.uk/projects/fastqc/). After, TrimGalore tool was used for trimming the adaptor sequences. Then, for the alignment of the sequences, BWA SAMtools View tool was used to create bam files. After utilization of the SAMtools Sort and SAMtools Index (54), GATK (74) MarkDuplicates tool used for removal of the duplicate sequences. Lastly, four different variant caller tools; Mosdepth (https://github.com/brentp/mosdepth), SAMtoolsIndex (54), Parliament2

(https://github.com/dnanexus/parliament2), DeepVariant (https://github.com/google/deepvariant) was utilized for creating the vcf files.

3.2.4 Variant filtering

For the filtering of the variants, created vcf files were uploaded to Gennext Platform by Geniva (gennext.bio) (https://github.com/GenivaInformatics/gennext-workflows/tree/main) for the annotation of the files. During the annotation step, care was taken to add tools like CADD, SIFT, PolyPhen2, and GERP++. After annotation of the vcf files, a comprehensive gene list (2500 genes) for cancer predisposition was used as well as different gene lists that were formed based on the patients' characteristics and diagnoses. The gene list consisting of 2500 genes was derived through the literature along with various databases (HPO, OMIM etc) by ACURARE team. After inputting the gene lists for the interested patient, allele frequency filter was set to <0.01. Later, functional filter was applied as exonic, splicing, SNVs and Indel variants, nonsense, and missense variants. After the functional filter, *in-silico* prediction tools were utilized for the further filtering. At this stage, since indel variants do not have CADD Phred scores, variants were sorted based on their CADD Phred scores in order not to skip any indel variants. Later, the variants that had CADD Phred score below 20 were eliminated. The remaining variants were classified based on the pathogenicity via ACMG guidelines (7). After classification of the variants by using Varsome (75), Franklin (https://franklin.genoox.com - Franklin by Genoox), Intervar (76) online databases; each variant was investigated in terms of clinical relevance for each patient. For the detection of the clinical relevance, based on the literature, each gene was investigated for their relevance with the clinical characteristics of the patients.

3.2.5 Sanger sequencing

After determining the candidate variants for each patient, DNA samples of the patients and their family members were studied with PCR and direct sequencing. For

the PCR reactions, primer pairs were designed for each candidate gene. Firstly, interested sequences was selected from Ensembl database (77) and run on Primer3Plus tool (https://github.com/primer3-org). At this stage, care was given to melting temperatures of the primer pairs. Primers pairs that have temperature difference of more than 1^{o}C were eliminated. Later, to check the hairpin formation of the primers, IDT OligoAnalyzer Tool (https://www.idtdna.com/pages/tools/oligoanalyzer) was utilized. Lastly, to check the SNP presence on the primer binding sites, SNPCheck tool (https://genetools.org/SNPCheck/snpcheck.htm) was used.

After synthesis of the primers, PCR reactions were formed based on the ingredients and conditions that are indicated in Table 2.1 and Table 2.2. After checking the amplicons via gel electrophoresis, samples were sent to Sanger Sequencing company as external service. After sequencing processes were completed, Sanger Sequencing data were analyzed by using CLC Main Workbench 21.

4 RESULTS

Here we determined 23 candidate variants in 10 patients, and 13 candidate variants were clinically relevant. Eight of these variants are classified as LP/P, whereas seven of them are VUS based on ACMG guidelines. Furthermore, eight of these variants were missense variants, whereas two of them were nonsense and five of them were frameshift variants. (Table 4.1) Also, to determine the quality of each fastq data, data were run on FastQC. FastQC result that represents the fastq quality is seen below (Figure 4.1).

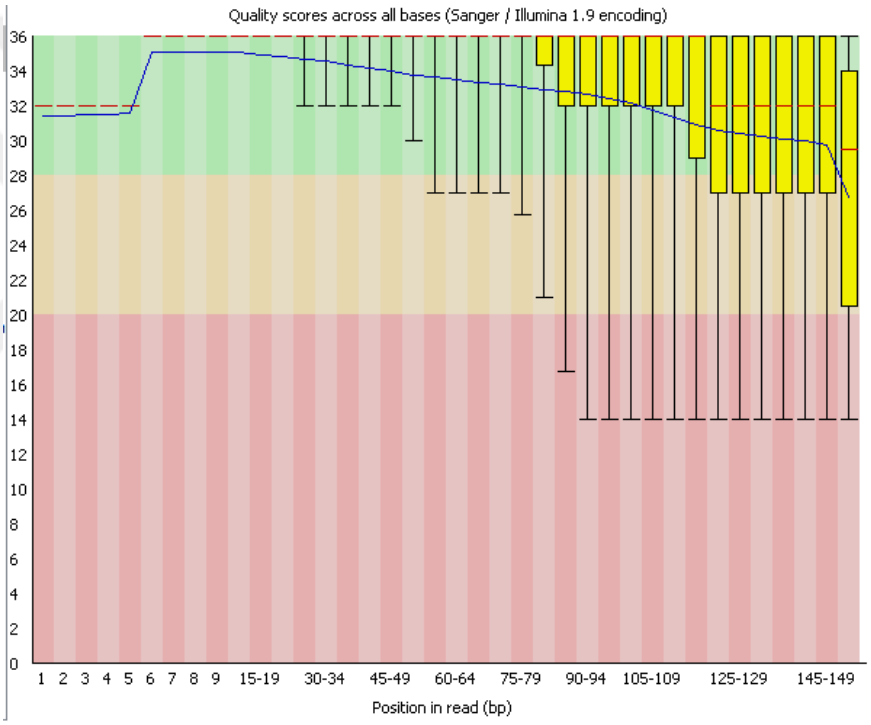

Figure 4.1. Representative FastQC quality from Patient 9's fastq data

When we perform pathway analysis of the candidate genes, we determined that there are three main pathways take part in the predisposition of childhood solid tumors. These pathways include DNA repair, intracellular processing, and transcription factor. (Figure 4.2)

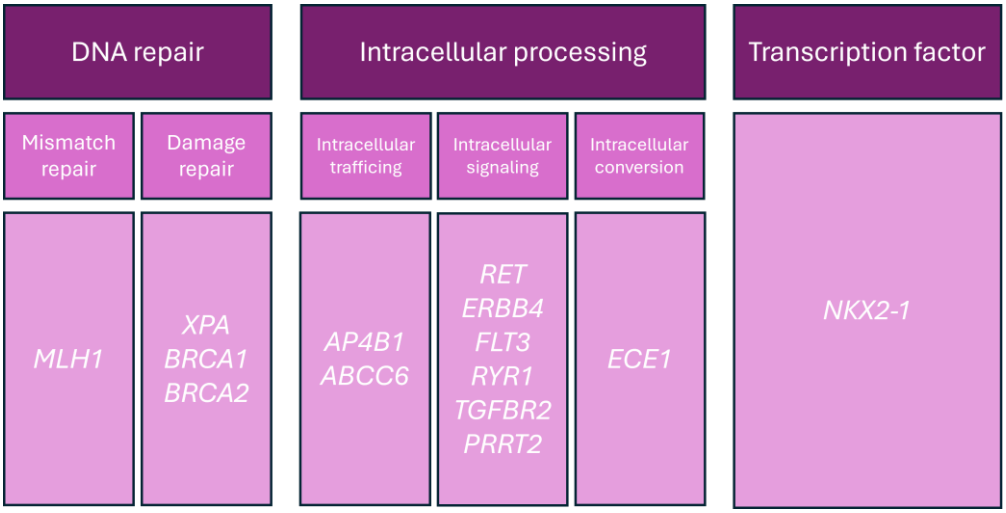

Figure 4.2. Three main classifications of the candidate variants

Table 4.1. Clinically relavant the candidate variants that are detected in the patients.

Patient Number	Gene	Genomic Location	Transcript	cDNA change	Protein change	Molecular Consequence	dbSNP	Varsome	Franklin	Intervar	ClinVar	CADD/SIFT/GERP++	gnomAD AF
1	*XPA*	9:97697171	NM_000380.4	c.122C>T	p.Ala41Val	Missense	rs778289269	VUS (PM2, PP3)	VUS (PM2)	VUS (PM2, PP3)	VUS (VCV001319452.3)	33 / D / 4.78	0.0000783846
1	*AP4B1*	1:113898770	NM_001253852.3	c.1146T>A	p.Cys382Ter	Nonsense	-	LP (PVS1, PM2)	LP (PVS1, PM2)	LP (PVS1, PP3, PM2)	-	36 / - / 3.16	-
2	*MLH1*	3:37028833	NM_000249.4	c.1459C>T	p.Arg487Ter	Nonsense	rs63749795	P (PVS1, PP5,	P (PVS1, PP5, PM2)	P (PVS1, PP5, PM2)	P (VCV000089744.29)	35 / - / 1.2	-
3	*RET*	10:43119548	NM_020975.6	c.2410G>A	p.Val804Met	Missense	rs79658334	P (PP5, PM5, PP3, PM2)	P (PP5, PM5, PP3, PM2)	LP (PS3, PM1, PM2, PP3)	P (VCV000037102.54)	29.3 / D / 3.36	0.00009859
4	*ECE1*	1:21225324	NM_001397.3	c.1966G>A	p.Gly656Arg	Missense	rs367812436	LB (PP5, BS2, BP4, BP1)	LP (PS4, PM2, PP3, BS2)	VUS (PM1, PM2, PP3)	LP (VCV001172769.1)	25.9 / D / 5.36	0.00002788
5	*BRCA1*	17:43057063	NM_007294.4	c.5266dup	p.Gln1756ProfsTer74	Frameshift	rs80357906	P (PVS1, PS3, PP5, PM2)	P (PVS1, PP5, PM2)	-	P (VCV000017677.91)	- / - / -	0.00005257
6	*ERBB4*	2:211383815	NM_005235.3	c.3727T>G	p.Trp1243Gly	Missense	-	LP (PP4, PP3, PM2, BP1)	LP (PP4, PP3, PM2)	P (PP4, PM1, PM2, PP1, PP2, PP3)	-	26.4 / T / 5.72	-
7	*BRCA2*	13:32337722	NM_000059.4	c.3367A>G	p.Ser1123Gly	Missense	rs80358581	VUS (PP3, PM2, BP1)	LP (PP3, PM2)	LP (PP3, PM1, PM2, BP1)	VUS (VCV000051455.10)	25.7 / D / 5.75	0.000001369
8	*ABCC6*	16:16154766	NM_001171.6	c.4070G>A	p.Arg1357Gln	Missense	rs201275608	VUS (PM5, PM1, PM2, PP3)	P (PM2, PM5, PP5, PP3)	VUS (PM2, PP3, BP1)	VUS (VCV001359254.5)	29.8 / D / 3.63	0.00007887
9	*FLT3*	13:28023414	NM_004119.3	c.2350_2354del	p.Thr784Ter	Frameshift	-	LP (PVS1, PM2)	LP (PVS1, PM2)	-	-	- / - / -	-
9	*PRRT2*	16:29813695	NM_145239.3	c.649dup	p.Arg217ProfsTer8	Frameshift	rs587778771	P (PVS1, PP5, PS3)	P (PP5, PM5, PP3, PM2)	-	P (VCV000065758.103)	- / - / -	0.0006636
10	*RYR1*	19:38499970	NM_000540.3	c.7277del	p.Tyr2426LeufsTer4	Frameshift	-	LP (PVS1, PM2)	LP (PVS1, PM2)	-	-	- / - / -	-
10	*TGFBR2*	3:30671728	NM_003242.6	c.545T>C	p.Val182Ala	Missense	-	VUS (PM2, PP3)	VUS (PM2, PP2)	-	-	22.5 / D / 5.28	-

4.1 Patient 1

Patient 1 was diagnosed with Wilms Tumor at the age of 11. Within the family there were individuals diagnosed with several cancers (Figure 4.3). A total of 12756 variations were determined and following the filtering processes, three candidate variants in *AP4B1, XPA and NOTH3* genes were determined (Table 4.2). The IGV findings showed that *NOTCH3* is false positive, so no further investigation was performed.

Table 4.2. List of candidate variants detected in Patient 1.

Genomic Location	Ref	Alt	Zygosity	Gene	Transcript	Molecular Consequence	AA Change	ACMG Classification	CADD Phred	GERP++	Gnomad AF	SIFT
1:113898770	A	T	Het	*AP4B1*	NM_001253852.3	Stopgain	p.Cys382Ter	LP (PVS1, PM2)	36	3.16	-	-
9:97697171	G	A	Het	*XPA*	NM_000380.4	Missense	p.Ala41Val	VUS (PM2, PP3)	33	4.78	0.0000784	Damaging
19:15170800	T	G	Het	*NOTCH3*	NM_000435.3	Missense	p.Asn1588His	VUS (PM2, PP3)	27.5	3.92	-	Damaging

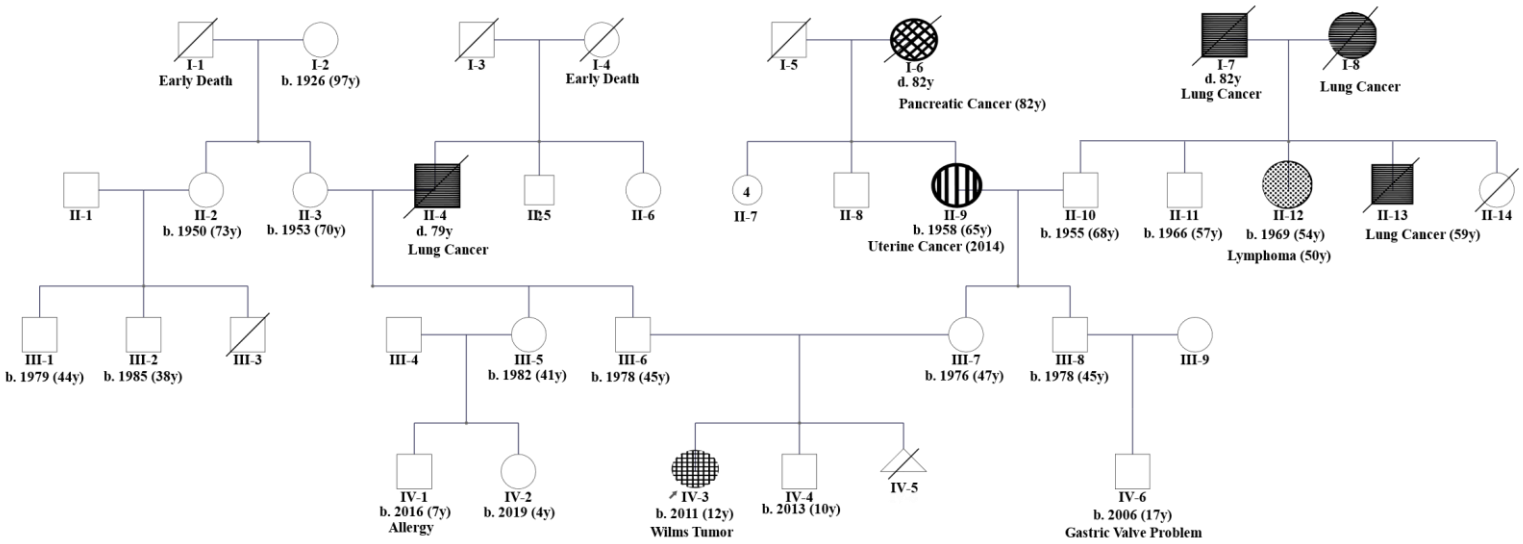

Figure 4.3. Pedigree of the Patient 1

The results were discussed within the molecular cancer studies group and the most clinically relevant variant was agreed to be XPA (*XPA:* c.122C>T: p.Ala41Val, rs778289269). The variant was determined in heterozygous state and classified as VUS with the evidence titles PM2, PP3. The segregation analysis revealed that the mother is heterozygous, and variant was inherited from mother. Since the DNA sample of the father was unavailable, carrier status of the father is unknown. Also, brother of the index case was sequenced, and observed as wildtype. (Figure 4.4)

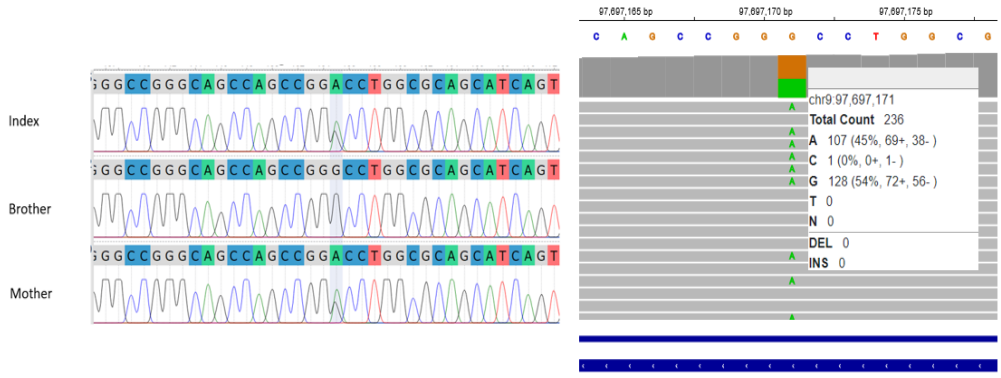

Figure 4.4. Segregation analysis of *XPA* by Sanger sequencing of the Patient 1

Furthermore, a heterozygous LP variant (PSV1, PM2) in *AP4B1* gene (*AP4B1:* c.1146T>A: p.Cys382Ter) was determined in the index case. However, based on the literature *AP4B1* gene was not found to be related with WT. In the segregation analysis, we observed that variant was not inherited from mother. Since the DNA sample of the father was unavailable, variant status of the father is unknown. Also, brother of the index case was sequenced, and found as heterozygous. (Figure 4.5) Since the indicated variant is LP, it was reported along with *XPA* variant.

