Arab American University

Faculty of Graduate Studies

Identification of a novel mutation in a Palestinian family with Hereditary Cancer

By

Dania Osama Najjar

Supervisor

Dr. Zaidoun Salah

Co-Supervisor

Dr. Nouar Qutob

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

Molecular genetics and genetic toxicology

Month/ 2021

Thesis Approval

Identification of a novel mutation in a Palestinian family with Breast Cancer

By

Dania Osama Najjar

This thesis was defended successfully on 29.07.2023 and approved by:

Committee members	Signature
1. Dr. Zaidoun Salah/ Supervisor	.....
2. Dr. Nouar Qotoub/ Co-Supervisor	.....
3. Dr. Siba Shanak/ Internal Examiner	.....
4. Dr. Rania Abu Seir/ External Examiner	.....

Declaration

Hereby, I declare that this thesis represents my own work which was written as a master's student in the molecular genetics and genetic toxicology program at the Arab American University of Palestine; and has not been previously submitted to this or any other universities for a degree, diploma, or other professional qualifications and the collaborative contributions have been indicated clearly and acknowledged.

Dania Osama Najjar

Date: 18/01/2024

Signature: 

Student ID:201912940

Dedication

I dedicate my thesis to my mentors and family members who taught me persistence. I especially thank my loving parents, Osama and Afaf who helped me in my journey and have never left my side. I also thank my sisters and brothers Lana, Zaina, Ahmad, Ameer, and Ayman who have been very supportive to me. I also dedicate this work and give special thanks to my beloved husband, Muath and my wonderful son Mazen for always supporting me and encouraging me to keep going. Finally, I dedicate this thesis to my friend Rawan who supported me and helped me all along the way. I will always be grateful for everything she did.

Acknowledgement

I would like to express my sincere gratitude to my supervisors, Dr. Nouar Qutob and Dr. Zaidoun Salah, for their invaluable guidance, support, and assistance throughout my master's program. Their expertise, encouragement, and distinguished knowledge in molecular genetics and scientific research helped me develop my research skills and knowledge and complete the degree requirements. I would also like to acknowledge Mr Husam Sallam and Mrs Rua' Thawabteh at the AAUP molecular genetics laboratory for their constant help and guidance. Finally, I would like to thank Dr. Mahmoud Ruzayqat for his assistance and guidance in performing the next-generation sequencing field.

Abstract

Next generation sequencing is becoming an important tool for cancer diagnosis. It presents an effective way to early detection and medical intervention to plan effective treatment that may increase the chance of survival. The purpose of this study is to utilize next generation sequencing to identify the molecular causes of breast cancer in a Palestinian family with hereditary breast cancer. We performed whole exome sequencing of DNA extracted from the blood of two sisters diagnosed with breast cancer and validates the results by sanger sequencing. We used segregation analysis to confirm the presence of the identified variant in the daughters of the targeted family members.

Our analysis resulted in the identification of a novel stop-gain mutation *IGSF8* c.C487T: p.R163X, which is located in exon 3 and may lead to unfunctional protein. This was subsequently validated by sanger sequencing. Furthermore, segregation analysis confirmed the presence of this variant in the daughters of the targeted family members.

Our study shows that WES provides a more conclusive genetics diagnoses compared to panel sequencing. It also confirms that molecular studies and NGS applications are needed to detect significant breast cancer relevant mutations, which may in return lead to the right medical interventions.

Table of Contents

List of tables

List of figures

List of abbreviations

#	Abbreviation	Definition
1	AFF2	Fragile XE mental retardation gene
2	AJCC	American Joint Committee on Cancer
3	AKT	Alpha Serine/Threonine-Protein Kinase
4	ALN	Axillary Lymph Node
5	APC	Adenomatous polyposis coli gene
6	ARID	AT-Rich Interactive Domain gene
7	ASXL1	Additional Sex Combs Like 1 gene
8	ATM	Ataxia Telangiectasia Mutated gene
9	BAP1	BRCA1 Associated Protein-1 gene
10	BARD1	BRCA1 Associated RING Domain 1 gene
11	BL-1	Basal Like -1
12	BL-2	Basal Like -2
13	BRCA1	Breast Cancer Gene 1
14	BRCA2	Breast Cancer gene 2
15	BRIP1	BRCA1 Interacting Protein 1 gene
16	CASP8	Caspase-8 gene
17	CBFB	Core Binding Factor Beta
18	CCND3	Cyclin D3 gene
19	CDH1	Cadherin 1 gene
20	CDKN1B	Cyclin Dependent Kinase Inhibitor 1B gene
21	CGAN	Cancer Genome Atlas Network
22	CHEK2	Checkpoint Kinase 2 gene
23	CNA’S	Copy Number Alterations
24	DCIS	Ductal Carcinoma in Situ
25	DNA	Deoxyribonucleic acid
26	dsDNA	double-stranded DNA
27	EGFR	Epidermal Growth Factor Receptor gene
28	ER	Estrogen Receptor
29	FDA	Food and Drug Administration
30	FOXA1	Forkhead box protein A1 gene
31	GATA3	GATA Binding Protein 3
32	GHSR	Growth Hormone Secretagogue Receptor gene
33	HBOC	Hereditary Breast and Ovarian Cancer
34	HER2	human epidermal growth factor receptor 2
35	IM	Immunomodulatory
36	KI-67	The nuclear protein Ki67 (pKi67
37	KRAS	Kirsten Rat Sarcoma Viral Oncogene Homolog
38	LABC	Locally Advanced Breast Cancer
39	LAR	luminal androgen receptor
40	M	Mesenchymal
41	MAP2K	Mitogen-Activated Protein Kinase Kinase gene

42	MAP3K13	Mitogen-Activated Protein Kinase Kinase Kinase 13 gene
43	McrBC	Methylation-dependent restriction enzymes
44	microRNA	Micro Ribonucleic acid
45	miR	MicroRNA
46	MLL2	Myeloid/Lymphoid or Mixed-Lineage Leukemia 2 gene
47	MRE11A	Meiotic Recombination 11 Homolog A
48	mRNA	Messenger Ribonucleic acid
49	MSH2	DNA mismatch repair gene
50	MSL	mesenchymal stem cell-like
51	MTAP	Methylthioadenosine Phosphorylase gene
52	mTOR	Mechanistic Target of Rapamycin Kinase
53	MYC	Myelocytomatosis Oncogen
54	NBN	Nibrin gene
55	NBS1	Nijmegen breakage syndrome gene
56	NCCN	National Comprehensive Cancer Network
57	NF1	Neurofibromatosis 1 gene
58	NGS	Next Generation Sequencing
59	NM23	Non-Metastatic clone 23 gene
60	PALB2	Partner And Localizer of BRCA2
61	PCR	Polymerase Chain Reaction
62	PI3K	Phosphoinositide 3-Kinases
63	PIK3R1	Phosphoinositide-3-Kinase Regulatory Subunit 1 gene
64	PMS2	Mismatch Repair System Component
65	PPP2R2A	Protein Phosphatase 2 Regulatory Subunit Balpha gene
66	PR	Progesterone Receptor
67	PTEN	Phosphatase And Tensin Homolog gene
68	PTPN22	Protein Tyrosine Phosphatase Non-Receptor Type 22 gene
69	PTPRD	Protein Tyrosine Phosphatase Receptor Type D gene
70	RAD50	RAD50 Double Strand Break Repair gene
71	RAD51	RAD51 Recombinase gene
72	RB1	Retinoblastoma 1 gene
73	RCT	Randomized Controlled Trial
74	RUNX1	Runt-related transcription factor 1
75	SERM'S	Selective Estrogen Receptor Modulator
76	SETD2	SET Domain Containing 2 gene
77	SF3B1	Splicing Factor 3b Subunit 1 gene
78	SLN	Sentinel Lymph Node
79	SMAD4	Mothers Against Decapentaplegic Homolog 4 gene
80	SMARC	SWI/SNF Related, Matrix Associated, Actin Dependent Regulator of Chromatin gene

81	SNAIL2	Snail Family Transcriptional Repressor 2 gene
82	STK11	Serine/Threonine Kinase 11 gene
83	SWI/SNF	SWItch/Sucrose Non-Fermentable complex
84	TBX3	T-Box Transcription Factor 3 gene
85	TGF	Transforming Growth Factor gene
86	TNBC	Triple Negative Breast Cancer
87	TP53	Tumor Protein P53 gene
88	UV	Ultraviolet
89	VEGF	Vascular endothelial growth factor
90	WES	Whole Exome Sequencing
91	WGD	Whole Genome Sequencing
92	WHO	World Health Organization
93	ZEB2	Zinc Finger E-Box Binding Homeobox 2 gene

Chapter 1: Introduction

According to the World Health Organization (WHO), cancer is one of the primary contributors to global mortality, with approximately 10 million fatalities by the year 2020, translating to roughly one in every six recorded deaths. (WHO et al., n.d.). Breast cancer is the most prevalent cancer in Palestine; however, not much has been done to identify the molecular basis of breast cancer in Palestine. Advancement in molecular genetics techniques holds potential for discovering the molecular basis of breast cancer among Palestinians, which may lead to improved clinical health, through better diagnosis and management.

1.1 Tumor Classification

A tumor is defined as an uncontrolled growth of cells. Tumors can be classified as either malignant or benign, varying in size from small nodules to massive masses, and can emerge in any part of the body (Klein & Hecker, 1977). Benign tumors are noncancerous in nature, displaying slow growth and lacking the ability to infiltrate or spread into neighboring tissues. Examples of benign tumors encompass adenomas, fibroids, hemangiomas, lipomas, and others. Generally, benign tumors do not elicit detrimental medical issues; however, their potential to grow significantly in size might lead to compression of neighboring tissues, resulting in discomfort or other medical complications. For example, a large benign lung tumor might restrict the trachea, impinging on breathing and necessitating immediate surgical intervention (Patel, 2020). Once removed, benign tumors are unlikely to return. Nonetheless, certain benign tumors have the potential to develop into malignant cancers, warranting continuous monitoring

and potentially requiring surgical excision. For instance, colon polyps can transform into cancerous growths and must be surgically removed (National Cancer Institute, 2021).

On the other hand, malignant tumors are cancerous in nature, capable of affecting nearby tissues and disseminating via the circulatory or lymphatic systems to distant locations within the body (Patel, 2020). This process, known as metastasis, can occur anywhere in the body, with common sites including the liver, lungs, brain, and bones. Malignant tumors include categories such as carcinomas, sarcomas, germ cell tumors, and blastomas (Valastyan & Weinberg, 2011). These tumors can spread quickly and demand immediate intervention. In cases of early detection, surgical treatment is often accompanied by chemotherapy or radiation. However, when cancer is diagnosed at an advanced stage, systemic treatments like chemotherapy or immunotherapy are employed (Moo et al., 2019).

Cancer arises due to genomic alterations disrupting fundamental biochemical processes governing cell proliferation and survival. These alterations are frequently induced by hazardous environmental agents, such as toxins in cigarette smoke and UV radiation from the sun (Klaunig, 2019). The genetic changes in each individual's cancer are distinctive, rendering cancers heterogeneous. Moreover, additional alterations manifest as malignancy progresses, leading to diverse genetic alterations within the same tumor (Sheikh et al., 2015). The principal types of gene alterations in cancer encompass proto-oncogenes, tumor suppressor genes, and DNA repair genes. These genes are affected by cancer driver mutations, which substantively contribute to cancer development (Saroj Chakraborty et al., 2017).

Proto-oncogenes are instrumental in regulating proper cell division and proliferation. Mutations or heightened activity in these genes (due to gain-of-function mutations) result

in accelerated cell growth and division, leading to aberrant cell survival and growth conducive to cancer development (Willecke & Schäfer, 1984). On the other hand, tumor suppressor genes are engaged in regulating cell division and growth. They function to limit uncontrolled cellular growth, promote DNA repair, and activate cell cycle checkpoints. Loss-of-function mutations in certain tumor suppressor genes culminate in uncontrolled cell growth, proliferation, and ultimately, cancer development (Baker et al., 2009). DNA repair mechanisms rectify damaged DNA. Mutations in these genes can facilitate mutations in other genes and chromosomal abnormalities such as duplications and deletions. If these alterations coincide, they can contribute to cellular malignancy (Cancer Staging NCI, 2022).

1.2 Breast cancer the most common cancer among women

Breast cancer affected 2.3 million women worldwide in 2020, resulting in 685,000 fatalities. According to the WHO, breast cancer has been diagnosed in 7.8 million women within the past five years as of 2020, solidifying its status as the most prevalent cancer type among women on a global scale.

According to the Palestinian Ministry of Health, breast cancer is the most common in Palestine. Notably, 876 new cases were registered in 2021. This translated to an incidence rate of 17.8 cases of breast cancer per 100,000 individuals across the population, while among females, the incidence rate was notably higher at 36.2 per 100,000. This underscores the prominence of breast cancer specifically among females, with colorectal cancer ranking as the second most common, followed by lung cancer (MOH Health Annual Report Palestine 2021).

1.2.1 Histological and molecular classification of Breast Cancer

The breast comprises three primary components: lobules, ducts, and connective tissue. Lobules are responsible for milk production, while ducts serve as conduits for milk transport to the nipple. These structures are supported by connective tissue composed of fatty and fibrous components. The emergence of different types of breast cancer is often contingent upon which of these cell types becomes cancerous. Generally, cancer originates within the ducts or lobules (Jesinger, 2014).

Histologically, breast tissue abnormalities span a spectrum from hyperplasia and atypical changes to ductal carcinoma in situ (DCIS) and invasive carcinoma. Pre-invasive alterations may not always progress to invasive cancer; this transition can take decades. Lobular carcinoma in situ pertains to abnormal tissue growth within the breast lobules. Although this type does not progress to invasive breast cancer, it rises the risk of later invasive breast cancer development by around 7% within ten years (Maughan et al., 2010). Lobular carcinoma, classified as E-cadherin negative lesions, constitutes the second most frequent breast cancer type, accounting for 10–20% of cases. Conversely, DCIS denotes the presence of malignant cells within milk ducts. As the most prevalent type, DCIS is highly heterogeneous, accounting for 80% of all breast cancers, with the potential to progress into invasive breast cancer (Baroni et al., 2003). Further subtypes, including tubular, mucinous, medullary, papillary, micropapillary, and metaplastic, contribute to 5% of all lesions. Notably, three markers—tubular development, nuclear pleomorphism, and mitotic activity—aid in further grading these lesions (Libson & Lippman, 2014).

