Democratic and Popular Republic of Algeria

Ministry of Higher Education and Scientific Research

University of Dr. Moulay Tahar University – Saida–

Faculty of Natural and Life Sciences

Department of Biology

Final Thesis

In fulfillment of the requirements for the

Master's Degree in Applicated Biochemistry

Subject

GENETIC ASSOCIATION STUDY AND AN IN SILICO ANALYSIS OF BREAST CANCER

Presented by:

Miss HALIMAOUI Randa

Miss MEKKI Aicha

In front of the jury composed of

President: Mr Ammam Abdelkader	Professor	SAIDA University
Examinator: Mrs HAssani Maya	AP	SAIDA University
Supervisor: Mrs Berber Naima	AP	SAIDA University

Academic Year: 2023-2024

Acknowledgments

I am profoundly grateful to all those who have supported and guided me throughout the journey of completing this thesis.

First and foremost, I extend my sincere appreciation to **Mrs Berber Naima** for their invaluable guidance, encouragement, and insightful feedback throughout the research and writing process. Their expertise and dedication have been instrumental in shaping this work.

I am grateful to the Professor, **Mr Ammam Abdekader** and Examinator **Mrs HAssani Maya**, for their thorough review and constructive feedback on this thesis. Their insights and suggestions have significantly contributed to improving the quality of this work.

I would like to express my sincere appreciation to the faculty members of the Department of Biology at Dr. Moulay Tahar University – Saida–for their outstanding guidance and training during my years of specialization. Their commitment has enhanced my academic journey.

Dedicate:

"First of all, i dedicate this work to my parents, who have been the cause of all that I have accomplished and my sister and my brother also my grandmother and Grandfather and also my aunts and uncles my dear friends Randa, wafaa, Leila and all my classmate, let's not forget my advisor Berber Naima and all the teachers of the department of Life and nature sciences."

Aicha

“To my dearest Mom,

Your love has been the foundation upon which I've built my dreams, and your untiring support has been my guiding star. As I graduate, I carry with me the lessons you've taught me, the strength you've instilled in me, and the love that fuels my every step. This achievement is as much yours as it is mine.

To My Beloved Dad,

As I celebrate this achievement, I am reminded of the countless sacrifices you've made to ensure my happiness and well-being. Thank you, Dad, for believing in me, for inspiring me, and for always being there when I needed you most.

To my dear sister and brother,

Your support and endless love have been the greatest gifts in my life. Thank you for always being there”

Randa

Abstract

Breast cancer is a critical public health issue in Algeria, with increasing incidence and substantial challenges in diagnosis, treatment, and management. Breast cancer incidence in Algeria has increased significantly, with over 2.1 million women diagnosed in 2018.

BRCA1, a pivotal gene in DNA repair and tumor suppression, is critically implicated in the predisposition to breast and ovarian cancers. *In silico* analysis has emerged as a powerful approach to understand the functional implications of BRCA1 mutations, offering insights into their structural and mechanistic effects.

This study leverages computational tools (I-Mutant 2.0), (SIFT), (Align-GVGD), (PolyPhen-2), (Project HOPE) software) to analyze the impact of various BRCA1 mutations on protein structure and function. By employing molecular modeling, sequence alignment, and predictive algorithms, we identify mutations that significantly disrupt BRCA1's role in maintaining genomic stability.

The analysis reveals key regions within the BRCA1 protein that are particularly susceptible to substituted alterations in two different *in silico* analysis. The (c. 135G>C) (p. Lys45Asn) mutation and (c. 122A>G) (p. His41Arg) mutation provide a framework for prioritizing mutations for experimental validation and clinical assessment. Additionally, we explore the potential of *in silico* findings to inform the development of targeted therapies.

Key words: breast cancer; BRCA1; mutations; *in silico* analysis.

ملخص

سرطان الثدي مشكلة صحية عامة حرجة في الجزائر، حيث تشهد زيادة في الاصابات وتحديات كبيرة في التشخيص والعلاج والإدارة. زادت حالات الاصابة بسرطان الثدي في الجزائر بشكل كبير، مع تشخيص أكثر من 2.1 مليون امرأة في عام 2018. الجينBRCA1 ، الذي يعد حجر الزاوية في إصلاح الحمض النووي وقمع الأورام، يشتبك بشكل حرج في الإمكانية المسبقة للإصابة بسرطان الثدي وسرطان المبيض. ظهر التحليل الحاسوبي باعتباره أداة قوية لفهم الآثار الوظيفية للطفرات فيBRCA1 ، مما يوفر رؤى في تأثيراتها الهيكلية والآلية. تستفيد هذه الدراسة من الأدوات الحاسوبية(I-Mutant 2.0) ،(SIFT)، (Align-GVGD)،(PolyPhen-2) ، (Project HOPE) لتحليل تأثير مختلف الطفرات في BRCA1على هيكل البروتين ووظيفته. من خلال استخدام النمذجة الجزيئية وموازنة التسلسل والخوارزميات التنبؤية، نحدد الطفرات التي تؤثر بشكل كبير على دور BRCA1 في الحفاظ على الاستقرار الجينومي. يكشف التحليل عن مناطق رئيسية داخل بروتين BRCA1تعرض بشكل خاص للتغييرات المستبدلة في اثنين من التحليلات الحاسوبية المختلفة. توفر الطفرة (p. Lys45Asn) (c. 135G>C) والطفرة .p) (c. 122A>G) (His41Arg إطارًا لتحديد الأولويات للطفرات من أجل التحقق التجريبي والتقييم السريري. بالإضافة إلى ذلك، نستكشف إمكانية استفادة النتائج الحاسوبية لإرشاد تطوير العلاجات المستهدفة.

Résumé

Le cancer du sein est un problème de santé publique critique en Algérie, avec une incidence croissante et d'importantes difficultés en matière de diagnostic, de traitement et de prise en charge. L'incidence du cancer du sein en Algérie a augmenté de manière significative, avec plus de 2,1 millions de femmes diagnostiquées en 2018. BRCA1, un gène essentiel dans la réparation de l'ADN et la suppression des tumeurs, est impliqué de manière critique dans la prédisposition au cancer du sein et de l'ovaire. L'analyse *in silico* est devenue une approche puissante pour comprendre les implications fonctionnelles des mutations de BRCA1, offrant des éclaircissements sur leurs effets structuraux et mécanistiques. Cette étude exploite des outils informatiques (logiciels I-Mutant 2.0, SIFT, Align-GVGD, PolyPhen-2, Project HOPE) pour analyser l'impact de diverses mutations de BRCA1 sur la structure et la fonction des protéines. En utilisant la modélisation moléculaire, l'alignement des séquences et des algorithmes prédictifs, nous identifions des mutations qui perturbent significativement le rôle de BRCA1 dans le maintien de la stabilité du génome. L'analyse révèle des régions clés au sein de la protéine BRCA1 qui sont particulièrement susceptibles aux altérations substitutives dans deux analyses in *silico* différentes. La mutation (c. 135G>C) (p. Lys45Asn) et la mutation (c. 122A>G) (p. His41Arg) fournissent un cadre pour prioriser les mutations en vue d'une validation expérimentale et d'une évaluation clinique. De plus, nous explorons le potentiel des découvertes *in silico* pour informer le développement de thérapies ciblées.

LES MOTS Clés : cancer de sein, BRCA1 ; mutation ; analyse *in silico*

Table of Contents

List of Tables

List of Figures

Abreviations list

Ki-67 The percentage of positively stained cells among the total number of cancer cells assessed

ERBB2 Erythroblastic oncogene B .The human protein is also frequently referred to as HER2

HER2 Human Epidermal Growth Factor Receptor-2

HR Hormone receptor

ER Estrogen Receptor

PR Progesterone Receptor

IDC Invasive ductal carcinoma

DCIS Ductal carcinoma in situ

TNBC Triple-negative breast cancer

IBC Inflammatory breast cancer

FNA Fine needle aspiration

MRI Magnetic resonance imaging

CT Computed tomography

DDR DNA damage response

DNA Double-strand breaks

Hr Homologous recombination

NHEJ Non-homologous end joining

HRR Homologous recombination repair

TNBC Triple-Negative Breast Cancer

AA Amino acid

DNA Deoxyribonucleic acid

RNA Ribonucleic acid

RHA RNA helicase A

POL2 RNA polymerase 2

BRCT BRCA1 carboxyl terminus

BRCA1 BRCH1 BRCA1 associated c-terminal helicase

PALB2 Partner and localizer of BRCA2

BASC BRCA1-associated genome surveillance complex

PARP Poly (ADP-ribose) polymerase

NGS Next-generation sequencing

PDX Patient-derived xenograft

SETX sSnataxin

BC Breast cancer

RING Short interesting new gene

HOPE Have (y) Our Protein Explained

Align-GVGD Grantham Variation Grantham Deviation

SIFT Sorting Intolerant from Tolerant

Introduction

1 Introduction

Cancer refers to one of a large number of diseases characterized by the development of abnormal cells that divide uncontrollably and have the ability to infiltrate and destroy normal body tissue. Cancer often has the ability to spread throughout your body. Cancer is caused by changes (mutations) to the DNA within cells. The DNA inside a cell is packaged into a large number of individual genes, each of which contains a set of instructions telling the cell what functions to perform, as well as how to grow and divide. Errors in the instructions can cause the cell to stop its normal function and may allow a cell to become cancerous (1).

Breast cancer is the result of a disorder of certain cells that multiply and most often form a mass called a tumor. There are different types that do not evolve in the same way. Some are “aggressive” and evolve very quickly, others more slowly. Cancer cells can remain in the breast. They can also spread to other organs, which is an even more threatening situation. Then we talk about metastases. In most cases, the development of breast cancer takes several months, or even years (2).

BRCA1 and BRCA2 genes are tumor-suppressing genes the changes that genes undergo are called mutations. BRCA mutations are transmitted by the father or mother, or both parents, which increase the risk of cancer (3).

Several research teams were interested in studying the BRCA1 gene in several populations; the *in silico* study focused on two mutations identified in Algerian patients with breast cancer

In order to evaluate the impact of mutations of the BRCA1 gene in two deferent exons (c. 135G>C) (p. Lys45Asn) and (c. 122A>G) (p. His41Arg), at the level of protein function and structure. We have carried out a complementary study between two fields that are biology and bioinformatics by the use of multiple predation tools and molecular modulation (I-Mutant 2.0, (SIFT), (Align-GVGD), (PolyPhen-2), (Project HOPE).

2 Definition and epidemiology

2.1 Definition

Breast cancer is a disease in which abnormal breast cells grow out of control and form tumours. If left unchecked, the tumours can spread throughout the body and become fatal (4).

2.2 Epidemiology

2.2.1 Patients disposition

Breast cancer (BC) is a major health issue threatening women's life. No reliable epidemiological data on BC diagnosed by oncologists/senologists are available in Algeria.

Overall, 1,437 patients were analysed: median age was 48 [41; 57] years and 337 (23.5%) women had a family history of BC. BC incidence was 22.3 (95% CI: 21.5; 23.2) cases per 100,000 inhabitants over 8 months. Delayed diagnosis was reported in 400 (29.2%) patients. First line of treatments were mainly chemotherapy and surgery. Twenty-eight serious adverse events were reported including 10 (37.0%) events, which led to death. Mortality rate reached 3.2% at 12 months (5).

2.2.2 Demographics, clinical and disease characteristics

Table 1: Tumor and molecular characteristics *(5)*

	<40 years(n=293)	≥40 years(n=1144)	Total(n=1437)
Histology of breast cancer at diagnosis			
Missing	14	52	66
Epithelial non-infiltrating tumours	0 (0.0%)	2 (0.2%)	2 (0.1%)
Ductal carcinoma in situ (DCIS)	6 (2.1%)	22 (2.0%)	28 (2.0%)
Epithelial infiltrating tumours	7 (2.5%)	21 (1.9%)	28 (2.0%)
Infiltrating ductal carcinoma (without any other indication)	223 (79.9%)	839 (76.8%)	1062 (77.5%)
Invasive ductal carcinoma with a predominant intraductal component	11 (3.9%)	36 (3.3%)	47 (3.4%)
Invasive lobular carcinoma	17 (6.1%)	133 (12.2%)	150 (10.9%)
Mucinous carcinoma (colloid carcinoma)	4 (1.4%)	15 (1.4%)	19 (1.4%)
Medullary carcinoma	3 (1.1%)	2 (0.2%)	5 (0.4%)
Papillary carcinoma	2 (0.7%)	6 (0.5%)	8 (0.6%)
Tubular carcinoma	0 (0. 0%)	2 (0.2%)	2 (0.1%)
Apocrine carcinoma	3 (1.1%)	2 (0.2%)	5 (0.4%)
Metaplastic carcinoma of epidermoid type	1 (0.4%)	5 (0.5%)	6 (0.4%)
Spindle cell metaplastic carcinoma	1 (0.4%)	3 (0.3%)	4 (0.3%)
Mixed metaplastic cancer	1 (0.4%)	2 (0.2%)	3 (0.2%)
Paget's disease of the nipple	0 (0.0%)	2 (0.2%)	2 (0.1%)
Histological type determined by			
Missing	14	52	66
Biopsy	249 (89.2%)	969 (88.7%)	1218 (88.8%)
Mastectomy	29 (10.4%)	113 (10.3%)	142 (10.4%)
Not determined	1 (0.4%)	10 (0.9%)	11 (0.8%)
Stage at diagnosis			
Missing	15	56	71
0	1 (0.4%)	19 (1.7%)	20 (1.5%)
I	20 (7.2%)	118 (10.8%)	138 (10.1%)
IIA	59 (21.2%)	200 (18.4%)	259 (19.0%)

IIB	56 (20.1%)	226 (20.8%)	282 (20.6%)
IIIA	41 (14.7%)	132 (12.1%)	173 (12.7%)
IIIB	41 (14.7%)	142 (13.0%)	183 (13.4%)
IIIC	11 (4.0%)	26 (2.4%)	37 (2.7%)
IV	29 (10.4%)	162 (14.99%)	191 (14.0%)
Not determined	20 (7.2%)	63 (5.8%)	83 (6.1%)
Histo-prognostic grade of the tumour at diagnosis (Scarff-Bloom-Richardson)			
Missing	15	55	70
I	12 (4.3%)	75 (6.9%)	87 (6.4%)
II	180 (64.7%)	727 (66.8%)	907 (66.3%)
III	66 (23.7%)	215 (19.7%)	281 (20.6%)
Not determined	20 (7.19%)	72 (6.61%)	92 (6.73%)
Positive hormone receptors			
Missing	45	194	239
No	67 (27.0%)	202 (21.3%)	269 (22.4%)
Yes	181 (73.0%)	748 (78.7%)	929 (77.5%)
Positive human epidermal growth factor receptor 2			
Missing	49	214	263
No	155 (63.52%)	627 (67.42%)	782 (66.61%)
Yes	89 (36.48%)	303 (32.58%)	392 (33.39%)
BRCA 1 / 2 mutation			
Missing	241	978	1219
No	51 (98.0%)	163 (98.2%)	214 (98.1%)
Yes	1 (1.9%)	3 (1.8%)	4 (1.8%)
Ki-67 antigen			
Missing	61	273	334
No	17 (7.3%)	78 (9.0%)	95 (8.6%)
Yes	215 (92. 7%)	793 (91.0%)	1008 (91.4%)

Breast cancer incidence in Algeria has increased significantly, with over 2.1 million women diagnosed in 2018 and 627,000 deaths. Algerian women are more affected than other types of cancer. Factors like histological grade, tumor type, oestrogen receptor, and HER2/neu affect cancer prognostics and therapy response. Classical immunohistochemistry plays a critical role in molecular subtyping. The main challenge is identifying the best therapy option for each patient. This study aims to explore the epidemiological profile of invasive breast cancer and prognostic factors expression among Algerian women to improve awareness and healthcare management (6).

This retrospective study analyzed the histologic reports of 1037 mammary biopsies from January 2011 to December 2019, focusing on women with breast cancer. Pathologists performed histology examinations and graded cases according to the World Health Organization classification and Nottingham modification of Scarff Bloom Richardson (SBR) grading system. Immunohistochemical staining was performed on confirmed cases to evaluate prognostic markers, including oestrogen receptors, progesterone receptor, human epidermal growth factor type 2 receptor, and Ki-67. The study found that 455 (44%) of cases were histologically confirmed as breast cancer, while 568 (56%) were benign diseases. The mean age of the patients was 50.29 ± 12.24 years old, with a peak incidence between 50 and 53 years old. Infiltrating ductal carcinoma was found in 85% of patients, and grade II was the most frequent grade in 355 (78%). A statistically significant difference was found between the distributions of the three modalities of SBR grade in different age groups. As shown in table 2.

Table 2: Distribution of invasive breast cases according to SBR grade and hormone receptor expression in the three age groups *(6)*.

Age group	N [%]	Grade I	Grade II	Grade III	p	ER+/PR+	ER+/PR-	ER-/PR+	ER-/PR-	p
		N (%)	N (%)	N (%)		N (%)	N (%)	N (%)	N (%)	
< 40 years	72 (17)	1 (4)	34 (14)	24 (31)		32 (14)	3 (7)	3 (20)	24 (35)	
≤ 40 age < 60 years	272 (62)	19 (70)	147 (63)	47 (60)	0.001	150 (65)	27 (66)	8 (53)	38 (55)	0.001
≥ 60 years	91 (21)	7 (26)	54 (23)	7 (9)		50 (21)	11 (27)	4 (27)	7 (10)	
Total	435 (100)	27 (100)	235 (100)	78 (100)		232 (100)	41 (100)	15 (100)	69 (100)	

The study evaluated hormone receptors in invasive breast cancer cases, revealing that ER, PR, and HER2 are co-expressed in 58% of samples. Oestrogen receptors are expressed in 19.9% of samples, while progesterone receptors are present in 62.2%. Young women tend to lose hormone receptors, while middle-aged women express all four phenotypes. Hormone receptor expression is correlated with patient age at diagnosis. Grade III tumors have a higher frequency of ER than grade II or grade I. as shown in table 3.

Table 3: Distribution of invasive breast cancer cases according to oestrogen and progesterone receptor status and SBR grade *(6)*.

Phenotype	Grade I	Grade II	Grade III	p
	N (%)	N (%)	N (%)	
ER+	25 (93)	204 (85)	34 (41)	0.000
ER-	2 (7)	37 (15)	48 (59)	
Total	27 (100)	241 (100)	82 (100)	
PR+	25 (93)	178 (74)	38 (46)	0.000
PR-	2 (7)	63 (26)	45 (54)	
Total	27 (100)	241 (100)	83 (100)	

HER2 overexpression was found in 22% of cases, while 61% of tumours lack HER2 expression. In 17% cases, HER2 expression is equivoque. The gene amplification test using FISH is recommended for HER2-positive tumors. Ki-67 proliferation index expression is positive in 90% of cases, with high proliferation activity mostly positive for oestrogen and progesterone receptors. Ki-67 expression has no significant correlation with age or oestrogen receptor expression (6).

Table 4: Distribution of invasive breast cancer cases according to Ki-67 expression status in correlation with histological grade, progesterone and human epidermal growth factor 2 receptor state (6).

Characteristics		Ki-67 -	Ki-67 +	p-value
		N (%)	N (%)	
Grade	I	5 (17)	18 (6)	0.030
	II	22 (73)	205 (69)	
	III	3 (10)	74 (25)	
	Total	30 (100)	297 (100)	
PR	Negative	4 (12.5)	96 (31)	0.029
	Positive	28 (87.5)	214 (69)	
	Total	32 (100)	310 (100)	
HER2	Negative (0 or +1)	25 (78)	177 (58)	0.023
	Equivoque (+2)	6 (19)	56 (18)	
	Positive (+3)	1 (3)	72 (24)	
	Total	32 (100)	305 (100)	

Algerian women with breast cancer face unfavorable outcomes, with early diagnosis resulting in better survival. The cancer screening organization is a key focus of Algeria's 2015-

2019 cancer plan. Currently, treatment and care for breast cancer patients are high, but the challenge lies in identifying aggressive tumors and the best therapeutic option for each patient. This requires successfully transferring information about tumor biology to the clinical axe, as it is crucial for achieving this goal (6).

3 Breast cancer localisation

Breast cancers can start from different parts of the breast. The breast is an organ that sits on top of the upper ribs and chest muscles. There is a left and right breast and each one has mainly glands, ducts, and fatty tissue. In women, the breast makes and delivers milk to feed newborns and infants. The amount of fatty tissue in the breast determines the size of each breast (7).