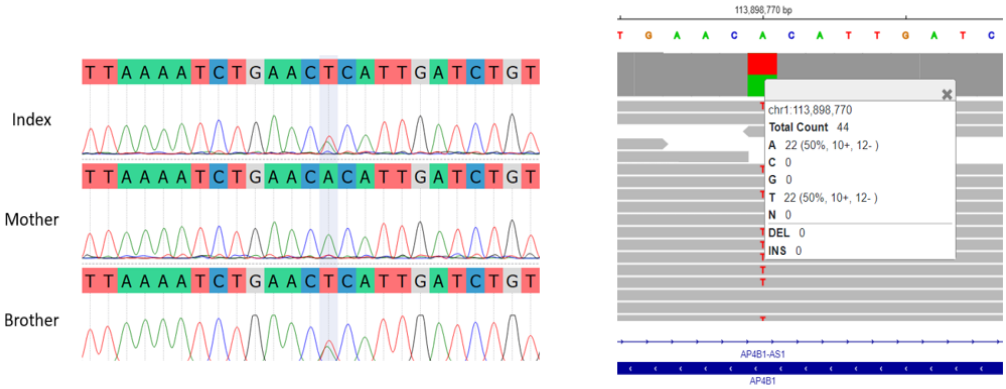

Figure 4.5. Segregation analysis of *AP4B1* by Sanger sequencing of the Patient 1

4.2 Patient 2

An 11-year-old girl diagnosed with lymphoma has shown to possess two variants that may be related with cancer predisposition. Clinical exome sequencing revealed a total of 9966 variants and stepwise variant filtering ended up with two candidate variants. The first variant was detected in *MLH1*, and the other variant was in *PRDM16* gene, and the variant details are listed in Table 4.3. The family history revealed multiple breast cancer cases in the paternal side (Figure 4.6).

Table 4.3. Raw list of candidate variants detected in Patient 2.

Genomic Location	Ref	Alt	Zygosity	Gene	Transcript	Molecular Consequence	AA Change	ACMG Classification	CADD Phred	GERP++	Gnomad AF	SIFT
3:37028833	C	T	Het	*MLH1*	NM_000249.4	Stopgain	p.Arg487Ter	P (PVS1, PP5, PM2)	35	1.2	-	-
1:3417910	T	C	Het	*PRDM16*	NM_022114.4	Missense	p.Leu925Pro	VUS (PM2, PP3)	24.8	4.58	-	Tolerated

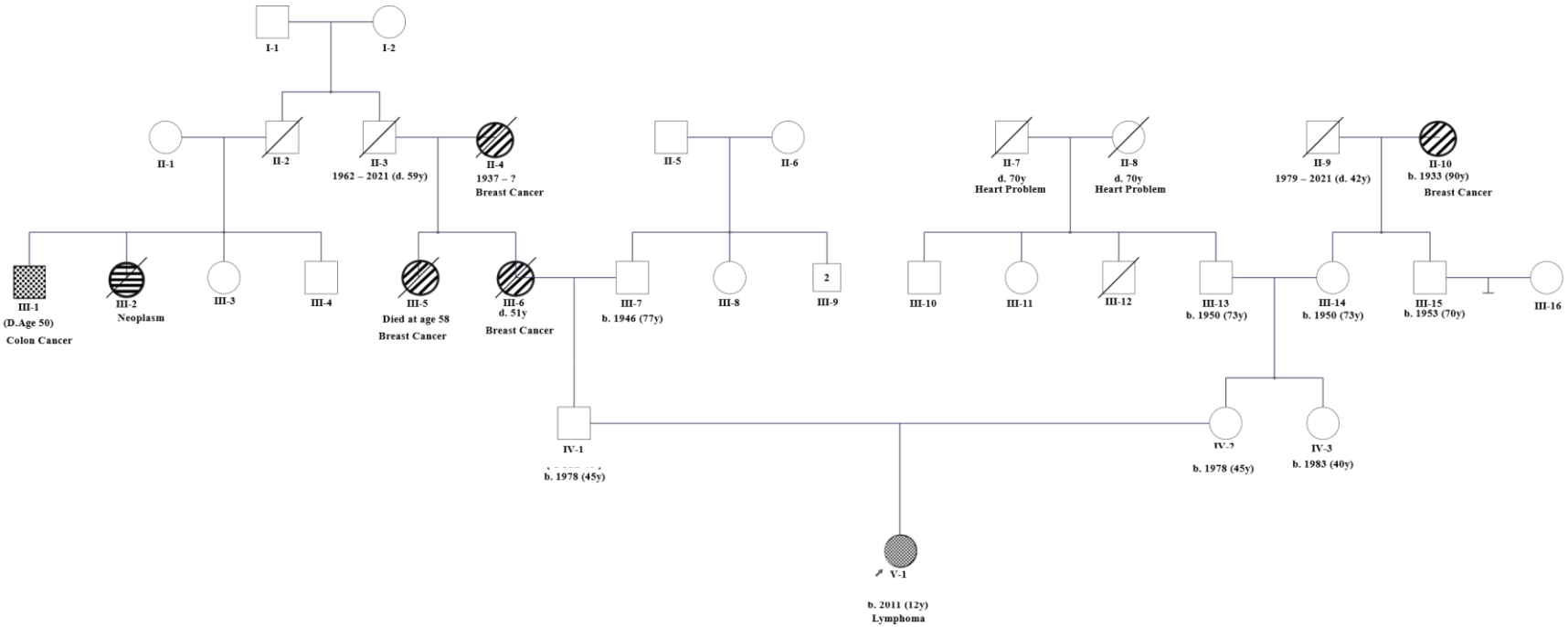

Figure 4.6. Pedigree of the Patient 2

After investigation of the raw candidate variant list based on the pedigree, it was determined a heterozygous pathogenic (PVS1, PP5, PM2) *MLH1* variant (NM_000249: c.1459C>T: p.Arg487Ter, rs63749795) as the cancer predisposition-causing variant. Based on the segregation analysis, it was detected that variant is inherited from the father, and the mother was wildtype (Figure 4.7).

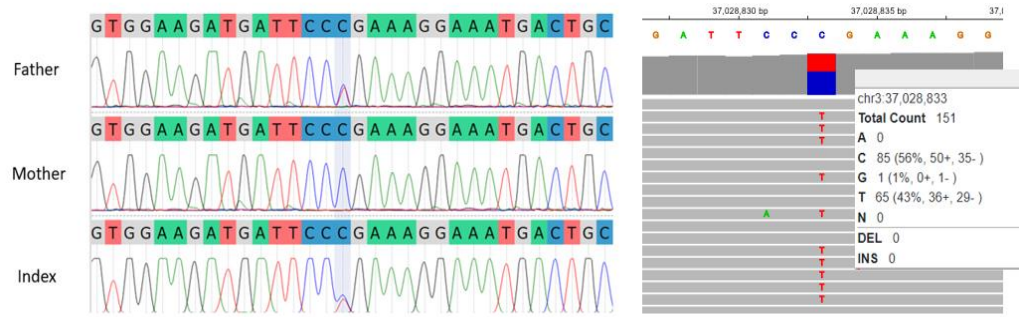

Figure 4.7. Segregation analysis of *MLH1* by Sanger sequencing of the Patient 2

4.3 Patient 3

Patient 3 was a three years old boy, born in a consanguineous family and diagnosed with glioneuronal tumor. Out of 9684 number of variants, with a stepwise filtering. We detected two variants that may cause cancer predisposition in the family. Initial candidate variants are listed in Table 4.4. Also, pedigree of the family is given in Figure 4.8.

Table 4.4. Raw list of candidate variants detected in Patient 3.

Genomic Location	Ref	Alt	Zygosity	Gene	Transcript	Molecular Consequence	AA Change	ACMG Classification	CADD Phred	GERP++	Gnomad AF	SIFT
10:43119548	G	A	Het	*RET*	NM_020975.6	Missense	p.Val804Met	P (PP5, PM5, PP3, PM2)	29.3	3.36	0.0001364	Damaging
21:34792308	A	G	Het	*RUNX1*	NM_001754.5	Missense	p.Ser424Pro	VUS (PM2, PP3)	28.5	4.58	-	Tolerated

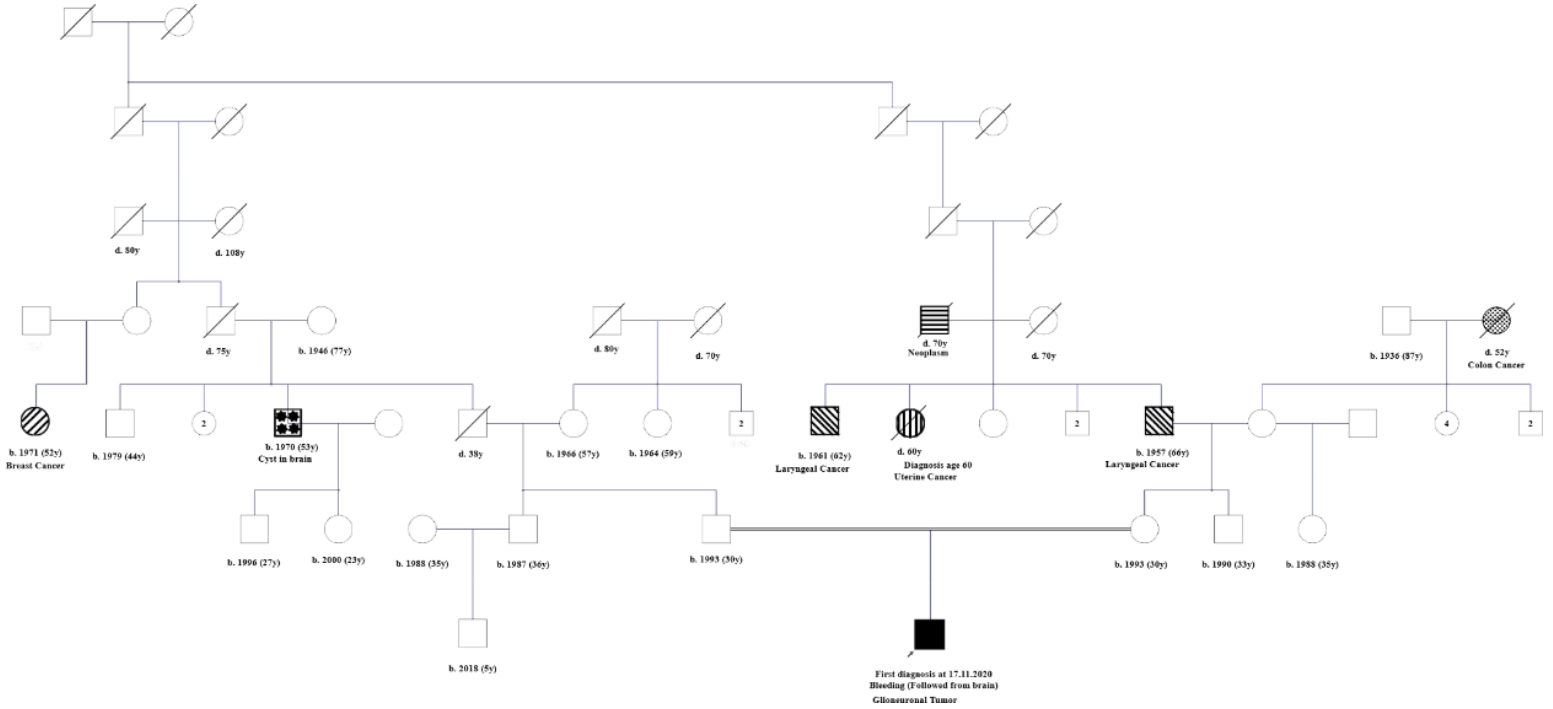

Figure 4.8. Pedigree of the Patient 3.

After investigation of the raw candidate variant list based on the pedigree, it was determined a heterozygous pathogenic (PM5, PP5, PP3, PM2) *RET* variant (NM_020975.6: c.2410G>A: p.Val804Met, rs79658334) as the cancer predisposition-causing variant. Based on the segregation analysis, it was detected that variant is inherited from the father, and the mother detected as wildtype (Figure 4.9). Since the family is consanguineous, we checked the maternal grandfather who was diagnosed with laryngeal cancer, however, he is detected as wildtype.

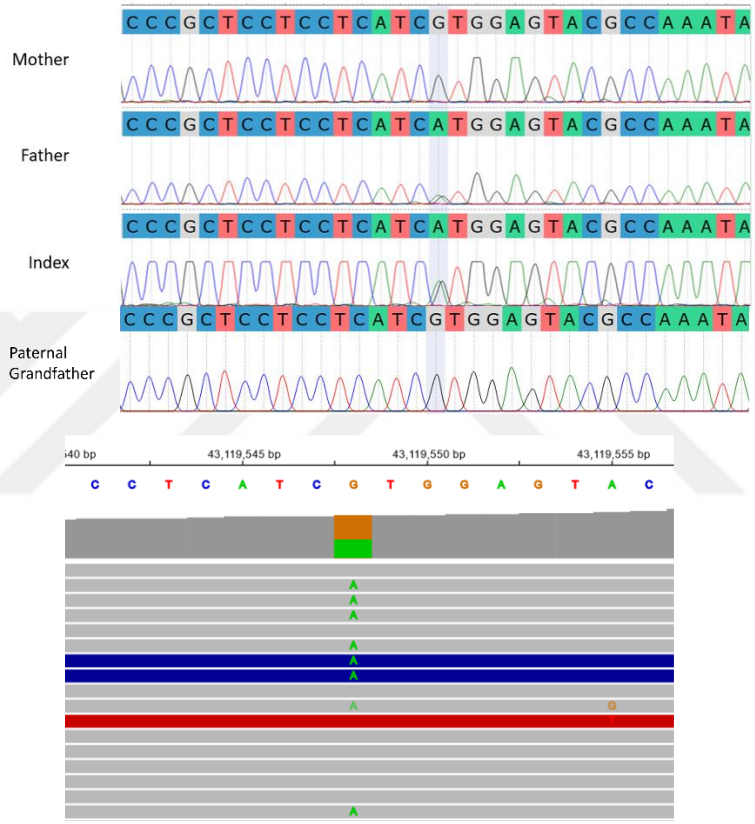

Figure 4.9. Segregation analysis of *RET* by Sanger sequencing of the Patient 3

4.4 Patient 4

In the female patient who is five years old and diagnosed with Wilms' tumor, after the stepwise filtering strategy, we detected three variants that may cause cancer predisposition in the family out of 10100 variants. Initial candidate variants are listed in Table 4.5. Also, pedigree of the family is given in Figure 4.10.