In addition to the aforementioned histological classification, an intrinsic or molecular classification utilizes molecular profiling and gene expression array analysis. A set of 496

genes, known as the intrinsic gene set, was discovered by Perou and his colleagues (Perou et al., 2000). This classification divides breast cancers into several subgroups: luminal A, luminal B, triple-negative breast cancer, basal-like breast cancer, HER2-enriched breast cancer, and luminal B-like breast cancer. The most common subtype, luminal A, is estrogen-receptor (ER) positive, progesterone-receptor (PR) positive, and HER2 negative. This subtype has low levels of the Ki-67 protein, indicating a slower pace of cancer cell proliferation, resulting in a favorable prognosis (Prat et al., 2015). Luminal B breast cancer, ER-positive and HER2-negative, exhibits either elevated Ki-67 levels (indicative of rapid cancer cell proliferation) or the absence of PR (Rakha & Ellis, 2011; Yeo et al., 2015). HER2 subtype breast cancers exhibit heightened HER2 gene expression and limited luminal gene expression, while basal subtype breast cancers display low expression of both luminal and HER2 genes alongside elevated expression of proliferative genes (Hayes et al., 2007). Luminal B-like breast cancer is ER-positive, HER2-positive, and Ki-67-positive, with possible PR positivity or negativity (Barzaman et al., 2020; Yeo et al., 2015). HER2-enriched breast cancer, devoid of ER and PR expression but HER2-positive, presents aggressive characteristics and a less favorable prognosis in comparison to luminal A and B cancers. Targeted therapy drugs addressing the HER2 protein are typically efficacious in their treatment (Erjan et al., 2021). Triple-negative breast cancer (TNBC), or basal-like breast cancer, lacks ER, PR, and HER2 expression (Weisman et al., 2016). This subtype is more prevalent among young women, women with a BRCA1 mutation, and black women, and is notably more aggressive compared to luminal A or B cancers. TNBC can be further categorized into six classifications: basal-like 1 (BL-1), basal-like 2 (BL-2), immunomodulatory (IM),

mesenchymal (M), mesenchymal stem cell-like (MSL), and luminal androgen receptor (LAR) (Gamble, 2013).

Therapeutic strategies vary across these subtypes. Hormonal therapy, predominantly Tamoxifen, is recommended for Luminal A and B breast cancers. This hormone blocker is complemented by aromatase inhibitors like letrozole, anastrozole, and exemestane, which curtail hormone production from the ovaries. HER2-expressing tumors can be treated with anti-HER2 monoclonal antibodies like trastuzumab, pertuzumab, and ado-trastuzumab emtansine, combining monoclonal antibody and microtubule inhibitor (Nagarajan & McArdle, 2018). Anti-HER2 monoclonal antibody therapy combined with chemotherapy is the usual treatment for breast cancer patients with positive HER2. TNBC poses greater treatment challenges, with chemotherapy constituting the standard approach. The combination of chemotherapy and bevacizumab, a monoclonal antibody against vascular endothelial growth factor (VEGF), might provide an alternative treatment (Barzaman et al., 2020; Berrada et al., 2010).

1.2.2 Predictive markers of breast cancer prognosis

Several predictive markers, including lymph nodes, tumor size, age, race, proliferation index, ER, PR, HER2 receptor status, and circulating tumor cells, play a pivotal role in evaluating sensitivity or resistance to specific therapies.

Lymph node presence and metastatic load hold crucial significance for tumor staging and function as prognostic indicators. The quantity of lymph nodes is directly associated with tumor prognosis, with overall survival diminishing as the number of affected lymph nodes grows. Furthermore, tumor size influence on prognosis in lymph nodes is noteworthy (Paoletti & Hayes, 2014). Tumor size emerges as the second predictive predictor, exerting

a direct influence on survival rates. Among these parameters, the size of the invasive component is particularly pertinent. Tumor sizes within the range of 0.2 to 2 mm are classified as micrometastases, while those smaller than 0.2 mm are classified as isolated tumor cells by the AJCC. Additionally, age and race contribute to prognostic outcomes; cancer diagnosis in younger women might yield poorer results compared to postmenopausal women. Importantly, even after accounting for other factors, individuals of black ethnicity exhibit lower survival rates in contrast to other ethnic groups (A Boyer-Chammard 1, T H Taylor, n.d.). The proliferation index, also regarded as a predictive predictor, employs diverse measurement methods. The earliest approach involves quantifying the number of mitotic bodies per high-powered field, whereas alternative methods encompass ki67 immunohistochemistry staining, evaluating a specific antigen in actively proliferating cells (Inwald et al., 2013).

ER has a mixed prognostic and predictive value, interacting with regulatory genomic components like cyclin D and MYC, potentially fostering cancer cell proliferation (Carroll, 2016). Patients diagnosed with ER-positive cancers, such as Luminal A and Luminal B tumors, typically have higher survival rates compared to those with ER-negative cancers such as TNBC over shorter periods (Sung et al., 2017). Often assessed in tandem with ER, PR's significance lies in its association with ER. Since ER stimulates PR expression and research indicates that PR-progesterone interaction modifies ER chromatin binding sites and alters it, an assessment of the ER-associated PR ratio has been established. Consequently, the presence of PR is considered a sign of functional ER (Barzaman et al., 2020; Hammond et al., 2010). Another prominent significant predictor is the HER2 receptor. In cases where the HER2 receptor is positive, monoclonal antibody treatment improves overall survival by approximately 30% in breast cancer patients

(Paoletti & Hayes, 2014). Early-stage breast cancer patients can also gain prognostic insights from circulating tumor cells; higher counts of circulating tumor cells correlate with increased risks of recurrence and mortality (Sharma et al., 2016).

1.2.3 Genetic and epigenetic changes in breast cancer

A recent study has unveiled substantial insights into the role of somatic mutations in breast cancer. It searched for somatic copy number alterations and mutations in coding exons of protein-coding genes across 100 cancer genomes. This study, spanning 21,416 protein-coding genes, identified 7,421 somatic point mutations (Byler et al., 2014). The prevalence of somatic mutations notably differed among various cancers. Prominent cancer genes associated with breast cancer pathogenesis, including *BRCA1, RB1, TP53, PTEN, AKT1, CDH1, GATA3, and PIK3CA,* exhibited somatic mutations. These genes regulate vital functions such as apoptosis, cell-cycle regulation, and transcriptional control. Additionally, signal transduction genes such as *APC, ARID1A, ARID2, ASXL1, BAP1, KRAS, MAP2K4, MLL2, MLL3, NF1, SETD2, SF3B1, SMAD4,* and *STK11* were found to possess somatic mutations. Furthermore, previously unlinked genes to breast cancer, such as *ARID1B, CASP8, MAP3K13, MAP2K13, NCOR1, SMARCD1, and CDKN1B,* were also discovered to be mutated (Stephens et al., 2012). Notably, the study highlighted the presence of at least 40 mutated cancer genes in the tumors studied, forming 73 distinct combinations of altered cancer genes. This underscores the vast genetic heterogeneity of breast cancer tumors and underscores the complexity of identifying cancers based solely on genetic subtypes (Byler et al., 2014).

Researchers further discerned variation in mutation frequency and type across different breast cancer subtypes such as Luminal A, B, basal-like, and HER2-enriched. For

example, luminal A and B types exhibited a preponderance of missense mutations, while the basal type primarily displayed nonsense and frameshift mutations (Koboldt et al., 2012). Distinct levels of copy number alterations were also observed, including localized gene amplifications involving *PIK3CA, EGFR, FOXA1,* and *HER2,* and targeted gene deletions in regions containing *MLL2, PTEN, RB1,* and *MAP2K4,* resulting in different levels of copy number changes (Koboldt et al., 2012).

In the realm of epigenetic alterations, breast cancer and several other cancer types hinge on epigenetic alterations for their genesis and progression. Recent studies have made significant headway in characterizing these changes, revealing how various epigenetic changes like histone modifications, DNA methylation, and miRNA expression influence gene expression (Byler et al., 2014). Investigating miRNA expression levels in 525 tumors and 22 tumor-adjacent normal tissues, the CGAN identified 3,662 differentially expressed genes and seven primary miRNA subtypes (Perou et al., 2000). Progress in understanding miRNA dynamics from healthy tissue to aggressive cancer has been made, highlighting numerous miRNAs exhibiting significant differential expression between DCIS and invasive breast cancer. Notably, miRNAs such as let-7d, miR-210, and miR-221 showed downregulation during the transition from normal tissue to DCIS, followed by upregulation from DCIS to aggressive cancer, with potential targets such as BRCA1 and E-cadherin (Volinia et al., 2012).

Epigenetic changes involving DNA methylation have also been implicated in breast cancer etiology. Based on methylation array data encompassing 802 breast cancers, the CGAN identified five distinct DNA methylation groups. Hyper-methylation phenotype predominated in luminal B breast cancer, while the lowest DNA methylation levels aligned with basal-like subtypes (Koboldt et al., 2012). Moreover, methylation alterations

have been linked to breast cancer progression, offering a means to differentiate between normal and malignant breast cells (Ibrahim et al., 2022).

Alterations in histone acetylation and methylation patterns could be an early indicators of breast cancer (Elsheikh et al., 2009). Moderate to low levels of lysine acetylation and lysine and arginine methylation have been linked to less favorable prognoses, as seen in basal carcinomas (triple-negative) and HER2-positive breast cancers (Kouzarides, 2007). In the study conducted by Ordway et al., the analysis of DNA methylation was performed on nine invasive ductal carcinomas along with adjacent normal tissues. This analysis utilized the cytosine methylation-dependent restriction enzyme McrBC combined with array hybridization, resulting in the identification of 220 differentially methylated genes. Notably, among these genes, *GHSR*, a G-protein-coupled receptor, stood out, and the methylation status of GHSR exhibited remarkable sensitivity and specificity in distinguishing invasive ductal carcinoma from normal and benign breast tissue (Ordway et al., 2007).

The methylation status of upstream promoter regions demonstrated links to genes involved in multiple pathways when comparing a highly metastatic breast cancer cell line to a less metastatic counterpart. Surprisingly, 20 to 30% of genes exhibited differing methylation statuses between the two cell lines (Rodenhiser et al., 2008). Notably, alterations in the methylation levels of the *SWI/SNF* gene, implicated in chromatin remodeling (Mandel & Gozes, 2007), were observed. These alterations could potentially be associated with the suppression of the TGF pathway (Hinshelwood et al., 2007) and the overexpression of *MYC*, thereby linking this gene to both carcinogenesis and metastasis (Jamerson et al., 2004). Moreover, the hypomethylation of *ZEB2* and *SNAIL2* was found to result in their overexpression, consequently suppressing E-cadherin

(*CDH1*), a gene pivotal in EMT. Within the integrin signaling pathway, distinct genes exhibited either hypermethylation or hypomethylation (Rodenhiser et al., 2008). This observed pattern aligns with the intricate roles of different integrin receptors in breast cancer progression, where certain genes associated with cell-to-cell attachment. For example, *CDH1*, necessitate repression through hypermethylation during metastasis, while others linked to cell motility need upregulation through hypomethylation (Caldeira et al., 2006). These findings demonstrate the intricacy of the metastatic process, suggesting that a unique genetic sequence is either silenced or re-expressed in more metastatic forms of breast cancer.

1.2.4 Hereditary Breast Cancer: An Overview of the most common genes mutated in hereditary breast cancer

Genetic alterations inherited from one or both parents and present at birth are known as germline mutations. These mutations impact every cell in the organism. Inherited mutations, along with other genetic changes, can increase an individual's risk of developing breast cancer. Familial breast cancer is identified by certain criteria, including the presence of a minimum of three cases of breast and/or ovarian cancer within a family. This designation is also met if there are two instances of breast cancer in close relatives, with at least one diagnosis occurring before age 50, or if two cases of breast cancer are diagnosed within a family before the age of 40. Additionally, familial breast cancer can be indicated by the occurrence of male breast cancer in conjunction with a family history of ovarian cancer or early-onset female breast cancer, or by the presence of Ashkenazi Jewish ancestry combined with breast cancer (Shiovitz & Korde, 2015). It is noteworthy that 90% of breast cancers are considered sporadic, arising due to somatic mutations that

manifest during an individual's lifespan (Thorat & Balasubramanian, 2020). In contrast, hereditary germline mutations contribute to the remaining breast cancer cases.

Up to 25% of hereditary breast cancers stem from germline mutations in the *BRCA1* and *BRCA2* genes. Carriers of *BRCA1* or *BRCA2* mutations often exhibit early-onset disease and have an elevated likelihood of developing other cancers like melanoma, prostate, pancreatic, gall bladder, stomach, or even male breast cancer (Larsen et al., 2014). While hereditary breast cancers can encompass various forms of the disease, certain pathological characteristics are more commonly observed in hereditary cases than in sporadic instances. For instance, *BRCA1*-associated cancers are typically of the medullary or atypical medullary type, displaying ER/PR negativity. On the other hand, *BRCA2*-associated cancers tend to be ER/PR positive (Da Silva & Lakhani, 2010). Beyond the realm of *BRCA* genes, a range of other uncommon gene variants and non-*BRCA* breast cancers have been linked to heightened breast cancer risk. This group includes high penetrance genes like *TP53, CDH1, PTEN, STK11, RAD51C,* and *RAD51D*, as well as low or moderate penetrance genes such as *ATM, CHEK2, BRIP1*, and *PALB2*. Most of these genes play crucial roles in DNA repair mechanisms and the preservation of genomic integrity (Antoniou et al., 2003).

While *BRCA1, BRCA2,* and *TP53* are commonly used in clinical practice, these genes do not account for the majority of families with a history of breast cancer. Additional breast cancer-predisposing genes have been identified; however, their pathogenic variants are relatively rare or possess low penetrance. Despite the widespread use of next-generation cancer gene panels, numerous studies indicate that a significant portion of diagnoses among women seeking testing still result from mutations in *BRCA1* or *BRCA2*. Remarkably, only 9–29% of individuals undergoing clinical assessment for hereditary

breast cancer exhibit pathogenic or suspected pathogenic variants in *BRCA1/2*. Globally, 4-16% of patients show such variants in the 20 to 40 other predisposition genes covered by hereditary cancer testing arrays (Keeney et al., 2017).

1.3 Next generation sequencing is the most recent technique used for cancer diagnostics

The assessment of genetic changes within cancers, both in a general context and particularly in breast cancer, has never been as accessible and straightforward as it is today. Next-generation sequencing (NGS), one of the most pivotal technologies in contemporary medicine, plays a critical role in sequencing genetic alterations in cancers (Behjati & Tarpey, 2013).