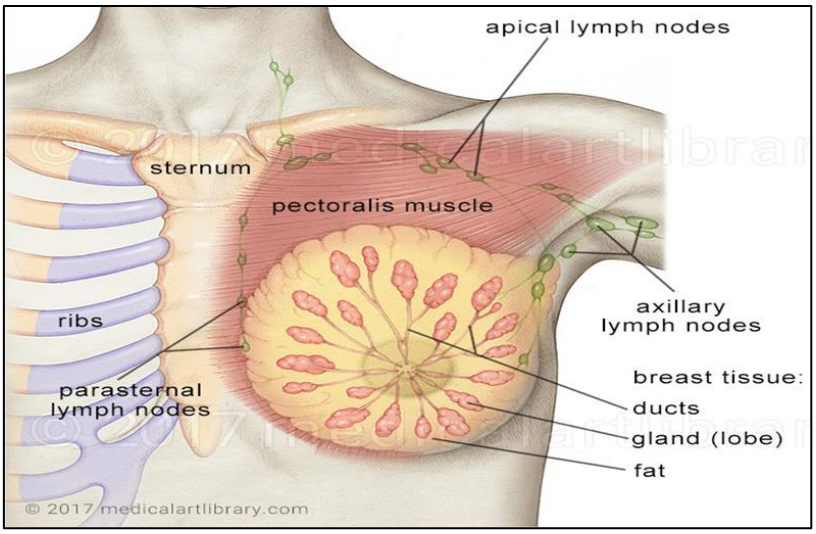


Figure 1: Breast constituent *(8)*

3.1 Breast parts that can be affected

- Lobules are the glands that make breast milk. Cancers that start here are called lobular cancers.
- Ducts are small canals that come out from the lobules and carry the milk to the nipple. This is the most common place for breast cancer to start. Cancers that start here are called ductal cancers.
- The nipple is the opening in the skin of the breast where the ducts come together and turn into larger ducts so the milk can leave the breast. The nipple is surrounded by slightly darker thicker skin called the areola. A less common type of breast cancer called Paget disease of the breast can start in the nipple.

- The fat and connective tissue (stroma) surround the ducts and lobules and help keep them in place. A less common type of breast cancer called phyllodes tumor can start in the stroma.
- Blood vessels and lymph vessels are also found in each breast. Angiosarcoma is a less common type of breast cancer that can start in the lining of these vessels. The lymph system is described below (7).

A small number of cancers start in other tissues in the breast. These cancers are called sarcomas and lymphomas and are not really thought of as breast cancers.

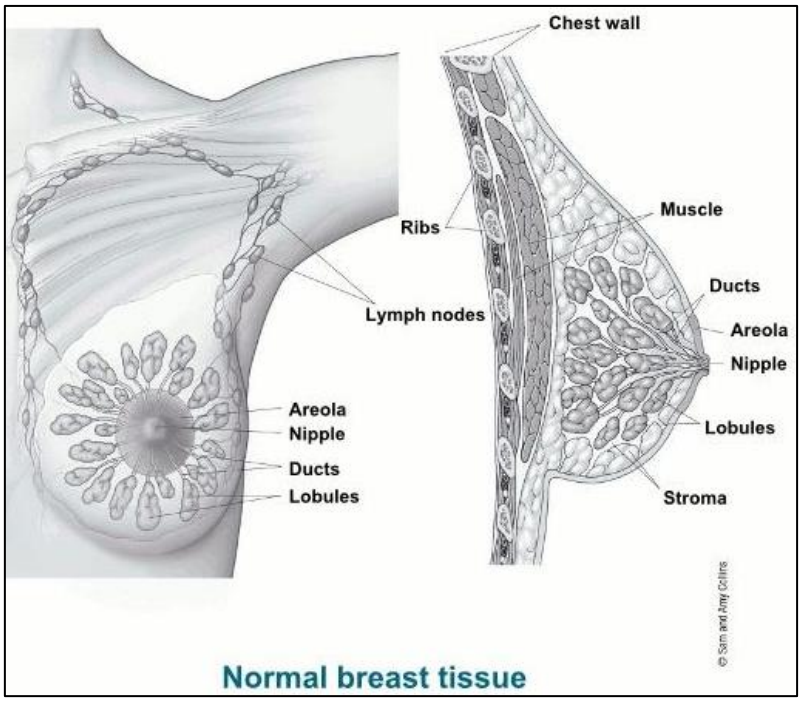


Figure 2: Normal breast tissue *(7)*

3.2 The diffusion of breast cancer

Breast cancer can spread when the cancer cells get into the blood or lymph system and then are carried to other parts of the body.

The lymph (or lymphatic) system is a part of your body's immune system. It is a network of lymph nodes (small, bean-sized glands), ducts or vessels, and organs that work together to collect and carry clear lymph fluid through the body tissues to the blood. The clear lymph fluid inside the lymph vessels contains tissue by-products and waste material, as well as immune system cells (7).

The lymph vessels carry lymph fluid away from the breast. In the case of breast cancer, cancer cells can enter those lymph vessels and start to grow in lymph nodes. Most of the lymph vessels of the breast drain into:

- Lymph nodes under the arm (axillary lymph nodes)
- Lymph nodes inside the chest near the breastbone (internal mammary lymph nodes)
- Lymph nodes around the collar bone (supraclavicular [above the collar bone] and infraclavicular [below the collar bone] lymph nodes)

If cancer cells have spread to your lymph nodes, there is a higher chance that the cells could have traveled through the lymph system and spread (metastasized) to other parts of your body. Still, not all women with cancer cells in their lymph nodes develop metastases, and some women with no cancer cells in their lymph nodes might develop metastases later (7).

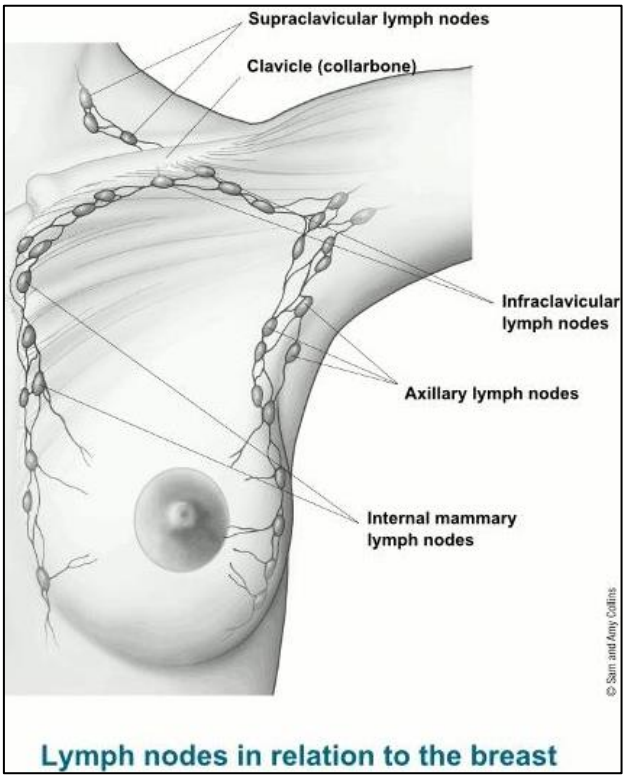


Figure 3: Lymph nodes in relation to the breast *(7)*.

4 Breast cancer types

Healthcare providers determine cancer types and subtypes so they can tailor treatment to be as effective as possible with the fewest possible side effects. Common types of breast cancer include:

- Invasive (infiltrating) ductal carcinoma (IDC): This cancer starts in your milk ducts and spreads to nearby breast tissue. It is the most common type of breast cancer in the United States (9).

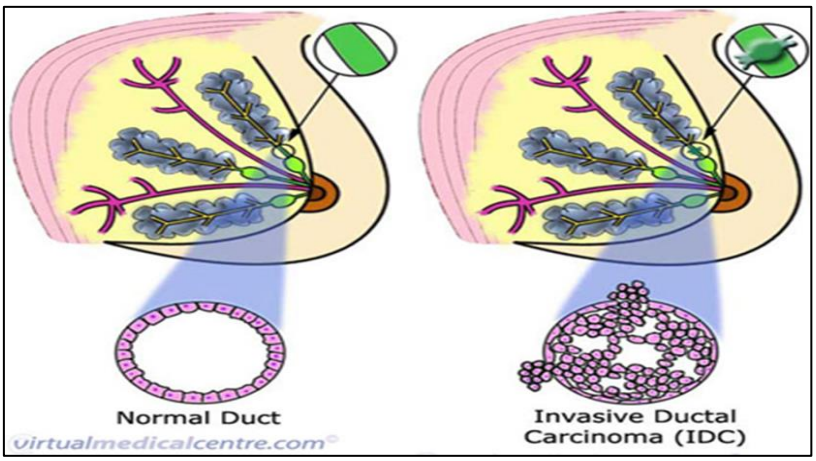


Figure 4: Normal and invasive ductal carcinoma (IDC) *(10)*.

- Lobular breast cancer: This breast cancer starts in the milk-producing glands (lobules) in your breast and often spreads to nearby breast tissue. It is the second most common breast cancer in the United States.

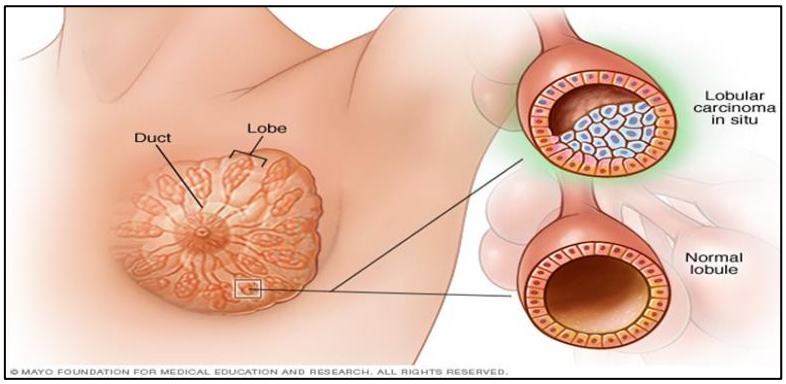


Figure 5: Lobular breast cancer *(11)*

- Ductal carcinoma in situ (DCIS): Like IDC, this breast cancer starts in your milk ducts. The difference is DCIS doesn't spread beyond your milk ducts (9).

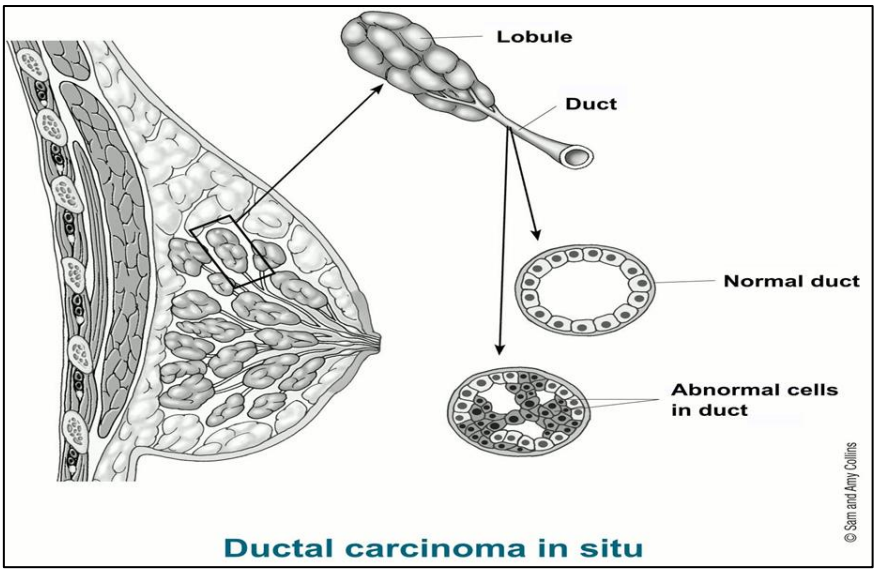


Figure 6: Ductal carcinoma in situ *(12)*

Less common breast cancer types include:

- Triple-negative breast cancer (TNBC): This invasive cancer is aggressive and spreads more quickly than other breast cancers.

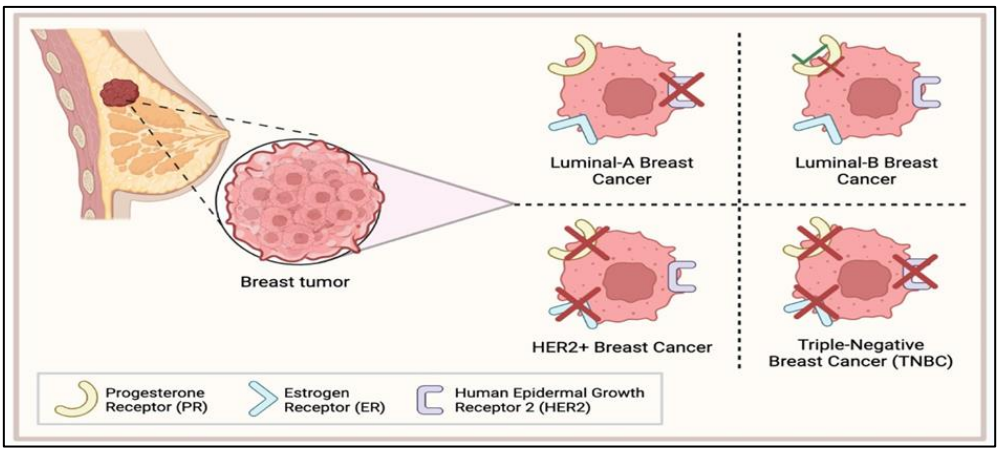


Figure 7: Triple negative breast cancer (TNBC) *(13)*

- Inflammatory breast cancer (IBC): This rare, fast-growing cancer looks like a rash on your breast. IBC is rare in the United States.
- Paget's disease of the breast: This are cancer affects the skin of your nipple and may look like a rash. Less than 4% of all breast cancers are Paget's disease of the breast (9).

4.1 Subtypes include

- ER-positive (ER+) breast cancers have estrogen receptors.
- PR-positive (PR+) breast cancers have progesterone receptors.

- HR-positive (HR+) breast cancers have estrogen and progesterone receptors.
- HR-negative (HR-) breast cancers don't have estrogen or progesterone receptors.
- HER2-positive (HER2+) breast cancers, which have higher than normal levels of the HER2 protein. This protein helps cancer cells to grow. About 15% to 20% of all breast cancers are HER2-positive (9).

5 Risk factors for breast cancer

Studies have shown that your risk for breast cancer is due to a combination of factors. The main factors that influence your risk include being a woman and getting older. Most breast cancers are found in women who are 50 years old or older (14).

Some women will get breast cancer even without any other risk factors that they know of. Having a risk factor does not mean you will get the disease, and not all risk factors have the same effect. Most women have some risk factors, but most women do not get breast cancer.

5.1 Risk factors that can't be changed

- Getting older**.** The risk for breast cancer increases with age. Most breast cancers are diagnosed after age 50.
- Genetic mutations**.** Women who have inherited changes (mutations) to certain genes, such as BRCA1 and BRCA2, are at higher risk of breast and ovarian cancer.
- Reproductive history. Starting menstrual periods before age 12 and starting menopause after age 55 expose women to hormones longer, raising their risk of getting breast cancer.
- Having dense breasts**.** Dense breasts have more connective tissue than fatty tissue, which can sometimes make it hard to see tumors on a mammogram. Women with dense breasts are more likely to get breast cancer (14).
- Personal history of breast cancer or certain non-cancerous breast diseases**.** Women who have had breast cancer are more likely to get breast cancer a second time. Some non-cancerous breast diseases such as atypical ductal hyperplasia or lobular carcinoma *in situ* are associated with a higher risk of getting breast cancer.
- Family history of breast or ovarian cancer. A woman's risk for breast cancer is higher if she has a mother, sister, or daughter (first-degree relative) or multiple family members

on either her mother's or father's side of the family who have had breast or ovarian cancer. Having a first-degree male relative with breast cancer also raises a woman's risk.

- Previous treatment using radiation therapy. Women who had radiation therapy to the chest or breasts (for instance, treatment of Hodgkin's lymphoma) before age 30 have a higher risk of getting breast cancer later in life.
- Exposure to the drug diethylstilbestrol (DES). DES was given to some pregnant women in the United States between 1940 and 1971 to prevent miscarriage. Women who took DES have a higher risk of getting breast cancer. Women whose mothers took DES while pregnant with them also may have a higher risk of getting breast cancer (14).

5.2 Risk factors that can be changed

- Not being physically active. Women who are not physically active have a higher risk of getting breast cancer.
- Being overweight or having obesity after menopause. Older women who are overweight or have obesity have a higher risk of getting breast cancer than those at a healthy weight.
- Taking hormones. Some forms of hormone replacement therapy (those that include both estrogen and progesterone) taken during menopause can raise risk for breast cancer when taken for more than five years. Certain oral contraceptives (birth control pills) also have been found to raise breast cancer risk.
- Reproductive history. Having the first pregnancy after age 30, not breastfeeding, and never having a full-term pregnancy can raise breast cancer risk.
- Drinking alcohol. Studies show that a woman's risk for breast cancer increases with the more alcohol she drinks (14).

6 Clinical aspects

6.1 Symptomes of breast cancer

The first symptom of breast cancer most people notice is a lump in their breast or some thickening (15).

6.1.1 Breast lump

Most breast lumps are not cancer. Lumps that aren't cancer are called benign lumps (15).

Most benign breast lumps are:

- Areas of normal lumpiness. These are usually more obvious just before a period
- Cysts. These are sacs of fluid in the breast tissue, which are quite common
- Fibroadenoma. This is a collection of fibrous glandular tissue which is more common in women younger than 40

6.1.2 Lump or sweeling in the armpit

You can't usually feel the lymph glands in your body. But they can become swollen when you have an infection or a cold.

A less common cause of swollen lymph glands in the armpit is breast cancer that has spread to this area.

6.1.3 Change in the size and shape

Breast cancer might cause your breast to look bigger or have a different shape than usual it might feel different.

It can help to be breast aware. This means getting to know the size, shape and feel of your breasts.

6.1.4 Skin changes

Skin changes of the breast include:

- Puckering
- Dimpling
- Rash
- Redness

The skin might look like orange peel or the texture might feel different. These skin changes can also be caused by other breast conditions.

6.1.5 Fluid leaking from nipples

Fluid leaking from a nipple in a woman who isn't pregnant or breastfeeding can be a sign of breast cancer. But other medical conditions can also cause this.

6.1.6 Change in the position the nipple

One nipple might turn in or sink into the breast. It might look or feel different from usual.

6.1.7 Breast pain

Breast pain is common and it's not normally due to cancer. You might get pain in one or both breasts. This usually goes away after some time. There might be no obvious reason for this pain, even if you have many tests (15).

6.2 Tests for breast cancer

You usually have a number of tests to check for breast cancer (16). These include:

- Breast examination
- Mammogram
- Taking a sample of tissue from your breast called a biopsy
- Scans

6.2.1 GP test

Most people with symptoms that could be due to breast cancer start by seeing their GP. Your GP might examine your breasts and feel the lymph nodes Open a glossary item under your arms and in your neck. If you have had a telephone appointment, your doctor might arrange for you to come in and see them.

Your GP might also ask about your family history. If you have a strong family history of cancer, there might be a faulty gene Open a glossary item in your family that increases your risk of breast cancer.

6.2.2 Breast cancer screening

Some women do not have any symptoms and are referred to a breast clinic for tests if their screening mammogram shows an abnormal area (16).

6.2.3 Mammogram

A mammogram is an x-ray of your breasts. X-rays use of high energy rays to take pictures of the inside of your body. It usually only takes a few minutes.

You need to remove your clothes from the waist upwards and put on a hospital gown. A health professional who specialises in taking mammograms (a mammography) helps position your breast between the 2 plates on the machine. The plates press your breast firmly for a few moments. This is likely to be uncomfortable or painful, but it is over quickly.

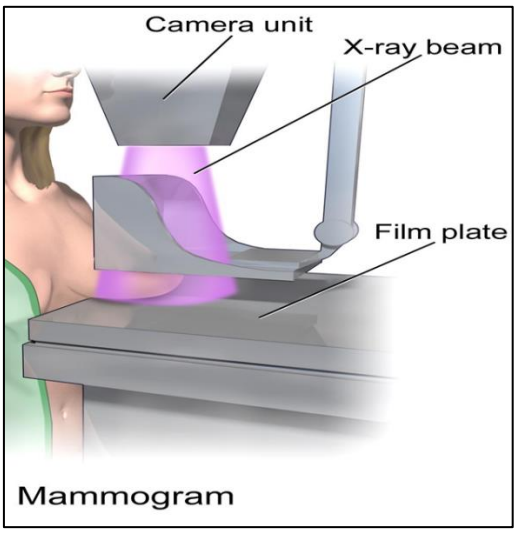


Figure 8: Mammogram *(17)*.

There are 2 types of mammography. They are done for different reasons.

Screening mammography is used to look for cancer in women who do not have any symptoms of breast cancer or breast problems. Both breasts are examined during a screening mammography.

Diagnostic mammography is done to diagnose a breast problem, such as a lump or suspicious area

6.2.4 Breast ultrasound scan

A breast ultrasound is a test that uses high frequency sound waves to create a picture of the inside of the breast. A radiologist Open a glossary item or a specialist called a sonographer puts some gel over your breast. This can feel cold. They then use an ultrasound probe to slide over the skin and give a clear picture on a screen.

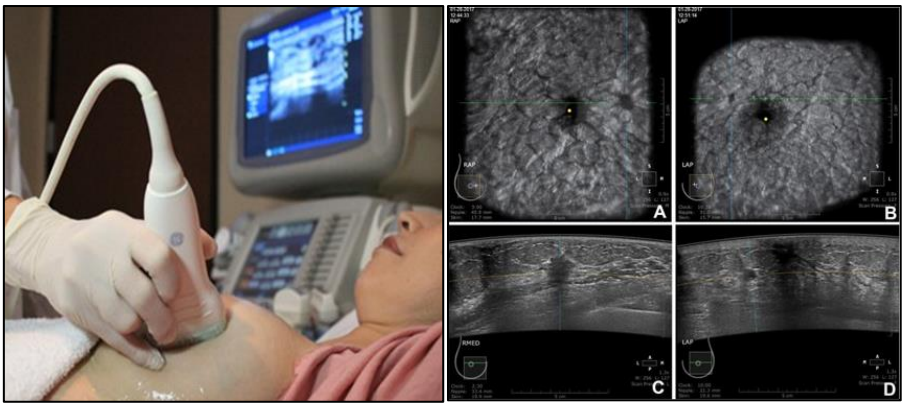


Figure 9: Breast ultrasound scan *(18) (19)*.