Table 4.5. Raw candidate variant list of the Patient 4.

Genomic Location	Ref	Alt	Zygosity	Gene	Transcript	Molecular Consequence	AA Change	ACMG Classification	CADD Phred	GERP++	Gnomad AF	SIFT
1:21225324	C	T	Het	*ECE1*	NM_001397.3	Missense	p.Gly656Arg	VUS (PM1, PP3, PP5, BS2)	28.5	5.36	0.000077818	Damaging
21:34792308	A	G	Het	*RUNX1*	NM_001754.5	Missense	p.Ser424Pro	VUS (PM2, PP3)	28.5	4.58	-	Tolerated

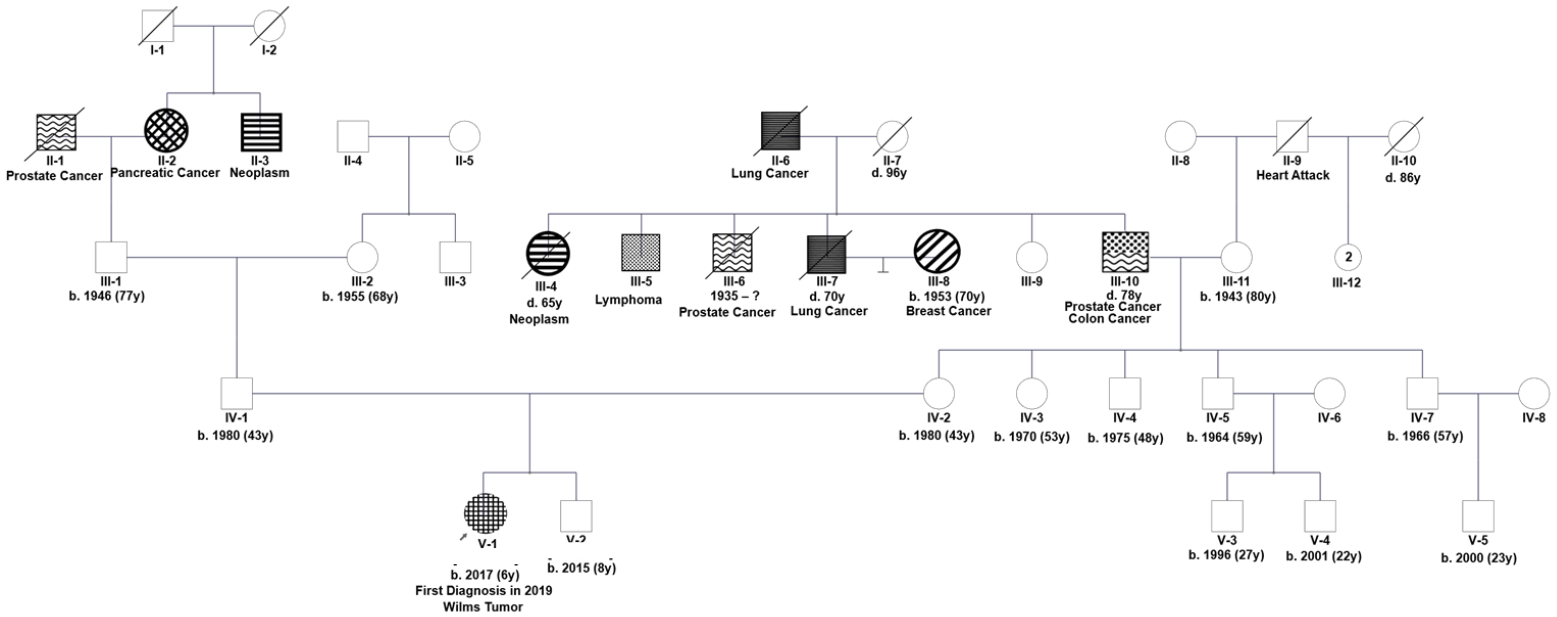

Figure 4.10. Pedigree of the Patient 4.

After investigation of the raw candidate variant list based on the diagnosis, none of the variants evaluated as the cancer predisposition-causing variant. Only *ECE1* VUS variant (NM_001397.3: c.1966G>A: p.Gly656Arg, rs367812436) was studied with Sanger sequencing for both confirmation and segregation analysis (Figure 4.11). Based on the segregation analysis, the index case and the mother detected as heterozygous, whereas father and brother detected as wildtype for the indicated variant.

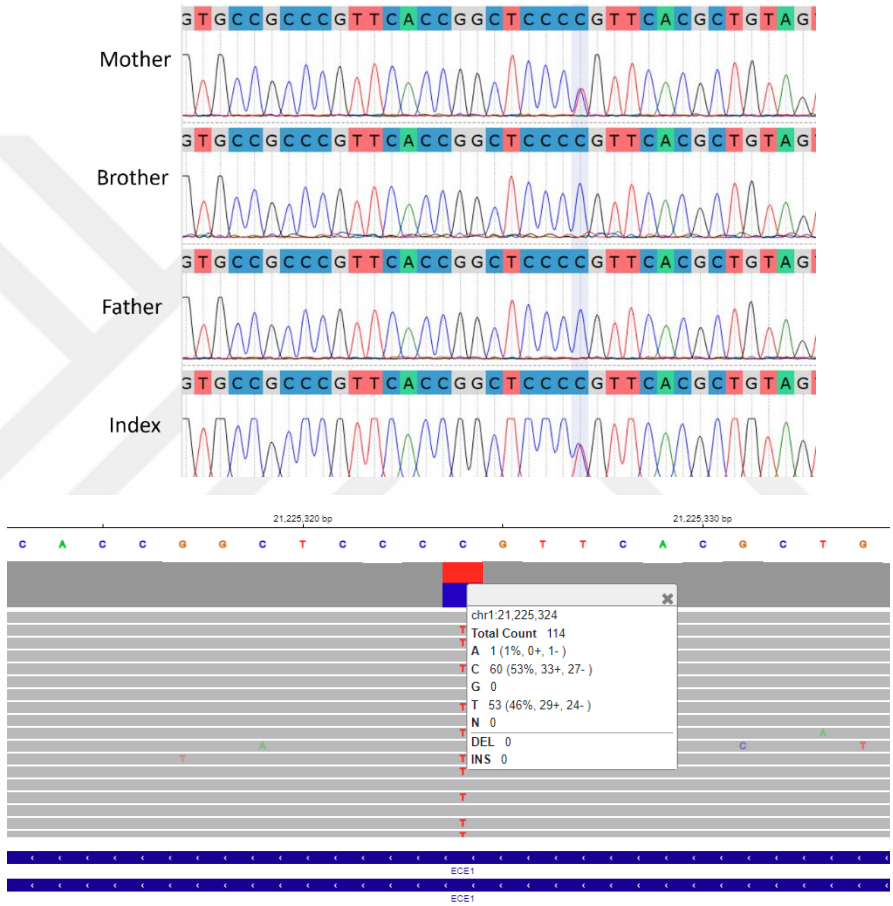

Figure 4.11. Segregation analysis of the Patient 4 for *ECE1* variant.

We did not perform any further experiments regarding the candidate variants as per the other candidate variants are not consistent with patient's clinical characteristics.

4.5 Patient 5

In the female patient who is eight years old and diagnosed with B-ALL, we detected two frameshift variants that may cause cancer predisposition in the family out of 9274 variants. Initial candidate variants are listed in Table 4.6. Also, pedigree of the family is given in Figure 4.12.

Table 4.6. Raw candidate variant list for the Patient 5.

Genomic Location	Ref	Alt	Zygosity	Gene	Transcript	Molecular Consequence	AA Change	ACMG Classification	CADD Phred	GERP++	Gnomad AF	SIFT
17:43057063	-	G	Het	*BRCA1*	NM_007294.4	Frameshift	p.Gln1756ProfsTer74	P (PVS1, PS3, PP5, PM2)	-	-	0.00018	-
13:32398608	C	GAATTATATCT	Het	*BRCA2*	NM_001754.5	Frameshift	p.Ser3366AsnfsTer4	P (PVS1, PP5, PM2)	-	-	-	-

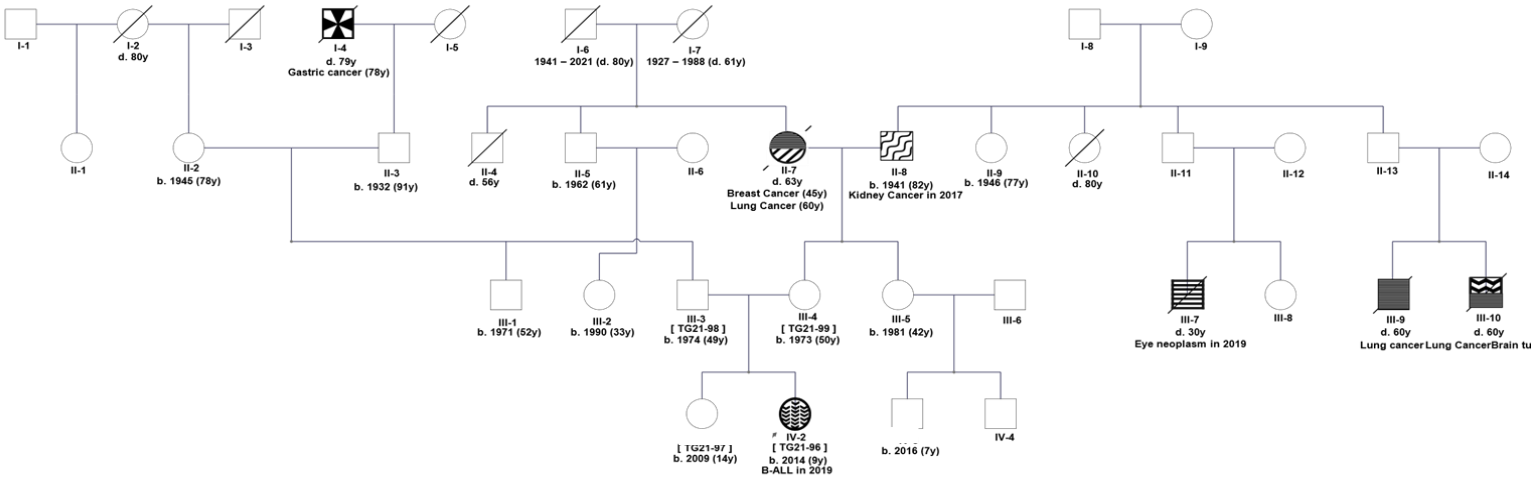

Figure 4.12. Pedigree of the Patient 5.

After investigation of the raw candidate variant list based on the pedigree, it was determined a heterozygous pathogenic (PVS1, PS3, PP5, PM2) *BRCA1* variant (NM_007294.4: c.5266dup: p.Gln1756ProfsTer74, rs80357906) as the cancer predisposition-causing variant. Based on the segregation analysis, it was detected that variant is inherited from the mother, and the father was detected as wildtype. Furthermore, we detected that the sister also inherited the variant (Figure 4.13). Lastly, we also Sanger sequenced the maternal grandfather who were diagnosed with kidney cancer; however, he detected as wildtype.

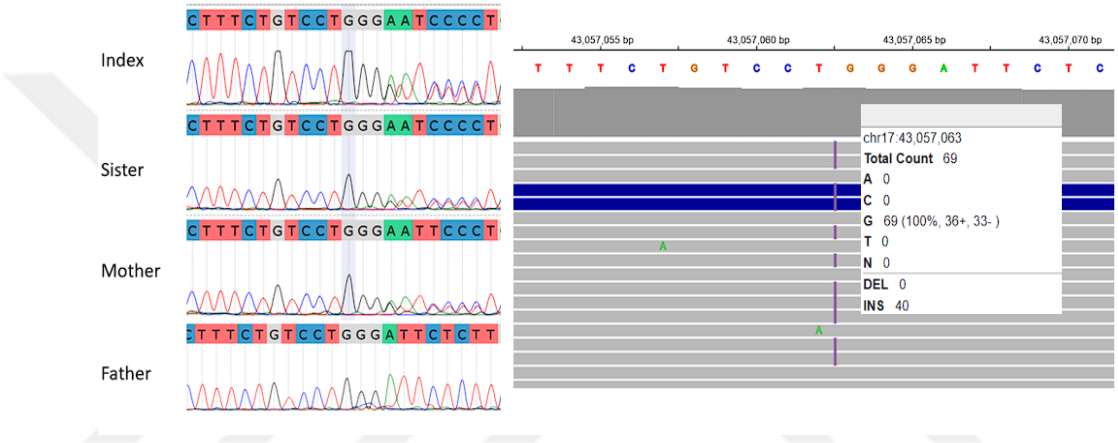

Figure 4.13. Segregation analysis of the Patient 5.

The other candidate variant is VUS (PVS1, PM2, BP6) *BRCA2* variant (NM_000059.4: c.10095delinsGAATTATATCT: p.Ser3366AsnfsTer4). Based on the segregation analysis, it was detected that variant is inherited from the father, furthermore, mother and sister of the index case detected as wildtype. (Figure 4.14)

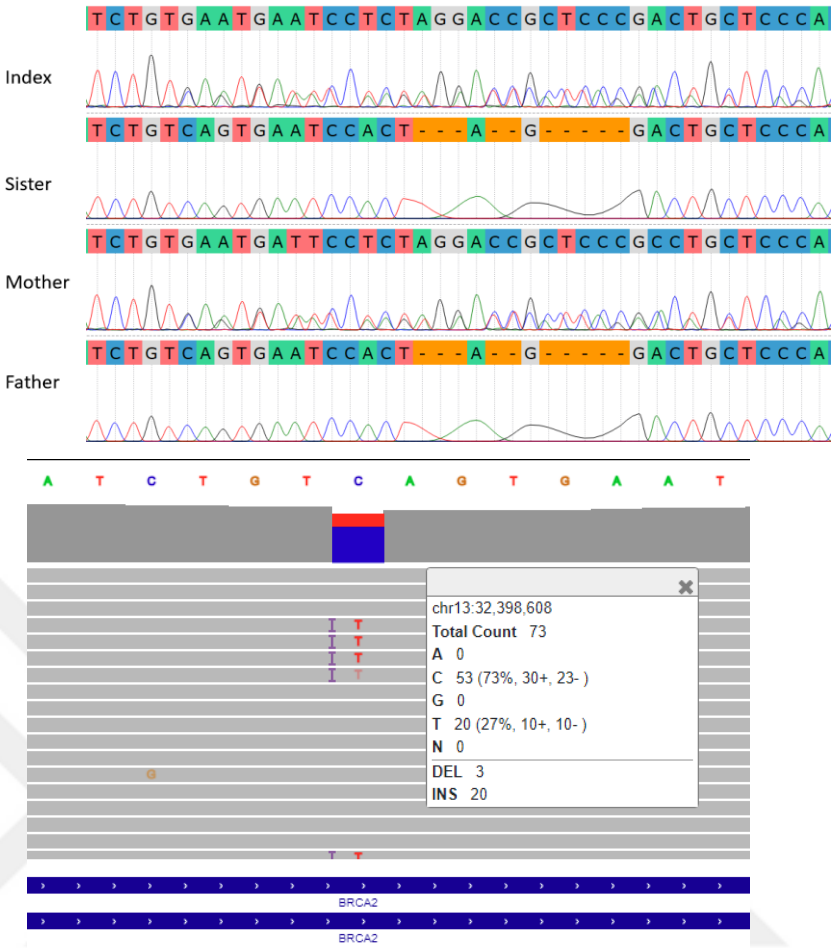

Figure 4. 14. Segregation analysis of the Patient 5 for *BRCA2* variant

4.6 Patient 6

In the female patient who is 16 years old and diagnosed with medulloblastoma, we detected two missense variants that may cause cancer predisposition in the family out of 9662 variants after stepwise filtering. Initial candidate variants are listed in Table 4.7. Also, pedigree of the family is given in Figure 4.15.