NGS, also known as massively parallel sequencing, emerged within the last decade, enabling the simultaneous sequencing of millions of DNA fragments without the need for prior sequence knowledge. In comparison to older sequencing methods where only one or a few relatively short DNA fragments, previously amplified through Polymerase Chain Reaction (PCR), could be sequenced per tube, this advanced technology has truly revolutionized the field. The impact of NGS in understanding the intricate biological processes of cancers has ushered in a new era of personalized patient care (personalized medicine) (Kamps et al., 2017). By delving into mutational profiles and uncovering genomic architectures, NGS investigations of various malignancies including prostate, breast, ovarian, pancreatic, and hematological cancers have unveiled novel cancer genes and provided insights into tumor progression. The NGS technology is classified into two main categories: second-generation sequencing and third-generation sequencing. Second-generation sequencing relies on short-read alignment techniques, whereas third-

generation sequencing is grounded in single DNA molecule sequencing, presenting the advantage of requiring less DNA input and facilitating single-cell sequencing (Sabour et al., 2017). NGS technology encompasses methods such as targeted gene-customized panels, whole exome sequencing, and whole genome sequencing, all of which contribute to cancer diagnostics and research.

1.4.1 Panel testing by NGS for gene mutations in breast cancer

Targeted NGS panels play a crucial role in detecting clinically relevant genetic changes, including somatic mutations, single nucleotide polymorphisms, copy number variations, and minor insertions/deletions, for cancer profiling, response monitoring, and therapy guidance (Martínez-Fernández et al., 2021). This highly optimized system leverages integrated unique molecular indices (UMIs) from cells, tissues, and biofluids to achieve ultrasensitive variant identification (Staadig et al., 2023). The approach of target enrichment enables the specific examination of disease-associated genes or other areas of interest, focusing sequencing efforts on selected regions rather than the entire genome. This approach enhances sequencing depth and sample throughput, reduces costs, and simplifies downstream bioinformatics analysis to achieve more comprehensive coverage. Originally designed as a research tool, targeted panels aim to uncover genetic abnormalities associated with hereditary breast and ovarian malignancies (Martínez-Fernández et al., 2021).

The Human *BRCA* Plus panel, for instance, was developed to identify genetic abnormalities linked to hereditary breast and ovarian cancers. This panel investigates mutations in six genes associated with breast and/or ovarian cancer (*BRCA1, BRCA2, CDH1, PALB2, PTEN, TP53*) (Qiagen *BRCA* Advanced DNA UMI Panel, 2023). A

notable focus is placed on breast cancer patients who tested negative for the Human *BRCA* Plus panel (non-*BRCA* patients). Research suggests that the prevalence of non-*BRCA1/2* mutations falls within the range of 4 to 16% (Keeney et al., 2017). Typically, whole exome sequencing is employed for these patients to pinpoint the responsible gene, aiding in early cancer detection or proactive measures that significantly enhance their chances of survival (Qiagen BRCA Advanced DNA UMI Panel, 2023).

1.4.2 Whole exome and whole genome sequencing

The evolution of each cancer results from the accumulation of diverse somatic mutations, copy number variations, epigenetic factors, and structural anomalies, rendering the cancer genome inherently dynamic (Byler et al., 2014). These alterations may occur in individuals with a genetic predisposition, such as hereditary malignancies, leading to distinct patterns within each tumor genome. Precision oncology is being shaped by recognizing that cancer is fundamentally a genetic ailment and coupling this insight with the expanding comprehension of targeted therapies. Rooted in the premise behind precision oncology, physicians employ both the patient's genome and tumor genome to direct them toward targeted treatments with anticipated effectiveness. This approach is facilitated by whole exome sequencing (WES) and whole genome sequencing (WGS) (Rossing et al., 2019), both conducted through NGS technology.

Whole genome sequencing entails the sequencing and mapping of the entire sequence back to the human genome. Its chief advantage is its coverage of the complete genome, encompassing promoters and regulatory regions. This approach is increasingly used to uncover novel and rare mutations (Sabour et al., 2017). On the other hand, WES targets the exons (protein coding regions) of all known genes and is commonly preferred over

WGS due to its cost-effectiveness and larger sequencing depth (Bonnefond et al., 2010; Shoubridge et al., 2010).

1.5 *IGSF8* Gene Structure, Expression, and Function

The Immunoglobulin superfamily member 8 (*IGSF8*), also known as the *EWI2* gene, consists of 12 exons and encodes a member of the immunoglobulin protein superfamily. IGSF8 comprises 613 amino acids, organized into five domains: Ig-like C2-type 1 domain, Ig-like C2-type 2 domain, Ig-like C2-type 3 domain, Ig-like C2-type 4 domain, and EWI motif, as illustrated in the schematic diagram below (Fig. 1) (https://www.uniprot.org/uniprotkb/Q969P0/entry).

Within this family, members possess an EWI (Glu-Trp-Ile) motif, a single transmembrane domain, and an array of immunoglobulin domains (Stipp et al., 2001). The IGSF8 protein interacts with CD81 and CD9 tetraspanins, which are part of a conserved group of four transmembrane domain proteins. In humans, 33 members of this family exhibit diverse tissue distributions (Zöller, 2009). IGSF8 appears to modulate the function of these molecules and their tetraspanins within cells, influencing processes such as cell migration, viral infection, sperm and oocyte fusion, hepatitis C virus cell infection, and keratinocyte differentiation and proliferation (Rocha-Perugini et al., 2008). Furthermore, *IGSF8* functions as a tumor suppressor by restraining the growth of certain malignancies. Tetraspanins, which also belong to the four transmembrane domain protein family, are known to have critical roles in eukaryotic cells. These proteins span the cell membrane four times and include 33 members in humans with varying tissue distributions. Research indicates that tetraspanins govern tumor growth, cell adhesion, invasion, and migration

(Zöller, 2009). IGSF8 has been linked to the regulation of cell spreading, motility, invasion functions, and filopodia production in an adverse manner (Stipp et al., 2001). Moreover, IGSF8 is implicated in maintaining the neuronal network within the adult brain and overseeing neurite development (Ray & Treloar, 2012).

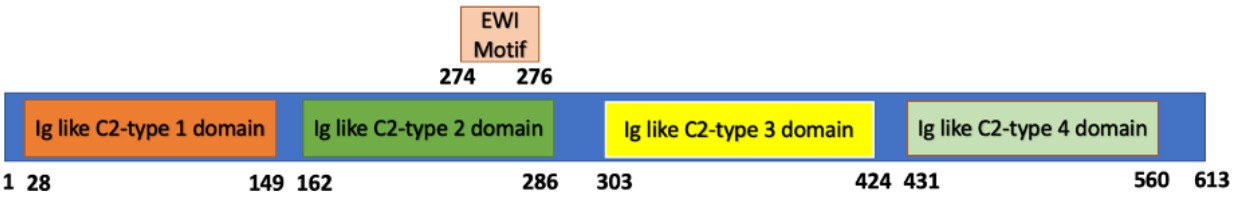


Figure 1: Structural and functional domains recognized in IGSF8. Schematic structure of the IGSF8 protein showing Ig like C2-type 1, Ig like C2-type 2, Ig like C2-type 3 d, and Ig like C2-type 4 domains and EWI motif. Ig like C2-type 3 and 4 required for the interaction with CD81 and CD9.

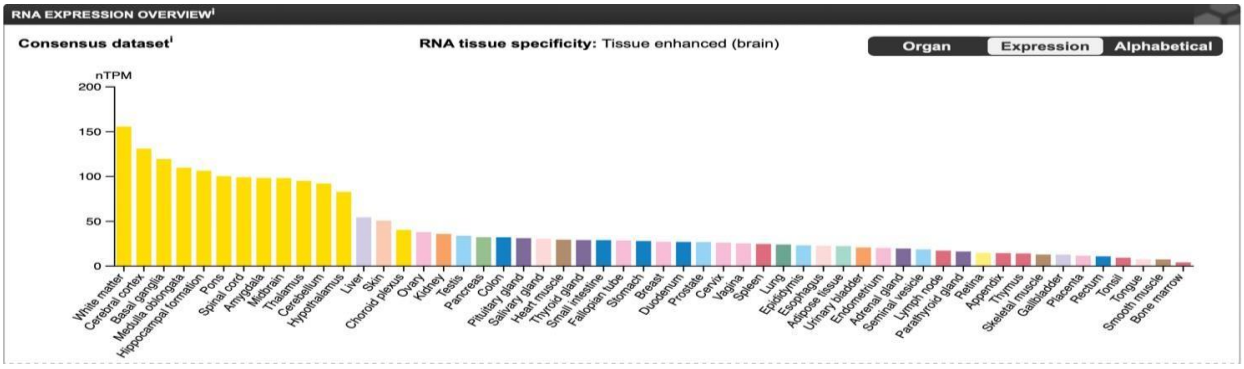


Figure 2: *IGSF8* expression pattern in different tissues. IGSF8 mRNA data showing high expression of IGSF8 in the brain, liver, and skin tissues. This bar graph screenshot from The Human Protein Atlas. *(https://www.proteinatlas.org/ENSG00000162729-IGSF8/ti*

1.5.1 IGSF8 in Cancers

Two tetraspanins proteins, CD9 and CD81, which are essential for the biology of astrocytes and gliomas, are significantly altered by IGSF8 at the biochemical level in terms of how they are organized (Longo et al., 2001). CD9 and CD81 on the cell surface directly interact with IGSF8, causing molecular reorganization (Longo et al., 2001). These interactions impact both the proliferation and motility of healthy cells and the metastatic potential and tumorigenicity of cancer cells. This influence primarily occurs

within CD9 and CD81-enriched microdomains on the cell surface, where they collaborate with IGSF8 to modulate these functions rather than acting as cell surface receptors (Hemler, 2005). Notably, the protein composition of tetraspanin-enriched microdomains can change as cells progress towards a malignant state (Termini & Gillette, 2017). *IGSF8* is a tumor suppressor gene that is highly expressed in normal brain cells. Glioblastoma tumors exhibit markedly reduced levels of *IGSF8*. Further evidence that IGSF8 may act as a natural inhibitor of glioblastoma comes from the correlation between reduction of *IGSF8* expression and a shorter overall survival time in patients with human gliomas. Conversely, overexpression of *IGSF8* in glioblastoma cell lines has been associated with decreased in vivo tumor growth (Kolesnikova et al., 2009). Studies comparing glioblastomas to normal brains reveal downregulation of IGSF8 expression. Furthermore, *IGSF8* expression has been shown to curtail three-dimensional growth, invasion, migration in vitro, and glioblastoma tumor growth in vivo. These effects coincide with changes in matrix metalloproteinases (MMP-2 and MT1-MMP), CD9, and CD81 expression on glioblastoma cell surfaces (Kolesnikova et al., 2009).

In prostate cancer, IGSF8 acts as an inhibitor of cell migration and exerts control over epidermoid cell reaggregation and motility on laminin 5—a protein promoting cell adhesion and migration. It achieves this by inhibiting T cell mobility through the association of CD81 with CD82 (Oren et al., 1990). This suggests that IGSF8 negatively regulates cell motility and influences integrin-dependent morphology. In vitro silencing of *IGSF8* has been associated with increased cancer metastasis potential, characterized by enhanced EGFR clustering, endocytosis, activation, and Erk-MAP kinase activation. These signaling alterations induce epithelial-mesenchymal transition (EMT) in cancer cells, promoting tumor progression and proliferation (Fu et al., 2021). Therefore,

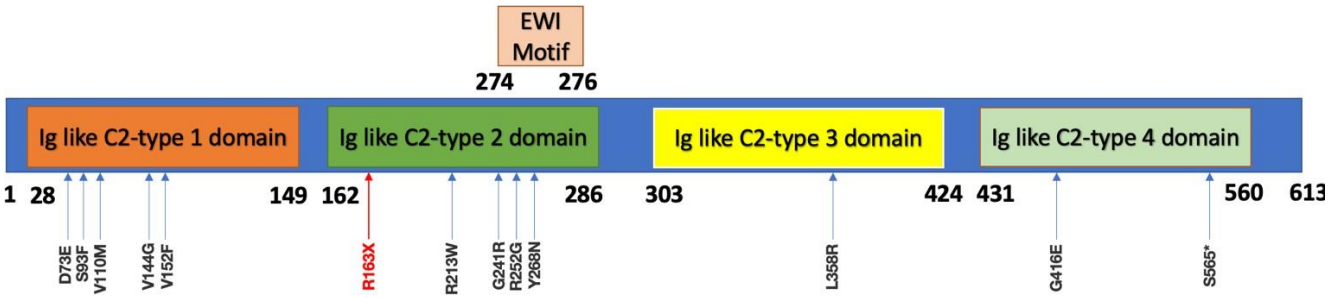


Figure 3: IGSF8 schematics illustrates different missense and nonsense mutations. Mutations scattered along the IGSF8 gene reported by different datasets from breast cancer patients. Highlighted in red the novel variant we have discovered (IGSF8 p. R163X).

inhibiting EGFR or Erk kinase may attenuate cancer cell development resulting from *IGSF8* silencing.

Additionally, changes in *IGSF8* expression can impact the transforming growth factor (TGF) signaling pathway. The presence of IGSF8 sequesters CD9 and CD81, disrupting the interaction between IGSF8 and TGF receptors. This negatively regulates TGF signaling, contributing to increased invasion and metastasis of melanoma cells lacking IGSF8 (Zöller, 2009). In vitro experiments manipulating *IGSF8* expression in various melanoma cell lines corroborate the adverse regulation of TGF signaling by *IGSF8* (Wang et al., 2015).

Furthermore, IGSF8 enhances lysosomal activity, promoting degradation of growth factor receptors and integrins. This leads to reductions in their levels and functions. Overexpression of *IGSF8* dominantly in cancer cells initiates motogenic and mitogenic signaling, facilitated by decreased total cellular and cell surface levels of integrins and growth factor receptors. This increase in lysosomal trafficking and proteolytic degradation of membrane receptors diminishes cell-matrix adhesion, cell motility, cell transformation in vitro, and tumor development, burden, and metastasis in vivo (Flannery et al., 2010).

However, IGSF8's efficacy in reducing glioblastoma, melanoma, and prostate cancer through its collaboration with tetraspanins contrasts with its relatively unexplored function in breast cancer (Vences-Catalán & Levy, 2018). Notably, datasets such as the

Metastatic Breast Cancer Project (MBC project), Breast Invasive Carcinoma (TCGA, Firehose Legacy), Breast Invasive Carcinoma (TCGA, PanCancer Atlas), Breast Invasive Carcinoma (TCGA, Nature 2012), Proteogenomic landscape of breast cancer (CPTAC, Cell 2020), and Breast Invasive Carcinoma (TCGA, Cell 2015) contain comprehensive summary data visualizations and clinical information from cancer patients. This clinical data encompasses mutation counts, details about mutated genes, patient demographics, sample types, and more. Among these datasets, nearly 22 samples with *IGSF8* nonsense, missense, and fusion mutations have been identified across the gene's sequence. These variations have been documented in diverse breast cancer patient types, as depicted in Figure 3. Additionally, the Ig-like C2-type 1 and Ig-like C2-type 2 domains showcase the most significant number of reported mutations in breast cancer patients analyzed in the mentioned datasets, as illustrated in Figure 4.

