ΙΑΤΡΙΚΗ ΣΧΟΛΗ

ΕΘΝΙΚΟ ΚΑΙ ΚΑΠΟΔΙΣΤΡΙΑΚΟ ΠΑΝΕΠΙΣΤΗΜΙΟ ΑΘΗΝΩΝ

Μεταπτυχιακό Πρόγραμμα Σπουδών
"Μοριακή & Εφαρμοσμένη Φυσιολογία"

ΔΙΠΛΩΜΑΤΙΚΗ ΕΡΓΑΣΙΑ

«HUMAN ENDOGENOUS RETROVIRUSES (HERVS) TRANSCRIPTOMICS IN HUMAN BREAST CANCER»

ΔΗΜΗΤΡΑ ΜΠΑΡΤΖΗ

ΙΑΤΡΟΣ

<u>ΤΡΙΜΕΛΗΣ ΕΠΙΤΡΟΠΗ</u>

ΜΙΧΑΗΛ ΚΟΥΤΣΙΛΙΕΡΗΣ, Καθηγητής

ΓΚΙΚΑΣ ΜΑΓΙΟΡΚΙΝΗΣ, Επίκουρος Καθηγητής

ΔΗΜΗΤΡΙΟΣ ΠΑΡΑΣΚΕΥΗΣ, Αναπληρωτής Καθηγητής

Αθήνα 2021

SCHOOL OF MEDICINE

NATIONAL AND KAPODISTRIAN UNIVERSITY OF ATHENS

MSc

IN MOLECULAR AND APPLIED PHYSIOLOGY

MASTER THESIS

«HUMAN ENDOGENOUS RETROVIRUSES (HERVS) TRANSCRIPTOMICS IN HUMAN BREAST CANCER»

DIMITRA BARTZI

MD

<u>THREE – MEMBER COMMITTEE</u>

MICHAEL KOUTSILIERIS

GKIKAS MAGIORKINIS

DIMITRIOS PARASKEVIS

Statue of Aphrodite (National Archaeological Museum)

«Γυναικί,ἐν Ἀβδήροισι, καρκίνωμα ἐγένετο περί το στῆθος, και διὰ τῆς θηλῆς ἔρρεν ἰχώρ ὒφαιμος έπιληφθείσης δε τῆς ῥύσιος, ἔθανεν.»

Ομήρου Ιλιάδα

Αφιερώνεται στη γιαγιά μου

18 Ιαν 21

Table of Contents

Περίληψη

Το ανθρώπινο γονιδίωμα φιλοξενεί σημαντική ποσότητα αλληλουχιών που προέρχονται από ρετροϊκές λοιμώξεις της γενετικής σειράς από την αρχαιότητα έως σήμερα, τους λεγόμενους ανθρώπινους ενδογενείς ρετροϊούς (HERVs). Οι ανθρώπινοι ενδογενείς ρετροϊοί (HERVs) πιθανολογείται ότι συμμετέχουν στην παθοφυσιολογία του καρκίνου του μαστού. Αυτή η ανάλυση εξετάζει την αλληλεπίδραση μεταξύ των ρετροϊών και του καρκίνου του μαστού αναλύοντας RNA ακολουθίες με σκοπό να μελετήσει υπερέκφραση των ενδογενών ρετροϊών σε ανθρώπινο ιστό μαστού κλινικού ενδιαφέροντος.

Abstract

The human genome harbours a significant amount of sequences that stem from retroviral infections of the germline in evolutionarily ancient times, so-called human endogenous retroviruses (HERVs). Human endogenous retroviruses (HERVs) are likely to take part in the physiopathology of breast cancer. This, to date, analysis examines the interaction between retroviruses and breast cancer, analyzing RNASeq datasets in order to study the overexpression of endogenous retroviruses in the human breast tissue of clinical interest.

Keywords: HERVs, breast cancer, overexpression, risk factors

1. Introduction

Historically, the oldest record of cancer was in Egypt in 3000 B.C. which has been noted down in an ancient textbook called the Edwin Smith papyrus [1]. Eight cases of tumours or ulcers of the breast were removed by cauterization with a so-called fire drill tool. The first physicians wrote about this disease:

"There is no treatment"

One more historical reference is the case of Empress Theodora's (wife of Justinian I) death which was recorded by Victor of Tonnena, on 28 June 548 A.C. at the age of 48. Her death was attributed to breast cancer [2].

Breast cancer is one of the most common cancers worldwide. In Greece, about 4,500 new cases are reported each year, while it is estimated that 1 in 8 women worldwide will develop breast cancer sometime in their lives. In Europe, 60% of breast cancers are diagnosed at an early stage. It is extremely rare in men, with about 100 times less frequency compared to women and the same survival rates.

However, breast cancer death rates are characterized by a declining trend, which is attributed to early diagnosis through mammograms as well as the improvement that have been made to the relevant treatments.

Initially, breast cancer has no symptoms. Later, a palpable formation, skin discolouration or discharge may occur. If the woman does not pay attention to the above symptoms then she may experience signs of advanced disease, such as hot and red breast (inflammatory cancer), bone pain, large swelling.

The main symptoms of breast cancer are:

- Lump or sclerosis in the breast area or in the armpit.
- Swelling of lymph nodes such as the armpit's lymph nodes.
- Secretion of fluid or blood from the nipple.

- Skin sagging.
- Redness, tenderness or chest pain.

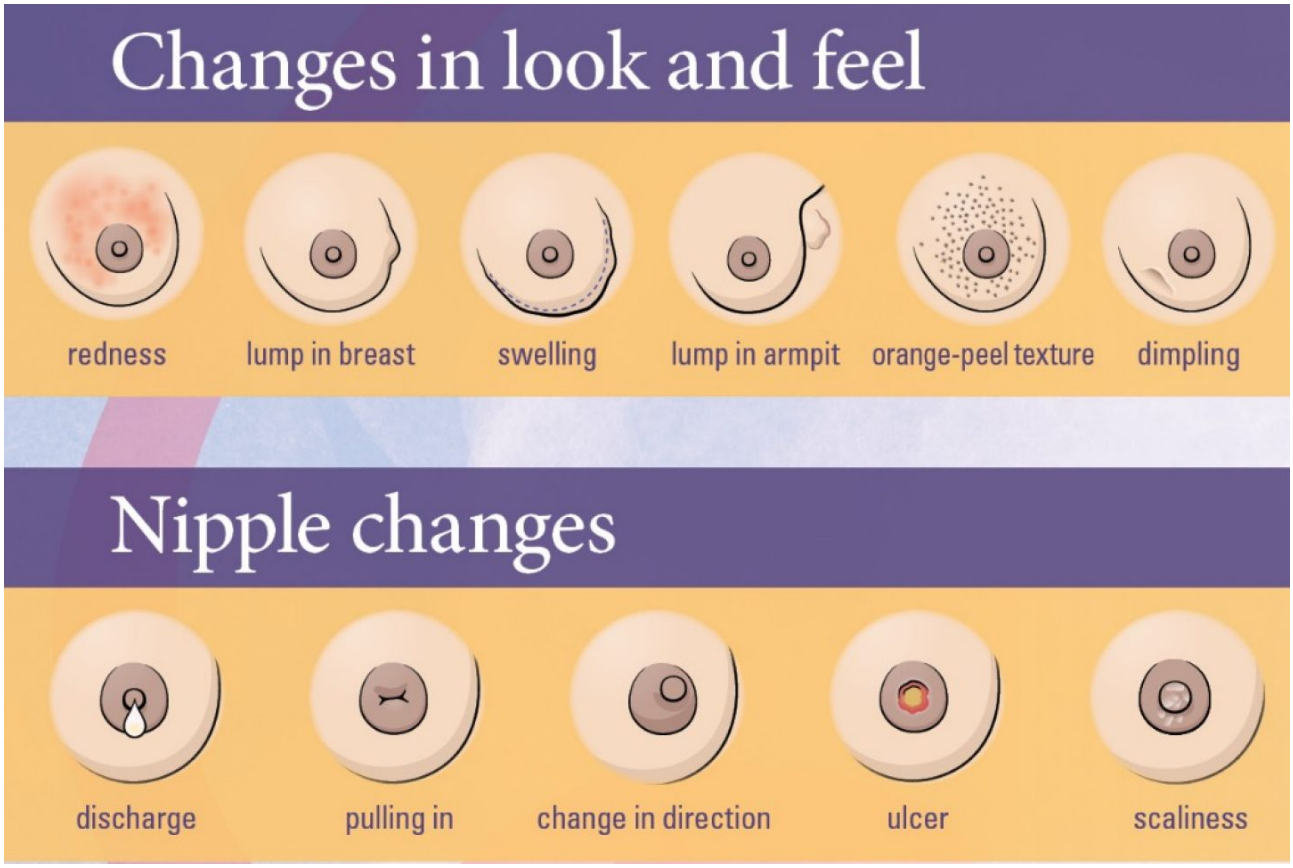


Figure 1. 11 Symptoms of breast cancer

The diagnosis of breast cancer at an early stage or even at a precancerous stage (non-invasive breast cancer) is increasing due to the awareness of women regarding the clinical palpation of breast, mammography and ultrasound, as well as self-palpation. Once a suspicious tumor is identified, the diagnosis is made by taking tissue from the tumor for microscopic examination. There are 4 diagnostic methods:

- FNA - Fine Needle Aspiration Cytological examination
- Histological biopsy
- Stereotactic biopsy
- Open surgical biopsy

2. General Part

2.1 About Breast

The breast is a bifurcated organ. Its boundaries extend from the second to the sixth rib and from the parasternal to the middle axillary line. Its shape is hemispherical with relatively conical configuration, which differs depending on the stage of the woman's life.

The breast is made of 3 types of tissue:

- ducts and lobes of the breast are made of glandular tissue
- fibrous, connective tissue includes the ligaments
- fatty tissue exists between glandular and fibrous tissue. Fatty tissue also forms the breast size.

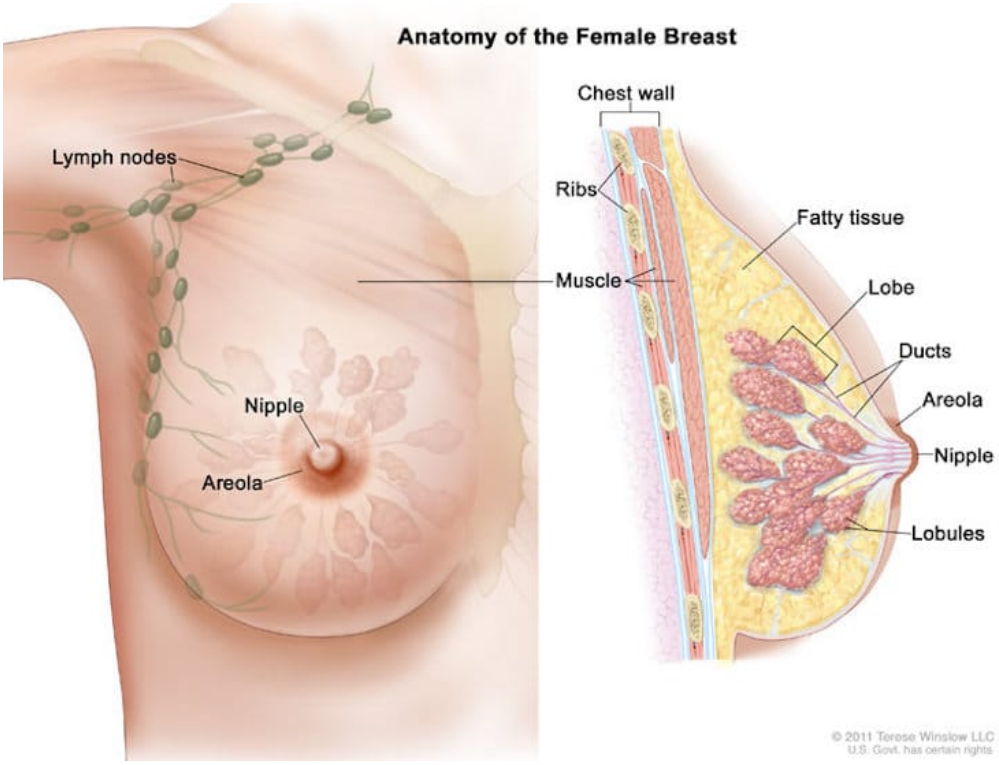


Figure 2 Anatomy of the female breast

2.2 Breast Development

In general, the size and shape of female breasts are due to quantity of adipose tissue. However, these characteristics are influenced by many factors, such as:

- Family history
- Age
- Weight loss or gain
- Pregnancy and breastfeeding
- Hormonal effect
- Menopause

Female breasts are rarely symmetrical. One may be larger than the other or higher or have a different shape. The nipple can be flat, round, cylindrical or inward. As the color of the nipple as the areola (colored area around the nipple) are determined by skin color. Both of them contain specialized muscles and fibers that respond to stimulation of straightening the nipple. The areola contains Montgomery gland which lubricates the halo. The shape and appearance of the breasts change during a woman's life. Young women have more glandular tissue than fat, which is responsible for thicker breasts [3].