Table 4.7. Raw candidate variant list of the Patient 6

Genomic Location	Ref	Alt	Zygosity	Gene	Transcript	Molecular Consequence	AA Change	ACMG Classification	CADD Phred	GERP++	Gnomad AF	SIFT
2:211383815	A	C	Het	*ERBB4*	NM_005235.3	Missense	p.Trp1243Gly	LP (PP4, PP3, PM2, BP1)	26.4	5.72	-	Tolerated
3:37025986	G	T	Het	*MLH1*	NM_000249.4	Missense	p.Gly463Val	VUS (PM2, PP3)	21.2	-	-	Tolerated

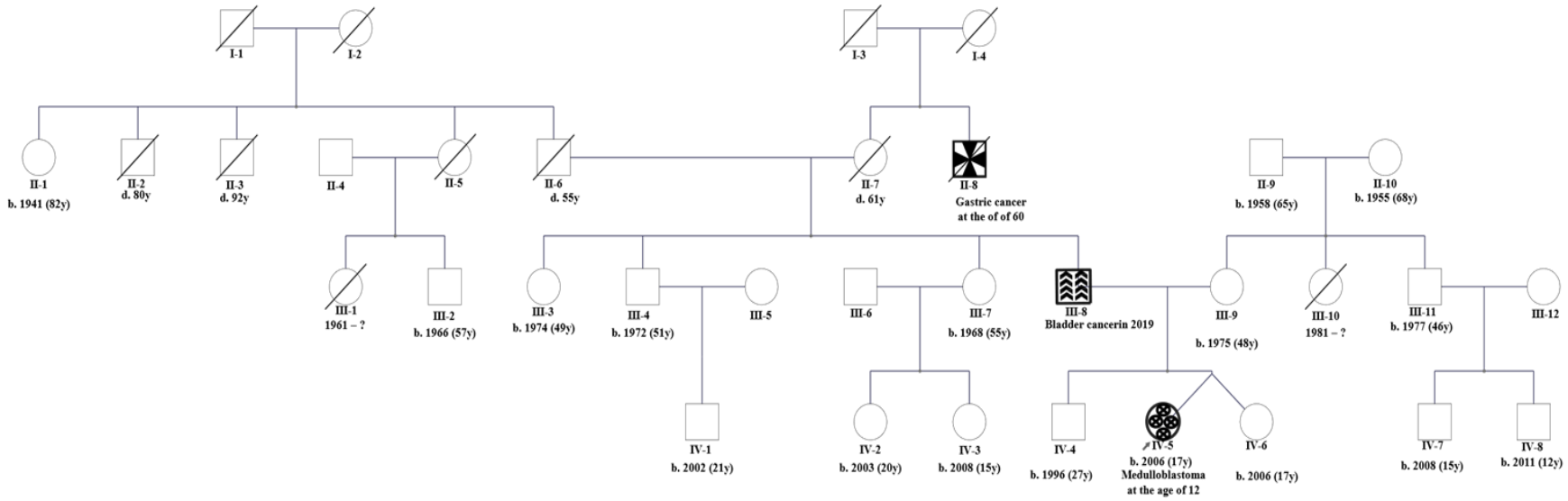

Figure 4.15. Pedigree of the Patient 6.

After investigation of the raw candidate variant list based on the pedigree, we determined a heterozygous LP (PP4, PP3, PM2, BP1) *ERBB4* variant (NM_005235.3: c.3727T>G: p.Trp1243Gly) as the cancer predisposition-causing variant. Based on the segregation analysis, it was detected that variant is inherited from the father, and the mother detected as wildtype. At this point, since the father was diagnosed with bladder cancer which is associated with *ERBB4* variants, PP1 parameter added into the variant classification. Therefore, pathogenic character of the variant was further improved. Furthermore, we detected that the twin sister and the brother of the index also inherited the variant (Figure 4.16).

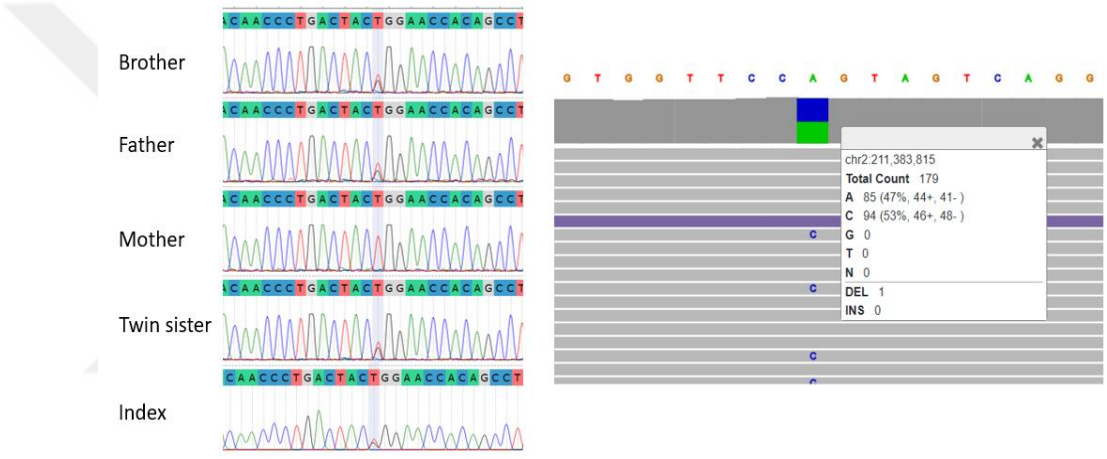

Figure 4.16. Segregation analysis of the Patient 6.

4.7 Patient 7

A 17 years old girl diagnosed with T-ALL has shown to possess three variants that may be related with cancer predisposition. Clinical exome sequencing revealed a total of 9920 variants and stepwise variant filtering ended up with two candidate variants. The first variant was detected in *BRCA2*, *LAMB3* and the other variant was in *FOXE1* gene, and the variant details are listed in Table 4.8. The family history revealed one breast cancer and one kidney cancer cases in the paternal side (Figure 4.17).

Table 4.8. Raw candidate variant list of Patient 7

Genomic Location	Ref	Alt	Zygosity	Gene	Transcript	Molecular Consequence	AA Change	ACMG Classification	CADD Phred	GERP++	Gnomad AF	SIFT
1:209629800	G	A	Het	*LAMB3*	NM_000228.3	Missense	p.Arg357Trp	VUS (PM2)	28.3	2.21	0.000003976	Damaging
13:32337722	A	G	Het	*BRCA2*	NM_000059.4	Missense	p.Ser1123Gly	VUS (PP3, PM2, BP1)	25.7	5.75	-	Damaging
9:97854422	GCCGCCG	-	Het	*FOXE1*	NM_004473.4	Frameshift	p.Ala170ProfsTer195	LP (PVS1,PM2)	-	-	-	-

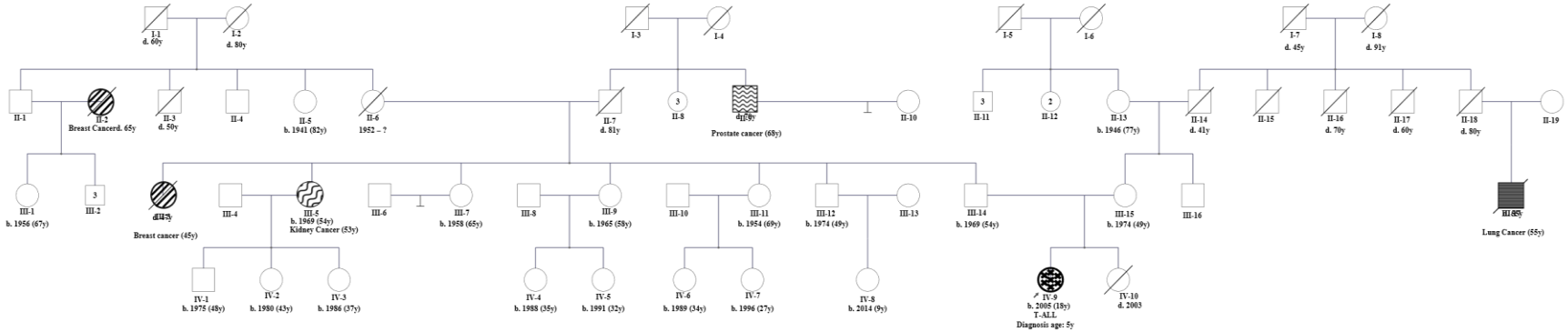

Figure 4.17. Pedigree of the Patient 7

After investigation of the raw candidate variant list based on the pedigree, it was determined a heterozygous VUS (PP3, PM2, BP1) *BRCA2* variant (NM_000059.4: c.3367A>G: p.Ser1123Gly, rs80358581) as the cancer predisposition-causing variant. Based on the segregation analysis, it was detected that variant is inherited from the father, and the mother detected as wildtype (Figure 4.18). Furthermore, paternal aunt of the index case who were diagnosed with kidney cancer was also Sanger sequenced for the indicated variant, however, she was detected as wildtype.

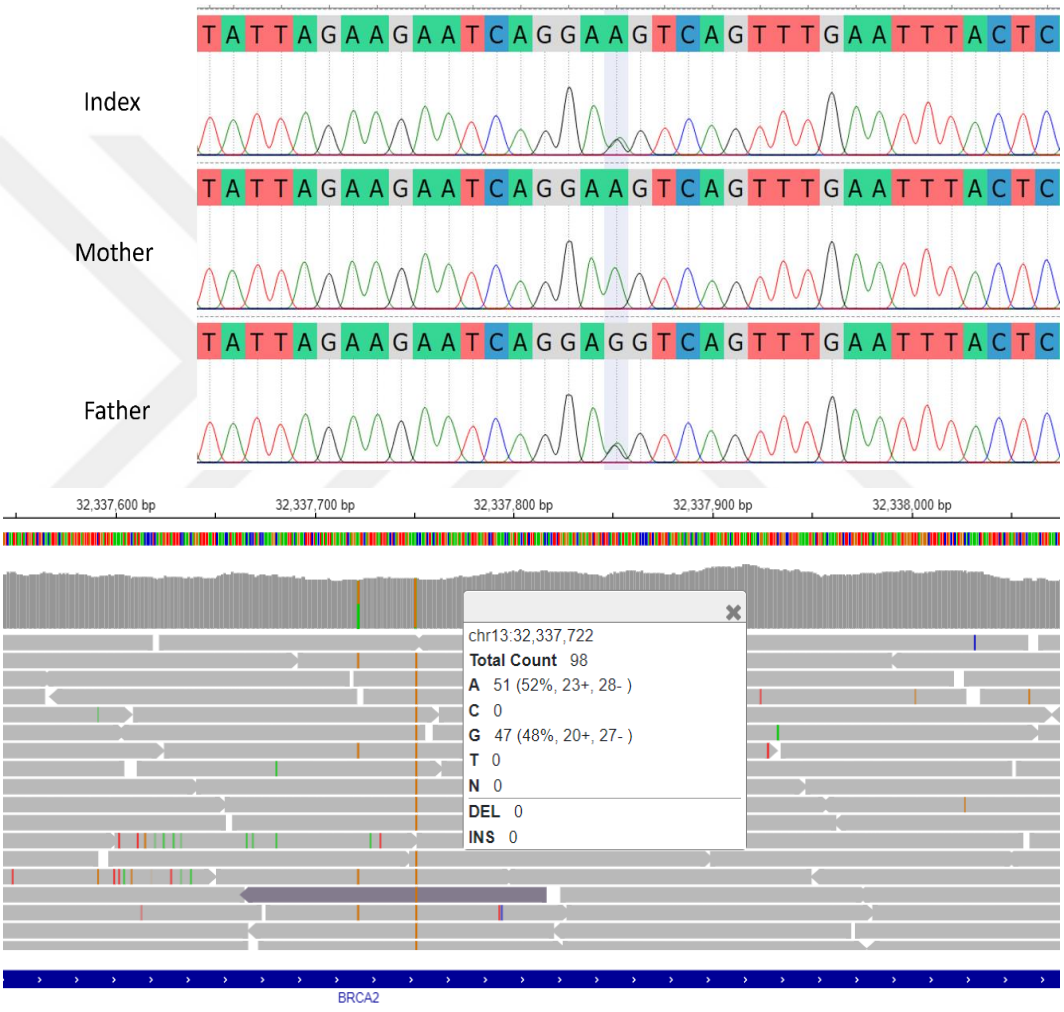

Figure 4.18. Segregation analysis of the Patient 7 for *BRCA2* variant

4.8 Patient 8

A 13 years old girl diagnosed with B-ALL has shown to possess two variants that may be related with cancer predisposition. Clinical exome sequencing revealed a total

of 9279 variants and stepwise variant filtering ended up with two candidate variants. The first variant was detected in *ABCC6* and the other variant was in *NKX2-1* gene, and the variant details are listed in Table 4.9. The family history revealed multiple cancer cases in both maternal and paternal sides (Figure 4.19).

Table 4.9. Raw variant list of the Patient 8

Genomic Location	Ref	Alt	Zygosity	Gene	Transcript	Molecular Consequence	AA Change	ACMG Classification	CADD Phred	GERP++	Gnomad AF	SIFT
16:16154766	C	T	Het	*ABCC6*	NM_001171.6	Missense	p.Arg1357Gln	VUS/LP	29.8	3.63	0.0000938	Damaging
14:36519338	T	G	Het	*NKX2-1*	NM_001079668.3	Missense	p.His37Pro	VUS/LP	27.3	5.12	0.0038975	Damaging

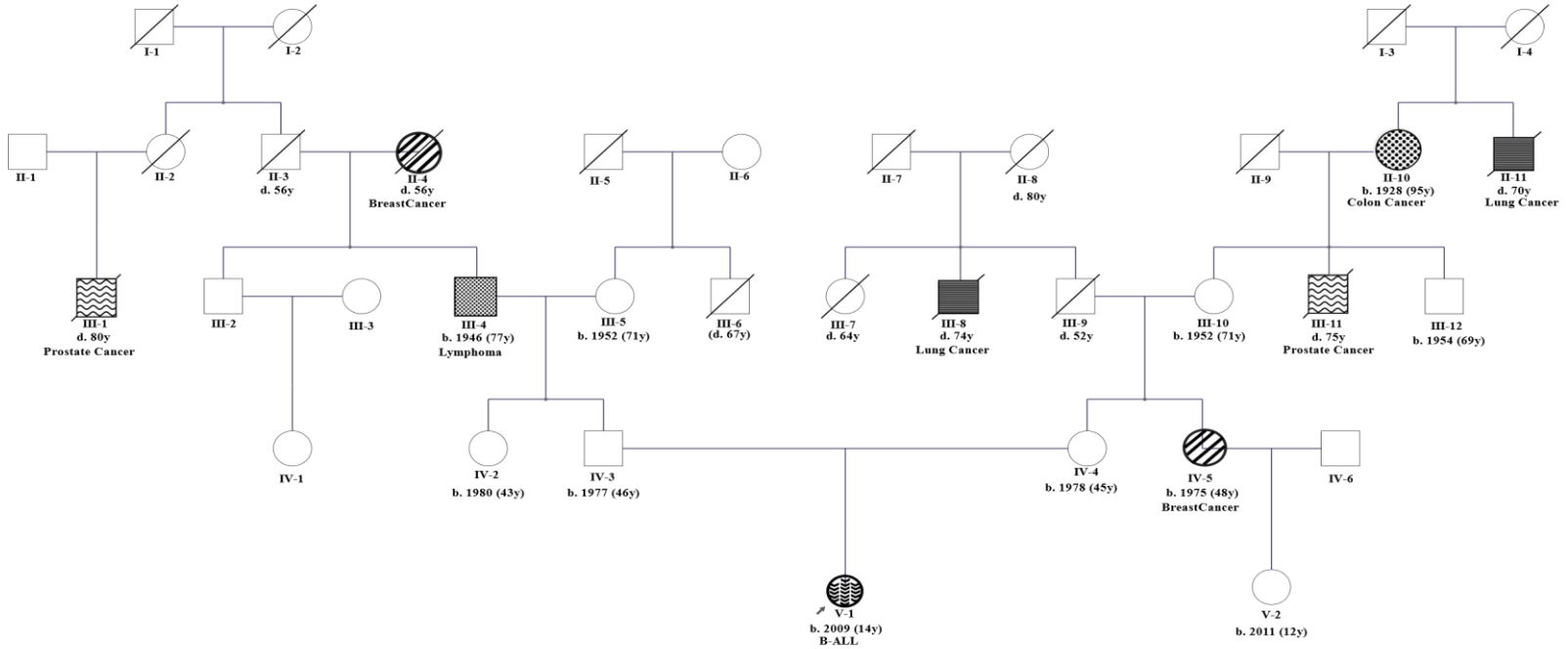

Figure 4.19. Pedigree of the Patient 8

After investigation of the raw candidate variant list based on the pedigree, it was determined a heterozygous LP (PP3, PM5 PM2, PP5) *ABCC6* variant (NM_001171.6: c.4070G>A: p.Arg1357Gln, rs201275608) as the cancer predisposition-causing variant. Based on the segregation analysis, it was detected that variant is inherited from the father, and the mother detected as wildtype (Figure 4.20).

