Ph.D. Thesis in Biomedicine and Biotechnology

Study of the epi- / genomic dysregulation of breast cancer in women under 35 years and evaluation of cellular models

Supervisors: Dra. Gloria Ribas Despuig
Dra. Ana Lluch Hernández

Tutor: Dr. Juan Saus Mas

Department of Biochemistry and Molecular Biology

Faculty of Biology

VNIVERSITAT DE VALÈNCIA

Sara Oltra Sanchis

Valencia, October 2018

The co-supervisors of the present doctoral thesis, Dr. Ana Lluch Hernández, Full Professor of Medicine at the Faculty of Medicine of the University of Valencia and Dr. Gloria Ribas Despuig, Senior Associate Researcher of the Cancer Functional Genomics Group within the Dept. of Medical Oncology at the Biomedical Research Institute – INCLIVA,

CERTIFY:

That, **Sara Oltra Sanchis**, graduated in Biology from the University of Valencia, has carried out the present Doctoral Thesis **"Study of the epi- / genomic dysregulation of breast cancer in women under 35 years and evaluation of cellular models"** and that, to their judgement, meets all the necessary requirements in order to qualify for the **PhD in Biomedicine and Biotechnology**, for which purpose it will be presented at the University of Valencia. The work has been done under their direction, authorizing its presentation before the Qualifying Court. And for this purpose, this certificate is extended.

Valencia, October 2018.

Thesis Co-supervisors:

Dra. Ana Lluch Hernández

Faculty of Medicine,
University of Valencia

Dra. Gloria Ribas Despuig

Cancer Functional Genomics Group
Biomedical Research Institute-INCLIVA

Thesis Tutor:

Dr. Juan Saus Mas

Department of Biochemistry and Molecular Biology,
University of Valencia

This thesis has been carried out in the Cancer Functional Genomics Group of the Biomedical Research Institute – INCLIVA, in collaboration with the Hospital Clinico of Valencia, under the supervision of Dr. Ana Lluch Hernández and Dr. Gloria Ribas Despuig.

This Doctoral Thesis was supported by grants from:

- Formación de profesorado Universitario (FPU) research training fellow from the Ministerio de Educación Ciencia y Deporte, FPU13/04976, (Spanish Government).

- Project from the The Ministry of Economy and Competitiveness and Carlos III Health Institute PI13/00606 and FEDER funding.

- Private patients Foundation LeCado-Proyecto Flor de Vida

MINISTERIO DE CIENCIA E INNOVACIÓN

Instituto de Salud Carlos III

le Cadó fundación contra el cáncer de mama

GOBIERNO DE ESPAÑA

MINISTERIO DE EDUCACIÓN, CULTURA Y DEPORTE

"*Hay que perseverar y, sobre todo, tener confianza en uno mismo*"

Marie Curie

Agradecimientos

Agradecer a todas las personas que han formado parte de esta etapa de mi vida tan maravillosa y a la vez con obstáculos, desilusiones e ilusiones, subidas y bajadas, alegrías y no tantas, pero que te hacen más fuerte para seguir peleando en este mundo de la investigación.

En primer lugar, agradecer a Gloria su acogida en el labo, que desde el primer momento me ha dado su apoyo y confianza. Con ella he aprendido muchísimo y por eso la admiro tanto. Gracias Gloria por tu lucha para dirigir un grupo de investigación tan maravilloso.

Agradecer a la Dra Ana Lluch la creación de este grupo y a la Fundación LeCadó su fiel aportación a nuestra investigación.

A lo largo del tiempo que he estado en el labo han pasado personas extraordinarias y de todas he aprendido algo. A Maider y Maria, por vuestra paciencia para enseñarme desde el primer día en que llegué, por vuestros consejos y ánimos. A Marta y Maite por vuestro apoyo y cariño incluso en la distancia. Porque somos las chicas del labo 3. A Fany, que apareciste hace unos años, pero es como si te conociese de toda la vida, gracias por tus consejos de vida, tus risas y tu cariño. Gracias Pepa por tus consejos y charlas de vuelta a casa.

Gracias a todos los que hacéis que el lab sea un sitio de trabajo y a la vez de encuentro con amigos como Regi, Fer, Paco, Ana… Vuestras risas son medicina.