The mammary gland after birth remains hypoplastic until puberty. Estrogen production and progesterone activate breast growth. The lactiferous ducts are elongated and branched. The nipple begins to project and the first lobes of the gland are formed. At the same time fat tissue is deposited between lobular units, shaping the size and shape of breast.

In the adult female, the breast changes in the various phases of the menstrual cycle. The breast begins to increase in size gradually about the 8^{th} day of the cycle. This increase is due to swelling of the lobes, vascular congestion and increase of parenchyma with the appearance of new lobes. The changes recede after the onset of menstruation, so the first 8-10 days of the cycle are the best time period for examination of breast.

During pregnancy and lactation, the breast shows important changes. The size of breast and the diameter of the nipple increase in the 2^{nd} month of pregnancy. The lactiferous ducts

are lengthened and new adenocytes are produced, so that the breast is ready for lactation at the end of the pregnancy.

At the end of pregnancy, colostrum is produced the first 2-3 days after childbirth. Then the milk production begins which is regulated by prolactin. At the end of lactation the breast begins to recede, but the return to its size and shape is not the same as breast was before pregnancy. After menopause, the mammary gland is atrophied progressively and replaced by fibrous connective tissue and fat tissue [4].

2.3 About Breast cancer

The disease is caused by the uncontrolled proliferation of pathological / cancerous cells, resulting in the formation of a malignant tumor. These abnormal cells have the potential to spread to adjacent tissues, or transport via hematogenous or lymphogenic spread to distant tissues with resulting in metastases [5, 6].

The transition from normal cells to cancer cells, and then to a solid tumor, is a complex process. This process involves a deregulation of cells by physiological mechanisms of cell proliferation and apoptosis due to genetic changes. The changes are hereditary (germ cell mutations), sporadic or due to viruses, radiation or "carcinogenic" substances. They involve oncogenes or loss of tumor suppressor genes, or they can also affect any region of genome. This leads to instability in the genetic material and finally induces the transformation of normal cells into cancer cells [7].

2.4 Molecular biology of breast cancer

Gene analysis has identified 4 distinct oncogenic subtypes with distinct clinical behavior (Luminal A, Luminal B, Basal and HER-2).

Luminal A or Tumor type A:

Cells express estrogenic (ER) and / or progesterone (PR) receptors. They do not express Ki-67 and do not have mutations in TP53. This kind of cancer usually has a better prognosis.

Luminal B or Tumor type B:

Cells express hormone receptors, ER and / or PR and Ki-67 They do not always respond to hormone therapy. Their biological behavior approaches to triple negative tumors.

Tumors that overexpress HER2:

HER2 is a transmembrane receptor of family of growth factors HER with tyrosine kinase activity. The overexpression of protein HER2 or multiple expression of HER2 gene is observed in 20% of breast cancer cases. Tumors, that overexpress HER2, are usually larger in size and coexist with infiltrated lymph nodes. About 50% of those tumors do not express hormone receptors [8, 9].

Basal or Triple negative tumors:

The cells of these tumors do not express hormone receptors and do not overexpress the HER2 protein. They often show mutations in the TP53 and have a poor prognosis. 60% of these cells express EGFR and high molecular weight cytokeratins.

In recent years many studies have been conducted on the molecular background of growth breast cancer. Among the genes that have been blamed for the appearance of the disease is the TP53 gene [10], topoisomerase IIa [11], epidermal growth EGF factor, the Akt1 protein gene [12, 13].

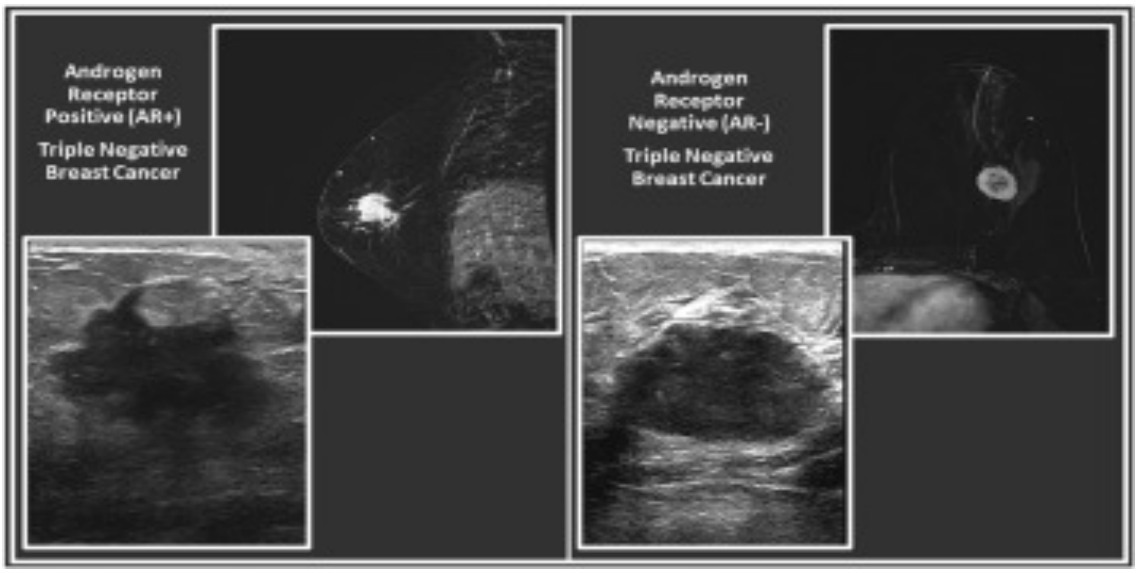


Figure 3 Triple negative Cancer

2.5 Risk factors

The aetiology of breast cancer is multifactorial. Risk factors could be categorized into biological, environmental and hereditary.

Biological factors

- The age of menarche. Women who have menstruation before the age of 11-12 years, have an increased risk of breast cancer [14]. According to studies, every year of delayed menarche is associated with a 10-20% reduction in risk of occurrence breast cancer, possibly due to reduced breast exposure to higher levels of circulating estrogen.

- The age of menopause. It is estimated that the risk of malignancy in breast is increased 3% for each year of delayed menopause, respectively due to increased breast exposure to estrogen, especially in cases of menopause after the age of 55 [14].

- Obstetric history is directly linked to the onset of cancer breast. Aggravating factors are infertility or incomplete gestation [15, 16], pregnancy over the age of 30 [15-17] and hormone therapy for delaying menopause [18, 19]. On the contrary, the younger age of a mother during the first full-term pregnancy reduces the risk of occurrence of breast cancer.

It is estimated that women with first full-term pregnancy before the age of 20 have a 30% lower risk compared to women with a first gestational age of 35 years [20].

- Miscarriage and abortion may act as an aggravating factor in appearance of malignancy in the context of abrupt changes in the hormonal status of a woman. [20].

- **Breastfeeding** may provide a protective effect against breast cancer. Studies have linked prolonged breastfeeding to a lower risk of occurrence of premenopausal breast cancer.

- The use of hormones affects the likelihood of carcinogenesis. Estrogens and progesterone appear to promote the risk of malignancy, while the Tamoxifen seems to have a beneficial effect due to its anti-estrogenic effect. A controversial risk factor is the use of contraceptives. Although contraceptives have been the subject of plethora of studies, the results are contradictory [21-25].

Environmental Factors

Environmental factors play an important role in carcinogenesis. Aggravating factors are:

- obesity [26-29]
- lack of physical activity [30]
- radiation exposure [31, 32]
- the history of radiotherapy [33],

Finally, they also play an important role in the development and progression of breast cancer other temperamental and anthropogenic factors, such as:

- the presence of dense breasts, mainly due to difficulty in examination and imaging control [34],
- the height above the average, [26],

- multiple breast biopsies [34] in the context of induction of excessive inflammation reaction

- the existence of a precancerous condition (atypical hyperplasia or other pre-malignant condition) [35].

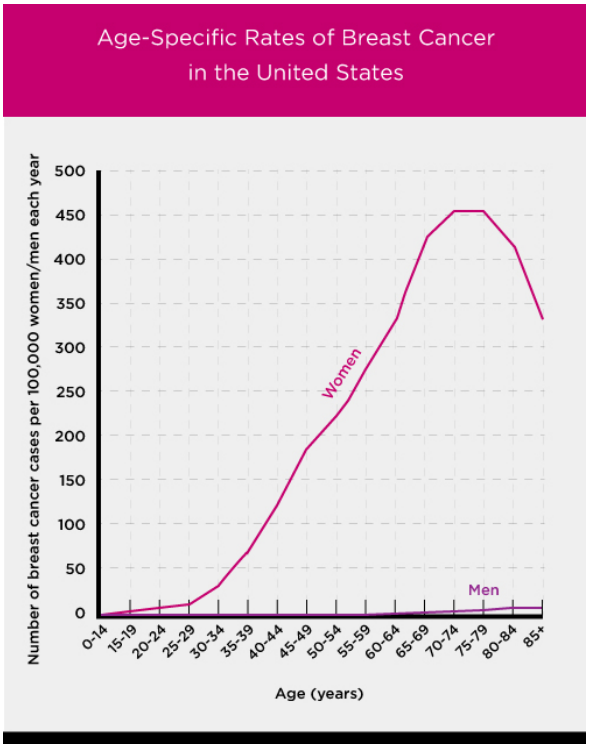


Figure 4 Age - Number of breast cancer cases

Age is an independent aggravating factor. According to bibliography, **a woman in the 7th decade presents a double risk of occurrence of breast cancer compared to the age of 50 years** [35].

- Diet is one of the most effective environmental factors

The daily intake of olive oil, vegetables and fruits have shown reducing endogenous estrogen levels which leads to lower breast cancer risk. Instead, total and particularly saturated fat and consumption of alcohol are associated with increased breast cancer risk because of increasing estrogen levels.[38]

Smokers also have increased risk of breast cancer [37,38]. Ionizing radiation is also a cause of breast cancer [36]. Total and free estradiol, estrone and estrone sulphate, androstenedione, testosterone and prolactin have been positively associated with breast

cancer risk in women after menopause [36-38] Case-control studies have shown positive association between estrogens and breast cancer risk and also present the increase of breast cancer risk because of high levels of androgens [37, 38]. In premenopausal women, blood insulin like growth factor 1 (IGF-1) is positively associated with breast cancer risk. [38]

Hereditary (genetic) Factors

Genetic predisposition is a major risk factor for occurrence of breast cancer [25]. In clinical practice, mutations in BRCA1 and BRCA2 are investigated. Also the Mutations in TP53 are positively correlated with the onset of the disease. Finally, rarer mutations in the population have been implicated, such as ATM whose heterozygous carriers are at increased risk of disease, and sufferers of Muir-Torre and Peutz-Jeghers syndromes [39].

In fact, genetic predisposition is responsible for only 5-10% of cases, with the frequency of mutation in the BRCA1 and BRCA2 genes 1: 500 in the general population. Among certain tribes such as the Eskenazy Jews, the frequency may reach 1:40 [25]. However, the biological behavior of tumors in patients, who are mutants of the BRCA1 and BRCA2 genes, appear to be different from breast carcinomas that occur sporadically in the population. The first one are associated with a worse prognosis and higher rates of bilateral breast carcinoma [40, 41]. It is estimated that the majority of cases are due to failures in genetic material arising during the life [39].

Family history plays a key role in the onset of the disease. Women with 1st degree relatives (or even 2nd degree) who have been diagnosed with the disease, especially at a young age, and women with a family history of occurrence in more than one member family are at increased risk [42, 43].