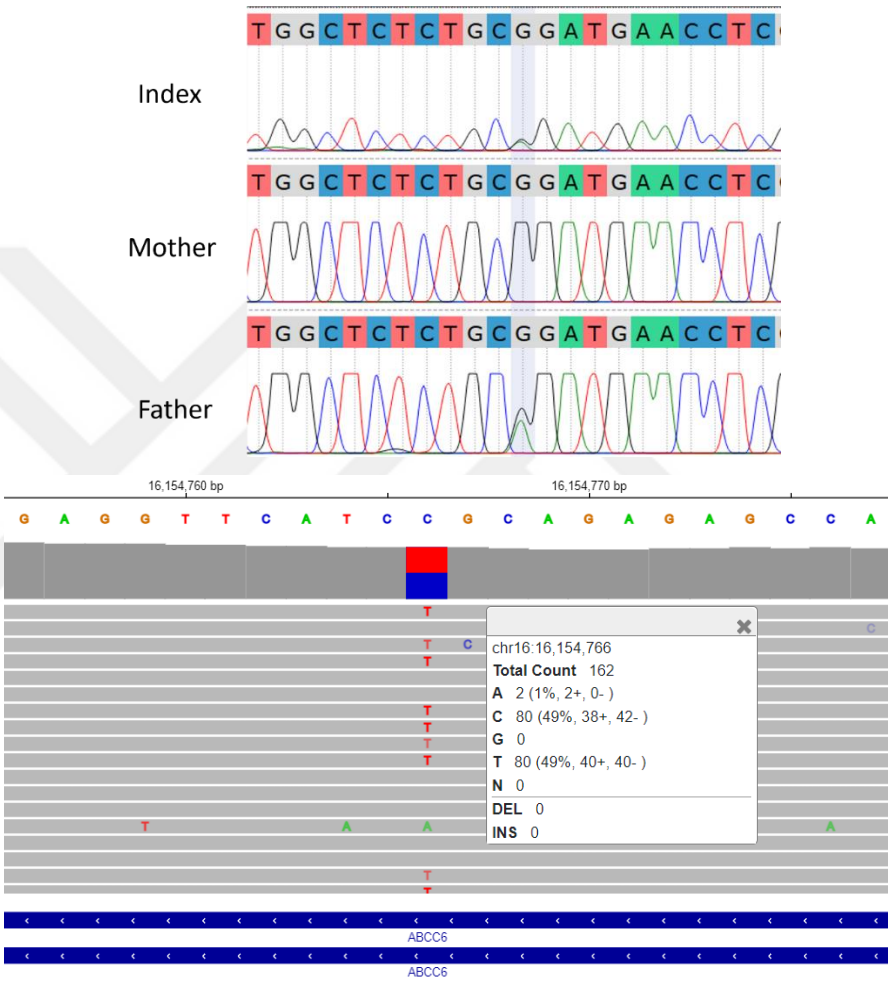

Figure 4.20. Segregation analysis of the Patient 8 for *ABCC6* variant.

It was also investigated that *NKX2-1* variant (NM_001079668.3. c.110A>C: p.His37Pro) may be the cancer predisposing variant in the patient as this variant is classified as VUS/LP (PM2, PP2). However, after the Sanger sequencing, it was observed that all three member of the family is wildtype. (Figure 4.21)

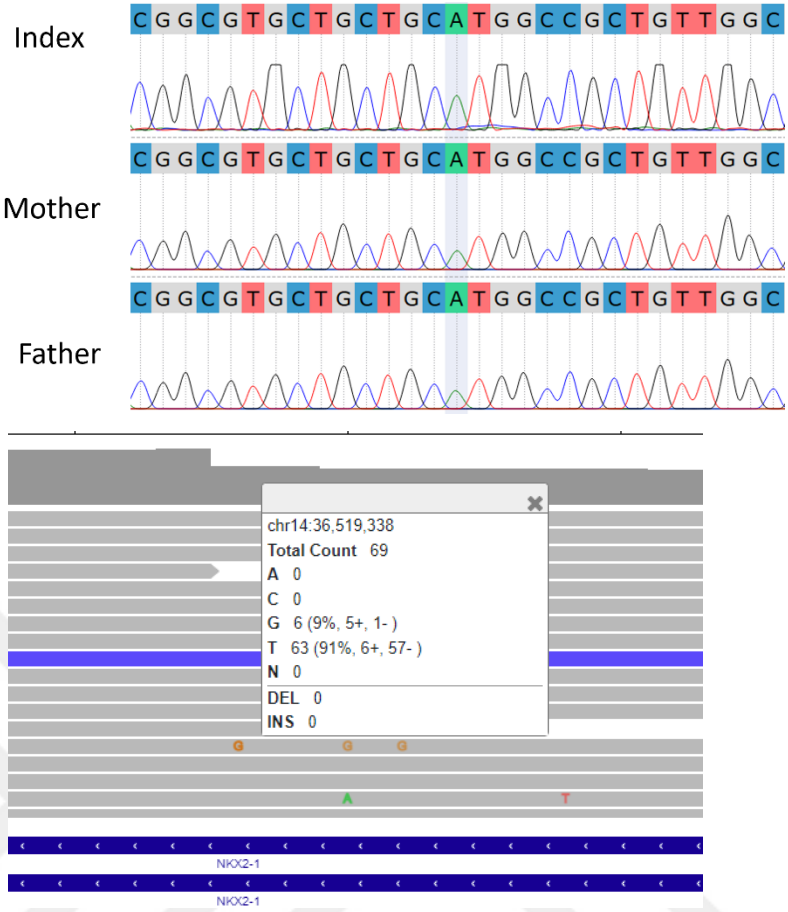

Figure 4.21. Segregation analysis of the Patient 8 for *NKX2-1* variant.

4.9 Patient 9

A 13 years old girl diagnosed with Pilocytic astrocytoma has shown to possess two variants that may be related with cancer predisposition. Clinical exome sequencing revealed a total of 11592 variants and stepwise variant filtering ended up with two candidate variants. The first variant was detected in *FLT3*, and the other variant was in *PRRT2* gene, and the variant details are listed in Table 4.10. The family history revealed multiple cancer cases in both maternal and paternal sides (Figure 4.22).

Table 4.10. Raw candidate variant list of the Patient 9.

Genomic Location	Ref	Alt	Zygosity	Gene	Transcript	Molecular Consequence	AA Change	ACMG Classification	ADD Phre	GERP++	Gnomad AF	SIFT
13:28023414	AATGT	-	Het	*FLT3*	NM_004119.3	Frameshift	p.Thr784Ter	P	-	-	-	-
16:29813695	-	C	Het	*PRRT2*	NM_145239.3	Frameshift	p.Arg217ProfsTer	P	-	-	0.00310262818	-

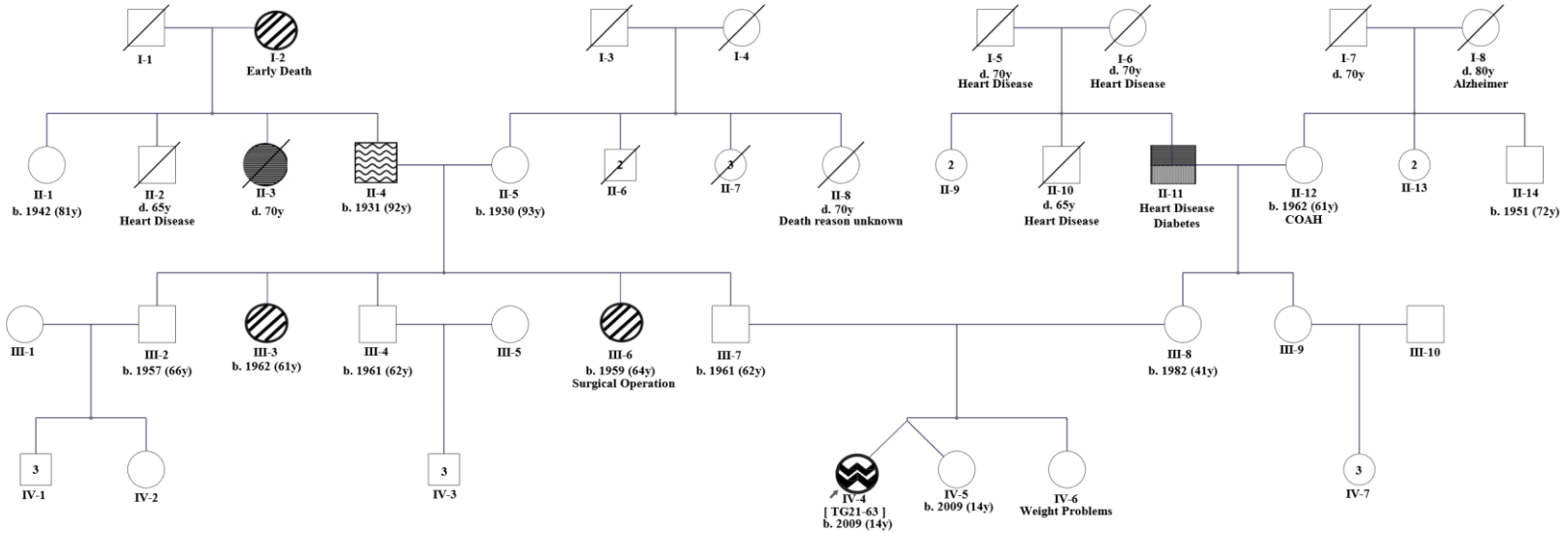

Figure. 4.22. Pedigree of the Patient 9.

After investigation of the raw candidate variant list based on the pedigree, it was determined a heterozygous P (PVS1, PM2) *FLT3* variant (NM_004119.3: c.2350_2354del: p.Thr784Ter) as the cancer predisposition-causing variant. Based on the segregation analysis, it was detected that variant is inherited from the mother, and the father detected as wildtype. Furthermore, it was observed that twin sister did not inherit the variant whereas older sister inherited. (Figure 4.23).

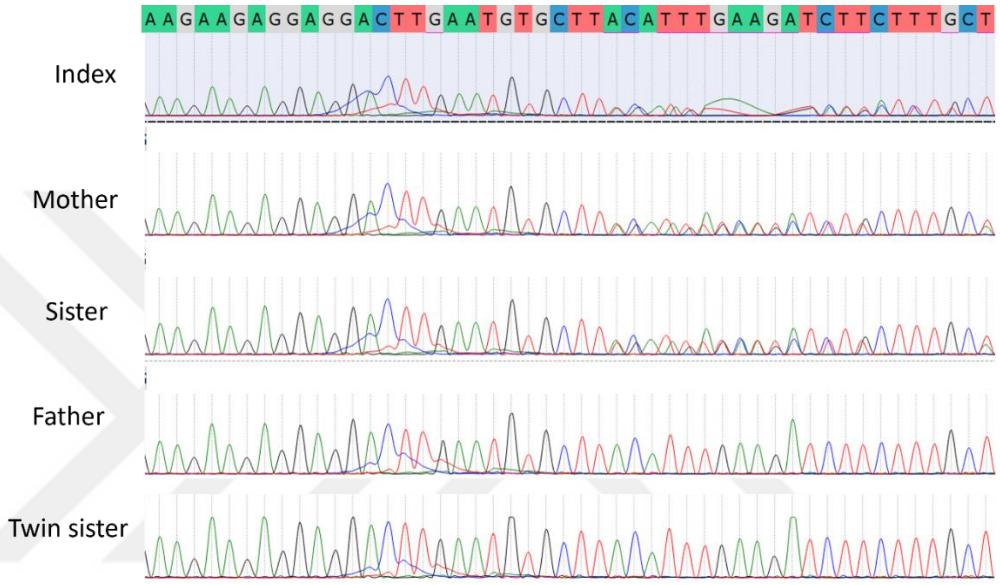

Figure 4.23. Segregation analysis of the Patient 9 for *FLT3* variant

After further investigation of the raw candidate variant list based on the pedigree, it was determined a heterozygous P (PVS1, PM2, PS3, PP1, PS4) *PRRT2* variant (NM_145239.3: c.649dup: p.Arg217ProfsTer8) as the potential cancer predisposition-causing variant. There was not clinical association with the *PRRT2* variant and we did not perform any segregation analysis.

4.10 Patient 10

A 13 years old boy diagnosed with glioblastoma has shown to possess two variants that may be related with cancer predisposition. Clinical exome sequencing revealed a total of 9918 variants and stepwise variant filtering ended up with two candidate variants. The first variant was detected in *TGFBR2*, and the other variant was in *RYR1* gene, and the variant details are listed in Table 4.11. The family history revealed multiple cancer cases in both maternal and paternal sides (Figure 4.24)

Table 4.11. Raw candidate variant list for the Patient 10

Genomic Location	Ref	Alt	Zygosity	Gene	Transcript	Molecular Consequence	AA Change	ACMG Classification	CADD Phred	GERP++	Gnomad AF	SIFT
19:38499970	A	-	Het	*RYR1*	NM_000540.3	Frameshift	p.Tyr2426LeufsTer4	LP	-	-	-	-
3:30671728	T	C	Het	*TGFBR2*	NM_003242.6	Missense	p.Val182Ala	VUS	22.5	5.28	-	Damaging

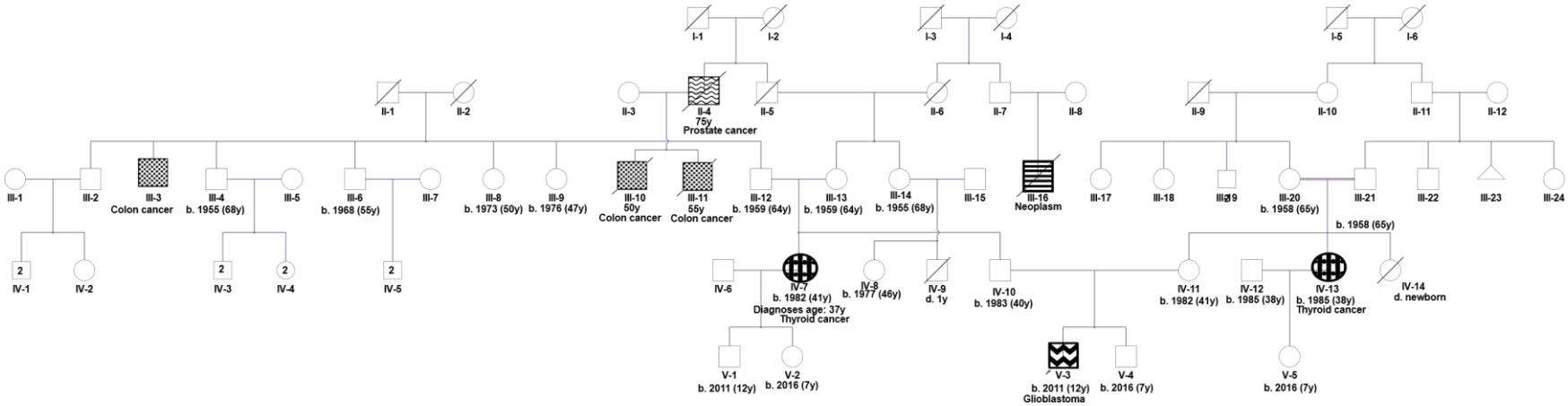

Figure 4.24. Pedigree of the Patient 10

After investigation of the raw candidate variant list based on the pedigree, it was suggested a heterozygous VUS (PP3, PM2) *TGFBR2* variant (NM_003242.6: c.545T>C: p.Val182Ala) might be the cancer predisposition-causing variant. Sanger sequencing was not performed for confirmation and segregation analysis. Another candidate variant is a novel *RYR1* variant (NM_000540.3: c.7277del: p.Tyr2426LeufsTer4) and was classified as LP (PVS1, PM2). Like the *TGFBR2* variant, Sanger sequencing was not performed for confirmation and segregation analysis. IGV image for both *TGFBR2* and *RYR1* variants are shown in Figure 4.25, respectively.