6.2.5 Breast biopsy

A breast biopsy means taking samples of tissue from a suspicious area. It is the only way to find out if you have breast cancer or other breast condition. The results of your mammogram and ultrasound help your doctor know if you need to have a biopsy and where to take the biopsy. It also helps your doctor decide the type of biopsy you need (16).

Your doctor sends the samples to a laboratory and a pathologist Open a glossary item looks at them under a microscope to check for cancer.

You might have a breast biopsy during your clinic appointment. There are different types of breast biopsy. These include :

- Needle core biopsy
- Fine needle aspiration
- Skin punch biopsy
- Vacuum assisted biopsy

The type of biopsy you have depends on a number of factors including how big the suspicious area is and where it is. Your doctor will tell you which type of biopsy you need. You may need more than one biopsy or different types of biopsies. (16)

The pathologist does various tests on the sample. The tests can:

- Diagnose breast cancer
- Show which type of breast cancer you have
- Look for proteins on the surface of the breast cancer cells

Fine needle aspiration (FNA) uses a very thin needle and syringe to remove a small amount of tissue from a lump. Doctors use it to find out if the lump is a cyst or solid tumour. FNA cannot tell doctors if the cancer is non-invasive or invasive.

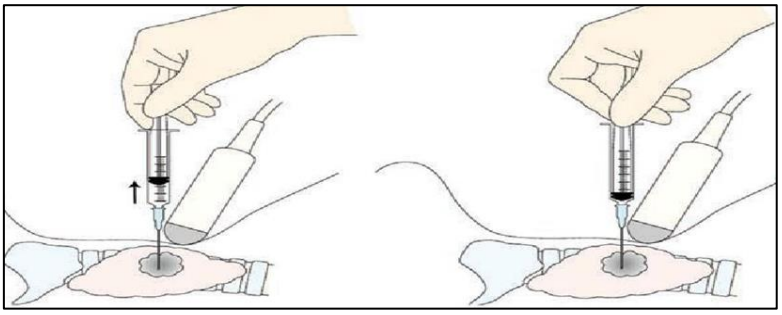


Figure 10: Fine needle aspiration *(20)*.

Core biopsy uses a special hollow needle to remove tissue from the body. Doctors use it to take a sample from a suspicious area in the breast. They may take several samples of the area during the procedure. Sometimes doctors use a special vacuum to remove more tissue through the hollow needle. This technique is called vacuum-assisted core

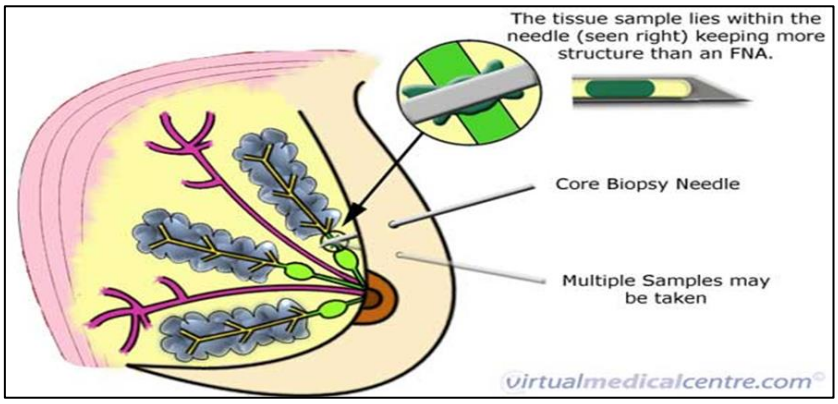


Figure 11: Core biopsy needle *(21)*.

Stereotactic core biopsy uses 3-D, or stereotactic, images to find the exact location of a tumour or suspicious area in the breast. Doctors use these images to guide them during a core biopsy. Stereotactic core biopsy is used to take samples from lumps that can be seen on an imaging test but that can't be felt in the breast.

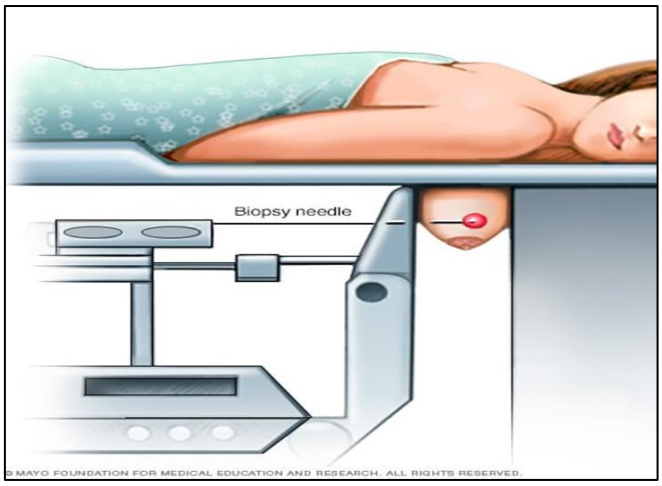


Figure 12: Stereotatic core biopsy *(22)*.

Surgical, or open, biopsy removes all or part of a lump or a suspicious area. An excisional biopsy removes the whole area with a margin of healthy tissue around it. An incisional biopsy removes only part of the area.

Punch biopsy uses a sharp hollow cutting instrument to remove an area of skin and the underlying tissue. It is used to diagnose inflammatory breast cancer, which means there are cancer cells in the lymph vessels of the skin.

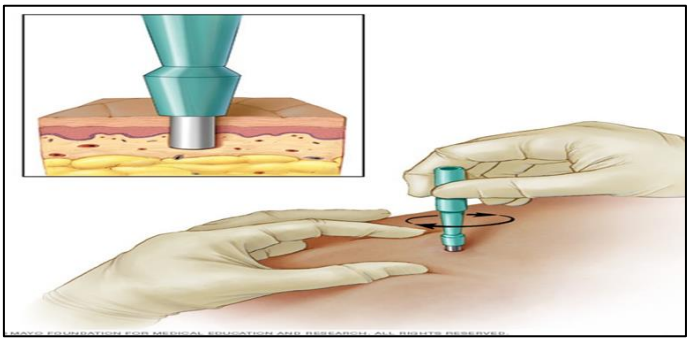


Figure 13: Punch biopsy *(23)*.

Lymph node biopsy a lymph node biopsy is a surgical procedure that removes lymph nodes so they can be examined under a microscope to find out if they contain cancer.

Breast cancer cells can break away from the tumour and travel through the lymphatic system. The first place they may spread is the lymph nodes under the arm (called the axillary lymph nodes).

Doctors use the number of lymph nodes that have cancer in them to help determine the stage of breast cancer (16).

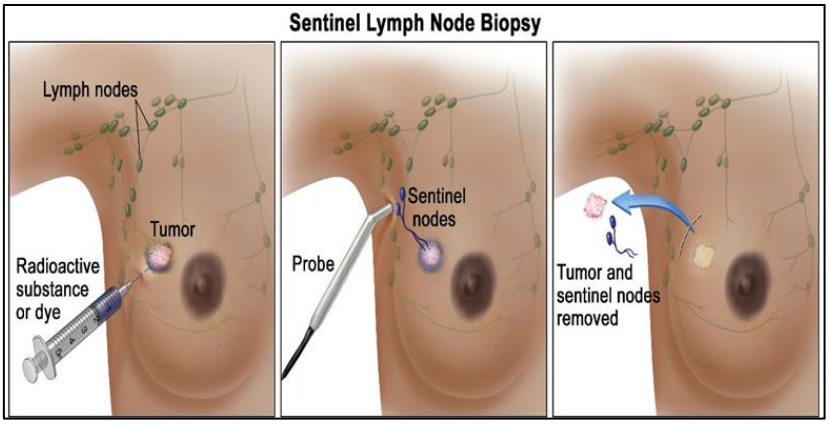


Figure 14: Lymph node biopsy *(24)*.

6.2.6 Procedure after biopsy

If you have a biopsy that shows that you have breast cancer, you might have more tests to work out where and how big the cancer is. This is called staging the cancer (25).

The tests you may have include:

- Checking for cancer cells in the lymph nodes in your armpit
- MRI scan
- CT scan
- Blood tests

6.2.7 Checking for cancer cells in the lymph nodes of the armpit

Breast cancer cells can sometimes spread into the lymph nodes under the arm (the axilla). To check for this, you may have an axilla lymph node ultrasound and biopsy.

The ultrasound scanner has a probe that gives off sound waves. The sound waves bounce off the organs inside your body, and the probe picks them up. The probe links to a computer that turns the sound waves into a picture on the screen.

If there are any areas that look abnormal, doctors can use the ultrasound to guide a needle and take a sample of cells (25).

6.2.8 Breast MRI (Magnetic Resonance Imaging) scan

MRI stands for magnetic resonance imaging. It uses magnetism and radio waves to take pictures of the inside of the body.

An MRI scan can tell your doctor:

- Where the cancer is and how big it is
- If there are any other suspicious areas in the breast
- Whether it has spread to the lymph nodes

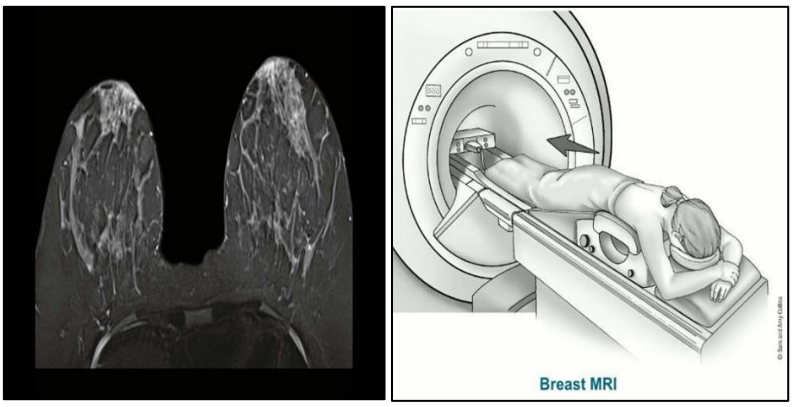


Figure 15: Breast MRI scan *(26) (27)*.

6.2.9 CT (Computed Tomography) scan

CT (or CAT) scan stands for computed (axial) tomography. It is a test that uses x-rays and a computer to create detailed pictures of the inside of your body (25).

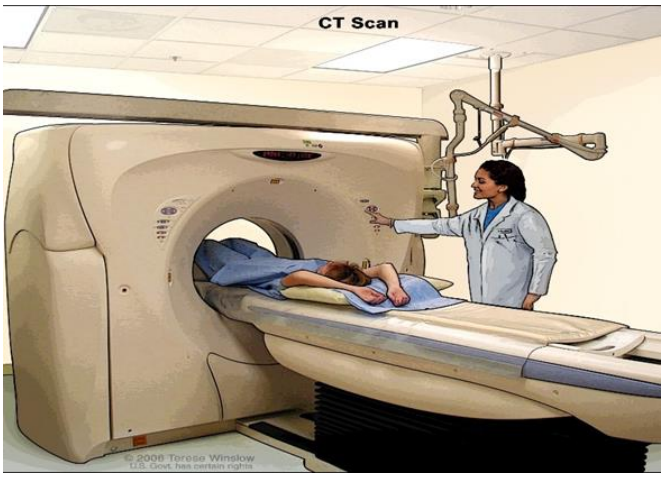


Figure 16: Computed axial topography (CT) *(28)*.

6.2.10 Blood tests

Blood tests can check your general health, including how well your liver and kidneys are working. Doctors can also check the number of blood cells Open a glossary item. You usually have blood tests to check that you are well enough to have treatment.

- Tumour marker blood test

Tumour markers are substances that might be raised if there is a cancer. They're usually proteins that can be found in the blood.

A tumour marker that is sometimes used in breast cancer is a protein called CA 15-3.

Doctors may use the CA 15-3, along with other tests, to check if treatment is working. A raised level may suggest that the cancer is spreading in some people with secondary breast cancer (25).

Tumour marker tests are not reliable enough to use on their own too:

- Diagnose breast cancer
- Make decisions about your treatment

6.2.11 Hormone receptors status testing

Estrogen and progesterone are hormones that can stimulate the growth of breast cancer cells. Hormone receptor status testing looks for estrogen receptors (ERs) and progesterone receptors (PRs) in the breast cancer cells. This information will help your healthcare team decide which treatments will work best for you (16).

HER2 status testing

ERBB2 is more commonly known as HER2 (or HER2/neu). HER2 stands for human epidermal growth factor receptor 2. A gene has changed (mutated), so it helps a tumour grow (called an oncogene).

HER2 status testing is done to find out if breast cancer cells are making more HER2 protein than normal (called overexpression). This information will help your healthcare team decide which treatments will work best for you

6.2.12 Bone scan

A bone scan uses bone-seeking radioactive materials (called radiopharmaceuticals) and a computer to create a picture of the bones. It is used to find out if breast cancer has spread to the bones (called bone metastasis).

A bone scan is done if a woman has bone pain or the level of alkaline phosphatase is higher than normal. Doctors may also order a bone scan if a breast tumour is larger than 5 cm or they can feel lymph nodes under the arm (called axillary lymph nodes). They do not do bone scans for women with stage I breast cancer (16) .

5 Treatments

Treatment for breast cancer depends on the subtype of cancer and how much it has spread outside of the breast to lymph nodes (stages II or III) or to other parts of the body (stage IV) (29).

Doctors combine treatments to minimize the chances of the cancer coming back (recurrence). These include :

- Surgery to remove the breast tumour
- Radiation therapy to reduce recurrence risk in the breast and surrounding tissues
- Medications to kill cancer cells and prevent spread, including hormonal therapies, chemotherapy or targeted biological therapies.

  Early initiation and completion of breast cancer treatments result in improved efficacy and tolerance. Surgical options include lumpectomy (removal of cancerous tissue) or mastectomy (removal of the entire breast), often coupled with lymph node removal to assess cancer spread. Radiation therapy targets residual microscopic cancers in breast tissue or lymph nodes, reducing the risk of recurrence.

  Advanced breast cancers may manifest as painless skin erosion, necessitating medical attention for biopsy if wounds persist. Treatment selection hinges on tumor biology, determined through specialized tests. Many breast cancer medications are listed in the WHO Essential Medicines List.

Historically, complete axillary lymph node removal was standard practice, but now sentinel node biopsy is preferred due to fewer complications. Medical treatments administered before (neoadjuvant) or after (adjuvant) surgery, are tailored to cancer subtype. Hormone-positive cancers respond well to endocrine therapies, significantly reducing recurrence risk. Hormone receptor-negative cancers typically require chemotherapy, which is often outpatient-based.

HER-2 positive breast cancers can be targeted with expensive biological agents like trastuzumab, usually combined with chemotherapy. Radiotherapy is vital, potentially averting mastectomy in early stages and lowering recurrence risk post-mastectomy. In advanced cases, radiation may improve survival chances. Completion of the full treatment regimen is crucial for optimal outcomes (29).

GENETIC ASPECTS

2 Genetic aspects

2.1 BRCA1 and BRCA2

BRCA1 (on chromosome 17q12-21) and BRCA2 (on chromosome 13q12-13) genes are responsible for 80% of early-onset female breast cancer cases in families. Germline mutations in BRCA1 are linked to ovarian cancer, while BRCA2 mutations increase the risk of male breast, ovarian, prostate, and pancreatic cancer. The isolation of these genes could help identify genetic defects and use genetic information for predictive testing (30).

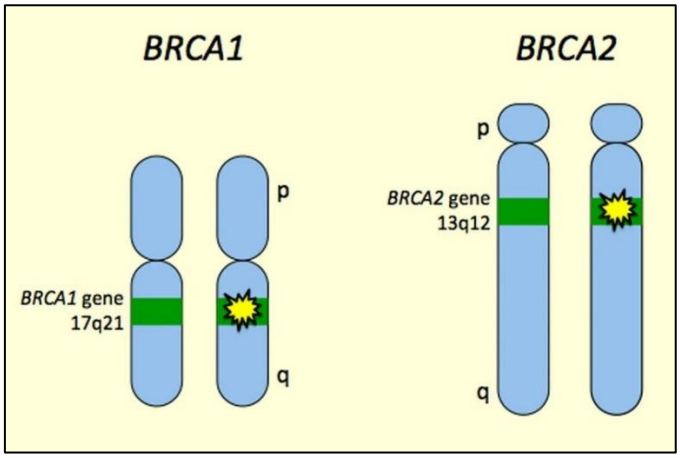


Figure 17: BRCA1 BRCA2 gene localization *(31)*.

BRCA1 and BRCA2 are genes that repair DNA are sometimes called tumor suppressor genes. When inherited with harmful variants, cancer can develop. People with these variants have increased risks of breast and ovarian cancer, and may develop cancer at younger ages. Inherited mutations are present from birth in all cells in the body. Even if a harmful variant is inherited from one parent, the normal copy can be lost or altered, leading to somatic alteration, causing cells without functioning proteins to grow out of control and become cancerous (32).

2.2 BRCA1 gene

2.2.1 Structure and Function of BRCA1 Gene

BRCA1 is an incomplete recessive gene located on an autosomal, positioned on chromosome 17q21. It encodes a 220 KD protein with multifunctional domains (33). The gene consists of 24 exons, with exons 2–5 responsible for the RING domain and exons 15–23

encoding, the BRCA1 C-terminal (BRCT) domain (Figure 19) (34) (35). The N-terminal RING domain demonstrates E3 ligase activity and forms a stable heterodimer with the BRCA1-associated RING domain protein 1 (BARD1) (36). The BRCT domain interacts with various phosphorylated proteins. Apart from these domains, there exists a coiled-coil domain upstream of the BRCT domains, binding another coiled-coil domain at the N-terminus of PALB2. PALB2 also interacts with BRCA2, forming the BRCA1-PALB2-BRCA2 complex (37) (38). In mammalian cells, homologous recombination (HR) and non-homologous end-joining (NHEJ) are the primary repair pathways for DNA double-strand breaks (DSBs) to ensure genome integrity. Both the RING and BRCT domains of BRCA1 play essential roles in HR. Many clinically significant mutations of the BRCA1 gene frequently target these two domains (39).

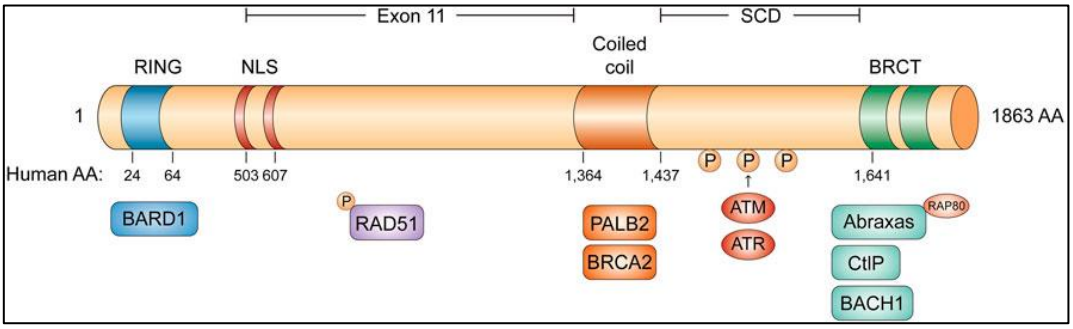


Figure 18: The domain structure of BRCA1 *(39)*.

The BRCT domain, present in various DNA damage response (DDR) proteins, allows BRCA1 to recognize a phospho-SPxF motif (S, serine; P, proline; x, varies; F, phenylalanine) (40) (41). BRCA1 forms four distinct complexes in cells by interacting with different adaptor proteins via the BRCT domain. These complexes include the BRCA1/RAP80/Abraxas complex, BRCA1/BACH1 complex, BRCA1/PALB2/BRCA2 complex, and BRCA1/CtIP complex (39). The BRCA1/RAP80/Abraxas complex is recruited to DNA double-strand breaks (DSBs) by RAP80, an ubiquitin-binding protein. It targets MDC1-rH2AX-dependent K6 and K63-linked ubiquitin polymers at DSBs, preventing excessive end resection and potentially harmful homology-directed DSB repair mechanisms (42) (43). The helicase catalytic function of BRCA1/BACH1 is vital not only for BRCA1-mediated DDR but also for maintaining the DNA damage-induced G2/M checkpoint (44). As previously noted, PALB2, acting as the partner and localizer of BRCA2, can bind directly to BRCA1, forming the BRCA1/PALB2/BRCA2 complex. This complex facilitates RAD51-mediated localization and repair at DNA breaks (45) (46). Additionally, the

BRCA1/CtIP complex promotes homologous recombination (HR) by facilitating DNA end resection (39).