Finally, the personal history of breast cancer and history of cancer ovaries or uterus are correlated. In the first case, there is a possibility of 1-2% per year for recurrence of the disease in the other breast [44], and in the second case the risk is higher due to the

increased likelihood of BRCA1, BRCA2 or other oncogenes being present [25, 39]. The individual characteristics associated with an increased probability of mutation in BRCA 1 and 2 and consequently the onset of the disease are:

- breast cancer at a young age
- Ovarian cancer at any age
- bilateral breast cancer
- a history of breast and ovarian cancer

Table 1 Factors evaluated in relation to breast cancer risk [45]

Risk factor	Category/change	Strength
Gender	Women vs. men	++++
Age	Increase	++++
Ethnic group	Caucasian vs. Asian	+++
Family history	Yes vs. no	+++
Specific genes	Yes vs. no	++++
Cancer in other breast	Yes vs. no	+++
Height	Increase	++
Postmenopausal obesity	Increase	++
Birth weight	Increase	+
Having been breastfed	No vs. yes	0
Growth in early life	Increase	+
Atypical hyperplasia	Present vs. absent	+++
Mammographic density (mammary gland mass)	High vs. low density (increasing mass)	+++
Age at menarche	Earlier	++
Age at menopause	Later	++
Type of menopause	Natural vs. artificial	++
Age at 1st full term pregnancy	Later	+++
Age at other pregnancies	Later	+
Parity overall	Lower	++
Pregnancy timing	Proximal vs. distant	+
Lactation	No vs. yes	+
Abortion	No vs. yes	0
Oral contraceptive use (recent)	Increase	+
Hormone replacement	Increase	++
Plant foods and olive oil	Reduced intake	+
Saturated fat	Increased intake	+
Physical activity	Reduced	+
Ethanol intake	Increase	+
Ionizing radiation	Increased	+
Magnetic fields	Increased	0
Organochlorines	Increased	0

Association: ++++ very strong, +++ strong, ++ modest, + weak, 0 null.

2.6 Histological classification of breast cancer

With regard to malignant breast diseases, there are two main types of the disease:

- Ductal carcinoma, which originates from the cells that cover the inside of the wall of the milk ducts
- the lobular, which comes from the cells of the pods (lactiferous glands)

Breast cancer is divided into invasive and non-invasive or in situ. When the Cancer infiltrates the basement membrane of the ducts or lobes of the mammary gland called infiltrative, while when cancer is confined to the epithelium without infiltrating the basement membrane, is called Ductal carcinoma in situ (DCIS).

Respectively, when it does not infiltrate basement membrane of mammary gland, it's called lobular carcinoma in situ (LCIS). DCIS is considered a precancerous condition. The occurrence of LCIS indicates high risk of development of invasive cancer. The invasive breast cancer has the ability to metastasize, while the non-invasive type does not metastasize potentially.

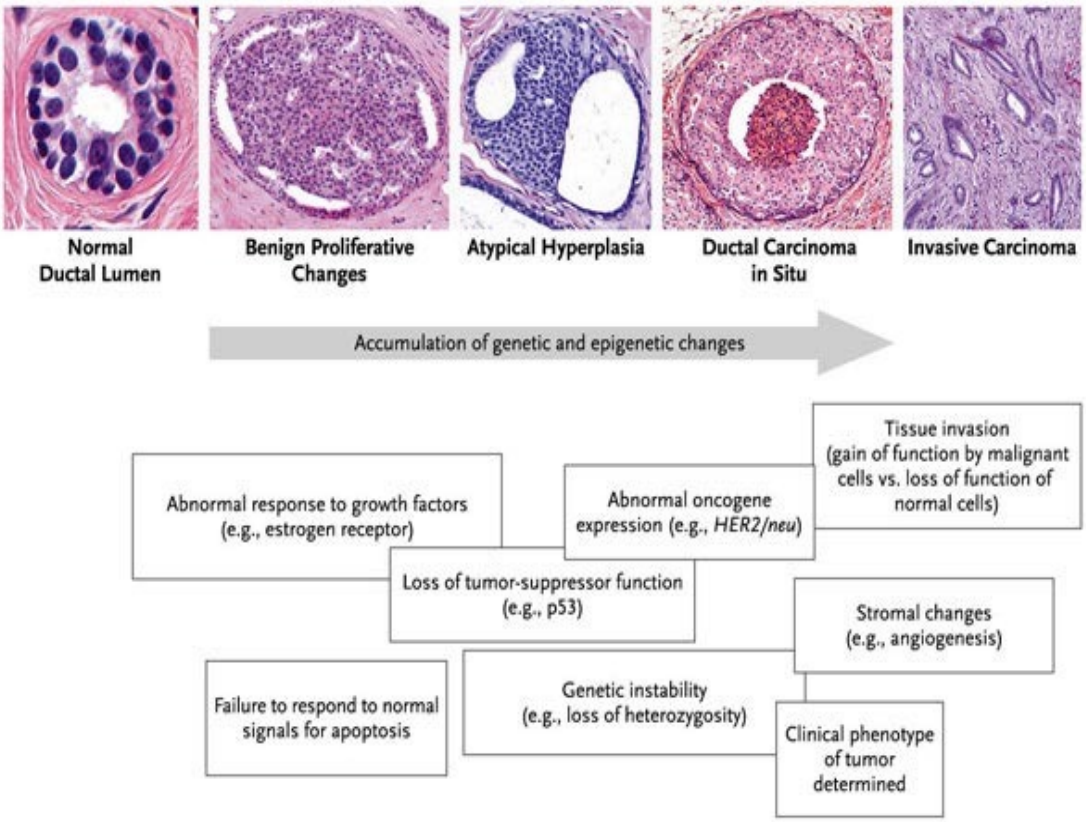


Figure 5 Accumulation of genetic and epigenetic changes

Other less common forms of breast cancer are inflammatory (1% -3%), Paget's disease (1%), and more rarely, angiosarcoma and other rare forms [46, 47]. More specifically, malignancies can be [48]:

- [Ductal Carcinoma in situ-DCIS]
- Lobular Carcinoma in situ-LCIS
- Invasive Ductal Carcinoma [IDC]
- Invasive Lobular Carcinoma [ILC]
- Tubular carcinoma [Tubular carcinoma]
- Myeloid Breast Carcinoma
- Mucinous Carcinoma of the Breast
- Papillary Carcinoma of the Breast

- Cribriform Carcinoma of the Breast
- Inflammatory Breast Cancer
- Paget's Disease of the Breast
- Phyllodes Tumor of the Breast
- Angiosarcoma of the Breast

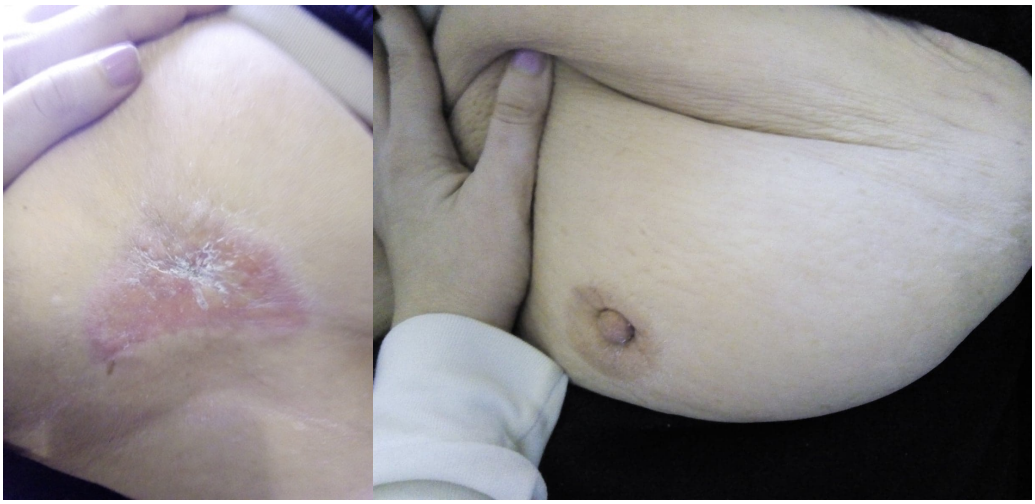


Figure 6. Inflammatory breast cancer

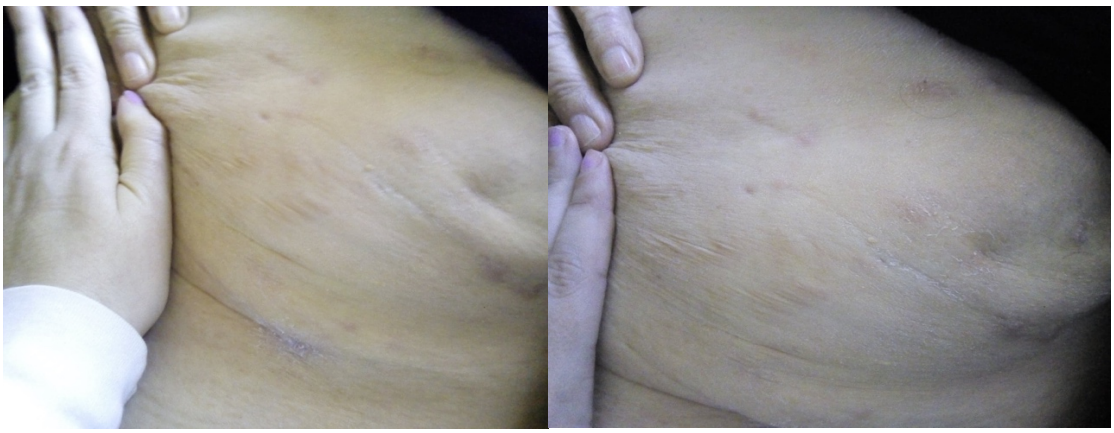


Figure 7. Inflammatory breast cancer

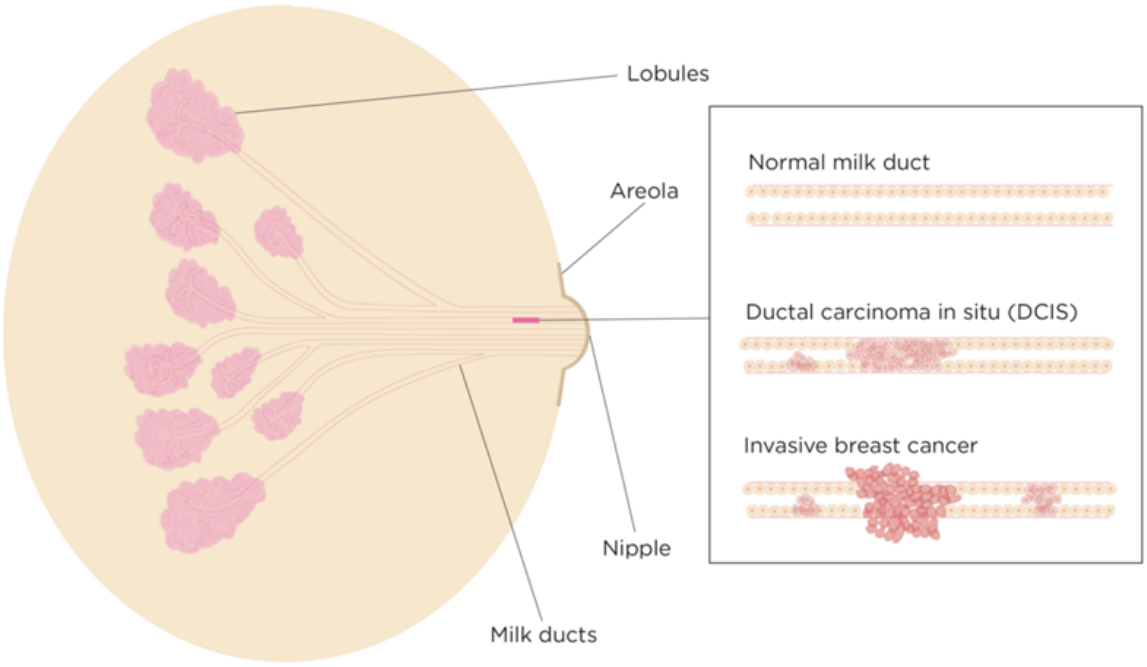


Figure 8 Normal milk duct & Cancerous milk duct

Ductal carcinoma in situ DCIS (noninvasive carcinoma)

Ductal carcinoma in situ is characterized by an increased number of epithelial cells, which show mild to severe cellular atypia. It is accompanied by the progression to invasive breast carcinoma [49]. Noninvasive carcinoma is more common than lobular carcinoma in situ. Deaths that occur from DCIS are associated with undiagnosed invasive carcinoma, which was present at the time of initial diagnosis of DCIS, or with de novo development of invasive carcinoma, at any other site in the breast [50]. The DCIS is distinguished in three degrees histologically, such as low, intermediate and high grade, mainly assessing the degree of nuclear atypia and the presence of ductal necrosis and less mitotic activity and presence of microcalcifications [49].