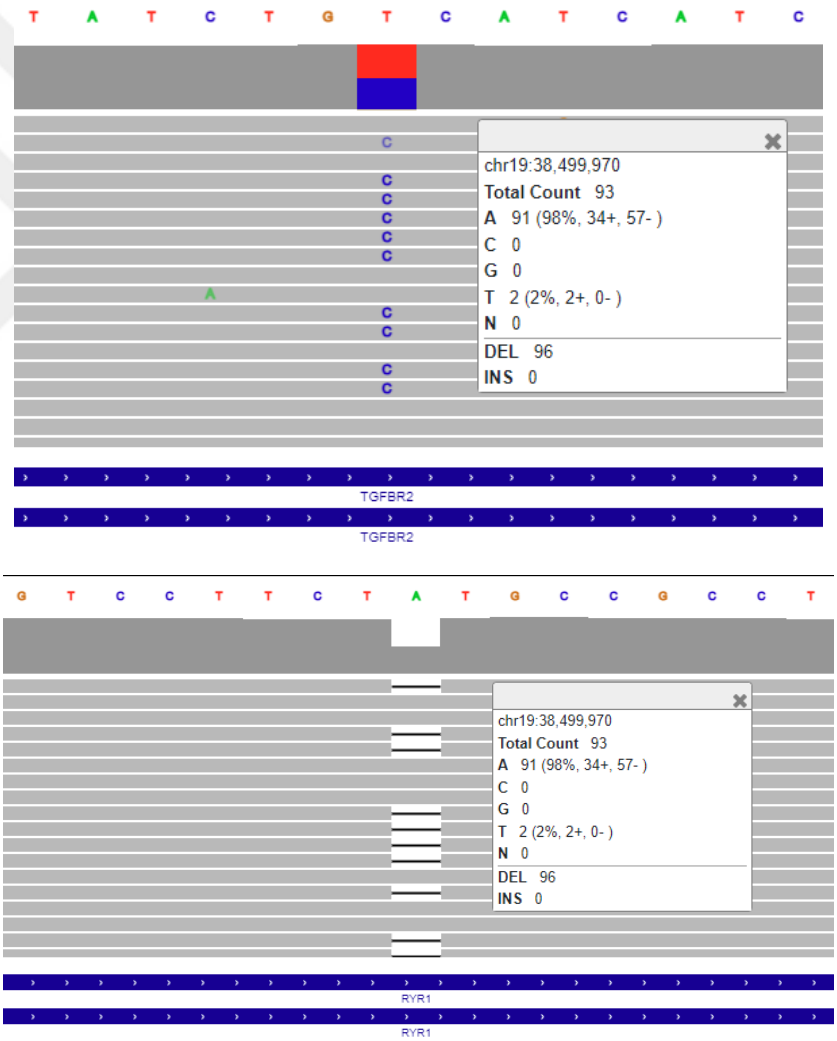

Figure 4.25. IGV images for both candidate variants. *(TGFBR2 on the top, RYR1 at the bottom*

5 DISCUSSION

Cancer is considered as a multifactorial disorder, developed due to genetic and environmental factors. On the other hand, 10-15% of all cancers are shown to be hereditary. Pediatric cancers were also considered as sporadic events whereas recent studies showed that around 10% of the cases harbor a germline variant in genes that are related to cancer predisposition (78, 79). Today, next generation sequencing based multigene panel testing is very effective in determining germline variants in cases with cancer predisposition indications. Pediatric oncologists prioritize achieving a cure as swiftly as possible, which can result in overlooking cancer predisposition during the intensive treatment decision-making process. On the other hand, practical guidelines can be employed to discern cases with a possible germline predisposing variant. For instance, Jongmann et al reported six main indications for cancer predisposition evaluation; 1) positive family history of the same or related cancers; 2) bilateral, multifocal, or multiple cancers; 3) cancer development at an earlier age than sporadic tumors of the same type; 4) accompanying symptoms of a predisposition syndrome; 5) occurrence of specific tumor types that frequently occur in the context of genetic predisposition; and 6) high toxicity during treatment (73). If there is even one of these indications, the patients should be referred to cancer genetics clinics for further investigation. Assessing pediatric cancers in accordance with these guidelines will not only serve to inform families about future occurrences but also has the potential to alter treatment protocols, thereby enhancing management and outcome.

Here we studied pediatric cancer patients evaluated according to Jongman's criteria and studied ten index cases via CES, to determine the genes/variants that cause cancer predisposition within the families. Based on the results, 23 different variants were detected in ten patients. Among these 23 variants; 14 of them were novel, whereas nine of them had extremely low allele frequencies.

Zhang et al. revealed that the *TP53, APC, BRCA2, NF1, PMS2, RB1* and *RUNX1* were the most frequently mutated genes in the affected patients (79). On the contrary, we did not detect any *TP53* variants in our study group, whereas we determined

candidate variants in *APC, BRCA2* and *RUNX1* genes. The main difference is of course the limited number of cases enrolled to our study. We were able to group candidate variants into three main cellular pathways: DNA damage repair, intracellular processing, and transcription factors. In our cohort, there were seven patients who had solid tumors. Although not solid based on the family history we also enrolled three hematologic malignancies to our study group. Based on the results, candidate variants were majorly in the genes which are responsible for intracellular processing specifically intracellular signaling. There is only one gene in the DNA damage repair pathway. We did not detect any genes in transcription factor pathway among solid tumor patients. DNA damage repair system deficiencies lead to a variety of different tumors. For example, *BRCA1* and *BRCA2* are well-known DNA repair system genes which are frequently associated with familial breast and ovarian cancer. Furthermore, *BRCA1* and *BRCA2* are also associated with prostate and pancreatic cancer as well. (80) Besides solid tumors, DNA damage incapacity leads to genomic instability and may cause hematologic malignancies like leukemia. In our patient cohort, *BRCA1* and *BRCA2* were observed in B-ALL and T-ALL patients. However, based on the pedigrees, we observed that there are different family members who have solid tumors. Another DNA damage repair gene is *MLH1*. In our cohort, deleterious *MLH1* variant was detected in the patient diagnosed with lymphoma. However, *MLH1* variants are frequently associated with colon cancer. On the other hand, tumors and malignancies might be caused by malfunctioned intracellular processes. Overexpressed proteins and/or malfunctioned intracellular transportation might increase the risk of cancer development. We identified eight candidate variants among seven solid tumor patients. These variants are abundantly in intracellular signaling pathway, including *ERBB4, FLT3, RET, RYR1 and TGFBR2.* The studies on the relation between intracellular signaling and tumor development provided significant insights into tumor development. *RET and ERBB4* are directly related with PI3K/AKT and JAK/STAT pathways, which are both now to be directly related with tumorigenesis. These findings are important especially from the targeted therapy point of view and the studies are fast growing in the field

The heterozygous *XPA* VUS variant was determined in a Wilms Tumor case, whose maternal relatives including the maternal grandmother were diagnosed with different tumor types including uterine, lung, pancreas cancer and lymphoma. XPA is crucial for nucleotide excision repair (NER) complex formation at DNA damage sites (81). NER pathway mainly works as excising the mutated part of the DNA, it has regulatory roles in the cell for cellular metabolism via facilitating genome stability (82, 83). Furthermore, the NER pathway plays a role in cancer predisposition since it repairs the damaged sites of the genome. Therefore, any defect on the NER pathway may initiate the tumorigenesis (84). Hence, variants on *XPA* gene may cause cancer predisposition as well. Based on the literature, variants in NER pathway genes significantly increase the risk for WT (83). Furthermore, in the segregation analysis, we observed that the variant was inherited from the mother whose mother was diagnosed with uterine cancer at the age of 56. It was shown that variants on NER pathway genes might increase the risk for endometrial cancer (85). However, we were not able to contact the other family members to take blood samples, therefore, further segregation analysis was not applicable. However, it was observed that all cancer cases are on the maternal side. During the analysis of the variants, genes that are related to WT predisposition were also covered (86) as well and no known WT related genes showed any P/LP or VUS variants. The other two candidate genes were *NOTH3* and *AP4B1*. NOTCH family is shown to be directly related with hematologic malignancies and NOTCH3 was found to be dysregulated in various cancers there was no indication of relation with WT or other tumors that were diagnosed within the family (87). On the other hand, *AP4B1* variant is located at the skirts of a known hereditary spastic paraplegia region which is highly expressed in central nervous system, and we did not further investigate the variant.

The pathogenic *MLH1* variant was detected in 11 years old lymphoma patient. *MLH1* is one of the genes that form DNA mismatch repair (MMR) complexes along with *MSH2, MSH6, and PMS2*. MMR complex works as recognizing and repairing the mismatched sequences as well as insertions and deletions (88, 89). Thus, deleterious variants on MMR complex genes cause Lynch Syndrome that is a type of cancer syndrome and inherited autosomal dominantly (90). Most frequent cancer seen in the

Lynch Syndrome is the colorectal cancer which is followed by endometrium cancer. Moreover, approximately 30% of the MMR deficit patients are being diagnosed with hematologic malignancies such as lymphomas or leukemias (91, 92). Our index patient was also diagnosed with lymphoma and the variant was previously reported in ClinVar (VCV000089744.28). Furthermore, in the literature, Lynch Syndrome with *MLH1* variants were shown to increase the breast cancer rate more than *MLH2* variants (93). The segregation analysis, showed that variant was inherited from father and the index patient had paternal grandmother who was diagnosed with breast cancer at the age of 45, later deceased at the age of 51. Also, the sister and mother of the paternal grandmother were diagnosed and later deceased due to breast cancer at the ages of 58 and 50, respectively. However, since all breast cancer cases were deceased, we were not able to screen those cases but *MLH1* variant can explain the lymphoma development and Lynch related breast cancer cases.

One of the interesting cases in this study was the three years old boy, born into a consanguineous family and diagnosed with glioneuronal tumor. Central nervous system (CNS) tumors are the most common cancer type among pediatric cancers and in 2021 WHO updated the classification of CNS tumors. The physician reported that tumor in this patient was radiologically and histologically presented a totally different clinic and most of the expected molecular markers like BRAF KIAA-1549 fusion, other MAPK/ERK pathway alterations, NTRK1/K2 fusions, alteration in 1p and/or 19q were not present in tumor tissue. Moreover, the pedigree indicated multiple cancer cases both from paternal and maternal side. Hence the case was referred to our clinic for further investigation. We determined a pathogenic *RET* variation which was also found in the tumor tissue. This is so far the only glioneuronal tumor case harboring a *RET* variation. *RET* is a proto-oncogene which encodes a transmembrane receptor tyrosine kinase. Deleterious *RET* variants increase the downstream signaling through multiple effectors, and affect multiple pathways including cell proliferation and cell growth (94). In children and young adults, *RET* alterations known to cause malignancies like glioma, inflammatory myofibroblastic tumor, lipofibromatosis, and majorly papillary thyroid carcinoma with the percentage of 22% to 45% (95). According to the literature, 40% to 50% of sporadic medullary thyroid carcinomas

have activating point *RET* variants. Predisposition to medullary thyroid carcinoma is one of the characteristics of multiple endocrine neoplasia type 2 (MEN2), a hereditary autosomal-dominant cancer condition caused by germline *RET* mutations. (96, 97). Additionally, paternal uncle of the mother and maternal grandfather of the patient were diagnosed with laryngeal cancer at the ages of 61 and 45, respectively. We have also checked the maternal grandfather for *RET* gene but he was wild type. Laryngeal cancer is reported to be developed due to tobacco use and no clear evidence was shown for genetic factors. On the other hand, recent studies determined higher frequencies of polymorphism regions which should be further investigated. (98) The other candidate variants were observed in *RUNX1* and *APC* genes. In the IGV analysis the *RUNX1* variant could not be detected which means that the read was false positive.

The heterozygous *ECE1* VUS variant was determined in a Wilms Tumor case, whose maternal relatives diagnosed with different tumor types including prostate, colon, lung cancers and lymphoma. *ECE1* is a gene that takes role in proteolytic processes of endothelin conversion, and frequently associated with Hirschsprung disease (99). *ECE1* gene is responsible for endothelin conversion and endothelin is responsible for vasoconstriction. Even though *ECE1* and cancer relationship is not enlightened enough, there are studies pointing to *ECE1* might be causing cancer due to vasoconstriction properties of the endothelin (100, 101). Furthermore, pathogenic variants of the *ECE1* are frequently associated with Hirschsprung disease. Also, *ECE1* variants are associated with increase in aggressiveness of prostate and colon tumors (102) which are observed in the maternal site. Last candidate variant was detected in *APC* gene which mainly important as tumor suppressor gene. Pathogenic *APC* variants may cause activation of WNT pathway which is a major characteristic of colon cancers. (103) Furthermore, it is also indicated that Wilms tumors are mainly related to kidney development. WNT pathway is crucial for kidney embryogenesis, therefore, *APC* variants may indirectly cause Wilms tumor (104). The other candidate variants were observed in *RUNX1* and *APC* genes. In the IGV analysis the *RUNX1* variant could not be detected which means that the read was false positive. Based on the results of this thesis, it must be highlighted that scientists must check the IGV image

of the variants in case of false positive results. In this thesis, some of the variants that were listed as candidate, were determined to be false positive.

A heterozygous pathogenic (PVS1, PS3, PP5, PM2) *BRCA1* variant (*BRCA1*: c.5266dup: p.Gln1756ProfsTer74, rs80357906) variant was detected in the Patient 3 who is diagnosed with B-ALL. *BRCA1* is a well-known cancer predisposition gene due to its tumor suppressor properties via DNA damage repair (105-107). Even though pathogenic *BRCA1* variants are frequently associated with breast cancer and ovarian cancer, they may cause different types of cancers such as bladder, melanoma, leukemia, lymphoma, colon and more as well (108). The variant we detected in the index case was previously reported in ClinVar (ClinVar Accession: VCV000017677.88). Furthermore, we observed that this variant was inherited from the mother whose parents are also diagnosed with cancer. The maternal grandfather of the index case who was diagnosed with kidney cancer studied and the result was wildtype for the same variant. However, the maternal grandmother was diagnosed with breast cancer at the age of 45, later diagnosed with lung cancer at the age of 60 and deceased. The mother of the index case most likely inherited the variant from her mother but, we were not able to perform further segregation analysis. We also detected a *BRCA2* variation, and the first classification was showing the variant as likely pathogenic. However, further investigation showed conflicting results in ClinVar, and benign classifications were abundant. Therefore, we left out the *BRCA2* variant as a candidate variant.