Gracias a mis amigas, por esas cervecitas recuperadoras en las que intentamos arreglar el mundo, nuestro mundo.

A vosotros Papa y Mama, por vuestro apoyo incondicional aun no sabiendo de qué os hablo cuando hablo de trabajo. Por acompañarme y confiar siempre en mí. Por animarme a no abandonar nunca y mostrar siempre vuestro orgullo de padres. Por ser los mejores padres del mundo. Vos vull molt.

A Maria, mi amiga, mi confidente, mi compañera científica, mi hermana. Gracias por esos momentos de desahogarnos de vuelta a casa e intentar solucionar nuestros experimentos fallidos. Por eso y por mucho más que no cabe ni en dos tesis.

A Javi, por estar siempre ahí apoyando mis decisiones y acompañándome en mi vida.

INDEX

Figure Index

Table Index

Annex Index

Abbreviations

3'UTR	3'-untranslated region
AIs	Aromatase inhibitors
AJCC	American Joint Committee on Cancer
AMR	Amplification Mix Restore reagent
ATCC	American Type Culture Collection
ATM	Ataxia telangiectasia mutated gene
BC	Breast cancer
BCO	Breast cancer affecting women older than 45 years old
BCVY	Breast cancer affecting women younger than 35 years old
BMI	Body mass index
bs-DNA	Bisulfite-converted DNA
BSA	Bovine serum albumin
CD340	340 differentiation cluster
CDH1	E-cadherin
CLO	Breast cancer cell lines from women older than 45 years old
CLVY	Breast cancer cell lines from women younger than 35 years old
cDNA	complementary DNA
CHEK2	Cell cycle checkpoint kinase 2
CNA	Copy Number Aberration
Ct	Threshold cycle
DCIS	Ductal carcinoma *in situ*
DFS	Disease-free survival
DMSO	Dimethylsulfoxide
DNAm	DNA methylation
DNMTs	DNA methyl transferases
DNP	2,4-dinitrophenol
dsDNA	Double-stranded DNA
ECM	Extracellular matrix
EGFR	Human epidermal growth factor receptor
EGFR / ERBB	Epidermal growth factor receptor family
EMT	Epithelial-mesenchymal transition
ENCODE	The ENCyclopedia Of DNA Elements consortium
EPICarray	Infinium MethylationEPIC BeadChip
ER	Oestrogen receptor
ERBB2	Homologous gene 2 of the avian viral erythrobatic leukemia oncogene
EtBr	Ethidium bromide

EUSOMA	European Society of Breast Cancer Specialists
FANTOM	Functional Annotation of the Mammalian Genome
FDR	False discovery rate
FFPE	Formalin fixed paraffin-embedded
GEO	Gene Expression Omnibus
GLM	Generalized linear models
HER2/EGFR2	Human epidermal growth factor receptor 2
HM450K	Infinium HumanMethylation450 BeadChip
HRT	Hormone replacement therapy
KDE	Kernel density estimation
LCIS	Lobular carcinoma *in situ*
LHRH	Luteinizing hormone-releasing hormone
lincARNs	Large intergenic non-coding RNAs
metEPICVal	Samples included in the Illumina Infinium MethylationEPIC BeadChip array
miRNAs	micro-RNAs
mRNA	messenger RNA
mTOR	Rapamycin mammary route
NO	Normal tissue samples from women older than 45 years old
NVY	Normal tissue samples from very young women (<35 years old)
OS	Overall survival
PARP	Poly-(adenosine diphosphate)-ribose polymerase
PCR	Polymerase chain reaction
PBS	Phosphate buffer saline
PI3K	Phosphatidyl inositol kinase
piRNAs	PIWI-interacting RNAs
PPR	Primer Pre Restore
PR	Progesterone receptor
PTHrP	Parathyroid hormone-related protein
qRT-PCR	Real-time quantitative PCR
RFS	Relapse-free survival
RIN	RNA Integrity Number
RMA	Robust Multichip Average
SAM	S-adenosyl-Lmethionine
SD	Standard deviation
snoRNAs	small nucleolar RNAs
SNP	Single Nucleotide Polymorphism
Taq	*Termus aquaticus*
TCGA	The Cancer Genome Atlas
TFBS	Transcription factor biding sites
TNM	Tumour size (T), the lymph nodes affectation (N), and the appearance of distant metastasis
TSS	Transcription start site
UTR	Untranslated region