2.7 Staging of breast cancer

The purpose of carcinoma staging is the accurate determination of the anatomical location and extent of the tumor and metastases, so the appropriate **treatment** could be

applied. This procedure requires an excellent recording of clinical and histopathological data. The cancer staging is a meticulous procedure.

In the past, the staging of cancer in various stages of its evolution was simplistic and almost arbitrary. The neoplasms were evaluated only by their clinical picture, if they were surgically exceptional or not, and whether the disease was local, regional or metastatic.

Today the TNM system devised at Goustave–Roussy Hospital of Paris, is widely applied and serves as prognostic system. It also accurately determines the therapeutic approach to breast cancer in each patient [51]. The system characteristics are analyzed in the table below:

Table 2. Staging of breast cancer

T	**is derived from the initial letter of the word tumor**
Tx:	The volume can not be estimated
T0:	No indication of tumor
T1:	Volume of size equal to or <of 2.5 cm in its largest diameter
T1a	<0.5cm
T1b	0.5-1cm
T1c	1-2cm.
T2	volume> 2-5 cm in its largest diameter
T3:	volume> 5cm in its largest diameter
T4	tumor of any size by fixation or infiltration of the chest wall of overlying skin.
T4a:	expansion in the chest wall.
T4b	swelling or ulceration of the skin, or satellite nodules

T4c	T4a and T4b together.
T4d	inflammatory carcinoma
N	**is derived from the initial letter of the word, nodes**
Nx	the condition of the lymph nodes cannot be determined.
No	without infiltration of the lymph nodes in the area
N1	metastasis to the axillary lymph nodes, which are agile, fixed between themselves or in the surrounding tissues.
N2	metastasis to the axillary lymph nodes, which are fixed to each other or in the surrounding tissues.
N3	metastasis to the internal mammary lymph nodes.
M	**is derived from the initial letter of the word, metastasis**
Mx	the presence of metastases in distant organs has not been assessed
Mo	there are no distant metastases
M1	there are distant metastases including metastasis in the supraclavicular lymph nodes.

Table 3. Stages of breast cancer

Stage	Definition
Stage 0	The abnormal cells are still confined to the ducts where they appeared at first.
Stage I	The diameter of the tumor is less than 2 cm and small islets of cancer cells could be found in the lymph nodes *. Stage I is divided into stages IA and IB.

Stage II	The diameter of the tumor is either less than 2 cm and has expanded to axillary lymph nodes *, either between 2 and 5 cm, but there is no extention to the axillary lymph nodes. Stage II is divided into stage IIA and IIB.
Stage III	The tumor can be of any size, but: • has spread to either the chest wall and / or the overlying skin of breast • has extended to at least 10 axillary lymph nodes * or the axillary lymph nodes are connected to each other or to other tissues. It has also spread to lymph nodes near the sternum • It has extended to lymph nodes below or above the clavicle. Stage III breast cancer is divided into stage IIIA, IIIB and IIIC.
Stage IV	The cancer has spread to other organs in the body, most often to the bones, the lungs, liver or brain. These are called metastases *

2.8 Epidemiology of breast cancer

More than 1,6 million women per year are diagnosed with breast cancer according to World Health Organization. In Europe more than 450.000 women develop breast cancer and 140.000 finally lose their lives [52].

In 2020, 7.772 women were diagnosed with breast cancer which represents 27,5% of total female cancer cases in Greece as shown in the figure below.

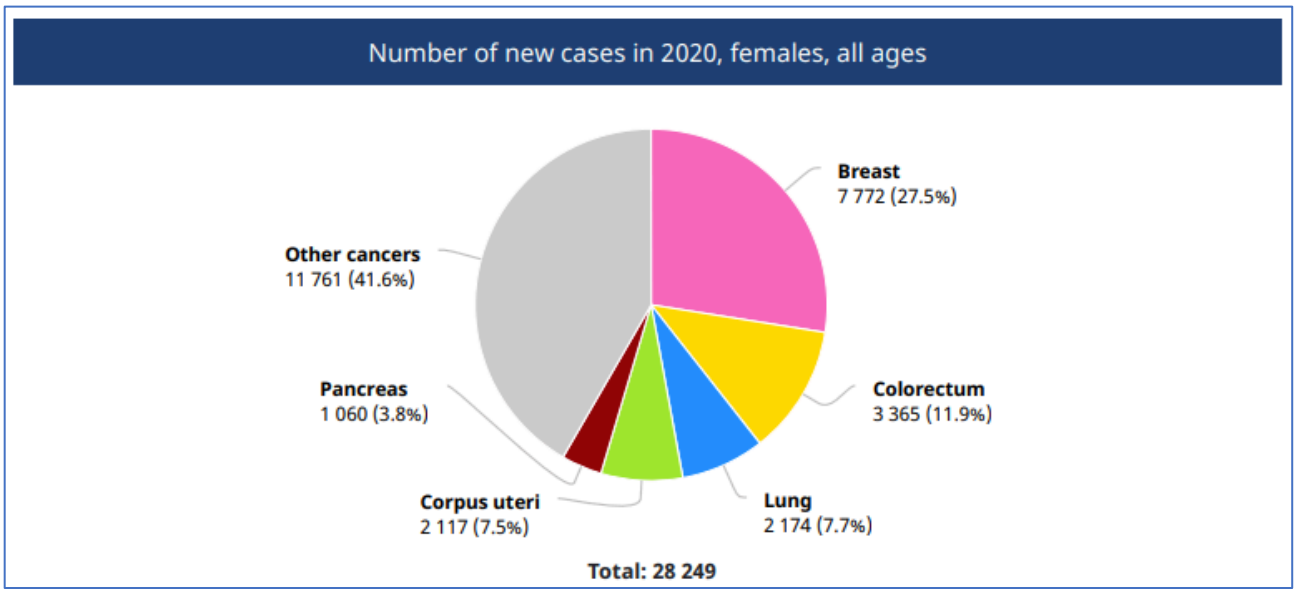


Figure 9 New female cases of cancer. Greece 2020

In 2020, 7.772 patients (both sexes) were diagnosed with breast cancer, which represents 12% of total cancer cases in Greece as shown in the figure below.

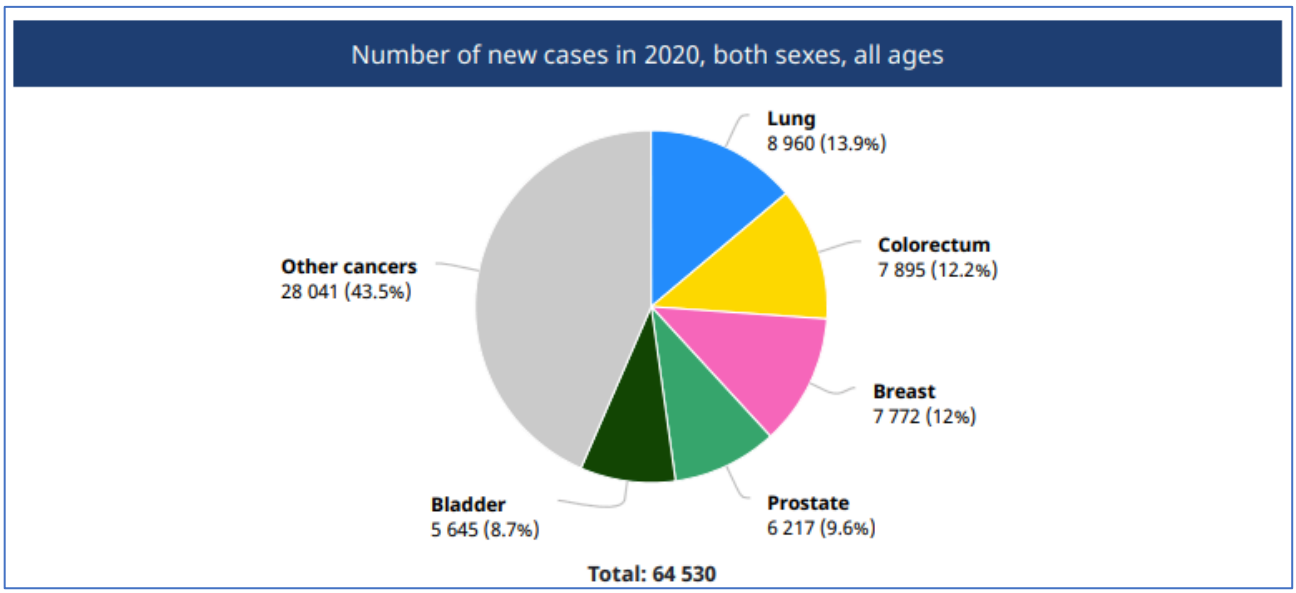


Figure 10 New cancer cases both sexes. Greece 2020

According to the following diagram, breast cancer is first rated in incidence and second in mortality rate.

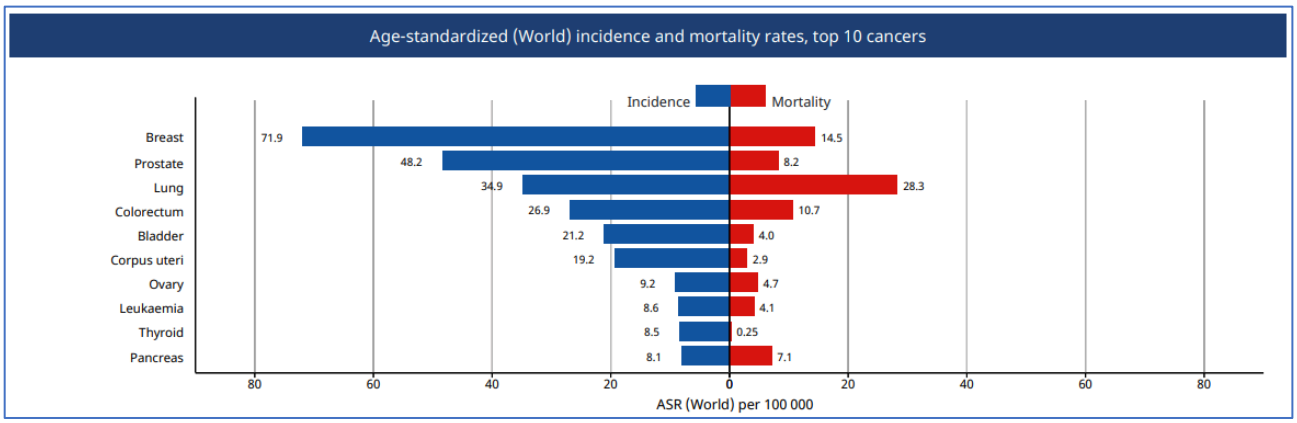


Figure 11. Incidence and mortality rates, top 10 cancers

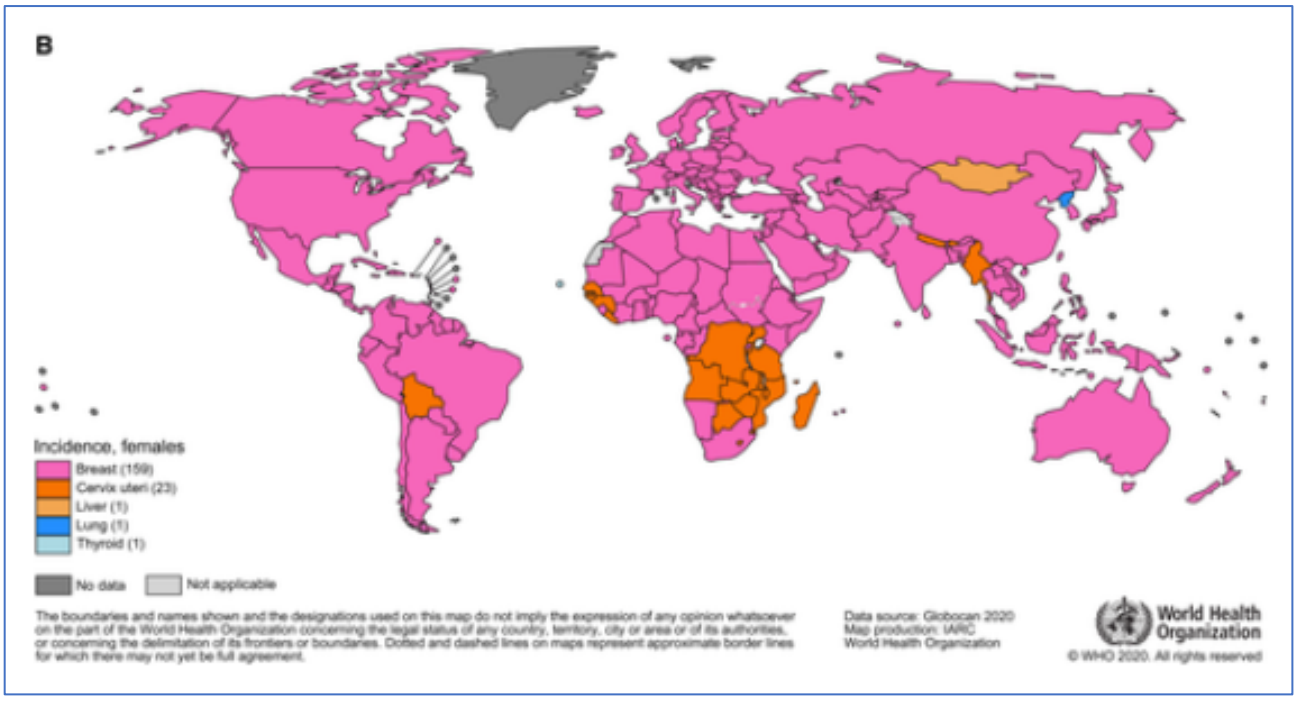


Figure 12 Global Maps Presenting the Most Common Type of Cancer Incidence in 2020 in Each Country Among (A) Men and (B) Women.