In the Patient 6 who were diagnosed with medulloblastoma, we observed a heterozygous VUS (PP3, PM2) in *ERBB4* (ERBB4: c.3727T>G: p.Trp1243Gly). ERBB4 is a member of a receptor tyrosine kinase family which is crucial for mitogenesis and differentiation of various tissues including the central nervous system. Hence, these properties are proposing ERBB receptor family as cancer susceptibility genes. Within the ERBB receptor family, *ERBB4* is crucial since it encodes the only protein that undergoes proteolytic procedures with ligand binding (109). Furthermore, it is indicated that ERBB receptor family, especially *ERBB4*, are implicated in cancer development (110, 111). For example, pathogenic variants on *ERBB4* are associated

with different cancers like lung cancer, melanoma, medulloblastoma and more (111-113). In our case, the index patient was diagnosed with medulloblastoma which is related to the candidate gene. However, the candidate variant was classified as VUS based on ACMG criteria. After segregation analysis, we observed the father as heterozygous for the variant. Since the father also has bladder cancer, PP1 criteria was added to the classification which upgrades the VUS classification into likely pathogenic. Furthermore, the maternal uncle of the father was also diagnosed with gastric cancer at the age of 60, later deceased at the same age due to cancer. Since *ERBB4* variants are associated with bladder and gastric cancers (114, 115), we concluded that the candidate variant can explain the clinical status of the patient.

We detected a previously reported (ClinVar accession: VCV000051455.7) heterozygous VUS in *BRCA2* gene in the Patient 5. *BRCA2* is a well-known cancer predisposition gene that works as a tumor suppressor gene and is responsible for genome stability via DNA damage repair pathway (116). Even though pathogenic *BRCA2* variants are mostly associated with breast and ovarian cancer, they may cause prostate, gastric, pancreatic, laryngeal cancers (117) and hematological cancers like lymphoma and leukemia as well (118). Here we identified a *BRCA2* variant in the ALL-diagnosed index case who inherited the variant from her father. In addition, the father of the patient has a deceased sister who was diagnosed with breast cancer at the age of 45. However, since she is deceased, Sanger sequencing was not applicable for the further segregation analysis. There were two other candidate variants *LAMB3* and a LP *FOXE1* variant. However, IGV images of both variants revealed that they were false positive. Therefore, we did not further investigate these two variants.

Two candidate variants were determined in the 13-year-old Patient 8 who was diagnosed with B-ALL. The first variant is a VUS/LP *ABCC6* variant. *ABCC6* gene is a member of ATP-binding cassette (ABC) transporter family which is responsible for extracellular and intracellular transportation of various molecules. There is still limited information about the relation between *ABCC6* and cancer development (119), however, studies show that ABC transporter family is associated with different cancers including hematological malignancies (120, 121). Another study showed *ABCC2-6*

expressions is increased in the pediatric relapsed ALL patients (122). Segregation analysis revealed that the variant is inherited from the mother and the father detected as wildtype. As the pedigree considered, maternal aunt of the index is diagnosed with breast cancer at early an age (34 years). Furthermore, there are different cancer cases in the maternal site including prostate, colon, and lung cancers. There is no clear evidence for *ABCC6* and breast cancer, however, one study in 2015 suggested that the downregulation of ABC transport proteins is observed in prostate cancer cells when compared to noncancerous cells (123). Another candidate variant was observed in *NKX2-1* gene and classified as VUS. *NKX2-1* is a thyroid specific transcription factor and plays crucial role in regulation of thyroid specific genes. There are studies that *NKX2-1* indirectly promotes the cancer progression through different pathways (124, 125). Another study in 2011 suggested that *NKX2-1* is in T-ALL oncogenic pathway (126). Even though our knowledge is limited regarding the *NKX2-1* and cancer development, we evaluated this variant as candidate. The IGV analysis revealed that this variant is also false positive read and do not exist in the patient.

In the 15-year-old Patient 9 who were diagnosed with pilocytic astrocytoma, we observed a heterozygous LP/P (PVS1, PM2) variant in the FLT3 gene. FLT3 is a receptor tyrosine kinase that is crucial for survival, differentiation, and proliferation of the hematopoietic cells and is frequently associated with acute myeloid lymphoma (AML) (127-129). In our case, we detected a deleterious frameshift variant. The *FLT3* gene is highly expressed in brain tissues (130), additionally, there are studies that show FLT3 is also associated with tumor formation in different brain tissues (131, 132). After the segregation analysis, we observed that this variant is inherited from the mother whose father was diagnosed and deceased due to lymphoma and lung cancer at the unknown ages. Even though *FLT3* variants are not frequently associated with ALL or lymphoma (133), it is closely related with AML. *FLT3* variants play a role in survival and proliferation of leukemic and lymphoblastic cells (134-136). On the other hand, on the paternal side two aunts of the index case were diagnosed with breast cancer at the ages of 60 and 63. Furthermore, the paternal grandfather, whose mother was also diagnosed with breast cancer was diagnosed with prostate cancer at an unknown age. Thus, based on the family history, we suspected that the paternal side

may have hereditary cancer syndrome. Therefore, it will be suggested further genetic tests like WES for the paternal side due to strong cancer history of the family if the family wants genetic counselling. Another candidate variant was a pathogenic *PRRT2* variant. *PRRT2* gene encodes a transmembrane protein, and studies on mice revealed that the gene is abundantly expressed in central nervous system during the embryonic and postnatal stages. Another study suggests that increased expression of *PRRT2* causes malfunctions in cellular processes such as viability and apoptosis in glioma cells (137). There is no clear evidence that suggests the *PRRT2* deleterious variants cause breast cancer. However, in a study in 2017, it was observed that expression of *PRRT2* with deleterious variants increases the proliferation, migration, and invasion in prostate and colorectal cell lines (138).

The VUS *TGFBR2* and LP *RYR1* variants were detected in 13 years old glioblastoma patient. Firstly, *TGFBR2* is a gene that encodes a transmembrane protein which regulates the transcription of the cell proliferation, tumor formation, cell cycle arrest related genes through kinase activities. Therefore, it is evaluated as cancer susceptibility gene. According to a study conducted, *TGFBR2*, along with *ERBB3* and *IGF1R* were observed highly expressed in the cells (139). Moreover, another study showed that variants on *TGFBR2* may cause thyroid cancers (140) which was observed in multiple family members of the patient. The segregation analysis is ongoing. The other candidate variant was detected on *RYR1* gene. *RYR1* encodes a protein which is responsible for calcium release, abundantly in skeletal muscles. Even though *RYR1* and cancer development is not investigated thoroughly, it is though that *RYR1* plays role via calcium signaling. (141) Even though the novel *RYR1* variant was classified as LP (PVS1, PM2), *RYR1* is not directly associated with the clinical characteristics of the patient. Yet, according to OMIM, deleterious *RYR1* variants are associated with autosomal dominant King-Denborough syndrome (Phenotype MIM number: 619542). Physical features of the patients with King-Denborough syndrome include myopathic features, low-set ears, ptosis, hypertelorism, hypotonia, muscle weakness and more with highly variable phenotype. Based on the OMIM information, primary physician of the patient was called to check if the patient shows any of these phenotypes. As of

the date of this thesis, the patient does not have any of these symptoms. Hence, the variant is not studied further investigated.

Although we were able to identify candidate variants in the majority of our cohort there were some limiting factors of our study. Firstly, our approach was based on clinical exome sequencing possessing around 4000 genes so the whole genomic profile is not covered. Moreover, it is not possible to perform structural variant analysis so the cases should be reevaluated for SVs as well. Another obstacle was that the affected family members were mostly deceased or were not cooperative, which made it impossible to determine the candidate variants in those cases. Most importantly, this is a very limited cohort mainly reflecting the patients from Marmara region and further investigations and larger cohort are needed to obtain better results for childhood cancer predisposition in Turkey.

This thesis study provides various novelties from different aspects.

-We were able to install a cancer predisposition awareness among physicians at pediatric hematology and oncology services.

-Strong novel candidate variants were identified in the majority of the families, which can be further investigated by functional studies.

- Families with a cancer history were able to receive genetic counseling regarding the cancer risks and prevention

-Patients' management and treatment was modified together with the findings of this thesis

-Even though WES or WGS are frequently used for detecting cancer predispositions, this thesis proved that CES can be a strong tool for cancer predisposition detection in places where higher technologies are not easy to reach.

6 CONCLUSION

In conclusion, even though whole genome sequencing (WGS) and WES are widely used in detection of cancer predisposition variants, the study shows that CES can provide an alternative and cost-effective technique for detecting cancer predisposition variants in all of the cases. For the cases that need further investigation, WES or WGS should be applied to the index case along with other effected members in the families. For the VUS variants, functional experiments will be useful for the reclassification of the variant pathogenicity, along with frequent checking of databases and literature. Furthermore, based on the detected variants and their clinical importance, necessary genetic counseling was given to the families for enlightening them in terms of risks and predisposition. Lastly, WGS and WES data need strong computing infrastructure and large data storage for data processing and storage. This is crucial for the countries which cannot afford WGS or WES, or groups that do not have strong hardware options for WGS or WES data processing.

7 REFERENCES

1.Rahner N, Steinke V. Hereditary Cancer Syndromes. Deutsches Ärzteblatt International. 2008/10;105(41).

2.Vetsch J, Wakefield CE, Tucker KM, McCarthy M, Signorelli C, Walwyn T, et al. Genetics-related service and information needs of childhood cancer survivors and parents: a mixed-methods study. European Journal of Human Genetics. 2020/01;28(1).

3.The landscape of genomic alterations across childhood cancers - PubMed. Nature. 03/15/2018;555(7696).

4.Clinical exome sequencing for genetic identification of rare Mendelian disorders - PubMed. JAMA. 11/12/2014;312(18).

5.Meldrum C, Doyle MA, Tothill RW. Next-generation sequencing for cancer diagnostics: a practical perspective. Clin Biochem Rev. 2011;32(4):177-95.

6.Nagahashi M, Shimada Y, Ichikawa H, Kameyama H, Takabe K, Okuda S, et al. Next generation sequencing- based gene panel tests for the management of solid tumors. Cancer Science. 2019/01;110(1).

7.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 - PubMed. Genetics in medicine : official journal of the American College of Medical Genetics. 2015 May;17(5).

8.Sha K, Boyer LA. The chromatin signature of pluripotent cells. 2009/05/31.

9.PDQ Cancer Information Summaries. 2002.

10.The complete sequence of a human genome - PubMed. Science (New York, NY). 2022 Apr;376(6588).

11.Ozkan E, Lacerda MP. Genetics, Cytogenetic Testing And Conventional Karyotype. StatPearls. Treasure Island (FL): StatPearls Publishing

Copyright © 2024, StatPearls Publishing LLC.; 2024.

12.The Human Genome Project: big science transforms biology and medicine - PubMed. Genome medicine. 09/13/2013;5(9).

13.Finishing the euchromatic sequence of the human genome - PubMed. Nature. 10/21/2004;431(7011).

14.Initial sequencing and analysis of the human genome - PubMed. Nature. 02/15/2001;409(6822).

15.Implications of the Human Genome Project for medical science - PubMed. JAMA. 02/07/2001;285(5).

16.A global reference for human genetic variation - PubMed. Nature. 10/01/2015;526(7571).

17.Alternative Splicing Regulatory Networks: Functions, Mechanisms, and Evolution - PubMed. Molecular cell. 10/17/2019;76(2).

18.The pathobiology of splicing - PubMed. The Journal of pathology. 2010 Jan;220(2).

19.Non-coding RNA: It's Not Junk - PubMed. Digestive diseases and sciences. 2017 May;62(5).

20.A repetitive elements perspective in Polycomb epigenetics - PubMed. Frontiers in genetics. 10/08/2012;3.
21.Deep intronic mutations and human disease - PubMed. Human genetics. 2017 Sep;136(9).
22.SVInterpreter: A Comprehensive Topologically Associated Domain-Based Clinical Outcome Prediction Tool for Balanced and Unbalanced Structural Variants - PubMed. Frontiers in genetics. 12/01/2021;12.
23.Realizing the significance of noncoding functionality in clinical genomics - PubMed. Experimental & molecular medicine. 08/07/2018;50(8).
24.Krieghoff-Henning E, Folkerts J, Penzkofer A, Weg-Remers S. Cancer – an overview. Med Monatsschr Pharm. 2017;40(2):48-54.
25.Cancer and apoptosis: The apoptotic activity of plant and marine natural products and their potential as targeted cancer therapeutics - PubMed. Frontiers in pharmacology. 08/10/2022;13.
26.Cancer Genetics - PubMed. The Surgical clinics of North America. 2020 Jun;100(3).
27.Delineating the role of binge eating in cancer research - PubMed. Eating and weight disorders : EWD. 2021 Oct;26(7).
28.Cancer statistics, 2024 - PubMed. CA: a cancer journal for clinicians. 2024 Jan-Feb;74(1).
29.Awareness and current knowledge of breast cancer - PubMed. Biological research. 10/02/2017;50(1).
30.Smoking and lung cancer - PubMed. Annals of cardiothoracic surgery. 2014 Mar;3(2).
31.Age of onset in familial cancer - PubMed. Annals of oncology : official journal of the European Society for Medical Oncology. 2008 Dec;19(12).
32.Compare and contrast: pediatric cancer versus adult malignancies - PubMed. Cancer metastasis reviews. 2019 Dec;38(4).
33.Unleashing the immune response against childhood solid cancers - PubMed. Pediatric blood & cancer. 2017 Oct;64(10).
34.Timeliness of diagnosis and treatment: the challenge of childhood cancers - PubMed. British journal of cancer. 2021 Dec;125(12).
35.Epidemiology of childhood cancer - PubMed. Cancer treatment reviews. 2010 Jun;36(4).
36.Classics in oncology. Heredity with reference to carcinoma as shown by the study of the cases examined in the pathological laboratory of the University of Michigan, 1895-1913. By Aldred Scott Warthin. 1913 - PubMed. CA: a cancer journal for clinicians. 1985 Nov-Dec;35(6).
37.Hereditary cancer predisposition syndromes - PubMed. Journal of clinical oncology : official journal of the American Society of Clinical Oncology. 01/10/2005;23(2).
38.Mutation and cancer: statistical study of retinoblastoma - PubMed. Proceedings of the National Academy of Sciences of the United States of America. 1971 Apr;68(4).
39.Pediatric Cancer Predisposition and Surveillance: An Overview, and a Tribute to Alfred G. Knudson Jr - PubMed. Clinical cancer research : an official journal of the American Association for Cancer Research. 06/01/2017;23(11).