Based on the current literature, the relationship between *IGSF8* and breast cancer remains inadequately explored. This study has uncovered a novel mutation in *IGSF8*, a member of a protein family implicated in various cancers, including breast cancer. Furthermore, given its accessibility on the cell surface and its dual inhibition effect on growth factor receptors and integrins, IGSF8 emerges as a promising therapeutic candidate (Montpellier et al., 2011).

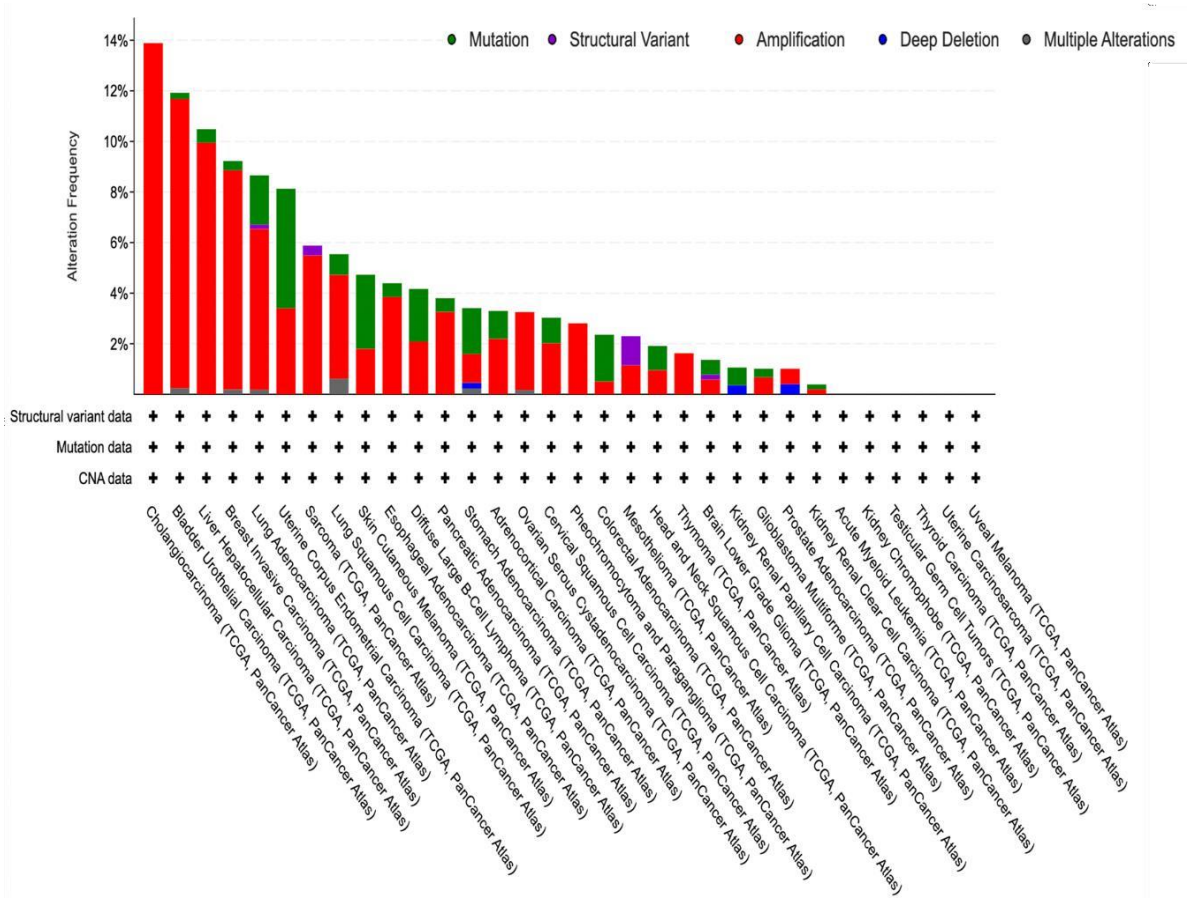


Figure 4: A summary of cancer types that *IGSF8* has been found mutated in from CBioportal. Different data sets reported *IGSF8* amplification in most of the cancer samples in the different studies while few samples have mutations and deletions in the *IGSF8* gene and the most common cancer types that have mutation in IGSF8 gene were cholangiocarcinoma, bladder, liver and breast cancers.

1.6 Thesis Statement

Breast cancer is recognized as the most prevalent malignancy among women globally, and it is statistically documented as the predominant cancer type in Palestine. Despite this prevalence, the genetics of breast cancer patients in Palestine have received limited attention. Given the escalating incidence of breast cancer and the growing identification of causal genes, this study aims to employ WES to investigate the genetic basis of hereditary breast cancer among Palestinians. The findings are anticipated to uncover the

genomic landscape of breast cancer in Palestine, thereby facilitating early diagnosis, screening, and targeted treatment approaches.

1.7 Study Objectives

Our study focuses on utilizing NGS to sequence the DNA of 7 members from a Palestinian family with a history of breast cancer, revealing a novel mutation *IGSF8* c.C487T that is present in all tested family members.

Our study aims to: (i) understand the genetic landscape of breast cancer in Palestinian patients by deciphering the mutation spectrum in genes associated with cancer, and (ii) contribute to the development of specialized targeted panels tailored to the Palestinian population, as universally applied panels such as the *BRCA* panel might not be universally applicable.

1.8 Study Significance

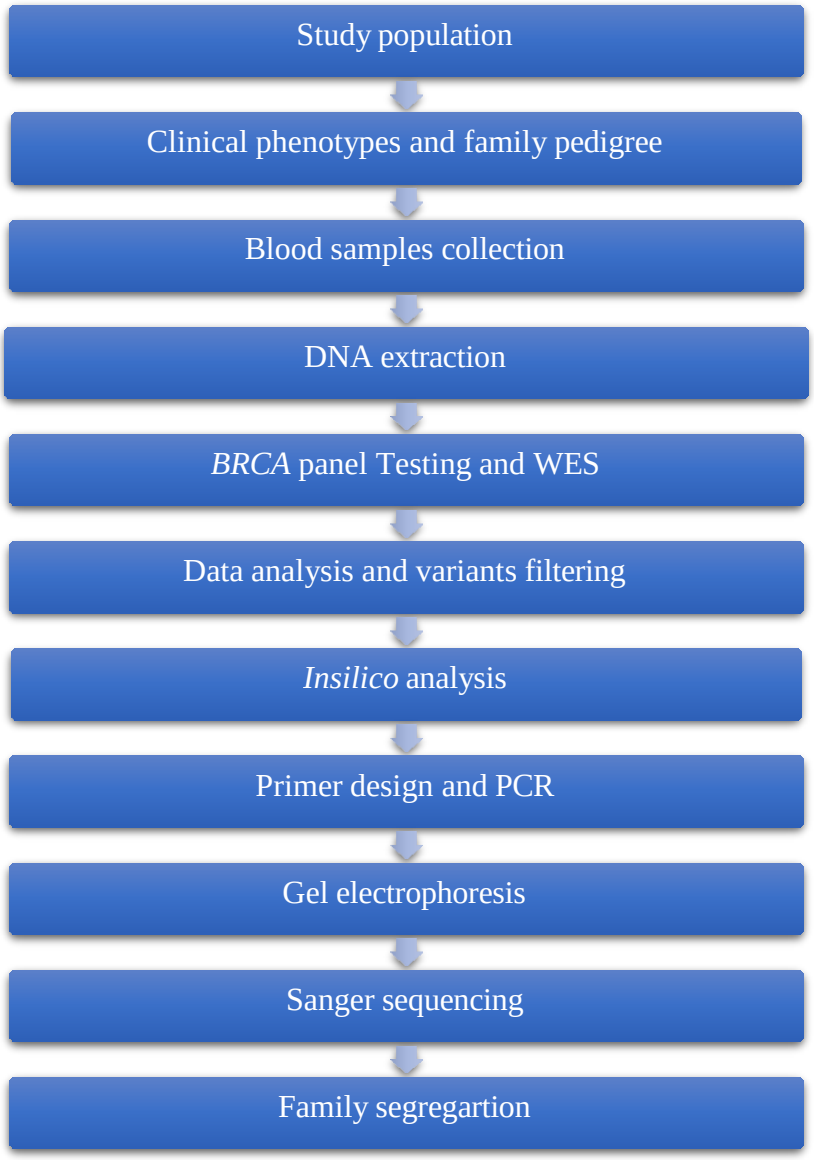
This project's core objective is to unravel the molecular underpinnings of breast cancer in Palestinian patients. Through the identification of novel germline mutations previously unexplored in breast cancer or hereditary breast cancer, the study seeks to unveil the genetic constitution of Palestinian breast cancer patients. Consequently, this insight will not only aid in diagnosis but also play a pivotal role in assessing cancer risk and susceptibility.

Chapter 2: Materials and Methodology

WES was performed in an attempt to identify germline pathogenic mutations within the family. Analysis of WES data resulted in the identification of a pathogenic mutation in the *IGSF8* gene. Family members were subsequently tested for the same pathogenic mutations carried by the proband by sanger sequencing.

2.1 Study Design

This study has been designed as shown in the schematics below:



This study was designed to target a Palestinian family with a history of hereditary cancers. The family consisted of two members who were diagnosed with breast cancer, and one with pancreatic cancer. Initially, an NGS targeted panel was conducted. However, no pathogenic variants were identified through this approach. Subsequently, whole exome sequencing was undertaken in an attempt to identify variants that could potentially elucidate the underlying phenotypes within the family.

2.2 Site and Population

The selected family for this study comprises 7 sisters, among whom two were diagnosed with breast cancer, one with pancreatic cancer, and all have undergone distinct therapeutic interventions (chemotherapy, radiotherapy, mastectomy, and hysterectomy). To explore the potential presence of germline mutations running in the family, samples were collected from both the sisters and their daughters.

2.3 Clinical Phenotyping and Family Pedigrees

The family pedigree was meticulously constructed, revealing the following information: The father was diagnosed with pancreatic cancer and passed away eight months later. On the paternal side, one uncle had a history of pancreatic, bone, and prostate cancers, another aunt had colon cancer, and yet another had breast cancer. Some cousins were diagnosed with thyroid, skin, prostate cancers, and leukemia. On the mother's side, the father had liver cancer, and one uncle had lung cancer.

2.4 Samples collection

Multiple field visits were organized to the families under study to facilitate the collection of samples. EDTA Whole-Blood samples (around 2.5 mL) were collected from the seven sisters and their daughters, both for the purpose of *BRCA* panel testing and whole exome sequencing.

2.5 Ethical Consideration

The individuals participating in this research, including both patients and family members, willingly provided their agreement and signed consent forms. To ensure the preservation of patient confidentiality, samples were assigned identification numbers. The Helsinki Committee for Ethical Approval granted its approval. The IRP number: RHRC/HC/1216/22.

2.6 DNA Extraction

DNA extraction was performed using the QIAamp Blood Mini Kit (QIAGEN, Germany). A 20 µl volume of QIAGEN Protease (proteinase K) was placed in a 1.5 ml microcentrifuge tube, followed by the addition of a 200 µl whole blood sample. Subsequently, 200 µl of lysis buffer (AL) was added into the mix. The tube was then subjected to a 15-second pulse-vortexing and incubated at 56°C for 10 minutes. Adding 200 µl of ethanol (96–100%) to the mixture, a 15-second pulse-vortexing was performed, and the tube was briefly centrifuged to eliminate any drops from the lid. The resultant mixture was poured into the QIAamp Mini spin column and centrifuged at 8000 rpm for 1 minute. The QIAamp Mini spin column was placed in a sterile 2 ml collection tube, and the tube containing the filtrate was discarded. Successively, 500 µl of wash 1 buffer

(AW1) was added to the mini spin column, which was again centrifuged at 8000 rpm for 1 minute. Placing the mini spin column in a sterile 2 ml collection tube, the collection tube housing the filtrate was discarded. 500 µl of wash 2 buffer (AW2) was then added to the mini spin column, and full-speed centrifugation (14,000 rpm) was conducted for 3 minutes. Once more, the mini spin column was transferred to a sterile 2 ml collection tube, and the collection tube with the filtrate was discarded. The mini spin column underwent a second full-speed centrifugation for 1 minute, followed by the discarding of the tube containing the filtrate. To elute the DNA from the mini spin column, 200 µl of elution buffer (AE) or distilled water was added and allowed to incubate at room temperature (15–25°C) for 1 minute. The final step involved centrifugation at 8000 rpm for 1 minute to facilitate DNA elution from the mini spin column. DNA concentration (ng/µl) was quantified using the nanodrop, revealing concentrations ranging between 30 to 50 ng/µl.

2.7 Initial assessment of the samples for *BRCA* mutations

DNA samples obtained from the sisters' blood were subjected to analysis for *BRCA* mutations using the *BRCA* plus panel test (Human *BRCA1* and *BRCA2* Plus Panel, Cat.no: DHS-103Z, QIAGEN, Germany). The testing procedure was carried out at the National Center of Cancer Diagnostics and Human Genetics within the Palestinian Ministry of Health, utilizing the Illumina Nextseq 550 NGS system (Illumina, Inc, USA).

The QIAseq Targeted DNA Panels are supplied in single-tube primer mixes containing up to 20,000 primers per panel. These panels employ 10 to 80 ng of fresh DNA to effectively enrich specific genes and regions. The key steps of the QIAseq Targeted DNA Panels are visually presented in (5) and Figure (6).