2.9 HERVs

Human endogenous retroviruses (HERVs) have infected human genome in the distant past, ongoing parallel to human evolution. Endogenous retroviruses (ERVs) are remnants of ancient active retroviruses that infected germline cells. ERVs are transmitted vertically through successive generations by Mendel's model. About 8% of human genome is composed of HERVs, most of which are nonfunctional because of epigenetic control or deactivating mutations [53.]

Retroviruses are a subgroup of viruses characterized by their ability to integrate their genome into host-cell DNA. Retroviruses are double-stranded positive-sense RNA viruses that use reverse transcriptase enzyme to transcribe their RNAs to DNAs. The resultant DNA is later integrated into host DNA through viral integrase enzyme. Viral-integrated DNA (proviral DNA) is translated and transcribed to proteins as part of host genome using the genetic machinery of infected cell [54-56].

The genomic structure of retroviruses is composed of gene gag, pol/pro, and env flanked by two long terminal repeats (LTRs) which open reading frames (ORFs) and encode structural and functional viral proteins. LTRs encode promoter and polyadenylation signals. LTR retrotransposon sequences, including HERVs, make up about 8% of the human genome. HERV sequences are created from integration events, which occurred in humans 100,000 years ago [57, 58]. Although there are currently no known infectious HERVs due to the accumulation of mutations over time, likely combined with selection against proviruses capable of giving rise to infectious virus, many of these sequences are still able to produce viral mRNA, protein, and even retrovirus-like particles [59–62]. Active LTRs affect host gene regulation up to 100 kb away, by giving alternative promoters, enhancers, splice sites, and termination signals. Additionally, HERV sequences are largely silenced in normal tissue through epigenetic effects, including DNA and chromatin modifications [63, 64].

During the retrovirus infection cycle, viral genomic RNA is reverse transcribed into a DNA copy that is permanently integrated into the genomic DNA of the host. The integration of retroviral cDNA into the DNA of a germ cell occasionally results in an endogenous retrovirus (ERV), a provirus that is transmitted vertically to that host's offspring, and which may become fixed in the host species over time [65].

The expression of HERV depends upon regulation at the level of the LTRs which function as promoters for HERV expression [66] and have strong RNA Polymerase II regulatory sequences [67, 68]. They also contain a plethora of transcription factor binding

sites [69]. Solo LTRs are present in the genome due to recombination that excises the rest of the provirus [70]. Indeed, up to 85% of HERVs have undergone this recombinatorial deletion [71], making most HERV loci solo LTRs. Solo LTRs can serve as promoters in both sense and antisense orientations [72] and can alter the expression of host genes. Further, the expression of very long intergenic RNAs (vlincRNAs) which control pluripotency and malignancy was HERV LTR-driven [73], suggesting a role for HERV LTRs in regulating not only protein-coding genes but also the expression of long non-coding RNAs. Thus, the LTRs are an important site for epigenetic modifications to control HERV and human gene expression [74].

Human endogenous retrovirus **type K** [HERV-K(HML-2)]**is expressed in subtypes of breast cancer**. HERV-K(HML-2) is a retrovirus integrated into the primate genome 55 million years ago. It ‘s a challenge for possible immunotherapy of breast cancer. A number of ERVs were recently reported to be re-activated in tumors, and several showed overexpression in the tumors but low or undetectable expression in normal tissues [75, 76].

3. Methods

In our study, it will be shown high expression of HERVs in malignant breast cancer and low expression in normal breast tissue. This search indicated the expression of HERVs around breast cancer types such as ductal carcinoma, triple negative cancer and adenocarcinoma.

This to date analysis, evaluated the expression of HERV-K in ductal carcinoma and triple negative breast cancer subtypes. In this study, the transcriptome data were analyzed from 11 invasive ductal carcinoma (IDC) breast cancer patients (***Table 4***), and 6 healthy patients Their characteristics are shown in (***Table 5)*** The data were downloaded from the Sequence Read Archive (SRA) database (https://www.ncbi.nlm.nih.gov/sra).

3.1 RNA-Seq data analysis

Data were accessed from SRA. Breast ductal carcinoma RNA-Seq data (BAM files) and their related clinical data were obtained from NCBI SRA. The paired-end FASTQ files for each sample were downloaded in sra format and were converted to fastq format using the command fastq-dump from sratoolkit.

Mapping/Alignment: The paired-end reads in FASTQ format were aligned to the human reference genome, hg19 using bowtie2 with default settings The paired-end reads in FASTQ format were also aligned to the pseudogenome which was made using the 5' and 3' LTR coordinated of HERV-K(HML-2) as described by Subramanian et al.

The resulting alignments were then transformed to the standard bam file format, using samtools with default settings. Expression of HERV-K-human junctions was extracted in raw reads using bedtools multicov with default settings. Then, the mapped reads were counted and were normalized by using the formula:

$$\frac{multicov\ pseudogenome\ reads}{alignment\ rate\ \% * reads}$$

A Pseudogenome based on coordinates of LTRS of HERV-K(HML-2) was created in order to detect expression of LTR-host junctions. [77] LTR coordinates used for the construction of the pseudogenome were described by Subramanian et al.. The integration of HERVs in human genome results in regions that were partial sequences of LTRs and also partial human sequences (LTR-host junctions). Several applications were used, in order to analyze these datasets such as SRA-toolkit, Samtools, Bowtie, IGV, BedTools (a software designed to detect how many reads correspond to every gene).

In order to include the integration sites for the pseudogenome, partial human sequences were used, which started and ended in specified distance of bases from the start and the end of LTRs (the coordinates of start and end of LTRs are attached to **excel)**. Hence, a Bed file was created adding and subtracting 66% of length of reads in order to include junctions. For example, the formula is described below:

chr(1,2,3,etc) LTR start-66% length LTR start+66% length

chr(1,2,3,etc) LTR end-66% length LTR end+66% length

for any LTR of HERV.

Command “bedtools getfasta” [78] was used with default settings, in order to create a fasta file including the LTR-host junction sequences that was used as the pseudogenome. As far as it concerns the alignment of pseudogenome, use the command:

“bowtie2 -t -x bowtie2/hg19 ncbi/public/sra/....fastq -S bowtie2/....sam”

Furthermore, downloading SAMtools was used for mapping the pseudogenome, via the following commands:

1. *”wget https://github.com/samtools/samtools/releases/download/1.3.1/samtools-1.3.1.tar.bz2”*
2. *“tar -jvxf samtools-1.3.1.tar.bz2”*

3. *"cd samtools-1.3.1"*

4. *"make"*

".sam file" was converted to ".bam file" by means of the commands:

"samtools view -S -b bowtie2/....sam > bowtie2/....bam"

After that, files were sorted with samtools by means of the commands:

"samtools sort bowtie2/....bam -o bowtie2/..._sorted.bam"

"samtools index bowtie2/..._sorted.bam"

3.2 RNA Seq Datasets.

The following RNA-seq datasets were obtained through SRA NCBI: (Healthy breast tissue) SRR7687780, SRR7687798, SRR7687782, SRR7687788, SRR7687781, SRR7687797, (breast cancer tissue) SRR8743332, SRR10502311, SRR10612305, SRR7509754, SRR7509758, SRR12180250, SRR7500823, SRR10247656, SRR7509721, SRR7509732, SRR7509733 and were analyzed with the methods described in previous chapter.

Table 4 RNA-seq data: Instrument ILLUMINA HiSeq 2500, strategy: RNA-seq, source: Transcriptomic, Selection: cDNA, Layout: Single (breast carcinoma)

SRR Experiment	Cell_Line	disease_state
SRR12180250	HCC1428-LTED	adenocarcinoma
SRR7509758	HCC38	ductal carcinoma
SRR7509754	HCC39	ductal carcinoma
SRR7500823	ZR75-1	ductal carcinoma
SRR10247656	HCC38	ductal carcinoma
SRR10612305	HCC38	ductal carcinoma

SRR10502311	MDA-MB-231	Triple-negative breast cancer
SRR8743332	MCF7	ductal carcinoma
SRR7509732	HCC38	ductal carcinoma
SRR7509733	HCC38	ductal carcinoma
SRR7509721	BT549	ductal carcinoma

Table 5 RNA-seq data: Instrument: Next seq 500 ,strategy: RNA-seq, source: Transcriptomic, Selection: PolyA, Layout: Single, (normal breast tissue)

SRR Experiment	Cell_Line	disease_state
SRR7687788	Normal left Stroma	normal breast tissue
SRR7687782	Normal left epi	normal breast tissue
SRR7687798	Normal left Stroma	normal breast tissue
SRR7687780	Normal left Stroma	normal breast tissue
SRR7687781	Normal left epi	normal breast tissue
SRR7687797	Normal right epi	normal breast tissue

All data were downloaded into secure, password-protected directories of the SRA. SRA toolkit was downloaded via Linux to get access to SRA data. SRA Toolkit allows to create next-generation sequencing files in the desired format and cloud bucket [79].

Table 6 *Raw Reads and overall alignment rate came from alignment of any RNA-seq to Human genome (hg19).*

Pseudogenome Reads came from alignment of any RNA sequences to Pseudogenome.

Normalized coverage of integration sites

$$\frac{multicov\ pseudogenome\ reads}{alignment\ rate\ \% * reads}$$

Health Controls	Alignment rate (%)	Raw Reads	Pseudogenome Reads	Normalized coverage of integration sites
SRR7687780	82,47	43714492	42	0,0000012
SRR7687798	80,74	43150648	44	0,0000013
SRR7687782	84,92	63226124	43	0,0000008
SRR7687788	82,18	59081485	54	0,0000011
SRR7687781	85,98	37692344	40	0,0000012
SRR7687797	85,58	46520403	57	0,0000014

Table 7 *Raw Reads and overall alignment rate came from alignment of any RNA-seq to Human genome (hg19).*

Pseudogenome Reads came from alignment of any RNA sequences to Pseudogenome.