40.Cancer predisposition genes: molecular mechanisms and clinical impact on personalized cancer care: examples of Lynch and HBOC syndromes - PubMed. Acta pharmacologica Sinica. 2016 Feb;37(2).
41.Is it time for BRCA1/2 mutation screening in the general adult population?: impact of population characteristics - PubMed. Genetics in medicine : official journal of the American College of Medical Genetics. 2015 Jan;17(1).
42.Genome-wide association studies of cancer: current insights and future perspectives - PubMed. Nature reviews Cancer. 2017 Nov;17(11).
43.Familial adenomatous polyposis - PubMed. Orphanet journal of rare diseases. 10/12/2009;4(1).x
44.PTEN Hamartoma Tumor Syndrome/Cowden Syndrome: Genomics, Oncogenesis, and Imaging Review for Associated Lesions and Malignancy - PubMed. Cancers. 06/22/2021;13(13).
45.Developmental abnormalities and cancer predisposition in neurofibromatosis type 1 - PubMed. Current molecular medicine. 2009 Jun;9(5).
46.DNA sequencing with chain-terminating inhibitors - PubMed. Proceedings of the National Academy of Sciences of the United States of America. 1977 Dec;74(12).
47.The sequence of sequencers: The history of sequencing DNA - PubMed. Genomics. 2016 Jan;107(1).
48.Overview of Next-Generation Sequencing Technologies - PubMed. Current protocols in molecular biology. 2018 Apr;122(1).
49.Targeted Sequencing Approach and Its Clinical Applications for the Molecular Diagnosis of Human Diseases - PubMed. Cells. 02/02/2023;12(3).
50.Repetitive DNA and next-generation sequencing: computational challenges and solutions - PubMed. Nature reviews Genetics. 11/29/2011;13(1).
51.The third generation sequencing: the advanced approach to genetic diseases - PubMed. Translational pediatrics. 2020 Apr;9(2).
52.The Sanger FASTQ file format for sequences with quality scores, and the Solexa/Illumina FASTQ variants - PubMed. Nucleic acids research. 2010 Apr;38(6).
53.SamQL: a structured query language and filtering tool for the SAM/BAM file format - PubMed. BMC bioinformatics. 10/02/2021;22(1).
54.The Sequence Alignment/Map format and SAMtools - PubMed. Bioinformatics (Oxford, England). 08/15/2009;25(16).
55.The variant call format and VCFtools - PubMed. Bioinformatics (Oxford, England). 08/01/2011;27(15).
56.A framework for variation discovery and genotyping using next-generation DNA sequencing data - PubMed. Nature genetics. 2011 May;43(5).
57.Sá P, Guimarães L, Graças D, Veras A, Barh D, Azevedo V, et al. Next-Generation Sequencing and Data Analysis. 2018. p. 191-207.
58.Large-scale validation of single nucleotide polymorphisms in gene regions - PubMed. Genome research. 2004 Aug;14(8).
59.Effects of short indels on protein structure and function in human genomes - PubMed. Scientific reports. 08/24/2017;7(1).

60.Copy Number Variation Disorders - PubMed. Current genetic medicine reports. 2017 Dec;5(4).
61.The genetic basis of disease - PubMed. Essays in biochemistry. 12/02/2018;62(5).
62.Synonymous Variants of Uncertain Silence - PubMed. International journal of molecular sciences. 06/23/2023;24(13).
63.When a Synonymous Variant Is Nonsynonymous - PubMed. Genes. 08/19/2022;13(8).
64.Analysis of missense variants in the human genome reveals widespread gene-specific clustering and improves prediction of pathogenicity - PubMed. American journal of human genetics. 03/03/2022;109(3).
65.Identification of pathogenic variant enriched regions across genes and gene families - PubMed. Genome research. 2020 Jan;30(1).
66.Nonsense Mutations in Eukaryotes - PubMed. Biochemistry Biokhimiia. 2022 May;87(5).
67.Therapeutic suppression of premature termination codons: mechanisms and clinical considerations (review) - PubMed. International journal of molecular medicine. 2014 Aug;34(2).
68.Frameshift and wild-type proteins are often highly similar because the genetic code and genomes were optimized for frameshift tolerance - PubMed. BMC genomics. 06/02/2022;23(1).
69.Identification of the nature of reading frame transitions observed in prokaryotic genomes - PubMed. Nucleic acids research. 2013 Jul;41(13).
70.Germline mutations and blood malignancy (Review) - PubMed. Oncology reports. 2021 Jan;45(1).
71.Miles B, Tadi P. Genetics, Somatic Mutation. StatPearls. Treasure Island (FL): StatPearls Publishing Copyright © 2024, StatPearls Publishing LLC.; 2024.
72.ACMG recommendations for standards for interpretation and reporting of sequence variations: Revisions 2007 - PubMed. Genetics in medicine : official journal of the American College of Medical Genetics. 2008 Apr;10(4).
73.Recognition of genetic predisposition in pediatric cancer patients: An easy-to-use selection tool - PubMed. European journal of medical genetics. 2016 Mar;59(3).
74.The Genome Analysis Toolkit: a MapReduce framework for analyzing next-generation DNA sequencing data - PubMed. Genome research. 2010 Sep;20(9).
75.VarSome: the human genomic variant search engine - PubMed. Bioinformatics (Oxford, England). 06/01/2019;35(11).
76.InterVar: Clinical Interpretation of Genetic Variants by the 2015 ACMG-AMP Guidelines - PubMed. American journal of human genetics. 02/02/2017;100(2).
77.Ensembl 2022 - PubMed. Nucleic acids research. 01/07/2022;50(D1).
78.Diagnostic Yield of Clinical Tumor and Germline Whole-Exome Sequencing for Children With Solid Tumors - PubMed. JAMA oncology. 05/01/2016;2(5).
79.Germline Mutations in Predisposition Genes in Pediatric Cancer - PubMed. The New England journal of medicine. 12/10/2015;373(24).
80.Petrucelli N, Daly MB, Pal T. BRCA1- and BRCA2-Associated Hereditary Breast and Ovarian Cancer. In: Adam MP, Feldman J, Mirzaa GM, Pagon RA, Wallace SE, Bean LJH, et al., editors. GeneReviews(®). Seattle (WA): University of Washington, Seattle

Copyright © 1993-2024, University of Washington, Seattle. GeneReviews is a registered trademark of the University of Washington, Seattle. All rights reserved.; 1993.
81.Role of the XPA protein in the NER pathway: A perspective on the function of structural disorder in macromolecular assembly - PubMed. Computational and structural biotechnology journal. 12/08/2015;14.
82.Nucleotide excision repair in eukaryotes - PubMed. Cold Spring Harbor perspectives in biology. 10/01/2013;5(10).
83.Association between NER Pathway Gene Polymorphisms and Wilms Tumor Risk - PubMed. Molecular therapy Nucleic acids. 09/07/2018;12.
84.Understanding nucleotide excision repair and its roles in cancer and ageing - PubMed. Nature reviews Molecular cell biology. 2014 Jul;15(7).
85.Polymorphisms in nucleotide excision repair genes and endometrial cancer risk - PubMed. Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology. 2011 Sep;20(9).
86.Identification of new Wilms tumour predisposition genes: an exome sequencing study - PubMed. The Lancet Child & adolescent health. 2019 May;3(5).
87.The Role of Notch3 in Cancer - PubMed. The oncologist. 2018 Aug;23(8).
88.Molecular pathology of Lynch syndrome - PubMed. The Journal of pathology. 2020 Apr;250(5).
89.SnapShot: DNA mismatch repair - PubMed. Cell. 05/14/2010;141(4).
90.Advances in the study of Lynch syndrome in China - PubMed. World journal of gastroenterology. 06/14/2015;21(22).
91.Acute lymphoblastic leukemia and lymphoma in the context of constitutional mismatch repair deficiency syndrome - PubMed. European journal of medical genetics. 2016 Mar;59(3).
92.Guidelines for surveillance of individuals with constitutional mismatch repair-deficiency proposed by the European Consortium "Care for CMMR-D" (C4CMMR-D) - PubMed. Journal of medical genetics. 2014 May;51(5).
93.Lynch syndrome caused by MLH1 mutations is associated with an increased risk of breast cancer: a cohort study - PubMed. Journal of medical genetics. 2015 Aug;52(8).
94.Pediatric, Adolescent, and Young Adult Thyroid Carcinoma Harbors Frequent and Diverse Targetable Genomic Alterations, Including Kinase Fusions - PubMed. The oncologist. 2017 Mar;22(3).
95.Activity of the Highly Specific RET Inhibitor Selpercatinib (LOXO-292) in Pediatric Patients With Tumors Harboring RET Gene Alterations - PubMed. JCO precision oncology. 04/14/2020;4.
96.Somatic mutations in the RET proto-oncogene in sporadic medullary thyroid carcinomas - PubMed. Molecular and cellular endocrinology. 03/12/2008;284(1-2).
97.Update on Multiple Endocrine Neoplasia Type 2: Focus on Medullary Thyroid Carcinoma - PubMed. Journal of the Endocrine Society. 07/13/2018;2(8).
98.Genetic Variation in the ZNF208 Gene at rs8103163 and rs7248488 Is Associated With Laryngeal Cancer in the Northwestern Chinese Han Male - PubMed. Frontiers in genetics. 04/11/2022;13.

99.A loss-of-function mutation in the endothelin-converting enzyme 1 (ECE-1) associated with Hirschsprung disease, cardiac defects, and autonomic dysfunction - PubMed. American journal of human genetics. 1999 Jan;64(1).
100.A non-canonical role of endothelin converting enzyme 1 (ECE1) in promoting lung cancer development via directly targeting protein kinase B (AKT) - PubMed. The journal of gene medicine. 2024 Jan;26(1).
101.Endothelin-1 and Its Role in Cancer and Potential Therapeutic Opportunities - PubMed. Biomedicines. 02/23/2024;12(3).
102.Endothelin-converting enzyme-1 in cancer aggressiveness - PubMed. Cancer letters. 06/28/2019;452.
103.Functions of the APC tumor suppressor protein dependent and independent of canonical WNT signaling: implications for therapeutic targeting - PubMed. Cancer metastasis reviews. 2018 Mar;37(1).
104.Gene Expression in Wilms Tumor: Disturbance of the Wnt Signaling Pathway and MicroRNA Biogenesis - PubMed. Wilms Tumor. 2016 Mar.
105.Casaubon JT, Kashyap S, Regan JP. BRCA1 and BRCA2 Mutations. StatPearls. Treasure Island (FL): StatPearls Publishing
Copyright © 2024, StatPearls Publishing LLC.; 2024.
106.Role of BRCA1 and BRCA2 as regulators of DNA repair, transcription, and cell cycle in response to DNA damage - PubMed. Cancer science. 2004 Nov;95(11).
107.The breast cancer susceptibility genes (BRCA) in breast and ovarian cancers - PubMed. Frontiers in bioscience (Landmark edition). 01/01/2014;19(4).
108.Cancers associated with BRCA1 and BRCA2 mutations other than breast and ovarian - PubMed. Cancer. 01/15/2015;121(2).
109.Presenilin-dependent ErbB4 nuclear signaling regulates the timing of astrogenesis in the developing brain - PubMed. Cell. 10/06/2006;127(1).
110.ErbB Receptors and Cancer - PubMed. Methods in molecular biology (Clifton, NJ). 2017;1652.
111.Function of ERBB4 is determined by alternative splicing - PubMed. Cell cycle (Georgetown, Tex). 08/15/2011;10(16).
112.ERBB receptors: from oncogene discovery to basic science to mechanism-based cancer therapeutics - PubMed. Cancer cell. 03/17/2014;25(3).
113.Erbb4 Is Required for Cerebellar Developmentand Malignant Phenotype of Medulloblastoma - PubMed. Cancers. 04/17/2020;12(4).
114.ERBB4 promotes the proliferation of gastric cancer cells via the PI3K/Akt signaling pathway - PubMed. Oncology reports. 2018 Jun;39(6).
115.ErbB receptor tyrosine kinase family expression levels in urothelial bladder carcinoma - PubMed. Pathology, research and practice. 02/15/2013;209(2).
116.Wang SC, Shao R, Pao AY, Zhang S, Hung MC, Su LK. Inhibition of cancer cell growth by BRCA2. Cancer Res. 2002;62(5):1311-4.
117.BRCA2 gene mutation in cancer - PubMed. Medicine. 11/11/2022;101(45).

118.The BRCA1/2 pathway prevents hematologic cancers in addition to breast and ovarian cancers - PubMed. BMC cancer. 08/06/2007;7(1).
119.The Expression Level of ABCC6 Transporter in Colon Cancer Cells Correlates with the Activation of Different Intracellular Signaling Pathways - PubMed. Pathophysiology : the official journal of the International Society for Pathophysiology. 05/12/2022;29(2).
120.Prognostic significance and molecular mechanisms of adenosine triphosphate-binding cassette subfamily C members in gastric cancer - PubMed. Medicine. 2019 Dec;98(50).
121.Expression of some ATP-binding cassette transporters in acute myeloid leukemia - PubMed. Hematology reports. 12/22/2017;9(4).
122.Mehrvar N, Abolghasemi H, Rezvany MR, Esmaeil Akbari M, Saberynejad J, Mehrvar A, et al. Pattern of ABCC Transporter Gene Expression in Pediatric Patients with Relapsed Acute Lymphoblastic Leukemia. Rep Biochem Mol Biol. 2019;8(2):184-93.
123.Frequent down-regulation of ABC transporter genes in prostate cancer - PubMed. BMC cancer. 10/12/2015;15(1).
124.NKX2-1 controls lung cancer progression by inducing DUSP6 to dampen ERK activity - PubMed. Oncogene. 2022 Jan;41(2).
125.Transcription factor NKX2-1 drives serine and glycine synthesis addiction in cancer - PubMed. British journal of cancer. 2023 May;128(10).
126.Integrated transcript and genome analyses reveal NKX2-1 and MEF2C as potential oncogenes in T cell acute lymphoblastic leukemia - PubMed. Cancer cell. 04/12/2011;19(4).
127.Isolation and chromosomal localization of a novel FMS-like tyrosine kinase gene - PubMed. Genomics. 1991 Feb;9(2).
128.FLT3 mutations in acute myeloid leukemia: Therapeutic paradigm beyond inhibitor development - PubMed. Cancer science. 2020 Feb;111(2).
129.FLT3 mutated acute myeloid leukemia: 2021 treatment algorithm - PubMed. Blood cancer journal. 05/27/2021;11(5).
130.The GTEx Consortium atlas of genetic regulatory effects across human tissues - PubMed. Science (New York, NY). 09/11/2020;369(6509).
131.Abundance of Flt3 and its ligand in astrocytic tumors - PubMed. OncoTargets and therapy. 05/24/2013;6.
132.Molecular genetic profiling reveals novel association between FLT3 mutation and survival in glioma - PubMed. Journal of neuro-oncology. 2020 Jul;148(3).
133.Activating FLT3 mutations in CD117/KIT(+) T-cell acute lymphoblastic leukemias - PubMed. Blood. 07/15/2004;104(2).
134.Absence of FLT3 mutations in Iranian adult T-cell leukemia/lymphoma patients - PubMed. Medical journal of the Islamic Republic of Iran. 06/09/2018;32.
135.The Cytokine Flt3-Ligand in Normal and Malignant Hematopoiesis - PubMed. International journal of molecular sciences. 05/24/2017;18(6).

136.DaSilva N, Hu ZB, Ma W, Rosnet O, Birnbaum D, Drexler HG. Expression of the FLT3 gene in human leukemia-lymphoma cell lines. Leukemia. 1994;8(5):885-8.
137.PRRT2 inhibits the proliferation of glioma cells by modulating unfolded protein response pathway - PubMed. Biochemical and biophysical research communications. 04/01/2017;485(2).
138.A mononucleotide repeat in PRRT2 is an important, frequent target of mismatch repair deficiency in cancer - PubMed. Oncotarget. 01/24/2017;8(4).
139.Song K, Yuan Y, Lin Y, Wang YX, Zhou J, Gai QJ, et al. ERBB3, IGF1R, and TGFBR2 expression correlate with PDGFR expression in glioblastoma and participate in PDGFR inhibitor resistance of glioblastoma cells. Am J Cancer Res. 2018;8(5):792-809.
140.Choe B-K, Kim SK, Park HJ, Park H-K, Kwon KH, Lim SH, et al. Polymorphisms of TGFBR2 contribute to the progression of papillary thyroid carcinoma. Molecular & Cellular Toxicology. 2012;8(1):1-8.
141.The ryanodine receptor mutational characteristics and its indication for cancer prognosis - PubMed. Scientific reports. 09/27/2022;12(1).
142.Narod SA, Stiller C, Lenoir GM. An estimate of the heritable fraction of childhood cancer. Br J Cancer. 1991;63(6):993-999. doi:10.1038/bjc.1991.216
143.Merks JH, Caron HN, Hennekam RC. High incidence of malformation syndromes in a series of 1,073 children with cancer. Am J Med Genet A. 2005;134A(2):132-143. doi:10.1002/ajmg.a.30603
144.Zhang J, Walsh MF, Wu G, et al. Germline Mutations in Predisposition Genes in Pediatric Cancer. N Engl J Med. 2015;373(24):2336-2346 doi:10.1056/NEJMoa1508054

8 CURRICULUM VITAE