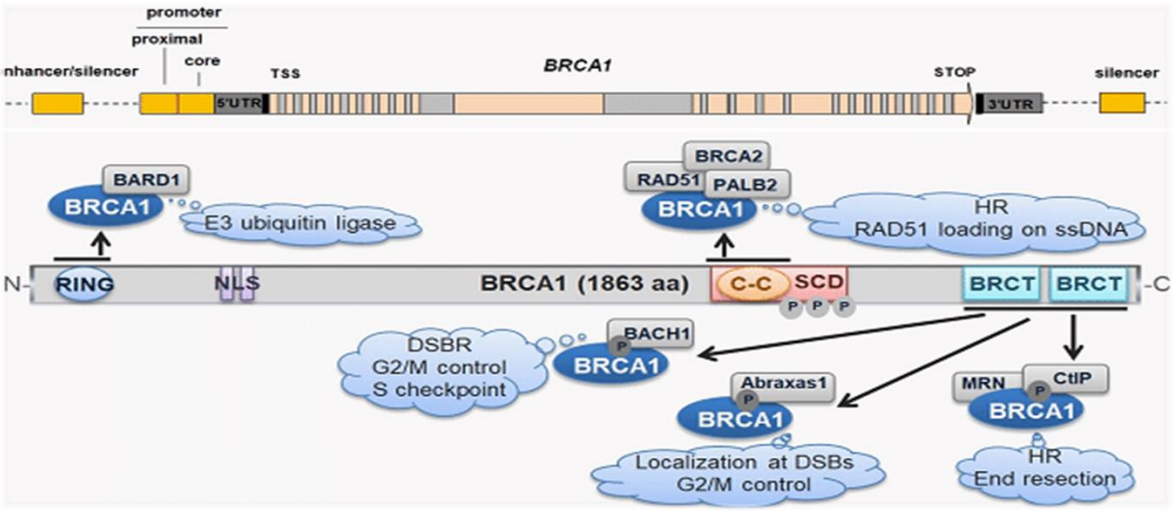


Figure 19: BRCA1 different domain complexes *(47)*

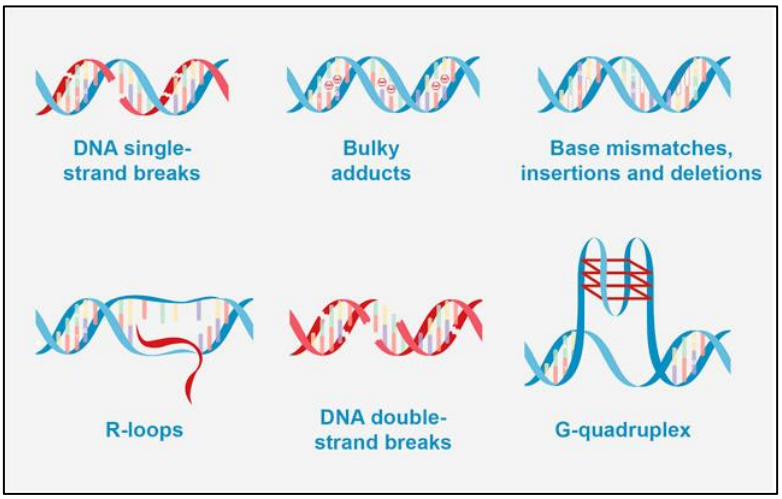


Figure 20: Types of DNA damage. DNA single-strand breaks: BER (base excision repair); bulky adducts: NER (nucleotide excision repair); base mismatches, insertions and deletions: MMR (Mismatch repair); R-loops caused double strand breaks (DSB): NHEJ (Non-homologous end joining) and HRR (homologous recombination repair); DNA double-strand breaks: NHEJ and HRR; G-quadruplex caused DSB: NHEJ, and HRR *(39)*.

BRCA1, known for its role in repairing DNA double-strand breaks (DSBs), also participates in DNA replication fork protection and transcriptional regulation (39). Independently of DSB repair, BRCA1 stabilizes RAD51 nucleofilaments during replication fork stress, inhibiting MRE11 exonuclease activity ((48); (49)). Upon replication fork stress, BRCA1 protects nascent DNA strands from degradation by stabilizing RAD51 nucleofilaments that affect the exonuclease activity of MRE11 (48). It facilitates RAD51-mediated fork restart after

short replication blocks (50) . Additionally, BRCA1 transactivates heterologous promoters and maintains chromatin structures through histone ubiquitylation ((51); (52); (53)) it influences gene transcription by modulating chromatin structure via the SWI/SNF complex. BRCA1 resolves RNA/DNA hybrids (R-loops) at transcriptional pause sites by recruiting senataxin (SETX) (54) . Elevated R-loop levels alter BRCA1 distribution, affecting transcription inhibition post-DNA damage and HR repair (55). As a result, these cells could not inhibit the transcription after DNA damage and showed a defect in HR repair. Further investigation in human neuroblastoma cells reveals additional insights into the connection between BRCA1, R-loops, and transcriptional regulation (56). In MYCN-amplified cells, BRCA1 prevents R-loop formation at transcriptional pause sites, stabilizing MYCN on chromatin (39).

2.2.2 BRCA1 DNA repair associated

The BRCA1 protein is crucial for repairing damaged DNA, interacting with other proteins in the nucleus of normal cells to mend DNA breaks caused by radiation or genetic exchange during cell division (57). Additionally, BRCA1 regulates gene activity and contributes to embryonic development by interacting with various proteins, including tumor suppressors and cell division regulators.

Mutations in the BRCA1 gene increase the risk of breast and other cancers and can be passed down through generations, although not everyone with these mutations develops cancer due to other factors (57). Most mutations lead to abnormal BRCA1 protein production, reducing DNA repair capacity and potentially triggering tumor formation.

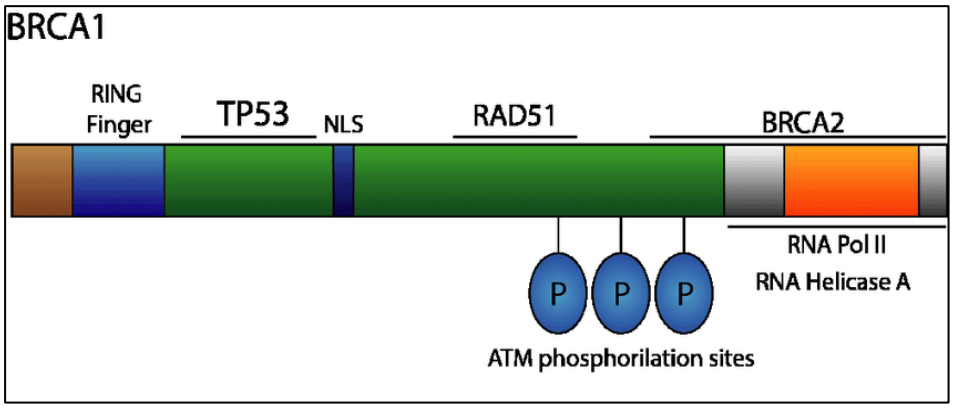


Figure 21: BRCA1 gene *(58)*.

2.2.3 The Mutation of BRCA1 and Breast Cancer

Since the cloning of the first BRCA1 gene in 1994 (59) , numerous studies have identified variable cut transcripts in both tumor and normal tissues (60). These alternative splicing transcripts, including BRCA1 exon 1a, exon 1b, exon 1c, BRCA1a (Δ11q, Δ11), BRCA1b (Δ9, 10), and BRCA1-IRIS, produce protein products with different molecular weights. Research has shown that some of these exon boundary variants may not result in loss of function, generating naturally occurring in-frame RNA isoforms, (61). Compared to the complete BRCA1 transcript, these variable-cut transcripts may exhibit similar or opposite functions and may even have unique functions (62). To date, approximately 1,800 mutations have been identified in human BRCA1, including intron mutations, missense mutations, nonsense mutations, frameshift mutations, and others. These mutations frequently occur in the RING and BRCT domains, critical for genome integrity (Supplementary Data Sheet). Missense mutations, such as BRCA1 c.5309G > T p. (Gly1770Val), pose significant challenges for patient prevention and treatment by inhibiting homologous recombination and serving as disease-causing mutations (63). Bioinformatic analysis has facilitated the discovery of additional BRCA1 variants through public databases like cBioPortal, ENIGMA, BRCA Exchange, and ClinVar (39).

Patients with BRCA1 mutations have a higher risk for cancer. The estimated lifetime risk of BC is about 80%, and the lifetime risk of ovarian cancer is 40%–65% ((64); (65)), which might alter according to the type and location of the mutations (66). BRCA1 gene deletion with or without p53 defect leads to a high incidence of basal-like BC and tends to form TNBC, which is the most aggressive type of BC (67). Some studies show that the TNBC in BRCA1 mutation carriers originated from luminal progenitor cells, not basal stem cells ((68); (69)). If BRCA1/p53 is perturbed in luminal progenitors, it could induce the abnormal alveolar differentiation premalignancy with pro-tumorigenic changes in the immune compartment. It belongs to cell autonomy and caused by the dysregulation of transcription factors (39). This study explains how BRCA1 aberration impacts the state of nascent tumor cells and their microenvironment. Bach et al. found that breast cells with BRCA1 mutations undergo changes similar to those common changes in women during pregnancy (70). Based on the data, they proposed a model in which BRCA1/p53-driven transcriptional and epigenetic changes inadvertently promote innate differentiation programs in luminal progenitors accompanied by protumorigenic

changes in the immune compartment, highlighting the decisive role of the origin cell and providing a potential explanation for the tissue specificity of BRCA1 tumors. Researchers have mapped early changes in seemingly healthy breast tissue before tumors appear which may have great significance for the early diagnosis of BC (70).

In addition to familial BC, BRCA1 gene silencing due to promoter methylation can also lead to sporadic BC (71). A study of tumor xenografts from TNBC patients (72) revealed a novel resistance mechanism in BRCA1-methylated PDX (patient-derived xenograft) tumors. Next-generation sequencing (NGS) data showed that the genome rearrangement places the BRCA1 gene under the transcriptional control of the heterologous promoter, which results in the re-expression of BRCA1 in a subset of BRCA1-mutated PDX tumors and leads to acquired resistance to PARP [poly (ADP-ribose) polymerase 1,2] inhibitor (PARPi) and cisplatin chemotherapy (72). This is a unique example of genomic plasticity that is caused by the treatment of BRCA1-deficient tumors, but it can lead to tumor regeneration (39).

2.2.4 BRCA1 protein

Breast cancer type 1 susceptibility (BRCA1) protein is a tumor suppressor that is expressed in breast tissue cells and other tissues where it primarily functions to sense and repair DNA damage. BRCA1 is an E3 ubiquitin-protein ligase that specifically mediates the formation of 'Lys6'-linked polyubiquitin chains and plays a central role in DNA repair by facilitating cellular responses to DNA damage (73). The E3 ubiquitin-protein ligase activity is required for its tumor suppressor function (73). BRCA1 associates with other tumor suppressors, DNA repair proteins, and signal transducers in a large complex known as the BRCA1-associated genome surveillance complex (BASC) that functions to identify double-strand DNA breaks and either repair them or destroy cells where DNA damage cannot be repaired. BRCA1 also associates with RNA pol II and histone deacetylase complexes to modulate transcription. In addition to its role in ubiquitination, DNA damage repair and transcriptional regulation, it is known to be required for normal cell cycle progression. Defects in BRCA1 are a cause of susceptibility to familial breast-ovarian cancer type 1 (BROVCA1), a condition associated with familial predisposition to cancer of the breast and ovaries, and pancreatic cancer type 4 (PNCA4) (73).

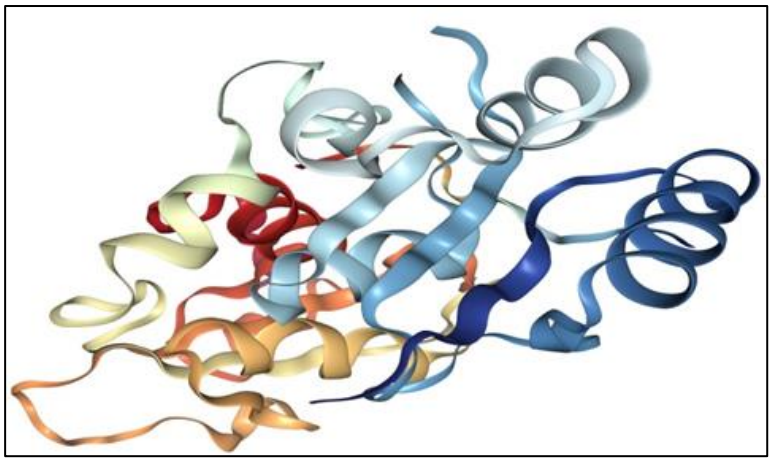


Figure 22: BRCA1 3D structure *(74)*.

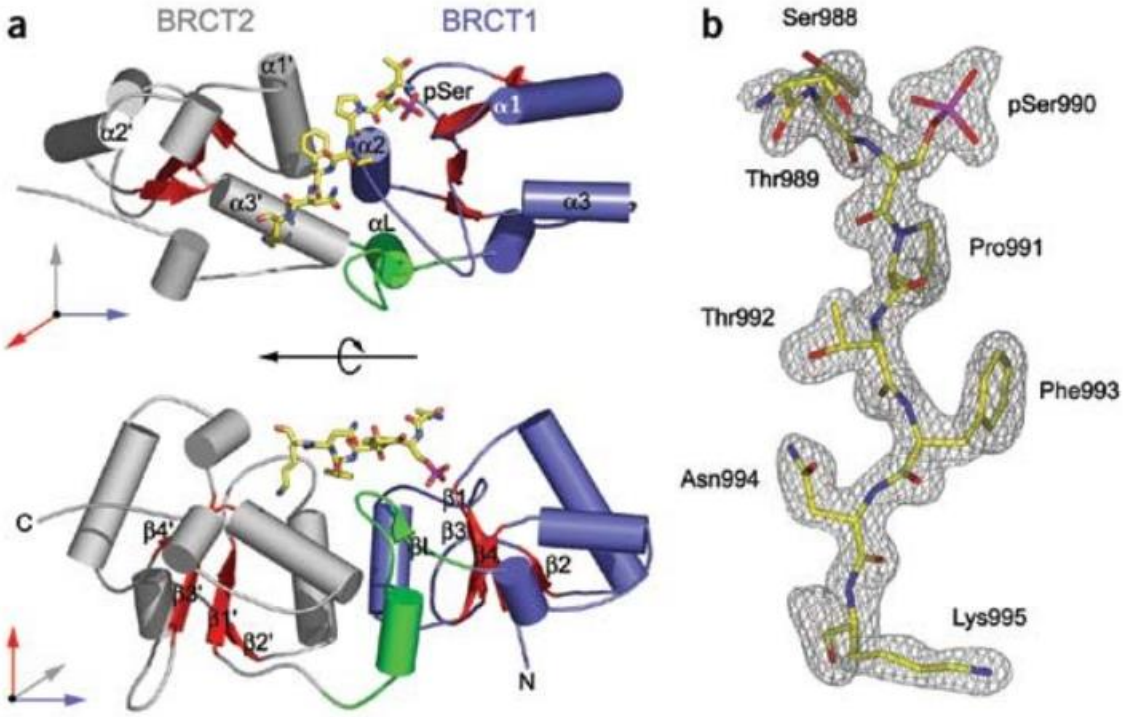


Figure 23: Structure and mechanism of BRCA1 BRCT domain recognition of phosphorylated BACH1 with implications for cancer *(75)* **.**

BIOINFORMATICS

3 Bioinformatics

Bioinformatics is a relatively new and evolving discipline that combines skills and technologies from computer science and biology to help us better understand and interpret biological data.

One field where bioinformatics is especially useful is genomics, which can generate vast quantities of information. Bioinformatics helps to give meaning to the data, which can be used to make a diagnosis for a patient with a rare condition, to track and monitor infectious organisms as they move through a population, or to identify the best treatment for a patient with cancer (76).

- Bioinformatics is the application of tools of computation and analysis to the capture and interpretation of biological data
- Bioinformatics is essential for management of data in modern biology and medicine
- The bioinformatics toolbox includes computer software programs such as BLAST and Ensembl, which depend on the availability of the internet
- Analysis of genome sequence data, particularly the analysis of the human genome project, is one of the main achievements of bioinformatics to date
- Prospects in the field of bioinformatics include its future contribution to functional understanding of the human genome, leading to enhanced discovery of drug targets and individualised therapy (77)

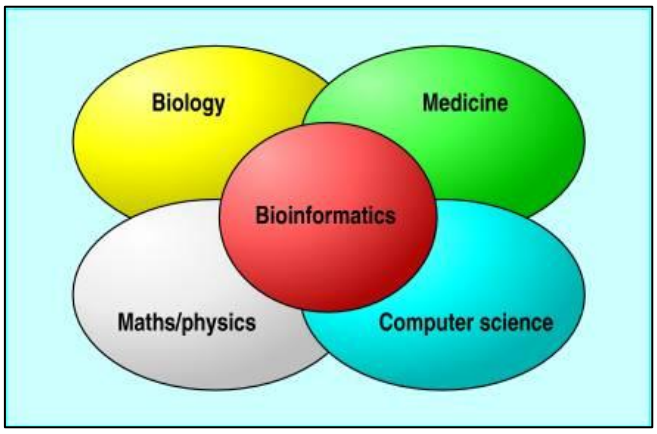


Figure 24: Relation between bioinformatics and other science *(77)*.

3.1 Bioinformatics history

It is easy for today's students and researchers to believe that modern bioinformatics emerged recently to assist next-generation sequencing data analysis. However, the very beginnings of bioinformatics occurred more than 50 years ago, when desktop computers were still a hypothesis and DNA could not yet be sequenced. The foundations of bioinformatics were laid in the early 1960s with the application of computational methods to protein sequence analysis (notably, de novo sequence assembly, biological sequence databases and substitution models) (78). Later on, DNA analysis also emerged due to parallel advances in (i) molecular biology methods, which allowed easier manipulation of DNA, as well as its sequencing, and (ii) computer science, which saw the rise of increasingly miniaturized and more computers that are powerful as well as novel software better suited to handle bioinformatics tasks. In the 1990s through the 2000s, major improvements in sequencing technology, along with reduced costs, gave rise to an exponential increase of data. The arrival of 'Big Data' has laid out new challenges in terms of data mining and management, calling for more expertise from computer science into the field. Coupled with an ever-increasing amount of bioinformatics tools, biological Big Data had (and continues to have) profound implications on the predictive power and reproducibility of bioinformatics results. To overcome this issue, universities are now fully integrating this discipline into the curriculum of biology students. Recent subdisciplines such as synthetic biology, systems biology and whole-cell modeling have emerged from the ever-increasing complementarity between computer science and biology (78).

3.2 Bioinformatic tools

The main tools of a bioinformatician are computer software programs and the internet (77). A fundamental activity is sequence analysis of DNA and proteins using various programs and databases available on the World Wide Web. Anyone, from clinicians to molecular biologists, with access to the internet and relevant websites can now freely discover the composition of biological molecules such as nucleic acids and proteins by using basic bioinformatic tools. This does not imply that handling and analysis of raw genomic data can easily be carried out by all. Bioinformatics is an evolving discipline, and expert bioinformaticians now use complex software programs for retrieving, sorting out, analysing, predicting, and storing DNA and protein sequence data.

Large commercial enterprises such as pharmaceutical companies employ bioinformaticians to perform and maintain the large scale and complicated bioinformatic needs of these industries. With an ever-increasing need for constant input from bioinformatic experts, most biomedical laboratories may soon have their own in-house bioinformatician. The individual researcher, beyond a basic acquisition and analysis of simple data, would certainly need external bioinformatic advice for any complex analysis.

The growth of bioinformatics has been a global venture, creating computer networks that have allowed easy access to biological data and enabled the development of software programs for effortless analysis. Multiple international projects aimed at providing gene and protein databases are available freely to the whole scientific community via the internet (77).

3.3 Bioinformatic analysis

The escalating amount of data from the genome projects has necessitated computer databases that feature rapid assimilation, usable formats and algorithm software programs for efficient management of biological data.13 Because of the diverse nature of emerging data, no single comprehensive database exists for accessing all this information. However, a growing number of databases that contain helpful information for clinicians and researchers are available. Information provided by most of these databases is free of charge to academics, although some sites require subscription and industrial users pay a licence fee for particular sites. Examples range from sites providing comprehensive descriptions of clinical disorders, listing disease susceptibility genetic mutations and polymorphisms, to those enabling a search for disease genes given a DNA sequence (box).

These databases include both “public” repositories of gene data as well as those developed by private companies. The easiest way to identify databases is by searching for bioinformatic tools and databases in any one of the commonly used search engines. Another way to identify bioinformatic sources is through database links and searchable indexes provided by one of the major public databases. For example, the National Center for Biotechnology Information (www.ncbi.nlm.nih.gov) provides the Entrez browser, which is an integrated database retrieval system that allows integration of DNA and protein sequence databases. The European Bioinformatic Institute archives gene and protein data from genome studies of all organisms, whereas Ensembl produces and maintains automatic annotation on

eukaryotic genomes (fig24). The quality and reliability of databases vary; certainly some of the better-known and more established ones, such as those above, are superior to others (77).