Normalized coverage of integration sites:

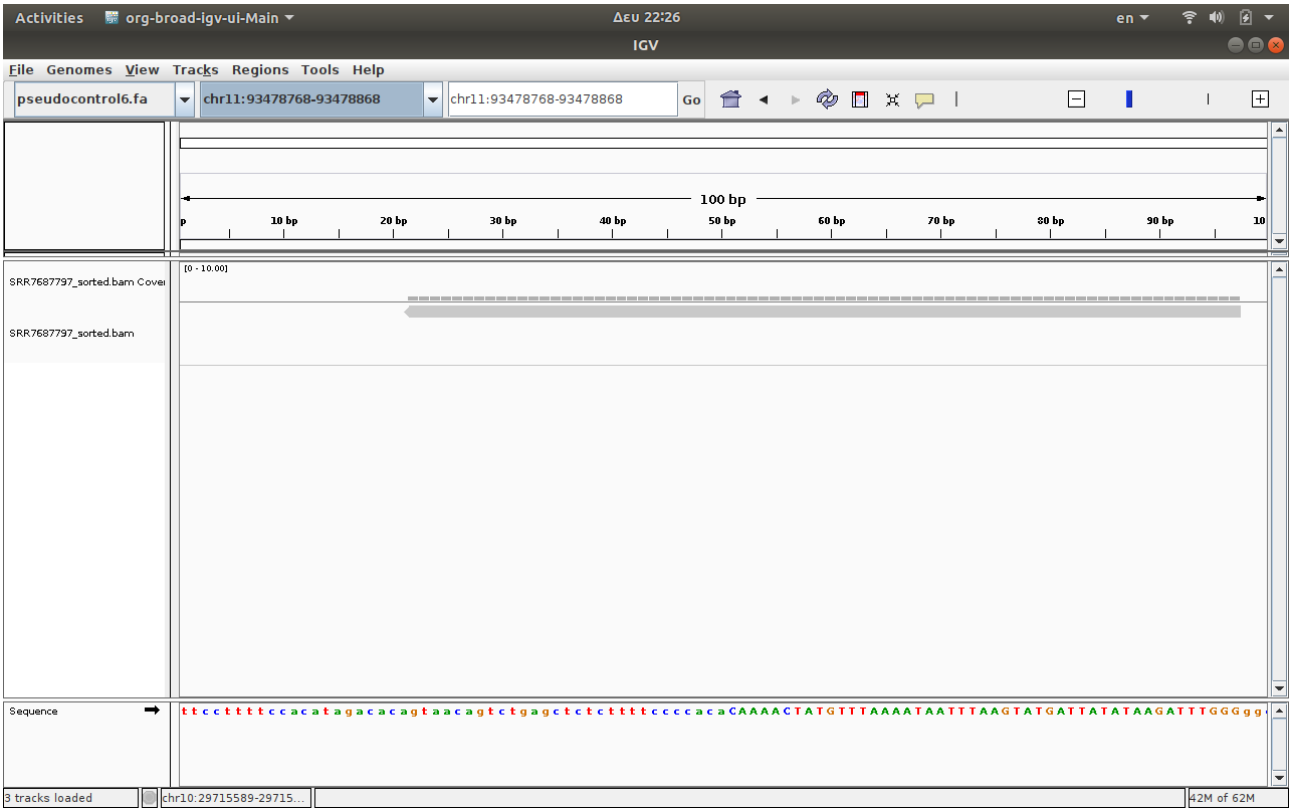
$$\frac{multicov\ pseudogenome\ reads}{alignment\ rate\ \% * reads}$$

Breast cancer	Treatment	Alignment Rate (%)	Raw Reads	Pseudogenome Reads	Normalized coverage of integration sites
SRR8743332 ductal	Yes	93,13	26333636	160	0,0000065

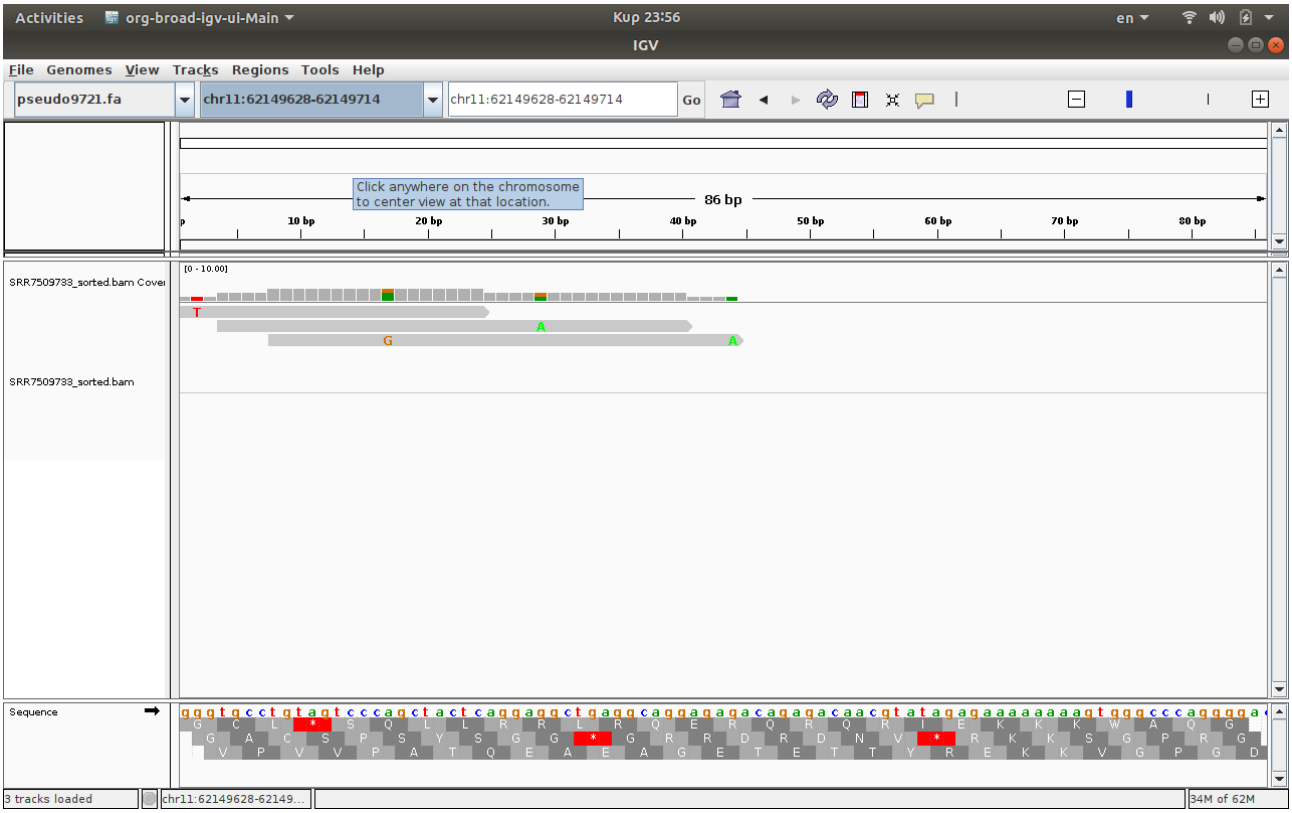
SRR10502311 Triple negative	Yes	91,25	8590444	39	0,0000050
SRR10612305 ductal	Yes	91	21682035	13	0,0000007
SRR7509754 ductal	No	95,58	1087054	23	0,0000221
SRR7509758 ductal	No	96,04	1347370	17	0,0000131
SRR12180250 adenocarcinoma	Yes	91,13	13538984	67	0,0000054
SRR7500823 ductal	Yes	85,92	13765655	12	0,0000010
SRR10247656 ductal	Yes	93,17	60339431	108	0,0000019
SRR7509721 ductal	Yes	93,2	2075305	52	0,0000269
SRR7509732 ductal	No	94,12	1464373	38	0,0000276
SRR7509733 ductal	No	94,32	1442350	47	0,0000345

The visualization of files was made by IGV, which is a program that allows to see the reads on chromosomes.

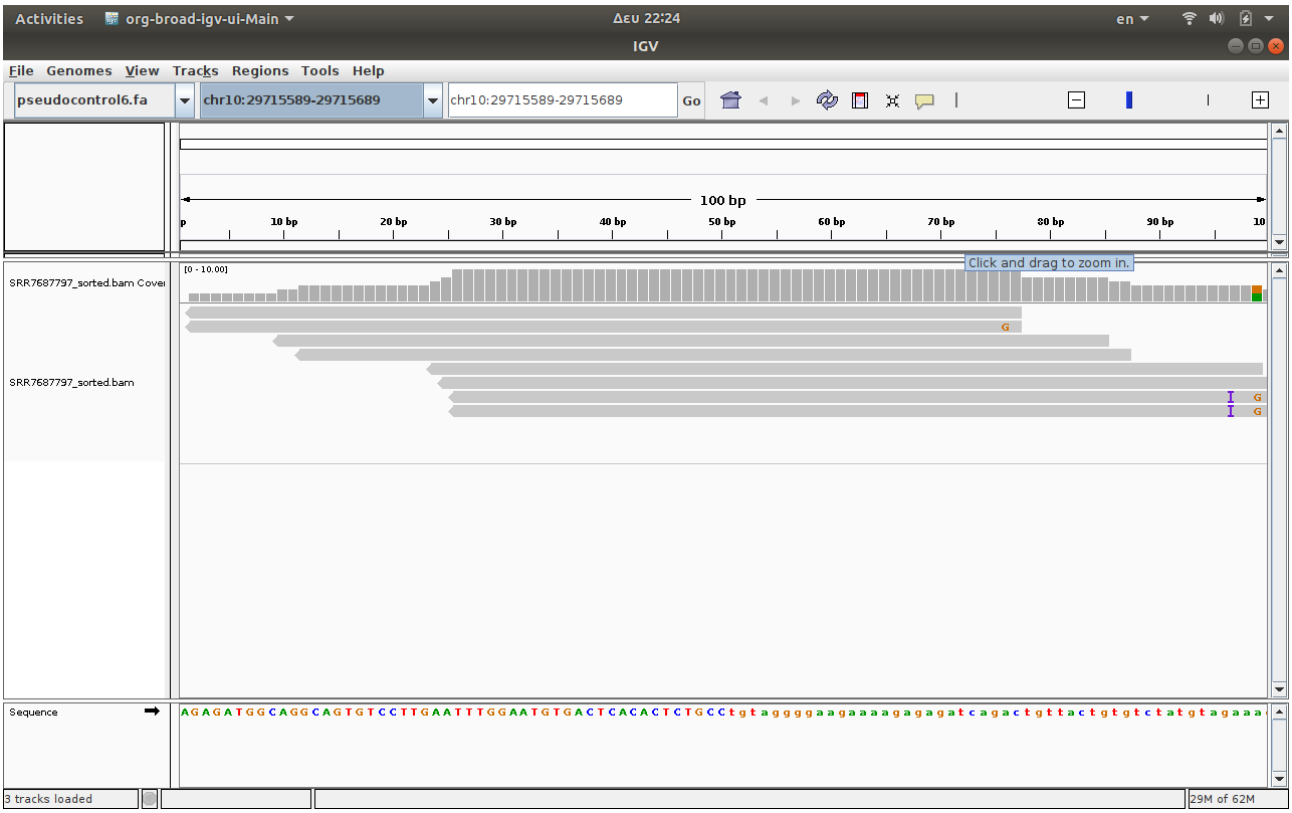
The accompanied IGV web tool allows users to upload FASTQ files and obtain visualized results.