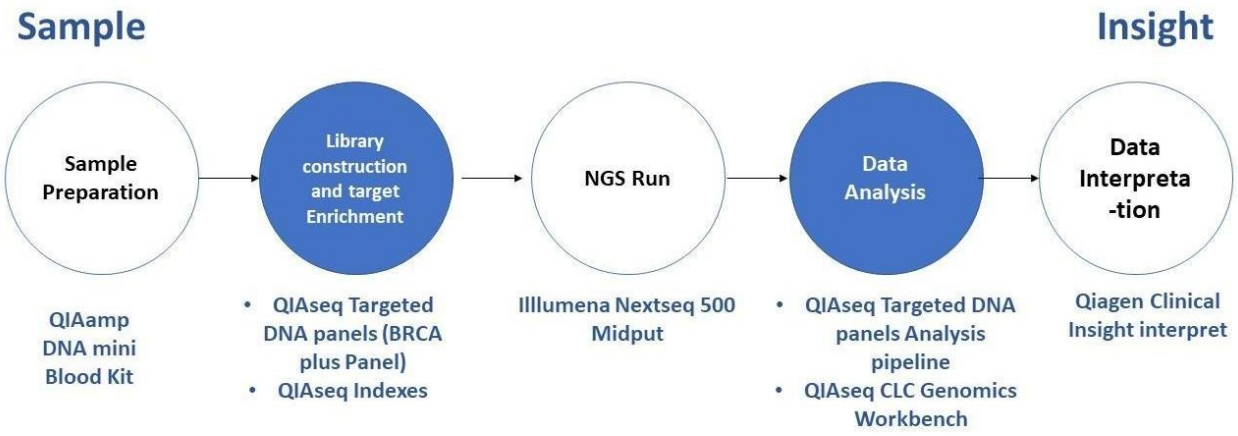


Figure 6: QIAseq targeted DNA panel library preparation. The major steps for library preparation using QIAseq targeted DNA panel (The *BRCA1* and *BRCA2* plus panel) includes Fragmentation of DNA first, then end-repaired and A-tailed. Ligation of the fragmented DNA at the 5′ ends using a platform-specific adaptor carrying UMIs and the sample index. Target enrichment to ensure UMI-containing DNA molecules are appropriately enriched in the sequenced library., and finally, Universal PCR amplification to expand the library using platform-specific adaptor sequences.

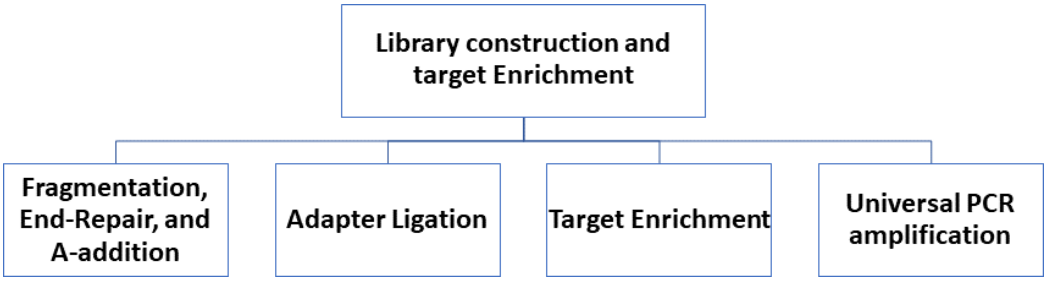


Figure 5: The QIAseq targeted DNA panel's major steps. From sample preparation, DNA extraction, library preparation using QIAseq targeted DNA panels (BRCA plus panel), NGS run on the Nextseq550 Midput machine, data analysis using QIAseq targeted DNA panels analysis pipeline or QIAseq CLC genomics workbench and Data and detected variants interpretation using Qiagen clinical insight interpret.

2.7.1 Fragmentation, End-Repair, and A-addition

The genomic DNA was diluted to a concentration of 2 ng/µl using nuclease-free water. For each sample, 5 µl of diluted DNA (10 ng in total) was placed in separate 0.2 ml PCR tubes for subsequent fragmentation. A mixture consisting of fragmentation buffer, FERA solution, and nuclease-free water (15 µl) was added to each sample, followed by 10 alternating upward and downward shaking movements. This series of steps was performed while maintaining the samples on ice. Subsequently, 5 µl of the fragmentation

enzyme mix was introduced to each reaction tube. The samples were then transferred to a thermal cycler and subjected to amplification using the following protocol: 1 minute at 4°C, 24 minutes at 32°C, and 30 minutes at 72°C. At the conclusion of the program, the samples were maintained at 4°C, and the adapter ligation protocol was initiated promptly.

2.7.2 Adapter Ligation

A total of 25 µl from the preceding step was combined with 25 µl of the ligation master mix (comprising ligation buffer, 5x IL-701NJ adapter, DNA ligase, and ligation solution) within 0.2 µl PCR tubes. After that, the reactions were placed in a PCR machine, incubated for 15 minutes at 20°C, and then moved to an ice bath.

2.7.3 Cleanup protocol

From the ligation PCR product, 50 µl was transferred to a LoBind Eppendorf tube, and then 50 µl of nuclease-free water and 100 µl of QIAact magnetic Beads were added to the mixture, followed by a 5-minute incubation at room temperature. The tubes were then positioned on a magnetic rack for 10 minutes to facilitate separation of the beads from the supernatant. After the solution turned clear, the supernatant was carefully removed. The addition of 200 µl of freshly prepared 80% ethanol was performed while the tubes were on the magnetic rack, and the process was repeated until the solution remained clear. Subsequently, the tubes were allowed to air dry to eliminate residual ethanol. For DNA elution from the beads, 52 µl of nuclease-free water was added. The cleanup procedure was repeated, but this time DNA was eluted using 12 µl of nuclease-free water, and 9.4 µl of the supernatant was transferred into 0.2 µl PCR tubes to proceed with the target enrichment protocol.

2.7.4 Target Enrichment

A total of 10.6 µl of the target enrichment master mix (comprising TEP-CR Buffer 5x, QIAseq Targeted DNA Panel, IL-Forward primer, and Hot Star Taq DNA Polymerase) was added to the 9.4 µl of cleaned DNA from the previous step. After that, the samples were placed in a thermal cycler and subjected to amplification using the following protocol: 13 minutes at 95°C, 2 minutes at 98°C as the initial denaturation, 8 cycles of 15 seconds at 98°C and 10 minutes at 68°C, followed by 5 minutes at 72°C. The program concluded with samples being held at 4°C. After the target enrichment step, another cleanup was performed as detailed in section 2.7.3 (Cleanup Protocol), using 16 µl of nuclease-free water for DNA elution, and 13.4 µl of the DNA was retained for the subsequent step.

2.7.5 Universal PCR amplification

A master mix containing 6.6 µl of UPCR buffer (5x), nuclease-free water, and Hot Star Taq DNA Polymerase was added to the 13.4 µl DNA from the previous step. The samples were then subjected to amplification using the following protocol: 13 minutes at 95°C, 2 minutes at 98°C for initial denaturation, 21 cycles of 15 seconds at 98°C and 2 minutes at 60°C, followed by 5 minutes at 72°C. At the end of the program, samples were maintained at 4°C. After the universal PCR step, cleanup was conducted as described in section 2.7.3, using 30 µl of nuclease-free water for DNA elution in this instance. Finally, 28 µl of the clean library was retrieved and stored at -20°C until further use.

2.7.6 Assessment of the Libraries

The yield (ng) of the PCR-enriched DNA library was determined using the Qubit dsDNA HS Assay Kit (Invitrogen, Thermo Fisher, US, lot no: 2483579). The product size (bp) was evaluated using the QIAxcel Advanced instrument and the QIAxcel DNA High-Resolution Kit (1200). Library concentrations measured on Qubit4 ranged between 20 to 25 ng/µl, while the product size assessed by QIAxcel fell within the range of 300 to 400 bp.

2.7.7 Denaturing and Diluting the Pooled Library

The libraries were first diluted to 10nM using 0.1% Tween 20. Subsequently, the 10nM samples were further diluted to 4nM by combining 2 µl of each sample with 5 µl of nuclease-free water. Equal volumes of the libraries were then mixed in a single tube for the next step.

For denaturation, 5 µl of the pool (mixed libraries) was treated with 5 µl of 0.2N NaOH and incubated for 5 minutes at room temperature. Following this, 5 µl of 200mM Tris-HCl pH 7 was added to the mixture. Prior to loading the library pool into the cartridge, it was diluted to 20 pM by adding 985 µl of prechilled HT1, and subsequently, it was further diluted to 1.5 pM by mixing 97 µl of the 20 pM library pool with 1203 µl of prechilled HT1 as the final step. The denatured and diluted library pool was then ready for loading into the NextSeq reagent cartridge (Cat no. 20024904).

2.7.8 Read 1 Primer (Custom Sequencing Primer)

The Read 1 Primer (Cat no. 333514) was prepared by adding 1994 µl of HT1 buffer to 6 µl of the primer, resulting in a final concentration of 0.3M. Subsequently, 2 ml of the diluted primer was loaded into the NextSeq reagent cartridge.

2.8. Whole Exome Sequencing by NGS using Illumina DNA Prep with Enrichment

Whole exome sequencing was performed at the Arab American University using NGS technology with the Illumina Nextseq 550 (Illumina, Inc, USA) NGS system. The Illumina DNA Prep with Enrichment Kit, along with tagmentation and the Illumina Exome Panel, was employed to prepare and enrich the exome. The prepared libraries were then placed onto a reagent cartridge before being subjected to sequencing using the NextSeq 550 System. Data analysis, including alignment and variant calling, was conducted either locally or in the cloud using the DRAGEN Enrichment Pipeline or other BaseSpace Sequence Hub programs. The workflow of the Illumina DNA Prep with Enrichment Kit is depicted in Figure (7).

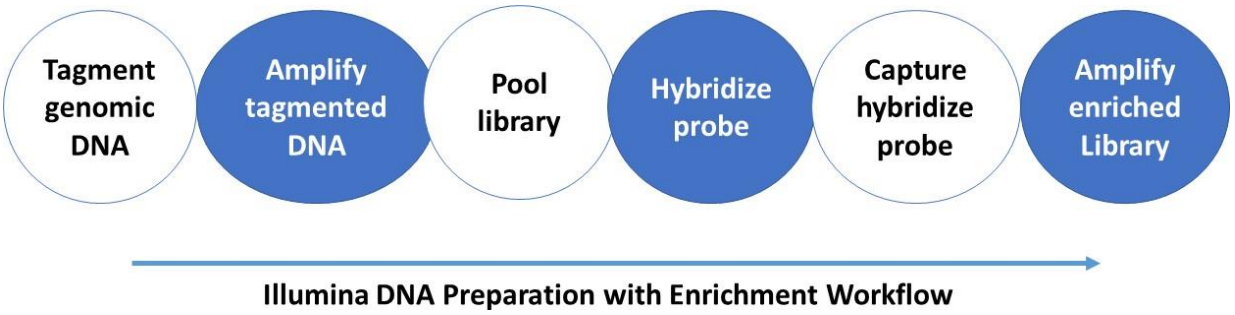


Figure 7: DNA preparation by Illumina. Illumina's DNA preparation workflow for whole exome sequencing by the NGS technique. Tagmentation of the genomic DNA is the first step, then amplifying the tagmented DNA to create libraries from each sample. Pool the libraries and hybridize the probe; this step involves binding capture probes to specific sections of the DNA. Then, Streptavidin Magnetic Beads (SMB3) were used to capture probes hybridized to the specified regions of interest and amplify enriched libraries as the last step before loading the libraries pool on the Illumina Nextseq 550.

2.8.1 Purification and Quality Control

The Qubit dsDNA HS assay kit #Q23850 (Thermo Fisher, USA) was utilized to determine the concentration of each library. Furthermore, the High Sensitivity DNA kit catalog #5067 (QIAGEN, Germany) was employed to measure the mean fragment size of the libraries. Sequencing primer preparation and cartridge loading were executed as described in step 2.2.8.

2.8.2 Variants Analysis for Human *BRCA* plus panel and WES

A VCF file containing annotations of variant effects was generated using the SnpEff tool. This tool annotates variants with their calculated impacts on known genomic characteristics. The VCF file includes various information, including chromosome number, position, reference and variant nucleotides, gene name, genomic context (such as exon, intron, splice site), synonymy (indicating whether the variant is synonymous or not), amino acid alteration, clinical significance (pathogenic or benign), phenotypes associated with the variant according to HPO and OMIM, as well as the average frequency of the variant within a population. Variants deemed unlikely to exert a pathogenic effect were subsequently filtered out based on several criteria:

1. High-quality metrics, including QD > 10, MQ > 59, SOR < 4, and coverage by at least 5 reads. 2. Frequency less than 0.01 (1%) according to GnomAD control and GnomAD genome datasets. 3. Variants not classified as benign or likely benign in ClinVar.

4. Variants not classified as benign or probably benign based on ACMG standards.

5. Missense variants with at least three damaging scores, such as CADD > 15, PolyPhen HVAR > 0.9, PolyPhen HDIV > 0.9, SIFT < 0.05, REVEL > 0.3, and MutationTaster = D.

6. Splicing non-indel variants with detrimental scores from ADA > 0.5, RF > 0.5, SpliceAI > 0.2, and Delta psi > 2.5 or -2.5

2.9. *In Silico* Analysis

Confirmed variants obtained from Human BRCA plus panel testing (*BRCA2* c.4563A>G, *BRCA2* c.6513G>C, *BRCA2* c.7242A>G, *BRCA2* c.3396A>G, *BRCA2* p.Val2466Ala, BRCA2 p.Asn372His, *BRCA1* p.Ser1613Gly, *BRCA1* p.Lys1183Arg, *BRCA1* p.Glu1038Gly, *BRCA1* p.Pro871Leu, *BRCA1* c.4308T>C, *BRCA1* c.2311T>C, *BRCA1* c.2082C>T, *TP53* c.993+12T>C, *TP53* c.215C>G and *TP53* c.108G>A) and from WES (*FAT4* c.C9494T: p.T3165M, *IGSF8* c.C487T: p.R163X, *PPP1R12C* c.C673T: p.R225C, *PSEN2* c.G166A: p.G56S, SLC6A12 c.G1247A: p.R416Q, *TRIM72* c.891delT: p. A298Rfs*64, *TSHR* c.C202T: p.P68S, *VSIG2* c.297dupC: p. T100Hfs*11, *WRN* c.C1457T: p.T486M, *ZFPM1* c.2018_2019insGGACGACGCGGA: p.D679_P680insAEDD) were subjected to in silico analysis using various bioinformatics tools to assess their pathogenicity and significance. Tools employed in this analysis included: COBALT Alignment tool for identifying locus conservation: https://www.ncbi.nlm.nih.gov/tools/cobalt. PolyPhen-2 (Polymorphism Phenotyping v2): http://genetics.bwh.harvard.edu/pph2/. PROVEAN (Protein Variation Effect Analyzer): http://provean.jcvi.org/index.php. FATHMM (Functional Analysis through Hidden Markov Models (v2.3)): http://fathmm.biocompute.org.uk/. Mutation Taster: https://www.mutationtaster.org/. SIFT (Sorting Intolerant from Tolerant): https://sift.bii.a-star.edu.sg/. GVGD (Grantham Variation, Grantham Difference): http://agvgd.hci.utah.edu/index.php. LIST-S2: https://precomputed.list-s2.msl.ubc.ca/ .

These tools collectively aid in evaluating the pathogenic potential and significance of the identified variants. Their conservation predicted effects on protein function, and potential consequences are analyzed to better understand their role in disease pathology.