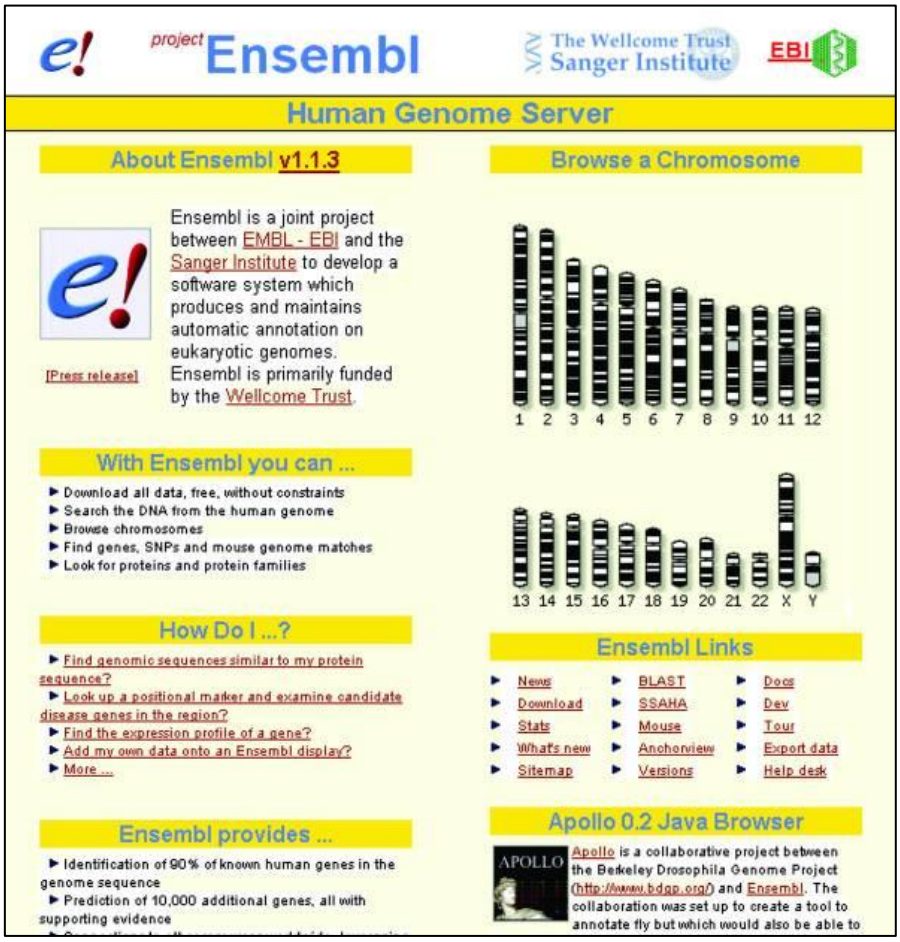


Figure 25: Ensembl website: a genomic data search facility freely available on the internet *(77)*.

One of the simplest and better-known search tools is called BLAST (basic local alignment search tool, at www.ncbi.nlm.nih.gov/BLAST/). This algorithm software is capable of searching databases for genes with similar nucleotide structure (fig25) and allows comparison of an unknown DNA or amino acid sequence with hundreds or thousands of sequences from human or other organisms until a match is found. Databases of known sequences are thus used to identify similar sequences, which may be homologues of the query sequence. Homology implies that sequences may be related by divergence from a common ancestor or share common functional aspects. When a database is searched with a newly determined sequence (the query sequence), local alignment occurs between the query sequence and any similar sequence in the database. The result of the search is sorted in order of priority on the basis of maximum similarity. The sequence with the highest score in the

database of known genes is the homologue. If homologues or related molecules exist for a query sequence, then a newly discovered protein may be modelled and the gene product may be predicted without the need for further laboratory experiments (77) .

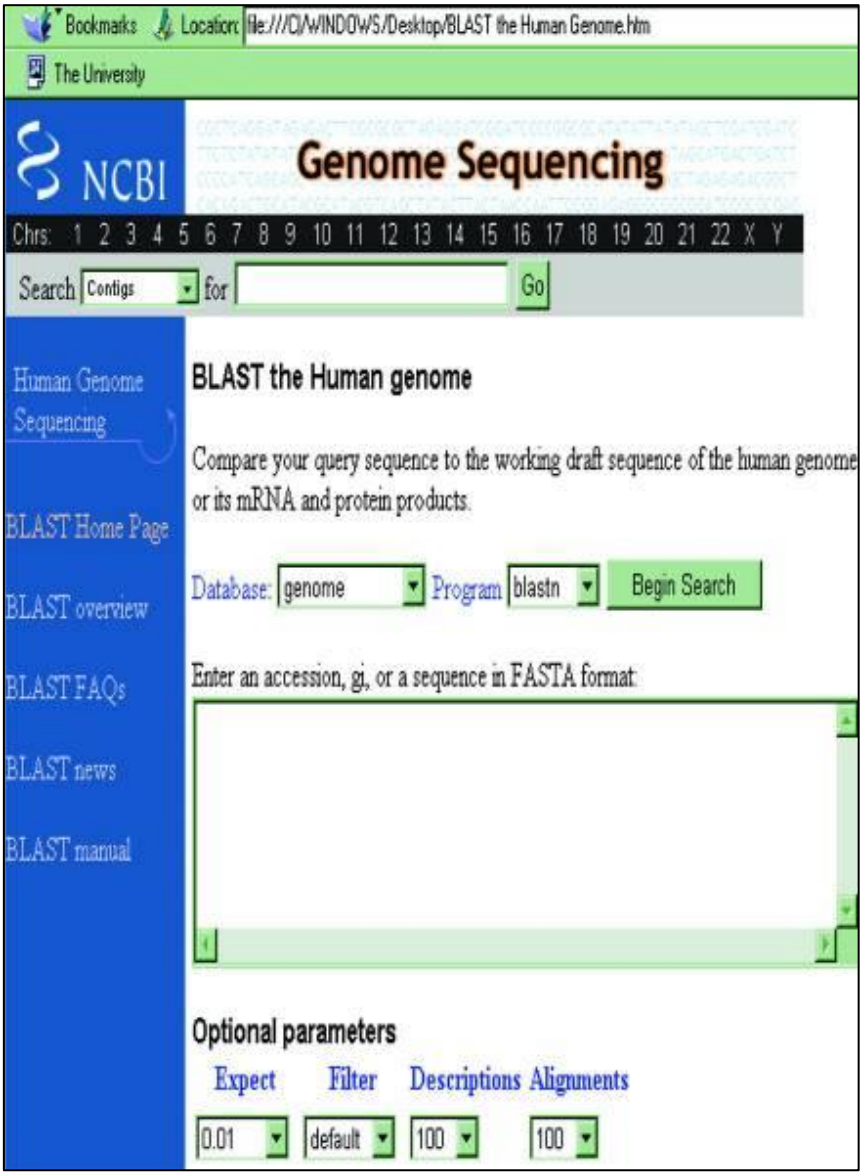


Figure 26: Web page illustrating freely available BLAST (basic local alignment search tool). Services run by the National Center for Biotechnology Information *(77)* .

POOULATION STUDY AND METHOD

4 Population of the study

The *in silico* study focused on two mutations identified in Algerian patients with breast cancer. The mutations are summarized in Table 5.

Table 5: List of mutations identified in the Algerian population (79)

Exon/Intron	Type of mutation	DNAc	Protein
Exon 2	missense	c.122A>G	p. His41Arg
Exon 2	missense	c.135G>C	p. Lys45Asn

5 Method

The aim of our study is to see the effect of this mutations c.122A>G and c.135G>C by using different software (I-Mutant 2.0, (SIFT), (Align-GVGD), (PolyPhen-2), (Project HOPE).

5.1 I-Mutant 2.0 software

I-Mutant2.0 is a support vector machine (SVM)-based tool for the automatic prediction of protein stability changes upon single point mutations. I-Mutant2.0 predictions are performed starting either from the protein structure or, more importantly, from the protein sequence (80). (https://folding.biofold.org/i-mutant/i-mutant2.0.html)

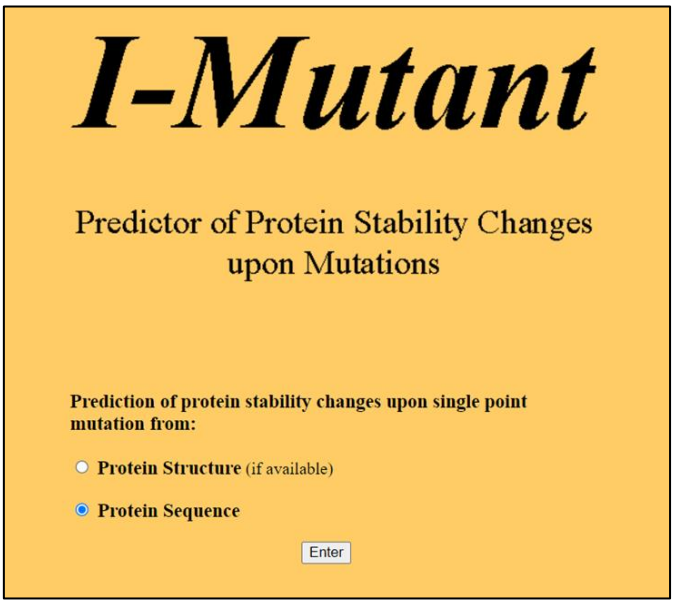


Figure 27: Interface of I-Mutant 2.0 software.

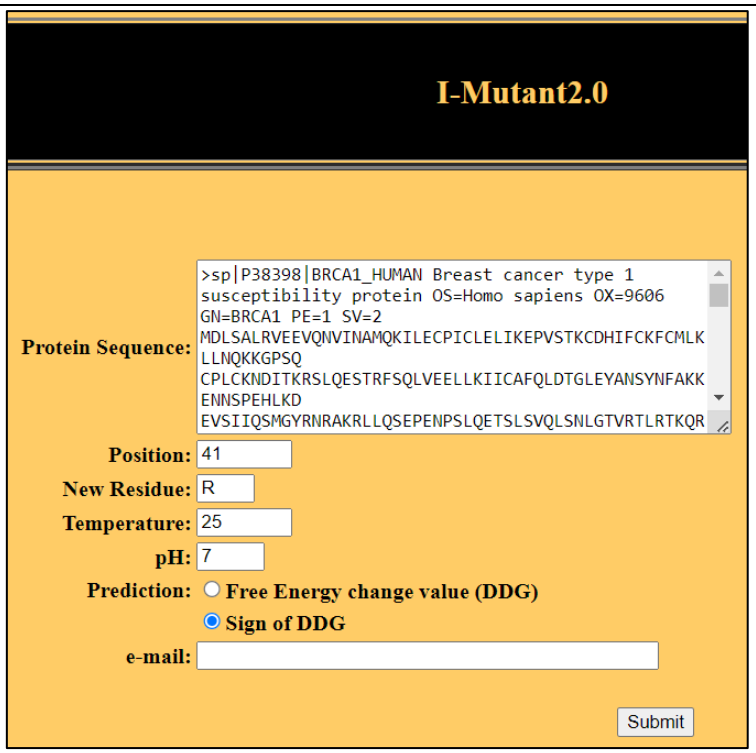


Figure 28: Input data of I mutant 2.0 software.

5.2 Project HOPE software

HOPE (Have (y) Our Protein Explained) collects structural information from a series of sources, including calculations on the 3D protein structure, sequence annotations in UniProt and prediction from the Report software. HOPE combines this information to give analyse the effect of a certain mutation on the protein structure. HOPE is an online web service where the user can submit a sequence and mutation. HOPE will show the effect of that mutation in such a way that even those without a bioinformatics background can understand it (81). (https://www3.cmbi.umcn.nl/hope/)

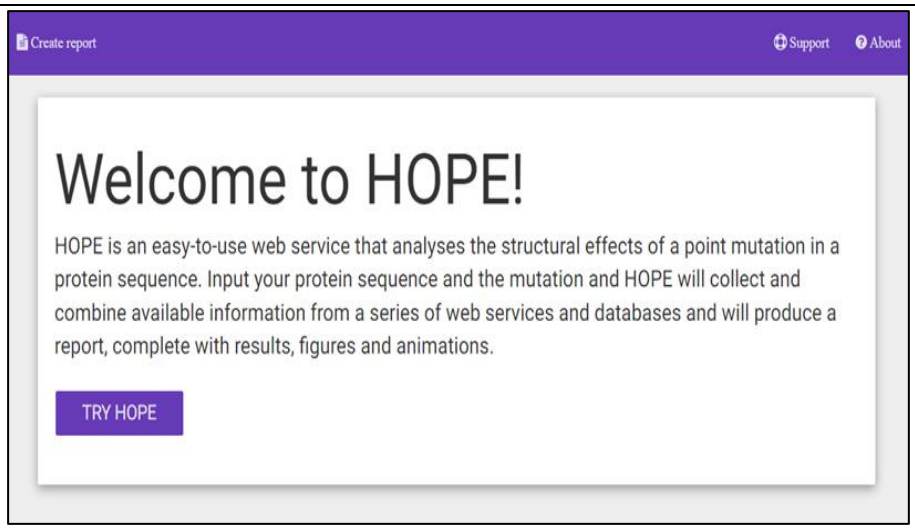


Figure 29: Project HOPE interface.

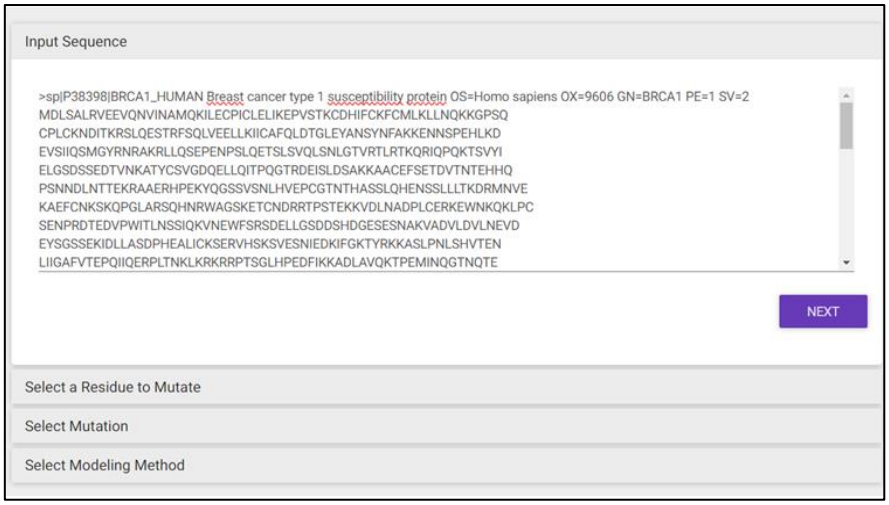


Figure 30: Input data of Project HOPE.

5.3 Align GVGD software

Abbreviations Align-GVGD: Grantham Variation (GV) which measures the degree of biochemical variation among amino acids found at a given position in the multiple sequence alignment, Grantham Deviation (GD), which reflects the 'biochemical distance' of the mutant amino acid from the observed amino acid at a particular position (82) (http://agvgd.hci.utah.edu/agvgd_input.php).

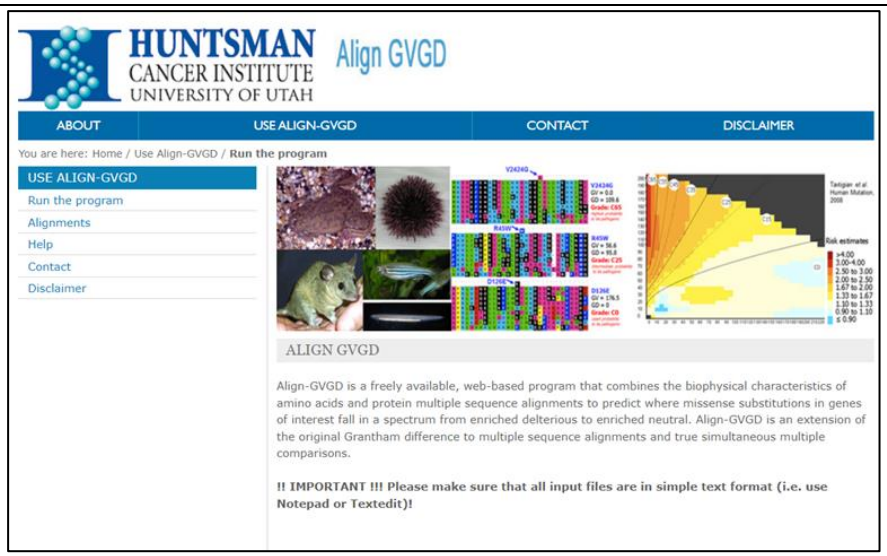


Figure 31: Align GVGD software interface.

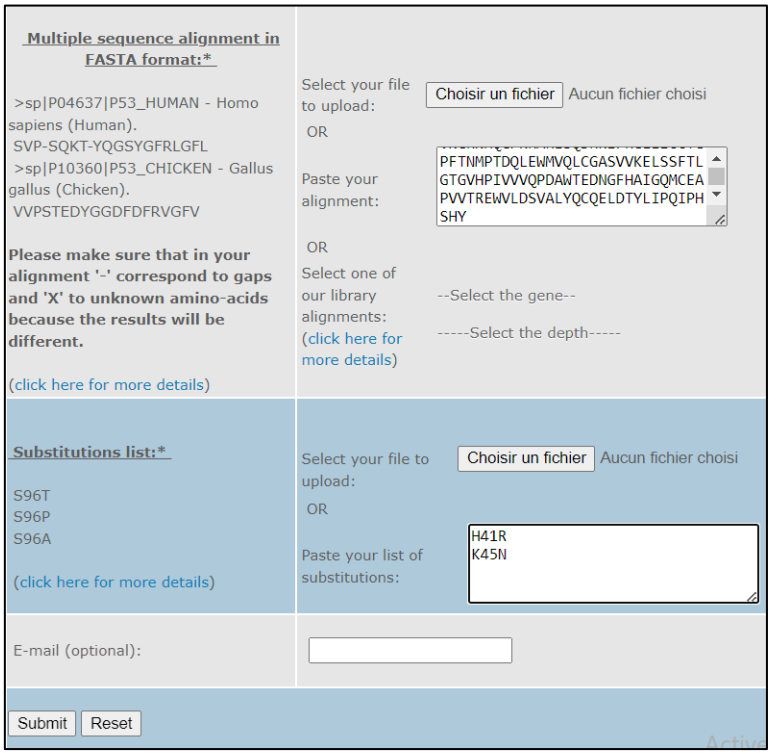


Figure 32: Input data of Align GVGD software.

5.4 PolyPhen-2 software

Polymorphism Phenotyping v2 is a tool which predicts possible impact of an amino acid substitution on the structure and function of a human protein using straightforward physical and comparative considerations. Please, use the form below to submit your query (83) (http://genetics.bwh.harvard.edu/pph2/)

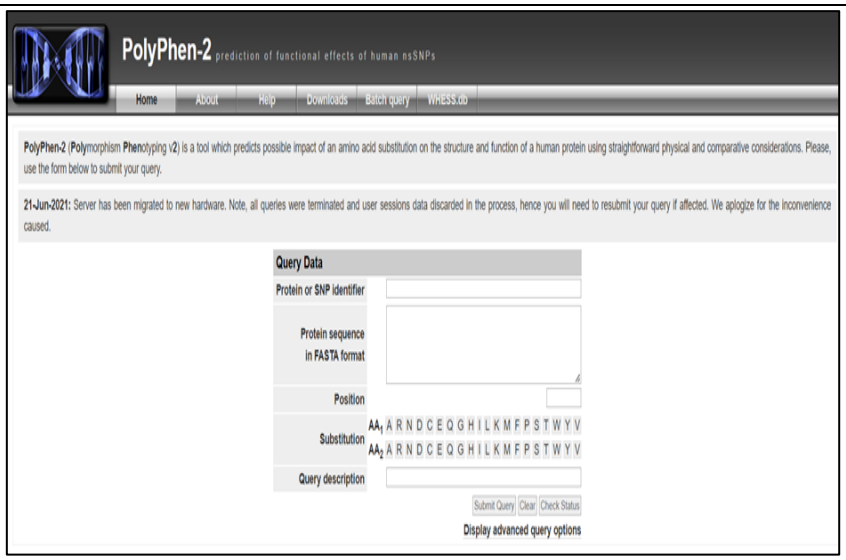


Figure 33: PolyPhen-2 software interface.

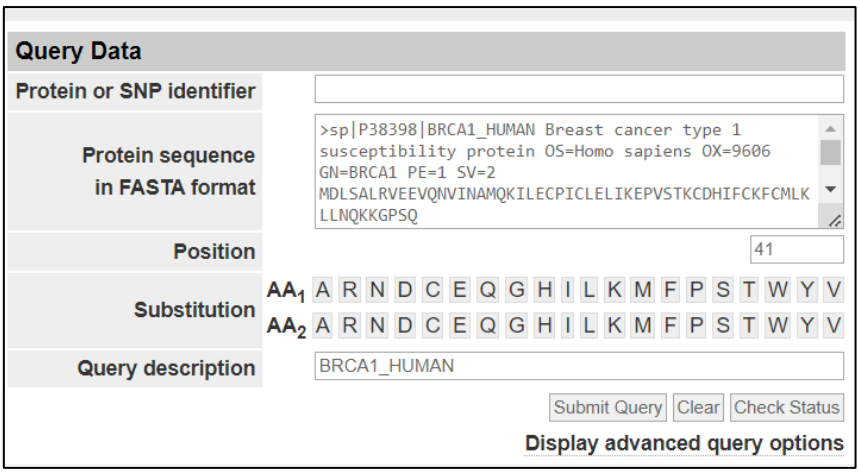


Figure 34: Input data of PolyPhen-2 software.

5.5 SIFT software

The Sorting Intolerant from Tolerant (SIFT) algorithm predicts the effect of coding variants on protein function. It was first introduced in 2001, with a corresponding website that provides users with predictions on their variants. Since its release, SIFT has become one of the standard tools for characterizing missense variation (84) (https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3394338/).