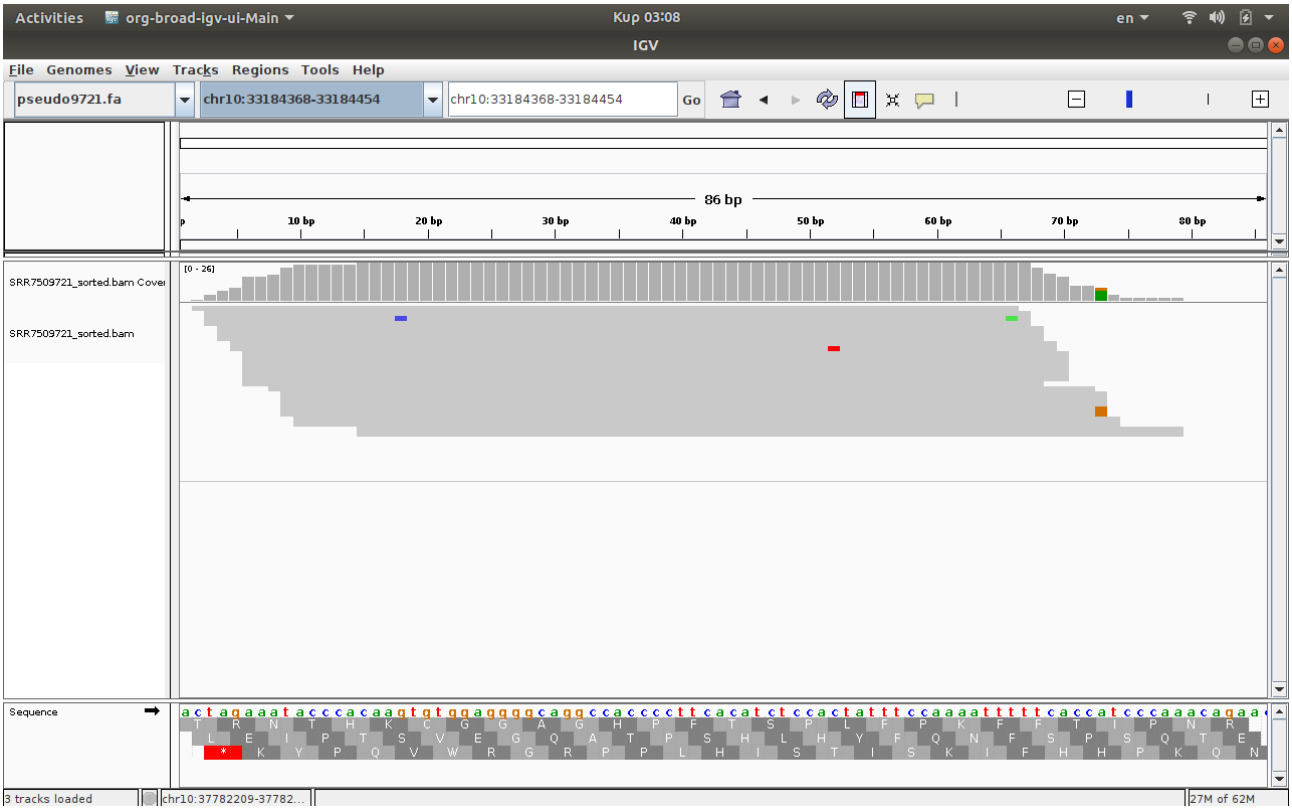
Pic.1 Visualization of a bam file of Healthy breast tissue (Health control) in chromosome 11 (chr11: 93478768-93478868) in the IGV viewer



Pic 2. Visualization of a bam file of breast cancer tissue in chromosome 11 (chr11:62149628-62149714) in the IGV viewer



Pic 3. Visualization of a bam file of healthy breast tissue (health control) in chromosome 10 (chr10:29715589-29715689) in IGV viewer



Pic 4. Visualization of a bam file of breast cancer tissue in chromosome 10 (chr10:33184368-33184454) in the IGV viewer

3.3 Method of statistical analysis

Data analysis was performed in Microsoft Office Excel 2016 and in IBM SPSS 24. Chromosomes were represented using frequencies. Independent samples t-test was used to compare mean differences between 2 independent samples that are normally distributed. Mann Whitney test was used to compare medians between 2 independent samples that are not normally distributed. Normality was tested using the Shapiro Wilk test. Significance was set at 5% (Field, 2013) [102].

3.4 Inferential statistics

3.4.1 Differences between health controls and breast cancers

Table 8 represents results of independent samples t-test for Normalized coverage of integration sites between health controls and breast cancers. Normalization mean value of health controls (M=0,000001) is statistically significant lower (t (10,010) =-3,210, p=0,009) than mean value of breast cancers (M=0,000013). **Figure 13** represents statistically significant differences

Table 8: *Independent samples t-test for normalized coverage of integration sites between health controls and breast cancers.*

Factor	Group	N	M	df	t	p-value
Normalized coverage of integration sites	Health controls	6	0,000001	10,010	-3,210	0,009
	Breast cancers	11	0,000013			

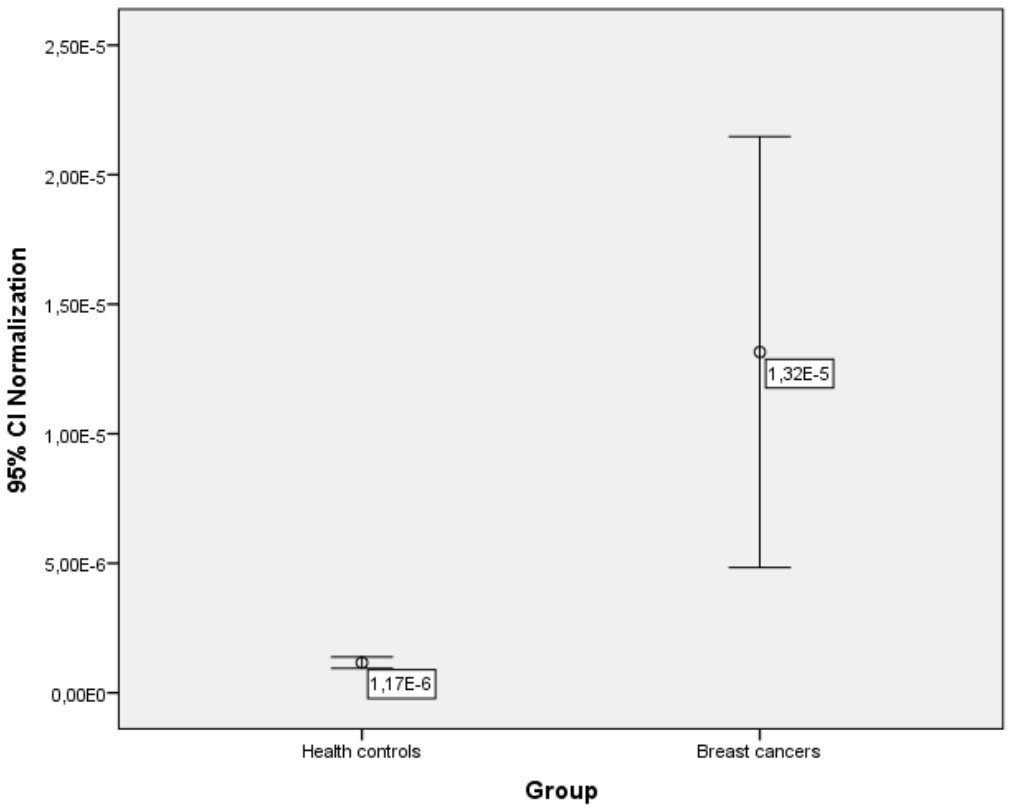


Figure 13: *Mean values for normalized coverage of integration sites between health controls and breast cancers*

3.4.2 Differences between treatment

Table 9 represents results of Mann Whitney test for normalized coverage of integration sites. Mean rank of normalized coverage of integration sites for participants who

did not receive treatment (9,00) was statistically significant higher (U=2, p=0,023) than mean rank of participants who received (4,29). **Figure 14** represents statistically significant differences.

Table 9: *Mann Whitney test for normalized coverage of integration sites in cases of treatment for breast cancers*

Factor	Treatment	N	Mean Rank	U	p-value
Normalized coverage of integration sites	No	4	9,00	2	0,023
	Yes	7	4,29		

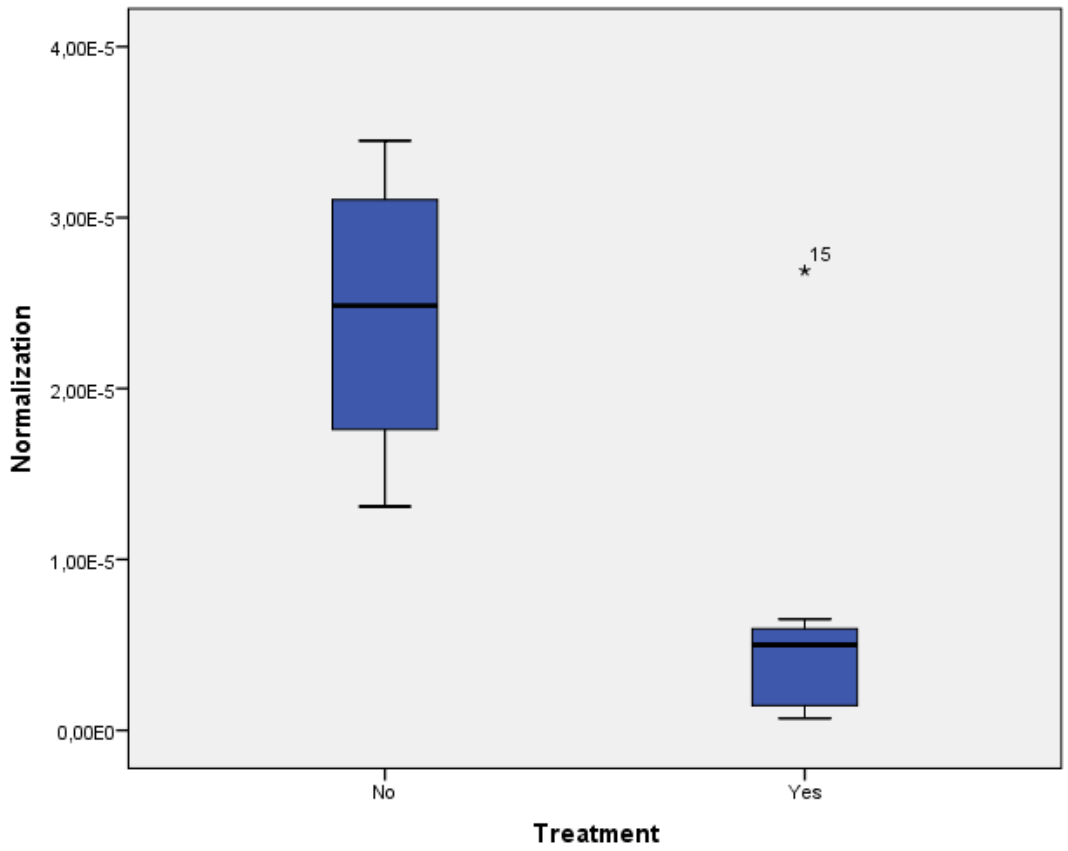


Figure 14: *Box plot for normalized coverage of integration sites in cases of treatment for breast cancers*

Discussion

Current analysis examined the expression of HERVs between breast cancer tissue and normal breast tissue. In addition, differences between breast cancers and health

controls in normalized coverage of integration sites were investigated as well, differences between breast cancers who received treatment and did not.

Breast cancers indicated higher levels of HERVs expression in normalized coverage of integration sites than health control participants. Examining breast cancers, participants who received treatment had lower values in normalized coverage of integration sites than those who did not .

To sum up, this study has indicated that breast cancer tissue has expressed higher levels of HERVs than normal breast tissue. Also, as far as it concerns breast cancer patients, those who received treatment had lower expression of HERVs than those who did not.

Historically, human retroviruses were for the first time associated to carcinogenesis in the early 1970s [80]. The HERVs have been utilized as biomarkers for staging and prognosis of malignant cancer [81] The HERV-K(HML-2) envelope protein is the most useful tool for the diagnosis of human breast cancer. HERV-K(HML-2) gene is present in breast cancer in contrast to its absence or very low levels in normal breast tissue [82]. The elevation of HERV-K(HML-2) antibodies and mRNA witnesses an initial stage breast cancer stage and further increase indicates metastasis [83].

The results of the present study showed that run-through transcripts over HERV integration sites are over-expressed in breast cancer samples in comparison with the normal tissues.

Conclusion

It has been recognized for many years that endogenous retroviruses and other retroelements contribute to malignant diseases. Targeting of the neoplastic cell will be an important issue to prevent jumping from the “frying-pan” into the fire. The expression as well as the over-expression of HERVs might become an important biomarker of breast cancer in the future.

References

1. https://acsjournals.onlinelibrary.wiley.com/doi/full/10.1002/cncr.25553
2. https://pubmed.ncbi.nlm.nih.gov/11624516/
3. Anbazhagan, R., et al., The development of epithelial phenotypes in the human fetal and infant breast. J Pathol, 1998. 184(2): p. 197-206.
4. Mammary gland development https://www.ncbi.nlm.nih.gov/pmc/articles/PMC3404495/?fbclid=IwAR1AEzbNakMc--YlOa9OlxlSGxhudsAo4zXlKb0PsaNdVDh3Atief4ba2kY
5. WHO. Cancer. 2015 [cited 2017 2017]; Available from: http://www.who.int/mediacentre/factsheets/fs297/en/.
6. Breast. 22 January 2017 01:57 UTC 29 January 2017 19:43 UTC]; Available from: https://en.wikipedia.org/w/index.php?title=Breast&oldid=761279505.
7. Korkola, J. and J.W. Gray, Breast cancer genomes--form and function. Curr Opin Genet Dev, 2010. 20(1): p. 4-14.
8. Gusterson, B.A., et al., Prognostic importance of c-erbB-2 expression in breast cancer. International (Ludwig) Breast Cancer Study Group. J Clin Oncol, 1992. 10(7): p. 1049-56.
9. Romond, E.H., et al., Trastuzumab plus adjuvant chemotherapy for operable HER2- positive breast cancer. N Engl J Med, 2005. 353(16): p. 1673-84.
10. Cui, X.S. and L.A. Donehower, Differential gene expression in mouse mammary adenocarcinomas in the presence and absence of wild type p53. Oncogene, 2000. 19(52): p. 5988-96.
11. Gelfand, V.I. and A.D. Bershadsky, Microtubule dynamics: mechanism, regulation, and function. Annu Rev Cell Biol, 1991. 7: p. 93-116.
12. Philp, A.J., et al., The phosphatidylinositol 3'-kinase p85alpha gene is an oncogene in human ovarian and colon tumors. Cancer Res, 2001. 61(20): p. 7426-9.
13. Takeda, A., et al., Role of the phosphatidylinositol 3'-kinase-Akt signal pathway in the proliferation of human pancreatic ductal carcinoma cell lines. Pancreas, 2004. 28(3): p. 353-8.
14. Brinton, L.A., et al., Menstrual factors and risk of breast cancer. Cancer Invest, 1988. 6(3): p. 245-54.
15. Brinton, L.A., R. Hoover, and J.F. Fraumeni, Jr., Reproductive factors in the aetiology of breast cancer. Br J Cancer, 1983. 47(6): p. 757-62.