2.10. Primer Design

Primer3web (v4.1.0) (https://primer3.ut.ee) was employed to design primers that flank the identified variants. Optimal primer pairs were selected based on parameters such as melting temperature (Tm), GC content percentage, and product size. To assess primer specificity, the UCSC *in-silico* PCR tool (https://genome.ucsc.edu/cgi-bin/hgPcr) was utilized. For genotyping the *IGSF8* variant, the following primer pair was used: forward primer: ctccactgatacccgctacc and reverse primer: tcccaccacttcctgcagag.

2.11. PCR Protocol

The *IGSF8* target region was amplified on the Veriflex thermal cycler (Applied Biosystems, US) using the following conditions: initial DNA denaturation for 5 min at 95°C, followed by 35 cycles of DNA denaturation for 30s at 95°C, annealing at 60°C for 30s, and extension at 72°C for 30s. A final extension was performed at 72°C for 7 min, followed by storage at 4°C.

The PCR reaction was prepared in a 20 µl final volume as follows: 0.5 µl of the forward primer (concentration), 0.5 µl of the reverse primer (concentration), 2 µl of DNA (31 ng/µl), 10 µl of GoTaq 2X green master mix, and 7 µl of nuclease-free water. A no-template control (NTC) was included to ensure no contamination. The PCR product was visualized using a 1X TBE 1.5% agarose gel.

2.12. Sanger Sequencing

First, PCR products were enzymatically cleaned up using the ExoSAP-IT kit (Applied Biosystem, Cat no.78201.1.ML). 2 µl of the ExoSAP reagent was added to 5 µl of the PCR product. The reaction mix was incubated for 15 min at 37°C, then for 15 min at 80°C, and finally held at 4°C.

Second, for end labeling, 4 µl of the BigDye™ Terminator v1.1 Cycle Sequencing Kit (Applied Biosystem, Cat no. 4337450), 4 µl of buffer, and 8 µl of free nuclease water were added to 1 µl of cleaned DNA and 1 µl of primer. The samples were then transferred to a thermal cycler and amplified using the following conditions: 1 min at 96°C, 25 cycles of 30s at 96°C and 5s at 50°C, followed by 4 min at 60°C. At the end of the program, samples were maintained at 4°C.

Before injecting the BigDye product into the SeqStudio genetic analyzer, the BigDye XTerminator™ Purification Kit (Applied Biosystem, Cat no. 4376486) was employed to clean the BigDye product. 7 µl of the cleaning beads with 30 µl of buffer were added to the BigDye product and mixed by vortexing for 30 min, then centrifuged for 2 min. The purified product was then ready to be injected into the SeqStudio. The long-run module was utilized since the product size is 970 bp.

Chapter 3: Results

Our study focused on identifying novel mutations in a Palestinian family with a history of hereditary breast cancer. The analysis of the enrolled patients' WES data resulted in the identification of *IGSF8* c.C487T mutation, as well as the segregation of the variant among family members and confirming the predicted inheritance pattern using Sanger sequencing.

3.1 Clinical Findings

The family pedigree, as illustrated in Figure 5, portrays multiple generations of the family with a diverse range of cancer occurrences. The proband of our study, denoted as III-6, received a breast cancer diagnosis at the age of 34. Within the following year, she underwent a total laparoscopic hysterectomy due to endometrial hyperplasia, the presence of small polyps on both ovaries, and owing to the family's history of pancreatic, colon, and breast cancer. This patient has two sisters, III-7 was diagnosed with breast cancer, while III-8 underwent splenectomy for pancreatic cancer. The proband's father, II-9, died from pancreatic cancer just eight months after diagnosis. Further back in the paternal family history, there was a notable incidence of cancer, including II-2 diagnosed with breast cancer, II-3 with colon cancer, II-7 with prostate, pancreatic, and bone cancer, III-2 with skin cancer, III-4 with thyroid cancer, and IV-1 with prostate cancer. Additionally, leukemia was diagnosed in IV-2 and IV-3. Meanwhile, III-1 exhibited a benign renal mass under regular observation. On the maternal side, I-3 (the proband's grandfather) was diagnosed with liver cancer, while II-11 was diagnosed with lung cancer.

3.2 *BRCA* Panel, Whole exome sequencing results and identification of candidate gene

Blood samples extracted from the proband of this study (III-6) and her sisters (III-7 to III-12), as depicted in Figure 8, underwent *BRCA* plus panel testing via NGS at the National Center of Cancer Diagnostics and Human Genetics within the Palestinian Ministry of Health. The analysis of variants using the Qiagen analysis software (QCI Secondary Analysis) is documented in Table 1. Notably, these samples yielded negative results for known pathogenic variants in *BRCA1*, *BRCA2*, *PTEN*, *P53*, and *CDH1* genes. Every variant that emerged was classified as benign in the ClinVar database. This led us to engage in trio whole exome sequencing, involving II-9, III-6, and III-7.

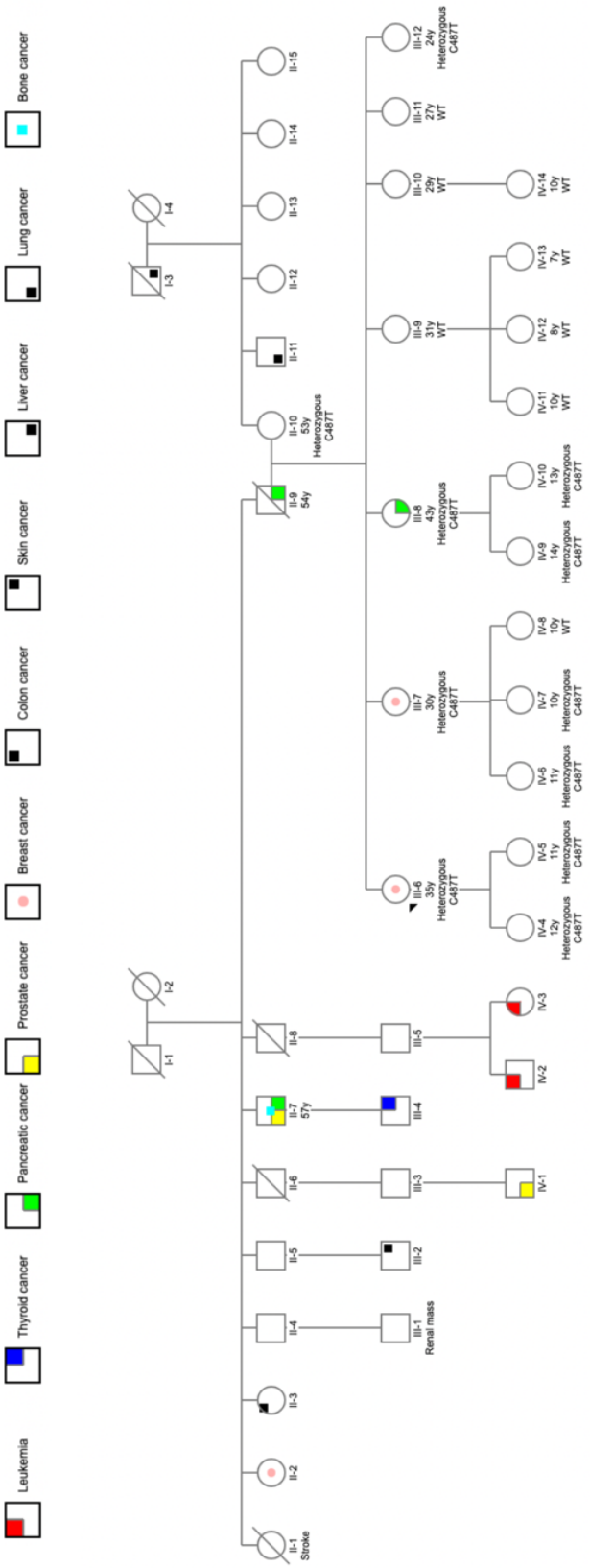


Figure 8: family pedigree. The pedigree structure of an Arab Palestinian family affected with different cancer types, including breast, pancreatic, prostate, colon, lung, liver, skin, and thyroid cancers, in addition to leukemias. The proband selected for this study is indicated by the arrow (III-6). II-10, III-6 to III-12, and IV-4 to IV-14 are genotyped for the *IGSF8* c.C487T mutation.

Table 1: Human *BRCA* plus sequencing panel filtered variants by QCI secondary analysis software for III-6 to III-12 samples.

Gene	Mutation name
BRCA 2	c.4563A>G
BRCA 2	c.6513G>C
BRCA 2	c.7242A>G
BRCA 2	c.3396A>G
BRCA 2	Val2466Ala
BRCA 2	Asn372His
BRCA 1	Ser1613Gly
BRCA 1	Lys1183Arg
BRCA 1	Glu1038Gly
BRCA 1	Pro871Leu
BRCA 1	c.4308T>C
BRCA 1	c.2311T>C
BRCA 1	c.2082C>T
TP53	c.993+12T>C
TP53	c.215C>G
TP53	c.108G>A

The outcomes of the whole exome sequencing were distilled into a list of ten distinct variants, as detailed in Table 2. Regrettably, none of these variants presented an explanatory correlation to the family's susceptibility to breast cancer and other malignancies. However, in our pursuit of identifying rare mutations linked to tumorigenesis, the *IGSF8* gene emerged as a potential candidate. Of particular interest, the *IGSF8* c.C487T variant manifested within a conserved region of the *IGSF8* protein

(Figure 9). Furthermore, MutationTaster prognosticated this variant as disease-causing and suggestive of a lost splice site.

Table 2: Whole exome sequencing filtered results for II-10, III-6, and III-7 samples.

Gene	Mutation name
FAT4	c.C9494T:p.T3165M
IGSF8	c.C487T:p.R163X
PPP1R12C	c.C673T:p.R225C
PSEN2	c.G166A:p.G56S
SLC6A12	c.G1247A:p.R416Q
TRIM72	c.891delT:p.A298Rfs*64
TSHR	c.C202T:p.P68S
VSIG2	c.297dupC:p.T100Hfs*11
WRN	c.C1457T:p.T486M
ZFPM1	c.2018_2019insGGACGACGCGGA:p.D679_P680insAEDD

Figure 9: IGSF8 arginine 163 conservation. Multiple amino acid sequence alignment of homo sapiens and other species was performed using Clustal Omega alignment tool of EMBL-EBI (https://www.ebi.ac.uk/Tools/msa/clustalo/). The arrow indicates the conserved Arginine amino acid at position 163.

3.3 Validation of *IGSF8* c.C487T and segregation analysis

For the validation of the presence of the *IGSF8* c.C487T mutation via Sanger sequencing, we designed primers flanking the mutation site, as elaborated in section 2.10. Subsequently, DNA extracted from samples of II-9, III-6, and III-7 were subjected to Sanger sequencing to validate *IGSF8* c.C487T. Concurrently, a segregation analysis was conducted to explore the distribution of the *IGSF8* c.C487T mutation within the family. This analysis encompassed the Sanger sequencing of samples from III-6, III-7, III-8, III-9, III-10, III-11, III-12 (Figure 8), and their daughters (IV-4 to IV-14) (Figure 8).

Preliminary to Sanger sequencing, the PCR products were visualized through electrophoresis on a 1.5% agarose gel to ascertain their quality. As evident in Figure 10, the PCR product measured 970 bp in length. Additionally, the 1000 bp DNA ladder was employed as a reference for size determination.

The PCR products underwent purification and preparation for Sanger sequencing, followed by analysis utilizing the Finch TV software.

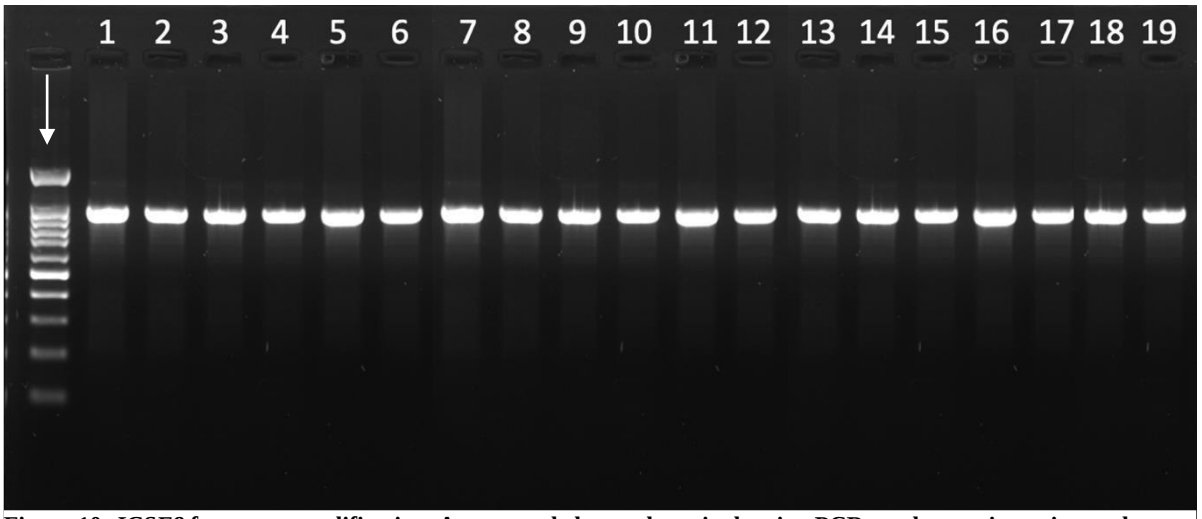


Figure 10: *IGSF8* fragment amplification. Agarose gel electrophoresis showing PCR products using primers that flank IGSF8 mutation. The arrow points to the 100bp DNA ladder. Lanes from 1 to 19 are the family member samples (1: III-6, 2: III-7, 3: III-8, 4: III-9, 5: III-10, 6: III-11, 7: III-12, 8: IV-4, 9: IV-5, 10: IV-6, 11: IV-7, 12: IV-8, 13: IV-9, 14: IV-10, 15: IV-11, 16: IV-12, 17: IV-13, 18: IV-14 and 19: II-10).