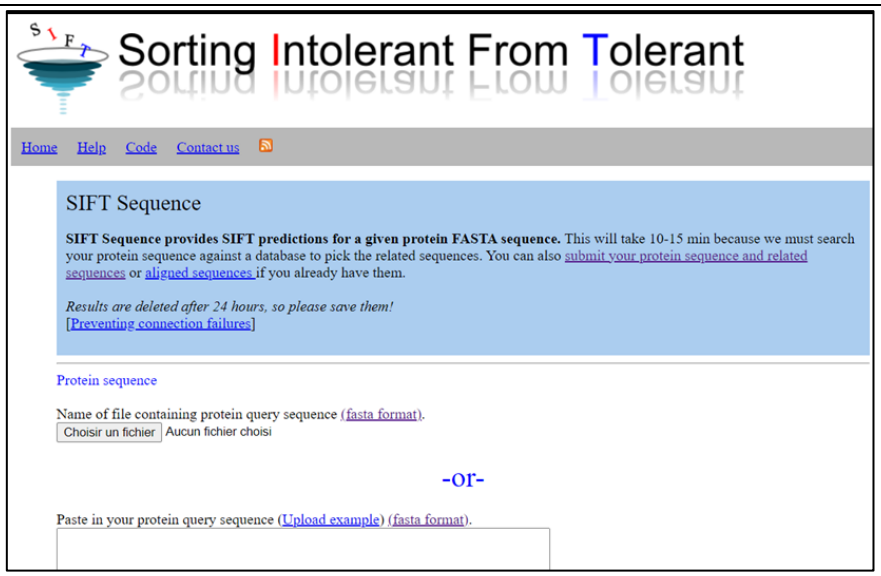


Figure 35: SIFT software interface.

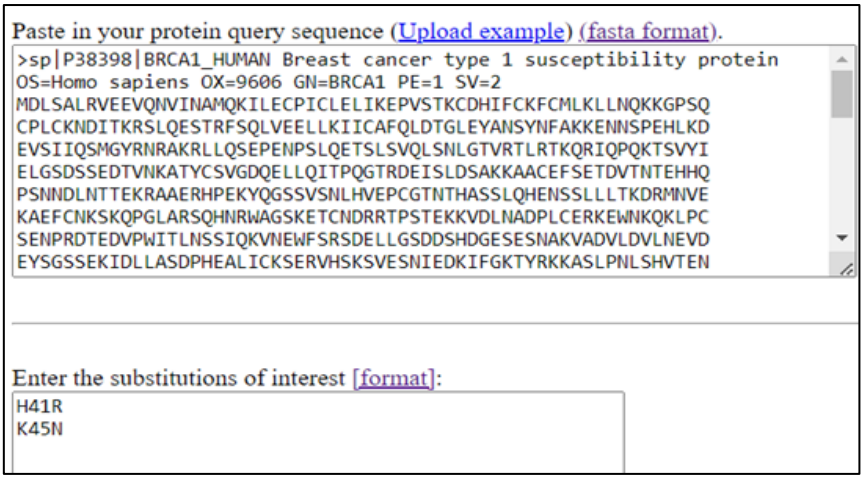


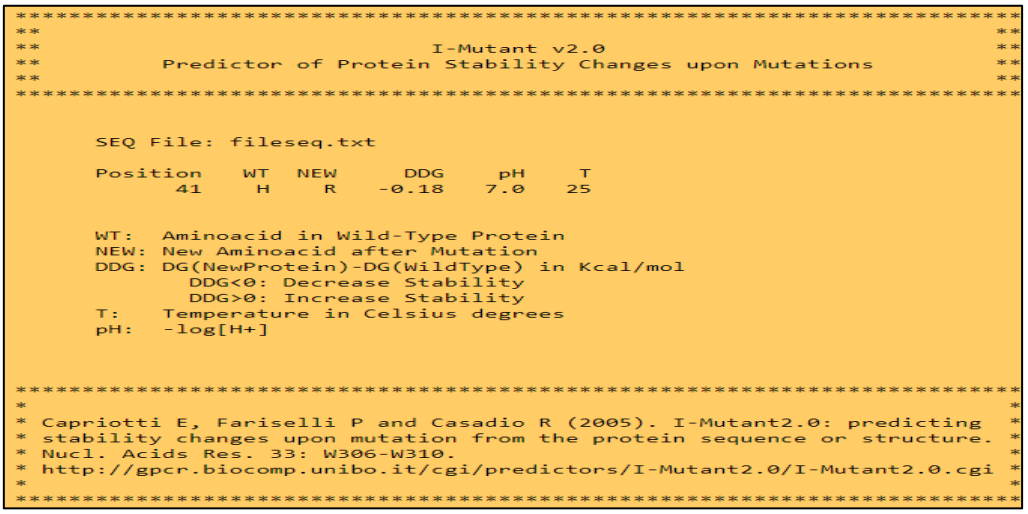
Figure 36: Input data of SIFT software.

RESULT AND DISCUSSION

6 Exonic mutation analysis of (c.122A>G) (p. His41Arg)

6.1 Result obtained by the I-Mutant 2.0 software

Analysis the effects of c.122A>G missense mutation on the stability of protein BRCA1 was predicted by I-Mutant 2.0 software. The result is shown in Figure 36.



```
**************************************************************************
**                                                                      **
**                            I-Mutant v2.0                             **
**          Predictor of Protein Stability Changes upon Mutations       **
**                                                                      **
**************************************************************************

      SEQ File: fileseq.txt

      Position    WT  NEW      DDG     pH     T
            41     H    R    -0.18    7.0    25

      WT:  Aminoacid in Wild-Type Protein
      NEW: New Aminoacid after Mutation
      DDG: DG(NewProtein)-DG(WildType) in Kcal/mol
             DDG<0: Decrease Stability
             DDG>0: Increase Stability
      T:   Temperature in Celsius degrees
      pH:  -log[H+]


**************************************************************************
*                                                                        *
* Capriotti E, Fariselli P and Casadio R (2005). I-Mutant2.0: predicting *
* stability changes upon mutation from the protein sequence or structure.*
* Nucl. Acids Res. 33: W306-W310.                                        *
* http://gpcr.biocomp.unibo.it/cgi/predictors/I-Mutant2.0/I-Mutant2.0.cgi*
*                                                                        *
**************************************************************************
```

Figure 37: Result obtained by the I-Mutant 2.0 software for the mutation (c.122A>G) (p. His41Arg).

The I-Mutant 2.0 software calculated the energy difference (DDG) between normal and mutated protein. The DDG value for this mutation is negative (-0.18) which means that the mutation is destabilizing for the structure of BRCA1 protein.

6.2 Result obtained by Project HOPE

The figure below shows the schematic structures of the original (Histidine) and the mutant (Arginine) amino acid. The backbone, which is the same for each amino acid, is colored red. The side chain, unique for each amino acid, is colored black.

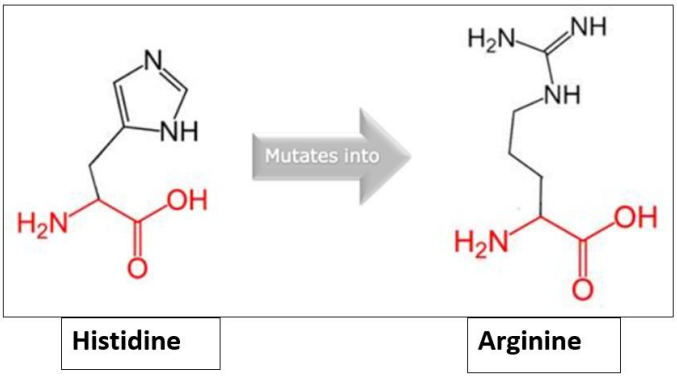


Figure 38: Schematic structures of Histidine and Arginine.

Table 6 : Phisiochemical properties of the residue before and after mutation.

RESIDUE	SIZE	CHARGE
HISTIDINE	155 KD	NEUTRAL
ARGININE	174 KD	POSITIVE

Structure

The differences in amino acid properties disturb a Zinc-finger domain. These type of domains are known to bind DNA that might be disturbed by this mutation.

Variants

We used dbNSFP to link this mutation to a genomic variant. This variant's MetaRNN score is (0.98977375). It can range from (0.0) to (1.0) the higher the score is the more likely to be pathogenic.

Conservation

The mutant and wild type residue are not very similar. Based on the conservation information this mutation is probably damaging to the protein. Our mutant residue is located near a highly conserved position.

Domains

Interpro Domain	Gene Ontology Term	Broad Gene Ontology Term
Zinc Finger, Ring-Type IPR001841	None	None
Zinc Finger, Ring/Fyve/Phd-Type IPR013083	None	None
Brca1-Associated IPR031099	None	None
Zinc Finger, C3Hc4 Ring-Type IPR018957	Metal Ion Binding GO:0046872	Ion Binding GO:0043167 Binding GO:0005488 Molecular_Function GO:0003674
Zinc Finger, Ring-Type, Conserved Site IPR017907	None	None

Figure 39: Domains affected by (c.122A>G) (p. His41Arg) mutation.

The mutated residue is located in a domain that is important for binding of other molecules. Mutation of the residue might disturb this function.

Amino Acid Properties

The difference in charge between the wild type and mutant amino acid .The mutation introduces a charge, which can cause repulsion of ligands or other residues with the same charge.

The wild type and mutant amino acids differ in size and since mutant residue is bigger, it might lead to bumps.

6.3 Result obtained by Align GVGD

The Align GVGD software was also used to predict the effects of c.122 A>G mutation on the BRCA1 protein.

The following classifiers, ordered from most likely to interfere with function to least likely, were used:

GD>=65+Tan (10) x (GV^2.5) => Class C65 <=> most likely

GD>=55+Tan (10) x (GV^2.0) => Class C55

GD>=45+Tan (15) x (GV^1.7) => Class C45

GD>=35+Tan (50) x (GV^1.1) => Class C35

GD>=25+Tan (55) x (GV^0.95) => Class C25

GD>=15+Tan (75) x (GV^0.6) => Class C15

Else (GD<15+Tan (75) x (GV^0.6)) => Class C0 <=> less likely

Table 7 : Result obtained by Align software of GV, GD values and the class for (c.122A>G) (p. His41Arg) mutation.

Substitution	GV	GD	Prediction
H41R	0.00	28.82	Class C25

The value of the variation of Grantham GV is equal to (0.00), which means that this substitution occurs at the level of a highly conserved region between species analyzed.

On the other hand, this same software allowed us to calculate the Grantham GD gap between wild type and mutant amino acid. The GD value is equal to (28.82) which means that the physicochemical properties between Histidine and Arginine are moderately preserved.

Also this software predicted that the class C25 indicating that the substitution does not affect the function of the protein.

6.4 Result obtained by polyphen-2 software

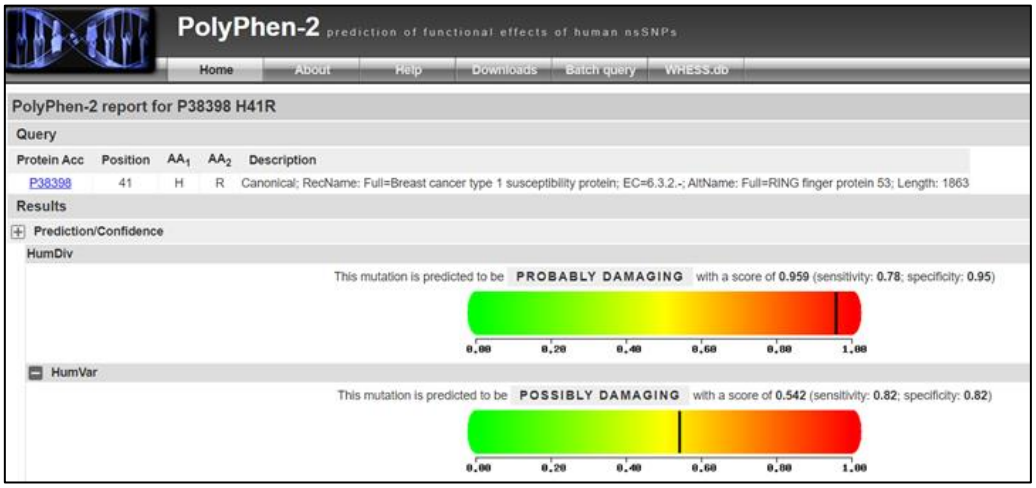


Figure 40 : Result obtained by polyphen-2 of the c.122A>G mutation.

Result and discussion

The prediction of mutation in polyphen-2 software was divided into three categories: benign probably damaging, benign possible damaging and by two models HimDiv and HimVar. The models are confirmed by a score (specificity and sensitivity) to measure the results of BRCA1 protein, which are expressed in the table. In this case HIS 41 ARG the results show that the mutation predicated to be probably domaging with score of (0,959) indicated that vareints is likely to affect the BRCA1 protein function and structure.

Table 8 : The result obtained by polyphen-2 software of HimDiv and HimVar values.

	Predaction	**Score**	**sensitivity**	**specificity**
HimDIV	Probably damaging	0,959	0,78	0,95
HimVar	Possible damaging	0,542	0,82	0,82

This software also made it possible to obtain the multiple alignment result by comparison of the BRCA1 protein with all species recorded on UniProtKB, which allowed us to confirm that the mutation is located in a conserved region between several species.

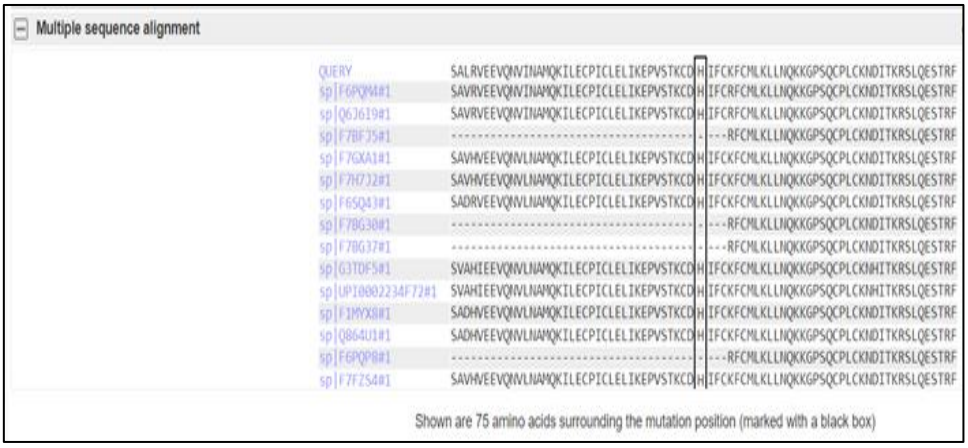


Figure 41: Result obtained by polyphen-2 software of the comparison between multiple sequence alignment from different species to (c.122A>G) (p. His41Arg) mutation.

6.5 Result obtained by SIFT software

Predict Not Tolerated	Position	Seq Rep	Predict Tolerated
ywvtsrqpnmlkigfedca	41H	0.94	H
whgmeqdfkpCYaLRNVTS	42I	0.94	I
hnkrqdgepctsamviwLY	43F	0.94	F
ywvtsrqpnmlkihgfeda	44C	0.94	C

Figure 42: Result obtained by SIFT software for (c.122A>G) (p. His41Arg) mutation.

The AA are colored according to their physicochemical properties:

- red :for Basic AAs
- green : for uncharged polar AAs
- black :for non-polar AAs
- blue : for acidic AAs

Mutation analysis shows that no amino acid can replace Histidine at position 41. The replacement of Histidine by Arginine is intolerant which means that the mutation could impact the BRCA1 protein function.

Probabilities

- Each row corresponds to a position in the reference protein. Below each position is the fraction of sequences that contain one of the basic amino acids. A low fraction indicates the position is either severely gapped or unalignable and has little information. Expect poor prediction at these positions.
- Each column corresponds to one of the twenty amino acids.
- Each entry contains the score at a particular position (row) for an amino acid substitution (column). Substitutions predicted to be intolerant are highlighted in red.

pos	A	C	D	E	F	G	H	I	K	L	M	N	P	Q	R	S	T	V	W	Y
26I 0.94	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.03	0.00	0.00
27C 0.94	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
28L 0.94	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.01	1.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
29E 0.94	0.00	0.00	0.04	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.01	0.00	0.00	0.00
30L 0.94	0.00	0.00	0.00	0.00	0.03	0.00	0.00	0.01	0.00	1.00	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.01
31I 0.94	0.00	0.00	0.00	0.00	0.02	0.00	0.00	1.00	0.00	0.07	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
32K 0.94	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.03	0.02	0.00	0.01	0.00	0.00	0.00
33E 0.94	0.01	0.00	0.01	1.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00
34P 0.94	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.01	0.01	0.00	0.00	0.00
35V 0.94	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.03	0.00	0.01	0.01	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00
36S 0.94	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.01	0.00	0.00	0.00	0.00	1.00	0.03	0.00	0.00	0.00	0.00
37T 0.94	0.01	0.00	0.00	0.00	0.00	0.00	0.00	0.02	0.00	0.02	0.00	0.00	0.00	0.00	0.00	1.00	0.03	0.00	0.00	0.00
38K 0.94	0.01	0.00	0.04	0.03	0.00	0.01	0.01	0.00	1.00	0.00	0.00	0.00	0.01	0.26	0.01	0.01	0.00	0.00	0.00	0.00
39C 0.94	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
40D 0.94	0.01	0.00	1.00	0.01	0.00	0.07	0.00	0.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00
41H 0.94	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00
42I 0.94	0.00	0.00	0.00	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.01	0.00	0.00	0.00	0.00	0.04	0.03	0.01	0.00
43F 0.94	0.00	0.00	0.00	0.00	1.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.00	0.01

Figure 43: Score values between normale and mutated amino acids (c.122A>G) (p. His41Arg)

Predictions

Substitution at pos 41 from H to R is predicted to AFFECT PROTEIN FUNCTION with a score of 0.00.

Median sequence conservation: 3.59

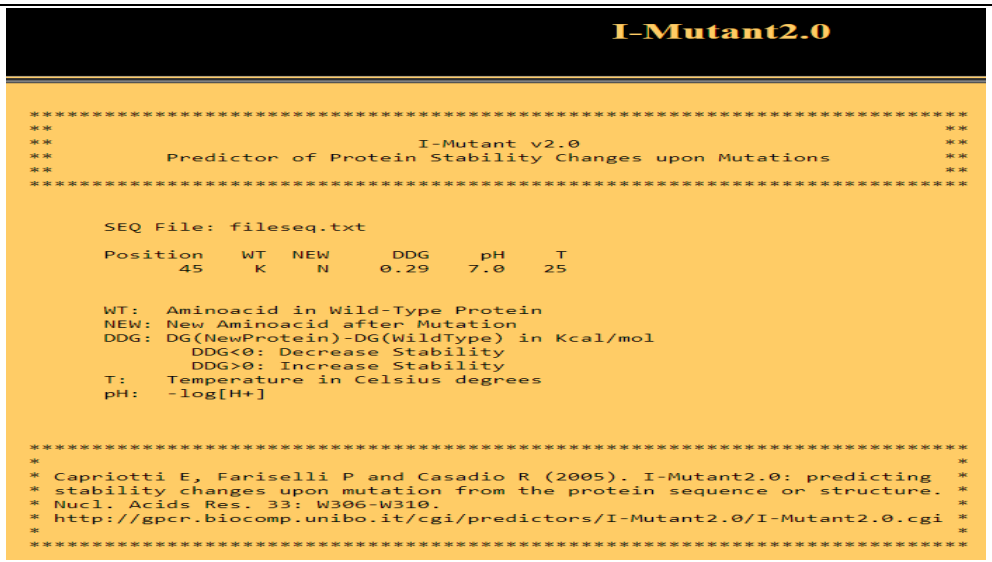
Sequences represented at this position: 97

This substitution may have been predicted to affect function just because the sequences used were not diverse enough. There is LOW CONFIDENCE in this prediction.

7 Exonic mutation analysis of (c.135G>C) (p. Lys45Asn)

7.1 Result obtained by the I-Mutant 2.0 software

Analysis the effects of (c.135G>C) missense mutation on the stability of protein BRCA1 was predicted by I-Mutant 2.0 software. The result is shown in Figure 43.



```
                                                                    I-Mutant2.0


********************************************************************************
**                                                                            **
**                              I-Mutant v2.0                                 **
**          Predictor of Protein Stability Changes upon Mutations             **
**                                                                            **
********************************************************************************

       SEQ File: fileseq.txt

       Position    WT   NEW      DDG     pH     T
             45     K     N     0.29    7.0    25


       WT:  Aminoacid in Wild-Type Protein
       NEW: New Aminoacid after Mutation
       DDG: DG(NewProtein)-DG(WildType) in Kcal/mol
              DDG<0: Decrease Stability
              DDG>0: Increase Stability
       T:   Temperature in Celsius degrees
       pH:  -log[H+]


********************************************************************************
*                                                                              *
* Capriotti E, Fariselli P and Casadio R (2005). I-Mutant2.0: predicting       *
* stability changes upon mutation from the protein sequence or structure.      *
* Nucl. Acids Res. 33: W306-W310.                                              *
* http://gpcr.biocomp.unibo.it/cgi/predictors/I-Mutant2.0/I-Mutant2.0.cgi      *
*                                                                              *
********************************************************************************
```

Figure 44: Result obtained by I mutant about (c.135G>C) (p. Lys45Asn) mutation.

The I-Mutant 2.0 software made it possible to calculate the energy difference (DDG) between the normal and mutated protein. The DDG value calculated for this mutation is equal (0.29). Since this value is positive, the mutation does not affect the stability of the BRCA1 protein structure.