16. *Brinton, L.A., R. Hoover, and J.F. Fraumeni, Jr., Epidemiology of minimal breast cancer. JAMA, 1983. 249(4): p. 483-7.*
17. *White, E., Projected changes in breast cancer incidence due to the trend toward delayed childbearing. Am J Public Health, 1987. 77(4): p. 495-7.*
18. *Breast cancer and hormone replacement therapy: collaborative reanalysis of data from 51 epidemiological studies of 52,705 women with breast cancer and 108,411 women without breast cancer. Collaborative Group on Hormonal Factors in Breast Cancer. Lancet, 1997. 350(9084): p. 1047-59.*
19. *Nelson, H.D., et al., Postmenopausal hormone replacement therapy: scientific review. JAMA, 2002. 288(7): p. 872-81*
20. *Τριχόπουλος, ed. Προληπτική Ιατρική και Δημόσια Υγεία. ed. Κ.Β. Πετρίδου Ε.*
21. *Gierisch, J.M., et al., Oral contraceptive use and risk of breast, cervical, colorectal, and endometrial cancers: a systematic review. Cancer Epidemiol Biomarkers Prev, 2013. 22(11): p. 1931-43.*
22. *Wolski, H., [Selected aspects of oral contraception side effects]. Ginekol Pol, 2014. 85(12): p. 944-9.*
23. *Bjelic-Radisic, V. and E. Petru, [Hormonal contraception and breast cancer risk]. Wien Med Wochenschr, 2010. 160(19-20): p. 483-6.*
24. *Poosari, A., et al., Hormonal contraceptive use and breast cancer in Thai women. J Epidemiol, 2014. 24(3): p. 216-20.*
25. *McPherson, K., C.M. Steel, and J.M. Dixon, ABC of breast diseases. Breast cancerepidemiology, risk factors, and genetics. BMJ, 2000. 321(7261): p. 624-8.*
26. *Tretli, S., Height and weight in relation to breast cancer morbidity and mortality. A prospective study of 570,000 women in Norway. Int J Cancer, 1989. 44(1): p. 23-30.*
27. *Goodwin, P.J., et al., Fasting insulin and outcome in early-stage breast cancer: results of a prospective cohort study. J Clin Oncol, 2002. 20(1): p. 42-51.*
28. *Del Giudice, M.E., et al., Insulin and related factors in premenopausal breast cancer risk. Breast Cancer Res Treat, 1998. 47(2): p. 111-20.*
29. *Suga, K., et al., Molecular significance of excess body weight in postmenopausal breast cancer patients, in relation to expression of insulin-like growth factor I receptor and insulin-like growth factor II genes. Jpn J Cancer Res, 2001. 92(2): p. 127-34.*
30. *Monninkhof, E.M., et al., Physical activity and breast cancer: a systematic review. Epidemiology, 2007. 18(1): p. 137-57.*

31. *Land, C.E., et al., Breast cancer risk from low-dose exposures to ionizing radiation: results of parallel analysis of three exposed populations of women. J Natl Cancer Inst, 1980. 65(2): p. 353-76.*
32. *Boice, J.D., Jr., et al., Frequent chest X-ray fluoroscopy and breast cancer incidence among tuberculosis patients in Massachusetts. Radiat Res, 1991. 125(2): p. 214-22.*
33. *Clemons, M., L. Loijens, and P. Goss, Breast cancer risk following irradiation for Hodgkin's disease. Cancer Treat Rev, 2000. 26(4): p. 291-302.*
34. *Nelson, H.D., et al., Risk factors for breast cancer for women aged 40 to 49 years: a systematic review and meta-analysis. Ann Intern Med, 2012. 156(9): p. 635-48.*
35. *Singletary, S.E., Rating the risk factors for breast cancer. Ann Surg, 2003. 237(4): p. 474-82.*
36. *Wang-Johanning F, Liu J, Rycaj K, Huang M, Tsai K, Rosen DG, Chen DT, Lu DW, Barnhart KF, Johanning GL. Expression of multiple human endogenous retrovirus surface envelope proteins in ovarian cancer. Int J Cancer. 2007;120:81–90. doi: 10.1002/ijc.22256.*
37. *Yu P. The potential role of retroviruses in autoimmunity. Immunol Rev. 2016;269:85–99.*
38. *Stoye JP. Studies of endogenous retroviruses reveal a continuing evolutionary saga. Nat Rev Microbiol. 2012;10:395–406.*
39. *Black, D.M., The genetics of breast cancer. Eur J Cancer, 1994. 30A(13): p. 1957-61.*
40. *Stoppa-Lyonnet, D., et al., BRCA1 sequence variations in 160 individuals referred to a breast/ovarian family cancer clinic. Institut Curie Breast Cancer Group. Am J Hum Genet, 1997. 60(5): p. 1021-30.*
41. *Pierce, L.J., et al., Effect of radiotherapy after breast-conserving treatment in women with breast cancer and germline BRCA1/2 mutations. J Clin Oncol, 2000. 18(19): p. 3360-9.*
42. *Newman, B., et al., Inheritance of human breast cancer: evidence for autosomal dominant transmission in high-risk families. Proc Natl Acad Sci U S A, 1988. 85(9): p. 3044- 8.*
43. *Skolnick, M.H., et al., Inheritance of proliferative breast disease in breast cancer kindreds. Science, 1990. 250(4988): p. 1715-20.*
44. *Board, C.N.E. (2012) Breast Cancer - Risk Factors.*
45. *Coffin JM, Hughes SH, Varmus HE. The Interactions of Retroviruses and their Hosts. In: Coffin JM, Hughes SH, Varmus HE, editors. Retroviruses. Cold Spring Harbor (NY): Cold Spring Harbor Laboratory Press; 1997.*
46. *Richie, R.C. and J.O. Swanson, Breast cancer: a review of the literature. J Insur Med, 2003. 35(2): p. 85-101.*
47. *Hankinson S, H.D., ed. Breast cancer. . Textbook of Cancer Epidemiology., ed. H.D. Adami HO, Trichopoulos D. 2002, Oxford University Press: New York. 301-339.*

48. *Denner J. Expression and function of endogenous retroviruses in the placenta. APMIS. 2016;124:31–43.*

49. *Tavassoli, F.A. and P. Devilee, Pathology and genetics of tumours of the breast and female genital organs. 2003, Lyon: International Agency for Research on Cancer ; Oxford : Oxford University Press [distributor].*

50. *Ernster, V.L., et al., Mortality among women with ductal carcinoma in situ of the breast in the population-based surveillance, epidemiology and end results program. Arch Intern Med, 2000. 160(7): p. 953-8*

51. *https://www.esmo.org/guidelines/breast-cancer*

52. *https://gco.iarc.fr/today/data/factsheets/populations/300-greece-fact-sheets.pdf*

53. *Weiss R. Spontaneous virus production from "non-virus producing" Rous sarcoma cells. Virology. 1967;32:719–23*

54. *Wang-Johanning F. et al.. Immunotherapeutic Potential of Anti-Human Endogenous Retrovirus-K Envelope Protein Antibodies in Targeting Breast Tumors. J Natl Cancer I 104, 189–210, doi: 10.1093/Jnci/Djr540 (2012).*

55. *Rycaj K. et al.. Cytotoxicity of human endogenous retrovirus K-specific T cells toward autologous ovarian cancer cells. Clinical cancer research: an official journal of the American Association for Cancer Research 21, 471–483, doi: 10.1158/1078-0432.CCR-14-0388 (2015).*

56. *Krishnamurthy J. et al.. Genetic Engineering of T Cells to Target HERV-K, an Ancient Retrovirus on Melanoma. Clinical cancer research: an official journal of the American Association for Cancer Research 21, 3241–3251, doi: 10.1158/1078-0432.CCR-14-3197 (2015).*

57. *Rooney M. S., Shukla S. A., Wu C. J., Getz G. & Hacohen N. Molecular and genetic properties of tumors associated with local immune cytolytic activity. Cell 160, 48–61, doi: 10.1016/j.cell.2014.12.033 (2015).*

58. ***https://www.ncbi.nlm.nih.gov/pmc/articles/PMC5292751/***

59. *Sarngadharan et al., 1972; Zhdanov et al., 1973*

60. *Chumakov et al., 1982; Repaske et al., 1983*

61. *Wang-Johanning et al., 2001, 2003, 2008*

62. *Burmeister et al. 2004*

63. *Contreras-Galindo et al., 2008*

64. *Golan et al., 2008*

65. *Zhou et al., 2016*

66. *Wang-Johanning et al., 2014*

67. *Wang-Johanning F, et al. Immunotherapeutic potential of anti-human endogenous retrovirus-K envelope protein antibodies in targeting breast tumors. J Natl Cancer Inst. 2012;104(3):189–210.*
68. *Johanning GL, et al. Expression of human endogenous retrovirus-K is strongly associated with the basal-like breast cancer phenotype. Sci Rep. 2017;7:41960.*
69. *Wang-Johanning F, et al. Human endogenous retrovirus K triggers an antigen-specific immune response in breast cancer patients. Cancer Res. 2008;68(14):5869–77.*
70. *Zhou F, et al. Chimeric antigen receptor T cells targeting HERV-K inhibit breast cancer and its metastasis through downregulation of Ras. Oncoimmunology. 2015;4(11):e1047582.*
71. *Zhao J, et al. Expression of human endogenous retrovirus type K envelope protein is a novel candidate prognostic marker for human breast cancer. Genes & cancer. 2011;2(9):914–22.*
72. *Naccarato AG, et al. Mouse mammary tumor virus (MMTV) - like exogenous sequences are associated with sporadic but not hereditary human breast carcinoma. Aging (Albany NY). 2019 Sep 13;11(17):7236–41.*
73. *Dossary R, et al. Prevalence of mouse mammary tumor virus (MMTV)-like sequences in human breast cancer tissues and adjacent normal breast tissues in Saudi Arabia. BMC Cancer. 2018;18:170.*
74. *Salmons B, et al. Revisiting a role for a mammary tumor retrovirus in human breast cancer. Int J Cancer. 2013 Oct 1;133(7):1530–5.*
75. *https://www.ncbi.nlm.nih.gov/pubmed/22067224*
76. *Field A. (2013) Discovering Statistics Using SPSS. London: Sage Publications Ltd*
77. *Belshaw R, Pereira V, Katzourakis A, Talbot G, Paces J, Burt A, Tristem M. Long-term reinfection of the human genome by endogenous retroviruses. Proc Natl Acad Sci. 2004;101:4894–4899. doi: 10.1073/pnas.0307800101.*
78. *https://bedtools.readthedocs.io/en/latest/content/tools/getfasta.html*
79. *https://www.ncbi.nlm.nih.gov/books/NBK158900/*
80. *Torre LA, Bray F, Siegel RL, et al. Global cancer statistics, 2012. CA Cancer J Clin 2015; 65:87.*
81. *Hulka BS and Moorman PG (2001). Breast cancer: hormones and other risk factors. Maturitas 38, 103–113; discussion 113–106.*
82. *Bittner J (1936). Some possible effect of nursing on the mammary tumor incidence. Science 84, 62.*
83. *https://gco.iarc.fr/today/data/factsheets/populations/300-greece-fact-sheets.pdf*

Figures