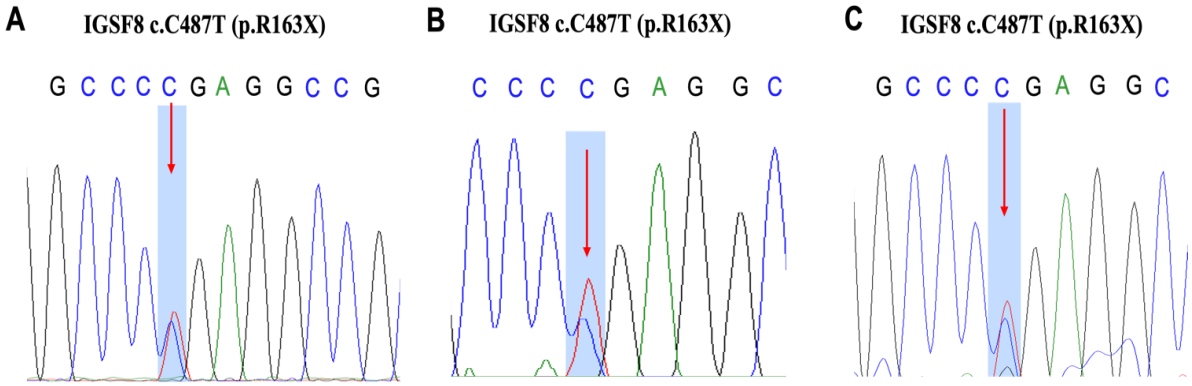


Figure 11: Electropherogram showing the DNA sequence to validate the presence of the *IGSF8* (c.C487T) stop mutation in II-9, III-6 and III-7. (A) II-9 (The mother is heterozygous for IGSF8 c.C487T. (B, C) III-6 and III-7 respectively, are heterozygous for IGSF8 c. C487T.

The findings unveiled that II-9 (the healthy mother) exhibited heterozygosity for the *IGSF8* c.C487T mutation. Moreover, III-6 and III-7 (the two-breast cancer-affected sisters) carried this mutation in a heterozygous pattern, as depicted in Figure 11. The electropherograms at position 487 displayed two distinctive peaks, one in blue (C) and another in red (T), implying a heterozygous sequence (C in one allele and T in the other). Furthermore, Figure 12-A illustrates the normal sequence, as depicted by the blue electropherogram (C nucleotide) of IGSF8 in position 487, which was observed in samples III-9, III-10, III-11, IV-8, IV-11, IV-12, IV-13, and IV-14. Conversely, Figure 12-B portrays the heterozygous sequence for the IGSF8 c.C487T mutation, as evidenced in samples III-6, III-7, III-8, III-12, IV-4, IV-5, IV-6, IV-7, IV-9, IV-10, and II-10.

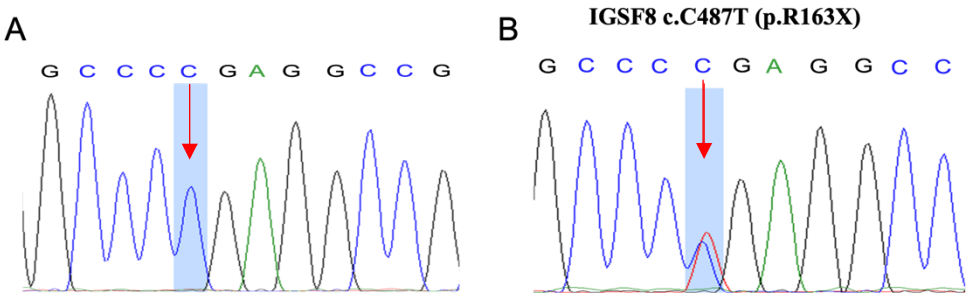


Figure 12: Electropherogram showing the DNA sequence of the *IGSF8* c.C487T stop mutation. (A) A wild-type state in an unaffected subject. (B) heterozygous state in affected subjects.

In conclusion, our comprehensive whole exome sequencing analysis unveiled a heterozygous stop-gain mutation within the *IGSF8* gene among the two breast cancer patients and their respective mothers. The subsequent segregation analysis indicated that this mutation was inherited in an autosomal dominant manner, as it was passed on to the daughters of the affected patients. Conversely, unaffected sisters exhibited a normal genotype for this mutation, which was mirrored in their daughters, lending credence to the autosomal dominant inheritance pattern.

Chapter 4: Discussion

Our investigation focused on the clinical presentation and genetic underpinnings of hereditary breast cancer within a cohort of three affected individuals from an extended Palestinian family in Hebron. Through rigorous bioinformatics analysis of the *BRCA* plus panel, we identified no known pathogenic variants associated with breast cancer within these individuals. However, an alternative approach involving the bioinformatics analysis of WES data revealed a significant association between breast cancer and a variant within the *IGSF8* gene.

The identified variant, *IGSF8* c.C487T, is located within exon 3 of the *IGSF8* gene. This variant results in the creation of a premature stop codon within the Ig-like C2-type2 domain of the IGSF8 protein. As this domain constitutes the protein's second domain, the mutation could potentially lead to the degradation of immature mRNA, resulting in the absence of functional protein or the production of a truncated nonfunctional protein. Notably, individuals afflicted with breast and pancreatic cancers within the family displayed this mutation in a heterozygous manner, a pattern that was consistently observed in their offspring. In contrast, unaffected family members and their descendants carried the normal gene variant.

IGSF8 functions as a negative regulator of the TGFβ signaling pathway (H. X. Wang et al., 2015). TGF beta serves as a tumor suppressor gene that curbs cell proliferation by inducing G1 phase cell cycle arrest. TGF signaling orchestrates a spectrum of biological processes, encompassing cell division, proliferation, apoptosis, plasticity, and migration. Dysregulation of this pathway can contribute to various disorders, including tissue fibrosis and cancer (Akhurst & Hata, 2012). Upon ligand binding, TGF interacts with its

serine and threonine kinase receptors, namely type 1 and type 2, situated on the cell membrane. This initiates the TGFβ signaling cascade. The binding event gives rise to a receptor heterocomplex formation, where the type 2 receptor phosphorylates specific serine and threonine residues in the TTSGSGSG motif of the type 1 receptor, consequently activating it. This activation prompts the recruitment and phosphorylation of Smad2/3, subsequently forming a heterocomplex with Smad4. Intriguingly, IGSF8 depletion in cells induces a 2 to 3-fold stimulation of Smad2, further amplifying TGFβ signaling (Wrighton et al., 2009). These Smad complexes are then translocated into the nucleus, where they collaborate with additional cofactors to regulate the transcription of target genes (Huang & Chen, 2012). TGFβ signaling holds the capacity to trigger epithelial-to-mesenchymal transition in both benign and malignant epithelial cells. This phenomenon is crucial for healthy embryonic development; however, exploiting it for cancer progression can potentially foster tumor invasion and metastasis (Barcellos-Hoff & Akhurst, 2009).

Conversely, Tetraspanins interact with transmembrane proteins and signaling molecules to govern signaling pathways. Two specific tetraspanins, CD9 and CD81, directly intersect with the TGFβ signaling pathway. When IGSF8 is present, it forms hetero-oligomeric complexes with CD9 and CD81. Consequently, CD9 and CD81 are unable to support the interaction between TGFβ receptors 1 and 2. In the absence of IGSF8, an abundance of CD9 homo-oligomers is facilitated by CD81, aiding the association of TGFβ receptors 1 and 2, thereby triggering TGF signaling (H. X. Wang & Hemler, 2015). In vitro experimentation using murine B16F10 and human SK-Mel-28 melanoma cell lines has convincingly demonstrated that stable *IGSF8* suppression markedly enhances cell migration and invasion. Similar outcomes were observed upon temporary IGSF8

depletion (H. X. Wang & Hemler, 2015). Conversely, *IGSF8* overexpression exerted a preventative effect on cell invasion. In murine models that received intravenous injections of melanoma cells with IGSF8 knockdown, an elevated number of metastatic lung colonies was noted after 6 weeks and 16 days (Javelaud et al., 2008; Lasfar & Cohen-Solal, 2010). While *IGSF8* knockdown significantly altered the inherent cell shape in line with increased invasion and metastasis, it is plausible that IGSF8 primarily restricts the metastasis of human and mouse melanoma cells by limiting tissue invasion rather than influencing colony growth (H. X. Wang & Hemler, 2015).

Additionally, IGSF8 can trigger tumor initiation through the regulation of EGFR signaling and cellular recycling. Chenying Fu et al. uncovered that miR-3934-5p downregulates the phosphorylation of EGFR. Interestingly, this miRNA exhibited significant downregulation in the exosomes of prostate cancer cells following *IGSF8* knockdown, consequently leading to the activation of EGFR signaling. The study demonstrated the upregulation of Tyr1068-phosphorylated EGFR along with increased AKT phosphorylation at Ser473 after IGSF8 silencing. Notably, the excessive activation of EGFR resulting from *IGSF8* knockdown was mitigated when Gefitinib (an EGFR inhibitor) was used. Furthermore, their findings solidified the notion that IGSF8 plays a pivotal role in the internalization of EGFR, as evidenced by increased EGFR clustering in *IGSF8*-deficient cells (Fu et al., 2021).

On one hand, this mutation in the *IGSF8* gene could explain the progression of cancer; however, it does not account for the transformation of normal cells into cancerous ones. Each significant mutation or set of changes could unveil the specific pathways that different cancer cells adopt to infiltrate and proliferate (Adjiri, 2017). The central inquiry pertains not to how cancer spreads, but rather how it originates. The alteration in driver

genes assumes paramount importance in the development of anti-cancer medications, particularly if the objective is to curtail tumor growth. Therefore, comprehending the biochemical mechanisms underlying IGSF8's role in cancer cells may pave the way for more effective diagnostic and therapeutic strategies (Sjöblom et al., 2013).

Furthermore, in conjunction with genetic alterations, cancers exhibit epigenetic modifications. This interaction between genetic changes and epigenetic alterations significantly influences the cancer phenotype. This phenomenon, now well-established, involves DNA methylation, histone readers and modifiers, chromatin remodelers, microRNAs, and other chromatin-related elements (Easwaran et al., 2014). Epigenetic changes can trigger gene mutations and vice versa. In the context of the two-hit or multiple-hit theory, Knudson posited that both copies of a tumor suppressor gene must be inactivated in malignant cell lines (Alfred G. Knudson, n.d.). It is currently recognized that three distinct categories of "hits" collaboratively silence tumor suppressor genes. These hits can manifest as direct changes to the coding sequence, loss of one or more gene copies, or epigenetic silencing. The notion that genetic and epigenetic anomalies synergistically contribute to cancer initiation and propagation is gaining prominence. Consequently, deeper investigation and comprehension of the intricate relationship between genetic and epigenetic alterations are imperative, particularly in regions like Palestine where targeted sequencing panels have yielded no known pathogenic variants among Palestinian cancer patients, hindering efficient cancer diagnosis and treatment.

Chapter 5: Conclusions

Our study shows that WES provides a more conclusive genetic diagnoses compared to targeted sequencing panels as a potent diagnostic and therapeutic approach for hereditary cancers in Palestine. While targeted sequencing panels, such as the *BRCA* panel, have successfully identified mutations associated with breast cancer in developed nations, the genetic landscape of breast cancer remains relatively uncharted in middle- and low-income countries like Palestine. With a prevalence of 7% for *BRCA1* or *BRCA2* mutations and 13% for pathogenic mutations in any breast cancer gene among patients with early-onset breast cancer or a family history of breast or ovarian cancer, the need for comprehensive genetic insights becomes increasingly apparent (Lolas Hamameh et al., 2017).

Considering the recent surge in breast cancer cases and the presence of variants with uncertain significance in breast cancer patients, understanding the patient's genotype holds paramount importance. This knowledge not only influences personalized therapy but also expands the spectrum of available follow-up options for their female relatives. Given the efficiency and accessibility of testing for all mutations within identified breast cancer genes, whole exome sequencing via next-generation sequencing emerges as the optimal approach to unveil elusive pathogenic variants in non-*BRCA* breast cancer patients and high-risk individuals. In our investigation, we unveiled the novel stop-gain mutation *IGSF8* c.C487T through WES, inheriting across the subjects in this study who tested negative for established *BRCA* mutations.

Moreover, we established a correlation between the depletion of *IGSF8* and its impact on the TGF signaling pathway, as well as its interplay with the CD9 and CD81 tetraspanins.

This relationship offers insights into the potential mechanisms underlying cancer initiation and progression. Furthermore, we highlighted the intriguing prospect of genetic and epigenetic alterations that might underlie the functional loss of IGSF8 in suppressing tumor initiation and advancement.

Recommendations

We recommend using WES for testing the genetic make-up of different types of cancers in the Palestinian population, especially breast cancer, because worldwide targeted panels, especially the BRCA panel, were found to be unreliable for the Palestinian population. Moreover, we recommend performing functional analysis for the IGSF8 gene in breast cancer and different cancer types.

Bibliography

Adjiri, A. (2017). DNA Mutations May Not Be the Cause of Cancer. *Oncology and Therapy*, *5*(1), 85–101. https://doi.org/10.1007/s40487-017-0047-1

Akhurst, R. J., & Hata, A. (2012). Targeting the TGFβ signalling pathway in disease. In *Nature Reviews Drug Discovery* (Vol. 11, Issue 10, pp. 790–811). https://doi.org/10.1038/nrd3810

Alfred G. Knudson. (n.d.). *Two genetic hits (more or less) to cancer*.