7.2 Result obtained by Project HOPE software

The figure below shows the schematic structures of the original (Lysine) and the mutant (Asparagine) amino acid. The backbone, which is the same for each amino acid, is colored red. The side chain, unique for each amino acid, is colored black.

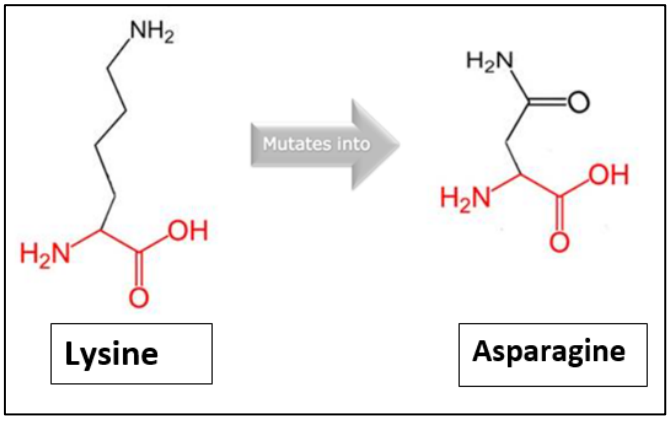


Figure 45: Schematic structures of Lysine and Asparagine.

Table 9 : Phisiochemical properties of the residue LYSINE after becoming ASPARAGINE.

RESUDUE	SIZE	CHARGE
LYSINE	146 DA	POSITIVE
ASPARAGINE	132 DA	Neutral

Structure

The differences in amino acid properties disturb a Zinc-finger domain. These type of domains are known to bind DNA, which might be disturbed by this mutation.

Variants

We used dbNSFP to link this mutation to a genomic variant. This variant's MetaRNN score is (0.6055749). It can range from 0.0 to 1.0. The higher, the more likely it is to be pathogenic.

Conservation

The wild-type residue is very conserved, but a few other residue types have been observed at this position too. Our mutant residue was among the residues at this position observed in other sequences. This means that homologous proteins exist with the same residue type as our mutant at this position and this mutation is possibly not damaging to the protein. The mutant residue is located near a highly conserved position.

Domains

Interpro Domain	Gene Ontology Term	Broad Gene Ontology Term
Zinc Finger, Ring-Type IPR001841	None	None
Zinc Finger, Ring/Fyve/Phd-Type IPR013083	None	None
Brca1-Associated IPR031099	None	None
Zinc Finger, C3Hc4 Ring-Type IPR018957	Metal Ion Binding GO:0046872	Ion Binding GO:0043167 Binding GO:0005488 Molecular_Function GO:0003674
Zinc Finger, Ring-Type, Conserved Site IPR017907	None	None

Figure 46 Domains affected by (c.135G>C) (p. Lys45Asn) mutation.

The mutated residue is located in a domain that is important for binding of other molecules. Mutation of the residue might disturb this function.

Amino Acid Properties

- There is a difference in charge between the wild type and mutant amino acid.
- The charge of the wild-type residue will be lost; this can cause loss of interactions with other molecules or residues.
- The wild type and mutant amino acids differ in size.
- The mutant residue is smaller; this might lead to loss of interactions.

7.3 Result obtained by Align GVGD software

The Align GVGD software was also used to predict the effects ofc.122 A>G mutation on the BRCA1 protein.

The following classifiers, ordered from most likely to interfere with function to least likely, were used:

GD>=65+Tan (10) x (GV^2.5) => Class C65 <=> most likely

GD>=55+Tan (10) x (GV^2.0) => Class C55

GD>=45+Tan (15) x (GV^1.7) => Class C45

GD>=35+Tan (50) x (GV^1.1) => Class C35

GD>=25+Tan (55) x (GV^0.95) => Class C25

GD>=15+Tan (75) x (GV^0.6) => Class C15

Else (GD<15+Tan (75) x (GV^0.6)) => Class C0 <=> less likely

Table 10 : Result obtained by Align software of GV, GD values and the class for (c.135G>C) (p. Lys45Asn) mutation.

Substitution	GV	GD	Prediction
K45N	0.00	93.88	Class C65

Analysis of the c.122A>G by Align GVGD mutation made it possible to calculate the values GV and GD. The value of the variation of Grantham GV is equal (0.00), which means that this substitution occurs at the level of a highly conserved region between species analyzed. On the other hand, this same software allowed us to calculate the Grantham GD gap between wild type and mutant amino acid. The GD value is equal to (93.88) which means that

the physicochemical properties between Lysine and Asparagine are moderately preserved. Also this software predicted that the class C65 indicating that the substitution does affect the function of the protein.

7.4 Result obtained by polyphen-2 software

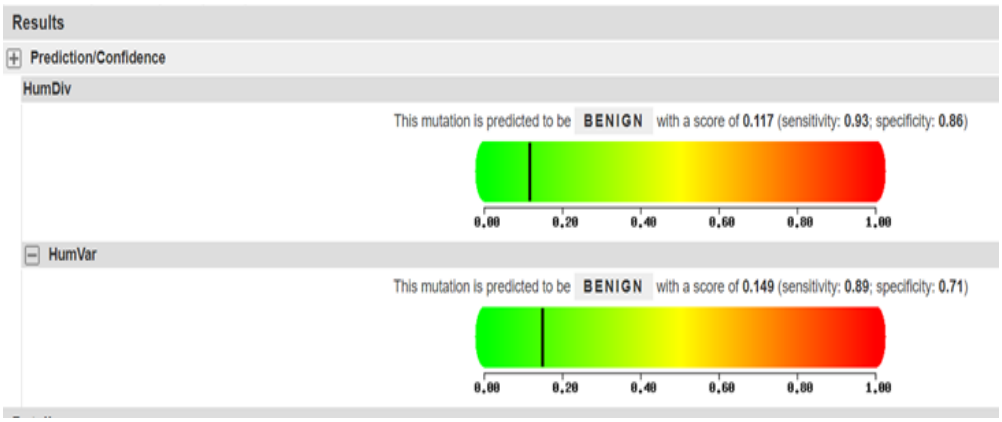


Figure 47: Result obtained by polyphen-2 of (c.135G>C) (p. Lys45Asn) mutation.

The c.135G>C (p.Lys45Asn) mutation was predicted by PolyPhen-2 software as benign. This means that it has no effect on the structure and function of the protein.

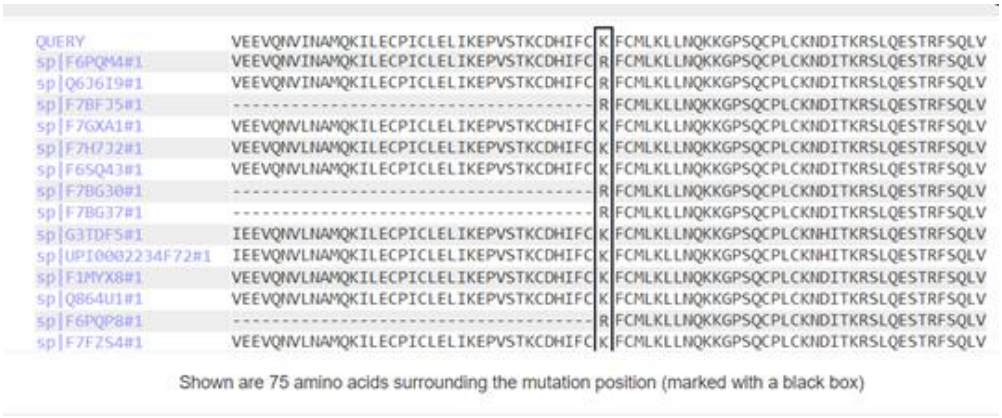


Figure 48: Result obtained by polyphen-2 software of the comparison between multiple sequence alignment from different species to (c.135G>C) (p. Lys45Asn) mutation.

This software also made it possible to obtain the multiple alignment result by comparison of the protein BRCA1 with all species recorded on UniProtKB, which allowed us to confirm that the mutation does not sit in a conserved region between several species.

Table 11: The result obtained by polyphen-2 software of HimDiv and HimVar values.

	Predaction	Score	sensitivity	specificity
HimDIV	Benign	0,117	0,93	0,86
HimVar	Benign	0,149	0,89	0,71

7.5 Result obtained by SIFT software

Predictions for positions 1 through 100

Threshold for intolerance is 0.05.

Amino acid color code: nonpolar, uncharged polar, basic, acidic.

Capital letters indicate amino acids appearing in the alignment, lower case letters result from prediction.

'Seq Rep' is the fraction of sequences that< contain one of the basic amino acids. A low fraction indicates the position is either severely gapped or unalignable and has little information. Expect poor prediction at these positions.

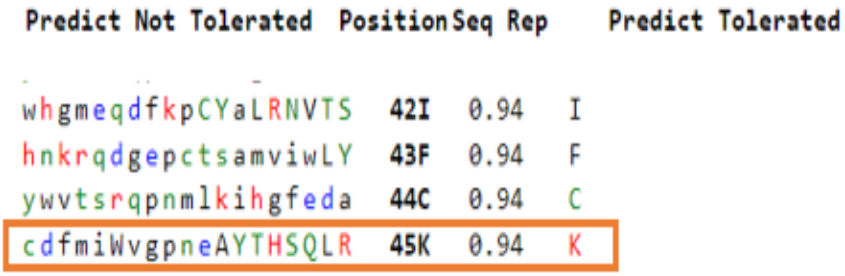
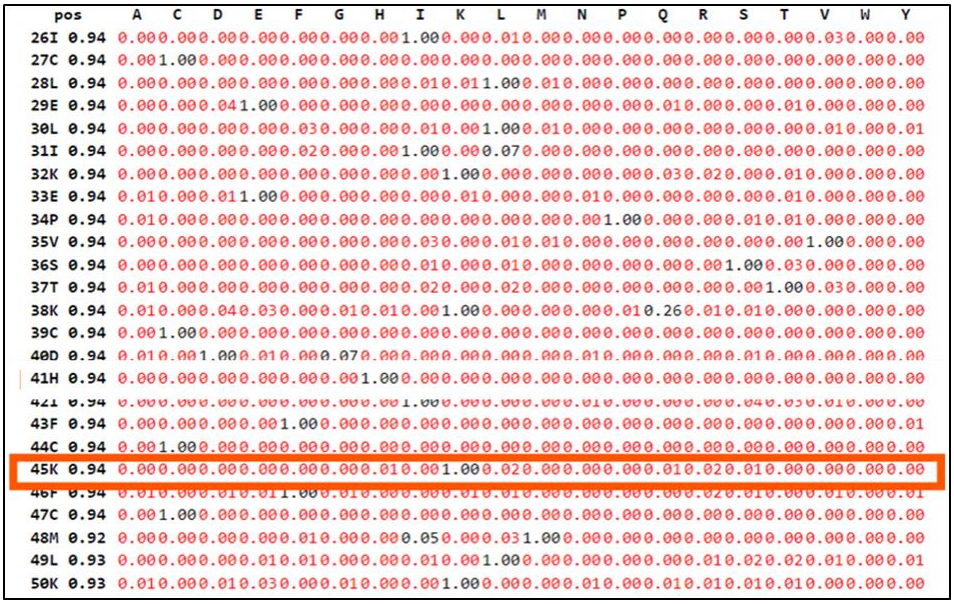


Figure 49: Result obtained by SIFT software for (c.135G>C) (p. Lys45Asn) mutation.

Probabilities

- Each row corresponds to a position in the reference protein. Below each position is the fraction of sequences that contain one of the basic amino acids. A low fraction indicates the position is either severely gapped or unalignable and has little information. Expect poor prediction at these positions.
- Each column corresponds to one of the twenty amino acids.

- Each entry contains the score at a particular position (row) for an amino acid substitution (column). Substitutions predicted to be intolerant are highlighted in red.



pos	A	C	D	E	F	G	H	I	K	L	M	N	P	Q	R	S	T	V	W	Y
26I 0.94	0.000	.000	.000	.000	.000	.000	.001	.000	.000	.010	.000	.000	.000	.000	.000	.000	.000	.030	.000	.00
27C 0.94	0.001	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.00
28L 0.94	0.000	.000	.000	.000	.000	.000	.000	.010	.011	.000	.010	.000	.000	.000	.000	.000	.000	.000	.000	.00
29E 0.94	0.000	.000	.041	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.010	.000	.000	.010	.000	.000	.00
30L 0.94	0.000	.000	.000	.000	.030	.000	.000	.010	.001	.000	.010	.000	.000	.000	.000	.000	.000	.010	.000	.01
31I 0.94	0.000	.000	.000	.000	.020	.000	.001	.000	.000	.070	.000	.000	.000	.000	.000	.000	.000	.000	.000	.00
32K 0.94	0.000	.000	.000	.000	.000	.000	.000	.001	.000	.000	.000	.000	.000	.030	.020	.000	.010	.000	.000	.00
33E 0.94	0.010	.000	.011	.000	.000	.000	.000	.000	.010	.000	.000	.010	.000	.000	.000	.000	.010	.000	.000	.00
34P 0.94	0.010	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.001	.000	.000	.010	.010	.000	.000	.00	
35V 0.94	0.000	.000	.000	.000	.000	.000	.000	.030	.000	.010	.010	.000	.000	.000	.000	.000	.001	.000	.000	
36S 0.94	0.000	.000	.000	.000	.000	.000	.000	.000	.010	.000	.010	.000	.000	.000	.000	.001	.000	.030	.000	.00
37T 0.94	0.010	.000	.000	.000	.000	.000	.000	.020	.000	.020	.000	.000	.000	.000	.000	.001	.000	.030	.000	.00
38K 0.94	0.010	.000	.040	.030	.000	.010	.010	.001	.000	.000	.000	.000	.010	.260	.010	.010	.000	.000	.000	.00
39C 0.94	0.001	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.00
40D 0.94	0.010	.001	.000	.010	.000	.070	.000	.000	.000	.000	.000	.010	.000	.000	.000	.010	.000	.000	.000	.00
41H 0.94	0.000	.000	.000	.000	.000	.001	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.00
42I 0.94	0.000	.000	.000	.000	.000	.000	.001	.000	.000	.000	.000	.010	.000	.000	.000	.040	.030	.010	.000	.00
43F 0.94	0.000	.000	.000	.001	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.01
44C 0.94	0.001	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.00
45K 0.94	0.000	.000	.000	.000	.000	.000	.010	.001	.000	.020	.000	.000	.000	.010	.020	.010	.000	.000	.000	.00
46F 0.94	0.010	.000	.010	.011	.000	.000	.010	.010	.000	.000	.000	.000	.000	.000	.020	.010	.000	.010	.000	.01
47C 0.94	0.001	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.000	.00
48M 0.92	0.000	.000	.000	.000	.010	.000	.000	.050	.000	.031	.000	.000	.000	.000	.000	.000	.000	.000	.000	.00
49L 0.93	0.000	.000	.000	.010	.010	.000	.000	.010	.001	.000	.000	.000	.000	.000	.010	.020	.020	.010	.000	.01
50K 0.93	0.010	.000	.010	.030	.000	.010	.000	.001	.000	.000	.000	.010	.000	.010	.010	.010	.010	.000	.000	.00

Figure 50: Score values between normale and mutated amino acids (c.135G>C) (p. Lys45Asn).

Predictions

The substitution at pos 45 from K to N is predicted to AFFECT PROTEIN FUNCTION with a score of (0.00).

Median sequence conservation: 3.59

Sequences represented at this position: 97

8 Discussion

8.1 Discussion of (c.122A>G) (p. His41Arg) mutation results

The c.122A>G mutation acts on the coding region of this gene and consists of a substitution of a Histidine by an Arginine at codon 41 thus affecting the protein synthesized. The results obtained by the I Mutant 2.0 software are in favore to tell us that the protein is destabilized. Also Polyphen2 software predicted that the substitution is Probably damaging Likewise, software SIFT predicted the substitution effects on the protein fonction. However, the result obtained by ALIGN GVGD shows that this variation didn't affect the protein since the class 25 is less likely to interfere with function. Last but not least, in Software hope, the substitious shows changes in the physicochemical properties of the two amino acids. The

amino acid Histidine is smaller than Arginine. These differences could lead to the loss or gain of interactions with other molecules, which change the structural level of the protein.

8.2 Discussion of (c.135G>C) (p. Lys45Asn) mutation results

Similarly, the c.135G>C mutation the results obtained by I mutant software predict that the BRCA1 protein is stable in relation to the DDG value and is not destabilizing, even by polyphen-2 software which shows that the change is benign. At the other cost the two programs SIFT and align GVGD are among of the mutation affecte the structure and function of the protein. However, in HOPE software the differences in chemical properties of amino acids explain the effect of mutations, concerning the size, the mutated residue Asparagine is a little bit smaller than the wild residue Lysine, The charge in the mutated residue is neutral while the wild one is positive. These differences disrupt the protein BRCA1.

8.3 General discussion

Using all this software helped us to collect more information about our mutated gene BRCA1 since each software have its own properties. Starting with I-Mutant 2.0 software it inform us about what if our mutated protein is destabilized or not while project HOPE tell us the effect of the mutation on the amino acids and its interactions. However, ALIGN GVGD informed us wether the protein interfere with function by using GV and GD values same to SIFT software which rate the mutation by scores. Last but not least, polyphen-2 software predict for us if the mutation will be damaging or benign also give us a comparison between multiple sequence alignment from different species to the mutation.

Both substitutes are located in the Zinc Finger Domain region. This pattern is part of a wider range containing antiparallel alpha propellers essential for protein interactions; this domain is involved in the ubiquitination process; the large number of BRCA1 mutations occur in N-terminal RING domain, which is responsible for the E3 ubiquitin ligase activity of BRCA1.

A growing body of evidence suggests that mutations in the BRCA1 RING domain inactivate BRCA1 ubiquitin protein ligase activity, may predispose to cancer development. Study of Nelson and Holt showed that BRCA1 protein with mutation in the RING domain fails to co-localize with BRCA1-interacting proteins BARD1 and BACH1, which are essential for DNA repair. Another study by Hashizume et al. showed that BRCA1-BARD1 heterodimeric complex can be inactivated by RING finger mutation, which results in the loss of ubiquitin ligase activity.

Taken together, the high rate of cancer-related mutations in specific domains of BRCA1 such as RING might explain the critical role of these domains for BRCA1 tumor suppressor activity (85).

Many researchers in various institutes, using different bioinformatics tools based on the two mutations that we previously studied, have reached similar results. For example, the National Library of Medicine found the same result with ClinVAR software . The BRCA1 (c.122A>G) (p.His41Arg) RING-domain variant was found from multifactorial likelihood analysis to have a posterior probability of pathogenicity of (0.995), a result consistent with existing protein functional assay data indicating lost BARD1 binding and ubiquitin ligase activity. Of the remaining variants, seven were determined to be not clinically significant (Class 1) (86).

In another study for sequence variants used for mammalian two-hybrid assay calibration, the BRCA1 (c.122A>G) (p.His41Arg mutation) was which we found the value of the probability of low activity to be (0.870). The vast majority of known pathogenic variants in BRCA1 are ultimately, either protein-truncating variants or large gene rearrangements. Nonetheless, about 10% of known pathogenic variants are either non-spliceogenic missense substitutions or in-frame insertion or deletion mutations (IFIs, IFDs) (87).

As for the (c.135G>C) (p. Lys45Asn) mutation there were also results; RAD51 is an important component of double-stranded DNA–repair mechanisms that interacts with both BRCA1 and BRCA2. A single-nucleotide polymorphism (SNP) in the 5′ untranslated region (UTR) of RAD51, 135G→C, has been suggested as a possible modifier of breast cancer risk in BRCA1 and BRCA2 mutation carriers. We pooled genotype data for 8,512 female mutation carriers from 19 studies for the RAD51 135G→C SNP (88).

CONCLUSION

9 Conclusion

In this study, we contributed to understand the molecular mechanisms that trigger the mutational status of BRCA1 genes, which is crucial information in defining therapeutic strategies for breast cancer patients. We've been studying the replacement effects of two mutations specific to the Algerian population by using a combination of *in silico* study software. These mutations correspond to two missense mutations at exon 2 (c.122A>G) and (C.135G>C).

The c.122A>G mutation (p.His41Arg) has been predicted as substituted at the structural level of the BRCA1 protein from the energy deference value in this mutation and even from Physico-chemical changes between native and mutant amino acids are responsible for loss of molecular interactions and conformational change having destabilizing effects on the structure of the protein. On the other hand, the variation c.135G>C (p.Lys45Asn) has been predicted as not substituted at the functional level of the protein. But the consequence of this false direction, mutation depends mainly on the position of the muted amino acid and its physiochemical properties on the protein BRCA1 so the functional study is very important for mutations evaluations.

The implementation of our work has demonstrated the effectiveness of the combined use of several predictive software to assess the effect of mutations at the protein level. It is interesting to compare the results of *in silico* studies with those of functional studies, in order to evaluate the characteristics of new mutations. Furthermore, using bioinformatics software predict the effect of mutations on the protein of interest remains an affordable choice, and offers the ability to quickly provide a predictive value to the results of genotyping, making them potentially useful for patient care and scientific research. However, it is necessary to verify them by laboratory studies in vitro/or in vivo and even for the diagnosis, prognosis and development of new therapeutic strategies.

BIBLIOGRAPHIC REFERENCES

Bibliographic references

1. Staff, Mayo Clinic. *Cancer.* s.l. : mayoclinic, 2022.

2. **Charra, Dr Antoine Arnaud Anne-Marie Brossard Dr Clément.** *LES TRAITEMENTS DES CANCERS DU SEIN.* 2013.

3. **[Online] https://cancer.ca/fr/cancer-information/resources/glossary/b/brca-gene-mutations.**

4. ***Breast cancer.*** **s.l. : world health organization, 2022. 1.**

5. ***Epidemiology of breast cancer in women based on diagnosis data from oncologists and senologists in Algeria.*** **F Smaili, A Boudjella,A Dib,S Braikia,H Zidane,R Reggad,A Bendib,A Abdelouahab,F Bereksi-Reguig,D Yekrou,A Bentouati,T Filali,A Chirouf,N Djehal,H Mahfouf,F Mechiat,H Tadjerouni,C Sedkaoui,M Hikem,A Amber,K Bouzid,A Ladjeroud,S Ghomari,H Aris et al. 100220, s.l. : Elsevier, 10 October 2020, Vol. Volume 25. 2.**

6. Amel Elbasyouni1, Leila Saadi1, 2, AbdelKarim Baha3. ***Epidemiological profile and distribution.*** **Blida 1 University, Algeria : s.n., 2021.**

7. team, The American Cancer Society. ***What Is Breast Cancer?*** **America : The American Cancer Society, 2021. 4.**

8. ***Breast Anatomy.*** **s.l. : medical art library, 2017. 5.**

9. ***breast cancer.*** **s.l. : cleveland clinic, 2023. 6.**

10. ***Breast cancer: Invasive/infiltrating ductal carcinoma (IDC).*** **s.l. : health engine, 07 April 2009. 7.**

11. staff, Mayo Clinic. ***Lobular carcinoma in situ (LCIS).*** **s.l. : mayo clinic, 2022. 8.**

12. team, The American Cancer Society. ***Ductal Carcinoma In Situ (DCIS).*** **America : American Cancer Society. 9.**

13. ***The Potential of Hormonal Therapies for Treatment of Triple-Negative Breast Cancer.*** **Wang, Melanie Kirkby Alyanna M. Popatia Jessie Lavoie Lisheng. 10.3390/cancers15194702, 24 September 2023, Vol. 15. 10.**

14. ***Breast Cancer Risk Factors.*** **s.l. : cdc, 2024. 11.**

15. ***Symptoms of breast cancer.*** **s.l. : cancer research uk, 2023. 12.**

16. ***Diagnosis of breast cancer.*** **s.l. : Canadian Cancer Society. 13.**

17. Dr.Samanthi. ***What is the Difference Between Screening Mammogram and Diagnostic Mammogram.*** **s.l. : differencebetweencom, 2022. 14.**

18. ***Breast Ultrasound.*** **s.l. : komen.org, 2022. 15.**

19. ***Automated breast ultrasound (ABUS) as a screening tool: initial experience.*** **Amera Abd Elsalam Mostafa, Mohamed Adel Eltomey, Ashraf Mohammed Elaggan & Amel A. Hashish. https://doi.org/10.1186/s43055-019-0032-9, Egypt : Springer Nature SharedIt, 2019. 16.**

20. ***US-guided Fine-Needle Aspiration of Thyroid Nodules: Indications, Techniques, Results.*** **Park, Min Jung Kim Eun-Kyung Kim S. I. Park Byung Moon Kim J. Kwak S. Kim J. Youk Sung Hee. 7, s.l. : RSNA, 1 November 2008, Vol. 28. 17.**

21. ***Core Biopsy.*** **s.l. : Virtual Medical Centre, 2015. 18.**

Bibliographic references

22. ***Stereotactic breast biopsy.*** **s.l. : Mayo Clinic. 19.**

23. ***Punch biopsy.*** **s.l. : mayoclinic. 21.**

24. ***sentinel lymph node biopsy.*** **s.l. : national cancer institute, 2022. 22.**

25. ***Tests for breast cancer.*** **s.l. : cancerresearchuk.org, 2023. 23.**

26. ***Breast MRI.*** **s.l. : cancer.org, 2022. 24.**

27. ***High sensetivity and specificity to support cancer imaging.*** **Wenkel, Dr Evelyn. Erlangen,Germany : s.n. 25.**

28. ***CT scan.*** **s.l. : National Cancer Institute, 2006. 26.**

29. ***Breast cancer.*** **s.l. : world health organization, 2024. 27.**

30. ***Mutations of the BRCA1 and BRCA2 genes and the possibilities for predictive testing.*** **S A Gayther 1, B A Ponder. 10.1016/S1357-4310(97)01017-4., Cambridge, UK. : National library of medcine, 1997, Vol. 3. 28.**

31. ***Potential interactions between Brca1 or Brca2 gene defects and DES exposure.*** **s.l. : DES DAUGHTER, 2017, Vol. 60. 29.**

32. ***BRCA Gene Mutations: Cancer Risk and Genetic Testing.*** **s.l. : National cancer institute, 2020. 30.**

33. ***A strong Candidate for the Breast and Ovarian Cancer Susceptibility Gene BRCA1.*** **Miki, Y., Swensen, J., Shattuck-Eidens, D., Futreal, P. A., Harshman, K., Tavtigian, S., et al. 10.1126/science.2270482, 1990.**

34. ***A strong Candidate for the Breast and Ovarian Cancer Susceptibility Gene BRCA1.*** **Miki, Y., Swensen, J., Shattuck-Eidens, D., Futreal, P. A., Harshman, K., Tavtigian, S., et al. 10.1126/science.7545954, 1994.**

35. ***Structure-Function of the Tumor Suppressor BRCA1. Comput.*** **Clark, S. L., Rodriguez, A. M., Snyder, R. R., Hankins, G. D. V., and Boehning, D. 10.5936/csbj.201204005, s.l. : Comput. Struct. Biotechnol. J. 1, e201204005, 2012.**

36. ***The RING Heterodimer BRCA1-BARD1 Is a Ubiquitin Ligase Inactivated by a Breast Cancer-Derived Mutation.*** **Hashizume, R., Fukuda, M., Maeda, I., Nishikawa, H., Oyake, D., Yabuki, Y., et al. 10.1074/jbc.c000881200, s.l. : J. Biol. Chem. 276 (18), 14537–14540., 2001.**

37. ***PALB2 Is an Integral Component of the BRCA Complex Required for Homologous Recombination Repair.*** **Sy, S. M., Huen, M. S., and Chen, J. 10.1073/pnas.0811159106, s.l. : Proc. Natl. Acad. Sci. U S A. 106 (17), 7155–7160., 2009.**

38. ***PALB2 Functionally Connects the Breast Cancer Susceptibility Proteins BRCA1 and BRCA2. . .*** **Zhang, F., Fan, Q., Ren, K., and Andreassen, P. R. 10.1158/1541-7786.mcr-09-0123, s.l. : Mol. Cancer Res. 7, 1110–1118., 2009.**

39. ***BRCA1 and Breast Cancer: Molecular Mechanisms and Therapeutic Strategies.*** **Xiaoyu Fu1, 2 Wei Tan3 Qibin Song2 Huadong Pei 4 Juanjuan Li1,2*. china : Frontiers, 2022, Vol. 10. 31.**

40. ***BRCA1 Tumor Suppressor Network: Focusing on its Tail.*** **Wang, B. :10.1186/2045-3701-2-6, s.l. : Cell Biosci 2, 6. , 2012.**

Bibliographic references

41. ***Targeting the DNA Damage Response in Immuno-Oncology: Developments and Opportunities. .*** **Chabanon, F., Lord, J. L., Soria, J.-C., Pasero, P., and Postel-Vinay, S. Nat. Rev. Cancer 21 (11), 701–717., 2021.**

42. ***DNA Damage Response Signaling Pathways and Targets for Radiotherapy Sensitization in Cancer. .*** **Huang, R.-X., and Zhou, P.-K. 10.1038/s41392-020-0150-x, s.l. : Sig Transduct Target. Ther. 5, 60., 2020.**

43. ***RAP80 and BRCA1 PARsylation Protect Chromosome Integrity by Preventing Retention of BRCA1-B/C Complexes in DNA Repair Foci. .*** **Vohhodina, J., Toomire, K. J., Petit, S. A., Micevic, G., Kumari, G., Botchkarev, V. V., et al. s.l. : Proc. Natl. Acad. Sci. USA , 2020.**

44. ***The BRCT Domain Is a Phospho-Protein Binding Domain. .*** **Yu, X., Chini, C. C. S., He, M., Mer, G., and Chen, J. 10.1126/science.1088753, s.l. : Science 302 (5645), 639–642, 2003.**

45. ***PALB2 Is an Integral Component of the BRCA Complex Required for Homologous Recombination Repair.*** **Sy, S. M., Huen, M. S., and Chen, J. 10.1073/pnas.0811159106, s.l. : Proc. Natl. Acad. Sci. U S A. 106 (17), 7155–7160., 2009.**

46. ***The Tumor Suppressor PALB2: Inside Out. .*** **Ducy, M., Sesma-Sanz, L., Guitton-Sert, L., Lashgari, A., Gao, Y., Brahiti, N., et al. 10.1016/j.tibs.2018.10.008, s.l. : Trends Biochem. Sci. 44 (3), 226–240. , 2019.**

47. ***Interplay between BRCA1 and GADD45A and Its Potential for Nucleotide Excision Repair in Breast Cancer Pathogenesis.*** **Pietrasik, & Zajac, & Morawiec, & Soszynski, & Fila, Michał & Blasiak. 21. 870. 10.3390/ijms21030870., s.l. : International Journal of Molecular Sciences, 2020.**

48. ***Schlacher, K., Wu, H., and Jasin, M. (2012). A Distinct Replication fork protection Pathway Connects Fanconi Anemia Tumor Suppressors to RAD51-Brca1/2. Cancer Cell 22, 106–116. doi:10.1016/j.ccr.2012.05.015. 2012.***

49. Long, D. T., Joukov, V., Budzowska, M., and Walter, J. C. (2014). BRCA1 Promotes Unloading of the CMG Helicase from a Stalled DNA Replication fork. Mol. Cel 56, 174–185. doi:10.1016/j.molcel.2014.08.012. 2014.

50. Petermann, E., Orta, M. L., Issaeva, N., Schultz, N., and Helleday, T. (2010). Hydroxyurea-stalled Replication forks Become Progressively Inactivated and Require Two Different RAD51-Mediated Pathways for Restart and Repair. Mol. Cel 37, 492–502. doi:10.1. 2010.

51. Zhu, Q., Pao, G. M., Huynh, A. M., Suh, H., Tonnu, N., Nederlof, P. M., et al. (2011). BRCA1 Tumour Suppression Occurs via Heterochromatin-Mediated Silencing. Nature 477, 179–184. doi:10.1038/nature10371. 2011.

52. Zhu, Q., Hoong, N., Aslanian, A., Hara, T., Benner, C., Heinz, S., et al. (2018). Heterochromatin-Encoded Satellite RNAs Induce Breast Cancer. Mol. Cel 70, 842–853. doi:10.1016/j.molcel.2018.04.023. 2018.

53. Bochar, D. A., Wang, L., Beniya, H., Kinev, A., Xue, Y., Lane, W. S., et al. (2000). BRCA1 Is Associated with a Human SWI/SNF-Related Complex. Cell 102, 257–265. doi:10.1016/s0092-8674(00)00030-1. 2000.

54. atchi, E., Skourti-Stathaki, K., Ventz, S., Pinello, L., Yen, A., Kamieniarz-Gdula, K., et al. (2015). BRCA1 Recruitment to Transcriptional Pause Sites Is Required for R-Loop-Driven DNA Damage Repair. Mol. Cel 57, 636–647. doi:10.1016/j.molcel.2015.01.011. 2015.

Bibliographic references

55. Gorthi, A., Romero, J. C., Loranc, E., Cao, L., Lawrence, L. A., Goodale, E., et al. (2018). EWS-FLI1 Increases Transcription to Cause R-Loops and Block BRCA1 Repair in Ewing Sarcoma. Nature 555, 387–391. doi:10.1038/nature25748. 2018.

56. Herold, S., Kalb, J., Büchel, G., Ade, C. P., Baluapuri, A., Xu, J., et al. (2019). Recruitment of BRCA1 Limits MYCN-Driven Accumulation of Stalled RNA Polymerase. Nature 567, 545–549. doi:10.1038/s41586-019-1030-9. 2019.

57. BRCA1 gene. Rockville Pike, Bethesda. s.l. : medline plus, 2020. 34.

58. [Online] https://www.researchgate.net/profile/Ines-Godet/publication/315981607/figure/fig1/AS:496918386012160@1495485983763/Schematic-representation-of-BRCA1-and-BRCA2-genes.png. 35.

59. Miki, Y., Swensen, J., Shattuck-Eidens, D., Futreal, P. A., Harshman, K., Tavtigian, S., et al. (1994). A strong Candidate for the Breast and Ovarian Cancer Susceptibility Gene BRCA1. Science 266, 66–71. doi:10.1126/science.7545954. 1994.

60. Li, D., Harlan-Williams, L. M., Kumaraswamy, E., and Jensen, R. A. (2019). BRCA1-No Matter How You Splice it. Cancer Res. 79, 2091–2098. doi:10.1158/0008-5472.CAN-18-3190. 2019.

61. Walker, L. C., Whiley, P. J., Houdayer, C., Hansen, T. V. O., Vega, A., Santamarina, M., et al. (2013). Evaluation of a 5-tier Scheme Proposed for Classification of Sequence Variants Using Bioinformatic and Splicing Assay Data: Inter-reviewer Variability. 2013.

62. Braunschweig, U., Gueroussov, S., Plocik, A. M., Graveley, B. R., and Blencowe, B. J. (2013). Dynamic Integration of Splicing within Gene Regulatory Pathways. Cell 152, 1252–1269. doi:10.1016/j.cell.2013.02.034. 2013.

63. Tudini, E., Moghadasi, S., Parsons, M. T., van der Kolk, L., van den Ouweland, A. M. W., Niederacher, D., et al. (2018). Substantial Evidence for the Clinical Significance of Missense Variant BRCA1 c.5309G>T p.(Gly1770Val). Breast Cancer Res. Treat. 172,. 2018.

64. King, M. C., Wieand, S., Hale, K., Lee, M., Walsh, T., Owens, K., et al. (2001). Tamoxifen and Breast Cancer Incidence Among Women with Inherited Mutations in BRCA1 and BRCA2. JAMA 286, 2251–2256. doi:10.1001/jama.286.18.2251. 2003.

65. Kuchenbaecker, K. B., Hopper, J. L., Barnes, D. R., Phillips, K. A., Mooij, T. M., Roos-Blom, M. J., et al. (2017). Risks of Breast, Ovarian, and Contralateral Breast Cancer for BRCA1 and BRCA2 Mutation Carriers. JAMA 317, 2402–2416. doi:10.1001/jama.201. 2017.

66. Rebbeck, T. R., Mitra, N., Wan, F., Sinilnikova, O. M., Healey, S., McGuffog, L., et al. (2015). Association of Type and Location of BRCA1 and BRCA2 Mutations with Risk of Breast and Ovarian Cancer. JAMA 313, 1347–1361. doi:10.1001/jama.2014.5985. 2015.

67. Tarsounas, M., and Sung, P. (2020). The Antitumorigenic Roles of BRCA1-BARD1 in DNA Repair and Replication. Nat. Rev. Mol. Cel Biol 21, 284–299. doi:10.1038/s41580-020-0218-z. 2020.

68. Lim, E., Vaillant, F., Vaillant, F., Wu, D., Forrest, N. C., Pal, B., et al. (2009). Aberrant Luminal Progenitors as the Candidate Target Population for Basal Tumor Development in BRCA1 Mutation Carriers. Nat. Med. 15, 907–913. doi:10.1038/nm.2000. 2009.

69. Molyneux, G., Geyer, F. C., Magnay, F.-A., McCarthy, A., Kendrick, H., Natrajan, R., et al. (2010). BRCA1 Basal-like Breast Cancers Originate from Luminal Epithelial Progenitors and Not from Basal Stem Cells. Cell Stem Cell 7, 403–417. doi:10.1016/j.stem. 2010.

70. Bach, K., Pensa, S., Zarocsinceva, M., Kania, K., Stockis, J., Pinaud, S., et al. (2021). Time-resolved Single-Cell Analysis of Brca1 Associated Mammary Tumourigenesis Reveals Aberrant Differentiation of Luminal Progenitors. Nat. Commun. 12, 1502. doi:10. 2021.

71. Esteller, M., Silva, J. M., Dominguez, G., Bonilla, F., Matias-Guiu, X., Lerma, E., et al. (2000). Promoter Hypermethylation and BRCA1 Inactivation in Sporadic Breast and Ovarian Tumors. J. Natl. Cancer Inst. 92, 564–569. doi:10.1093/jnci/92.7.564. 2000.

72. Ter Brugge, P., Kristel, P., van der Burg, E., Boon, U., de Maaker, M., Lips, E., et al. (2016). Mechanisms of Therapy Resistance in Patient-Derived Xenograft Models of BRCA1-Deficient Breast Cancer. JNCI J. Natl. Cancer Inst. 108, djw148. doi:10.1093/jn. 2016.

73. [Online] https://www.anticorps-enligne.fr/protein/2669409/Breast+Cancer+1+BRCA1+protein+His+tag/. 36.

74. Shen, Y., Tong, L. 2008. 37.

75. Structure and mechanism of BRCA1 BRCT domain recognition of phosphorylated BACH1 with implications for cancer. . Clapperton, J., Manke, I., Lowery, D. et al. s.l. : Nat Struct Mol Biol 11, 2004, Vol. https://doi.org/10.1038/nsmb775.

76. [Online] https://www.genomicseducation.hee.nhs.uk/education/core-concepts/what-is-bioinformatics/. 38.

77. Bioinformatics. Bayat, Ardeshir. 10.1136/bmj.324.7344.1018, s.l. : BMJ, 2002. 39.

78. A brief history of bioinformatics. Jeff Gauthier 1, Antony T Vincent 2, Steve J Charette 3 4, Nicolas Derome 1. 10.1093/bib/bby063., Canada : PubMed, 2019. 41.

79. Hammes, F. 2006.

80. I-Mutant2.0: predicting stability changes upon mutation from the protein sequence or structure. Emidio Capriotti, Piero Fariselli, and Rita Casadio. 10.1093/nar/gki375, s.l. : Oxford University Press, 2005. 44.*

81. Protein structure analysis of mutations causing inheritable diseases. An e-Science approach with life scientist friendly interfaces. Venselaar, H., te Beek, T.A., Kuipers, R.K. et al. 10.1186/1471-2105-11-548, s.l. : BMC Bioinformatics, 2010.

82. Germline mutations in BRCA1 and BRCA2 in epithelial ovarian cancer patients in Brazil. Simone Maistro, Natalia Teixeira, Giselly Encinas, Maria Lucia Hirata Katayama, Vivian Dionisio Tavares Niewiadonski, Larissa Garcia Cabral, Roberto Marques Ribeiro, Nelson Gaburo Junior, Ana Carolina Ribeiro Chaves de Gouvêa, Dirce Maria Carraro, Ester C. 10.1186/s12885-016-2966-x, s.l. : BMC Cancer, 2016.

83. [Online] http://genetics.bwh.harvard.edu/pph2/.

84. SIFT web server: predicting effects of amino acid substitutions on proteins. Ngak-Leng Sim, 1 Prateek Kumar,2 Jing Hu,3 Steven Henikoff,4 Georg Schneider,5 and Pauline C. Ng1,2. 10.1093/nar/gks539, s.l. : Oxford University Press, 2012.

Bibliographic references

85. BRCA Genes: The Role in Genome Stability, Cancer Stemness and Therapy Resistance. Ielizaveta Gorodetska1, Iryna Kozeretska2, Anna Dubrovska3. 10.7150/jca.30410, 2019.

86. Multifactorial likelihood assessment of BRCA1 and BRCA2 missense variants confirms that BRCA1:c.122A>G(p.His41Arg) is a pathogenic mutation. Phillip J Whiley 1, Michael T Parsons 2, Jennifer Leary 3, Kathy Tucker 4, Linda Warwick 5, Belinda Dopita 5, Heather Thorne 6, Sunil R Lakhani 7, David E Goldgar 8, Melissa A Brown 9, Amanda B Spurdle 2. 10.1371/journal.pone.0086836, 2014 .

87. Comprehensive evaluation and efficient classification of BRCA1 RING domain missense substitutions. Kathleen A. Clark, 1,11 Andrew Paquette,2,11 Kayoko Tao,1,8,11 Russell Bell,1 Julie L. Boyle,1 Judith Rosenthal,1 Angela K. Snow,1 Alex W. Stark,2,9 Bryony A. Thompson,1,10 Joshua Unger,1 Jason Gertz,1,2 Katherine E. Varley,1,2 Kenneth M. Boucher,1,3 David. 10.1016/j.ajhg.2022.05.004, 2022, Vol. 109.

88. RAD51 135G→C Modifies Breast Cancer Risk among BRCA2 Mutation Carriers: Results from a Combined Analysis of 19 Studies. Sinilnikova, Antonis C. Antoniou Olga M. 10.1086/522611, s.l. : The American Society of Human Genetics, 2007.

89. Pao, G. M., Janknecht, R., Ruffner, H., Hunter, T., and Verma, I. M. (2000). CBP/p300 Interact with and Function as Transcriptional Coactivators of BRCA1. Proc. Natl. Acad. Sci. 97, 1020–1025. doi:10.1073/pnas.97.3.1020. 2000.