Align GVGD. (n.d.). (n.d.). *Align GVGD. (n.d.)*. Http://Agvgd.Hci.Utah.Edu/about.Php. Retrieved 25 August 2023, from http://agvgd.hci.utah.edu/about.php

Barcellos-Hoff, M. H., & Akhurst, R. J. (2009). Transforming growth factor-β in breast cancer: Too much, too late. In *Breast Cancer Research* (Vol. 11, Issue 1). https://doi.org/10.1186/bcr2224

COBALT:Multiple Alignment Tool. (n.d.). *COBALT:Multiple Alignment Tool*. Https://Www.Ncbi.Nlm.Nih.Gov/Tools/Cobalt. Retrieved 25 August 2023, from https://www.ncbi.nlm.nih.gov/tools/cobalt

Easwaran, H., Tsai, H. C., & Baylin, S. B. (2014). Cancer Epigenetics: Tumor Heterogeneity, Plasticity of Stem-like States, and Drug Resistance. In *Molecular Cell* (Vol. 54, Issue 5, pp. 716–727). Cell Press. https://doi.org/10.1016/j.molcel.2014.05.015

Flannery, A. R., Czibener, C., & Andrews, N. W. (2010). Palmitoylation-dependent association with CD63 targets the Ca2+ sensor synaptotagmin VII to lysosomes. *Journal of Cell Biology*, *191*(3), 599–613. https://doi.org/10.1083/jcb.201003021

Fu, C., Zhang, Q., Wang, A., Yang, S., Jiang, Y., Bai, L., & Wei, Q. (2021). EWI-2 controls nucleocytoplasmic shuttling of EGFR signaling molecules and miRNA sorting in exosomes to inhibit prostate cancer cell metastasis. *Molecular Oncology*, *15*(5), 1543–1565. https://doi.org/10.1002/1878-0261.12930

Hemler, M. E. (2005). Tetraspanin functions and associated microdomains. In *Nature Reviews Molecular Cell Biology* (Vol. 6, Issue 10, pp. 801–811). https://doi.org/10.1038/nrm1736

Huang, F., & Chen, Y. G. (2012). Regulation of TGF-β receptor activity. In *Cell and Bioscience* (Vol. 2, Issue 1). https://doi.org/10.1186/2045-3701-2-9

Javelaud, D., Alexaki, V. I., & Mauviel, A. (2008). Transforming growth factor-β in cutaneous melanoma. In *Pigment Cell and Melanoma Research* (Vol. 21, Issue 2, pp. 123–132). https://doi.org/10.1111/j.1755-148X.2008.00450.x

Kolesnikova, T. V., Kazarov, A. R., Lemieux, M. E., Lafleur, M. A., Kesari, S., Kung, A. L., & Hemler, M. E. (2009). Glioblastoma inhibition by cell surface immunoglobulin protein EWI-2, in vitro and in vivo. *Neoplasia*, *11*(1), 77–86. https://doi.org/10.1593/neo.81180

Lasfar, A., & Cohen-Solal, K. A. (2010). Resistance to transforming growth factor β-mediated tumor suppression in melanoma: Are multiple mechanisms in place? In *Carcinogenesis* (Vol. 31, Issue 10, pp. 1710–1717). Oxford University Press. https://doi.org/10.1093/carcin/bgq155

LIST-S2. (n.d.). (n.d.). LIST-S2. (n.d.). *LIST-S2. (n.d.). Https://List-S2.Msl.Ubc.ca/;Jsessionid=74DD638E0B54ACF13B588FC959FFE2BC?Session=74DD638E0B54ACF13B588FC959FFE2BC*. Retrieved 25 August 2023, from https://list-

s2.msl.ubc.ca/;jsessionid=74DD638E0B54ACF13B588FC959FFE2BC?session=74DD638E0B54ACF13B588FC959FFE2BC

Lolas Hamameh, S., Renbaum, P., Kamal, L., Dweik, D., Salahat, M., Jaraysa, T., Abu Rayyan, A., Casadei, S., Mandell, J. B., Gulsuner, S., Lee, M. K., Walsh, T., King, M. C., Levy-Lahad, E., & Kanaan, M. (2017). Genomic analysis of inherited breast cancer among Palestinian women: Genetic heterogeneity and a founder mutation in TP53. *International Journal of Cancer*, *141*(4), 750–756. https://doi.org/10.1002/ijc.30771

Longo, N., Yáñ ez-Mó, M., Mittelbrunn, M., de la Rosa, G., Muñ oz, M.-L., Sá nchez-Madrid, F., & Sá nchez-Mateos, P. (2001). *Regulatory role of tetraspanin CD9 in tumor-endothelial cell interaction during transendothelial invasion of melanoma cells*. www.bloodjournal.org

Montpellier, C., Tews, B. A., Poitrimole, J., Rocha-Perugini, V., D'Arienzo, V., Potel, J., Zhang, X. A., Rubinstein, E., Dubuisson, J., & Cocquerel, L. (2011). Interacting regions of CD81 and two of its partners, EWI-2 and EWI-2wint, and their effect on hepatitis C virus infection. *Journal of Biological Chemistry*, *286*(16), 13954–13965. https://doi.org/10.1074/jbc.M111.220103

MutationTaster. (n.d.). (n.d.). *MutationTaster. (n.d.)*. Https://Www.Mutationtaster.Org/. Retrieved 25 August 2023, from https://www.mutationtaster.org/

Oren, R., Takahashi, S., Doss, C., Levy, R., & Levy, S. (1990). TAPA-1, the Target of an Antiproliferative Antibody, Defines a New Family of Transmembrane Proteins. In *MOLECULAR AND CELLULAR BIOLOGY*.

PolyPhen-2. (n.d.). *PolyPhen-2: prediction of functional effects of human nsSNPs.* PolyPhen-2: Prediction of Functional Effects of Human NsSNPs. (n.d.).

Http://Genetics.Bwh.Harvard.Edu/Pph2/. Retrieved 25 August 2023, from http://genetics.bwh.harvard.edu/pph2/

PROVEAN. (n.d.). J. Craig Venter Institute. (n.d.). *PROVEAN. (n.d.). J. Craig Venter Institute.* . Https://Www.Jcvi.Org/Research/Provean. Retrieved 25 August 2023, from https://www.jcvi.org/research/provean

Shihab, H. A. (n. d.). fathmm-H. (n.d.). *FATHMM* . Http://Fathmm.Biocompute.Org.Uk/. Retrieved 25 August 2023, from http://fathmm.biocompute.org.uk/

Sjöblom, T., Jones, S., Wood, L. D., Parsons, D. W., Lin, J., Barber, T. D., Mandelker, D., Leary, R. J., Ptak, J., Silliman, N., Szabo, S., Buckhaults, P., Farrell, C., Meeh, P., Markowitz, S. D., Willis, J., Dawson, D., Willson, J. K. V, Gazdar, A. F., … Velculescu, E. (2013). *The Consensus Coding Sequences of Human Breast and Colorectal Cancers* (Vol. 6). www.sciencemag.org

StackPath. (n.d.). (n.d.). *SIFT*. Https://Sift.Bii.a-Star.Edu.Sg/. Retrieved 25 August 2023, from https://sift.bii.a-star.edu.sg/

Termini, C. M., & Gillette, J. M. (2017). Tetraspanins function as regulators of cellular signaling. In *Frontiers in Cell and Developmental Biology* (Vol. 5, Issue APR). Frontiers Media S.A. https://doi.org/10.3389/fcell.2017.00034

Vences-Catalán, F., & Levy, S. (2018). Immune targeting of tetraspanins involved in cell invasion and metastasis. In *Frontiers in Immunology* (Vol. 9, Issue JUN). Frontiers Media S.A. https://doi.org/10.3389/fimmu.2018.01277

Wang, H., & Hemler, M. E. (2015). *in melanoma*. *December*, 4–6.

Wang, H. X., & Hemler, M. E. (2015). Novel impact of EWI-2, CD9, and CD81 on TGF-β signaling in melanoma. In *Molecular and Cellular Oncology* (Vol. 2, Issue 4). Taylor and Francis Ltd. https://doi.org/10.1080/23723556.2015.1030536

Wang, H. X., Sharma, C., Knoblich, K., Granter, S. R., & Hemler, M. E. (2015). EWI-2 negatively regulates TGF-β signaling leading to altered melanoma growth and metastasis. *Cell Research*, *25*(3), 370–385. https://doi.org/10.1038/cr.2015.17

WHO, 2023, & World health statistics. (n.d.). *World health statistics 2023: monitoring health for the SDGs, Sustainable Development Goals. Geneva: World Health Organization; 2023. Licence: CC BY-NC-SA 3.0 IGO.* https://www.who.int/news- room/fact-sheets/detail/cancer

Wrighton, K. H., Lin, X., & Feng, X. H. (2009). Phospho-control of TGF-β superfamily signaling. In *Cell Research* (Vol. 19, Issue 1, pp. 8–20). https://doi.org/10.1038/cr.2008.327

Zöller, M. (2009). Tetraspanins: Push and pull in suppressing and promoting metastasis. In *Nature Reviews Cancer* (Vol. 9, Issue 1, pp. 40–55). https://doi.org/10.1038/nrc2543

APPENDIX A: CONSENT FORMS

الجامعة العربية الأمريكية
ARAB AMERICAN UNIVERSITY

موافقة للمشاركة في دراسة بحثية للمخصصات البيولوجية والطفرات الجينية التي لها دور في حدوث السرطان

الهدف من هذا المشروع هو تجميع عينات الدم من عائلة في فلسطين لديها تاريخ سابق في الإصابة بسرطان الثدي لتحقيق الأهداف التالية:

1- دراسة طفرة جينية (تم ملاحظة وجودها عند بعض افراد العائلة) وربط هذه الطفرة بسرطان الثدي والمبيض.
2- فحص الأجيال اللاحقة لوجود هذه الطفرة وهل تم تناقلها من جيل الى جيل بالإضافة الى وضع حجر الأساس للبدء بخطة وقاية للأجيال الجديدة من سرطان الثدي والمبيض.

المطلوب منك

1- سيقوم موظف من وزارة الصحة الفلسطينية بأخذ عينات دم من افراد العائلة.
2- نأمل موافقتك ان نراجع ملفك الطبي في حين بدء البحث ولاحقا بهدف تجميع المعلومات الضرورية للبحث مع العلم انا سنحافظ على السرية **التامة**.
3- يحق للفرد الرفض في المشاركة بهذا البحث و لن يترتب عليه أية عواقب.

المخاطر المترتبة على المشاركة بهذه الدراسة

المخاطر المترتبة على سحب عينة دم من ذراعك تتضمن انزعاج لحظي فإن المخاطر الزائدة المترتبة على مشاركتك في البحث تكون ضئيلة.

الخصوصية والسرية

ستحفظ المعلومات البحثية والطبية الخاصة بهذه الدراسة تحديداً في ملف خاص بالبحث.

للمحافظة على الخصوصية، فإن اسمك أو أية معلومات قد تقوم بالتعريف بك، ستزال من عينة الدم ومن كل المعلومات عن وضعك الطبي وسيرتك المرضية التي ترافق عينتك. كل المعلومات الخاصة بالتعريف بك ستكون عن طريق إعطاء رمز خاص. هذا الرمز سيكون معروفا فقط لدى الباحثين واضاء الطاقم الطبي المشارك في البحث.

فوائد للمشاركة في هذه الدراسة؟

1- اكتشاف الطفرة الجينية الجديدة التي تجري في العائلة ويتم تناقلها عبر الأجيال وربط هذه الطفرة بحدوث المرض سيساعد لاحقا في العلاج للمرض وإيجاد خطة وقاية للأفراد المعرضين للإصابة بالمرض بالإضافة الى إمكانية تقليل احتمالات الإصابة بالمرض عند الأجيال اللاحقة.

التكاليف والتعويضات

لن يكون هناك أية تكاليف إضافية عليك إثر مشاركتك بهذه الدراسة. كل التكاليف المتعلقة بأخذ عينة الدم وعمل التحليلات الوراثية اللازمة سيتم تغطيتها من الموازنة المتوفرة للبحث من الجامعة. لن نقوم بالدفع لك لمشاركتك بهذه الدراسة.

الحق في رفض المشاركة بهذه الدراسة أو الانسحاب منها

المشاركة بهذه الدراسة هي مشاركة **طوعية بحتة**. يمكنك رفض المشاركة في هذه الدراسة أو في حال وافقت، يمكنك الانسحاب منها إن رغبت في أي وقت. قرارك بعدم المشاركة أو الانسحاب من المشاركة بهذه الدراسة لن يترتب عليه أية عقوبة أو خسارة لأية امتيازات تتمتع بها، ولن يؤثر ذلك على مقدرتك في الحصول على الرعاية الصحية.

بمن أتصل في حال كانت لدي أسئلة أو واجهتني مشاكل؟

للسؤال عن الدراسة أو عن إصابة ناتجة عن المشاركة بالدراسة، أو إذا كانت لديك مشاكل، مخاوف، أسئلة أو اقتراحات حول البحث، الرجاء الاتصال ب **دانيا اسامه نجار على رقم الهاتف 0598253120**

نص الموافقة

تم شرح لي الهدف من هذه الدراسة، الخطوات التي سيتم اتباعها، المخاطر والفوائد المترتبة على المشاركة بها. لقد تم السماح لي بسؤال أسئلة، وتمت الإجابة عن تساؤلاتي لحد يرضيني. لقد تم إخباري بمن أتصل إذا كانت لدي تساؤلات، أو لمناقشة مشاكل، أو مخاوف، أو اقتراحات متعلقة بالبحث، أو للحصول على معلومات أو إعطاء أية إضافات حول البحث. لقد قمت بقراءة وثيقة الموافقة هذه وأوافق على المشاركة بهذه الدراسة، مع العمل أنه بإمكاني الانسحاب أنّى شئت.

☐ أوافق على أن يتم أخذ عينة دم مني لفحص المادة الوراثية.

الاسم الثلاثي للمشارك/ة_____

توقيع المشارك/ة_____ التاريخ _____

اسم الباحث/ة: **دانيا أسامة نجار** توقيع الباحثين _____ التاريخ _____

اسم الدكتور/ة المشرف/ة:

الدكتور زيدون صلاح / الدكتورة نوار قطب

الملخص

السرطان هو أحد الأسباب الرئيسية للوفيات في العالم وفلسطين. في فلسطين ، سرطان الثدي هو النوع الأكثر انتشارًا حيث تم الإبلاغ عن 934 حالة جديدة من سرطان الثدي في عام 2022, مع معدل الوقوع الإجمالي هو 18.5 حالة لكل 100000 شخص ومعدل الإصابة للإناث هو 37.4 حالة لكل 100000شخص، جاء سرطان الرئة في المرتبة الثانية ، ثم سرطان القولون والمستقيم.

أصبحت تقنية تسلسل الجيل القادم NGS أداة مهمة لتشخيص السرطان وتعتبر وسيلة فعالة للكشف المبكر و التدخل الطبي لتخطيط العلاج الفعال الذي قد يزيد من فرصة البقاء على قيد الحياة.

الغرض من هذه الدراسة هو استخدام تقنية تسلسل الجيل القادم NGS لتحديد الأسباب الجزيئية لـ سرطان الثدي في عائلة فلسطينية مصابة بسرطان الثدي الوراثي حيث أجرينا تسلسل إكسوم كامل WES للحمض النووي المستخرج من دم شقيقتين تم تشخيص إصابتهما بسرطان الثدي والتحقق من صحة النتائج عن طريق تقنية sanger sequencing.

استخدمنا تحليل الفصل لتأكيد وجود المتغير المحدد في بنات أفراد الأسرة المستهدفين حيث أدى تحليلنا إلى تحديد طفرة توقف الكسب الجديدة*IGSF8*c.C487T، والتي تقع في إكسون 3 وقد تؤدي إلى بروتين غير وظيفي.

تم التحقق من صحة ذلك لاحقًا عن طريق تقنيةsanger sequencing وعلاوة على ذلك ، أكد تحليل الفصل وجود هذا المتغير في بنات أفراد الأسرة المستهدفين.

تظهر دراستنا أن WES يوفر تشخيصات وراثية أكثر حسماً مقارنة بتسلسل اللوحة. ويؤكد أيضًا أن الدراسات الجزيئية وتطبيقات NGS ضرورية للكشف عن الطفرات المهمة ذات الصلة بسرطان الثدي، والتي قد تؤدي في المقابل إلى التدخلات الطبية الصحيحة.