Abstract

Breast cancer has the highest incidence rate of all cancers in women worldwide. This cancer usually has an excellent prognosis with an 80% of survival rates. Although women under 35 years of age represent 3-4% of the total cases diagnosed, their tumours are characterized by being larger, more proliferative, and with an overrepresentation of more aggressive subtypes and worse prognosis. Neoplastic transformations are associated with alterations in DNA methylation, including both global hypomethylation and gene-specific hypermethylation. On the other hand, miRNAs are small non-coding molecules of RNA with an important role in the regulation of gene expression as part of the epigenetic machinery, but they can also be epigenetically modified by DNA methylation.

One of the aims of the present work is to establish a miRNA expression profile characteristic of women younger than 35 years using breast cancer cell lines commercially available and identify a good cellular model for the study of breast cancer in young women. Moreover, we want to evaluate methylation differences (global methylation and methylation of genes encoding miRNAs) between young and old women with breast cancer and their DNA epigenetic age.

The miRNA expression profile was carried out using Affymetrix Genechip miRNA 2.0 microarray that includes 1 100 human miRNAs whose expression was analysed in 5 breast cancer cell lines: HCC1500, HCC1937, MDA-MB-231, MDA-MB-468 and HCC1806. Additionally, DNA methylation studies were performed using the Illumina Human Methylation 850k Beadchip, analysing 21 samples from young women and 13 from older counterparts.

We identified a different miRNA expression profile between breast cancer cell lines from old and young women that was compared with the miRNA profile identified in breast cancer patients. Our study has allowed us to characterize the two breast cancer cell lines, HCC1937 and HCC1500, both from young women, as good models for the study of breast cancer in patients younger than 35 years. On the other hand, we observed a different methylation profile between age groups, also in regulatory regions of gene encoding miRNAs. Specifically, our results showed a global hypomethylation in tumours from young women as well as a specific hypermethylation, centred mainly in open-sea regions of the genome in younger patients. We also found an acceleration in the epigenetic age of tumours from young women that could be increasing their genome instability. *HDAC5* was significantly hypomethylated and overexpressed in young women with breast cancer. Functional studies with cell lines showed a significant reduction in the proliferation and migration when the *HDAC5* gene was inhibited by the HDACi inhibitor, LMK-235. This inhibition was evident in the triple-negative molecular subtypes of both age groups. The hypomethylation of miR-124-2 was associated with poor survival in patients with breast cancer younger than 35 years old compared to older women.

The differences observed in young patients with breast cancer in our work and others, reinforce the hypothesis that breast cancer that occurs in very young women has potentially unique, aggressive and complex biological features and more efforts should be done to achieve a more personalized treatment.

Resumen

El cáncer de mama tiene la tasa de incidencia más alta de todos los cánceres en mujeres en todo el mundo. Este cáncer suele tener un excelente pronóstico con un 80% de supervivencia. Aunque las mujeres menores de 35 años representan el 3-4% del total de casos diagnosticados, sus tumores se caracterizan por ser más grandes, más proliferativos y con una mayor representación de subtipos más agresivos y de peor pronóstico. Las transformaciones neoplásicas se asocian con alteraciones en la metilación del ADN, que incluyen tanto la hipometilación global como la hipermetilación específica. Por otro lado, los miRNAs son pequeñas moléculas no codificantes con un papel importante en la regulación de la expresión génica como parte de la maquinaria epigenética, pero también pueden modificarse epigenéticamente mediante la metilación del ADN.

Uno de los objetivos del presente trabajo es establecer un perfil de expresión de miRNAs característico de mujeres menores de 35 años utilizando líneas celulares de cáncer de mama disponibles comercialmente e identificar un buen modelo celular para el estudio del cáncer de mama en mujeres jóvenes. Además, queremos evaluar las diferencias de metilación (metilación global y metilación de genes que codifican miRNAs) entre mujeres jóvenes y mayores con cáncer de mama y evaluar su edad biológica a través del análisis de la metilación del ADN.

El perfil de expresión de miRNAs se llevó a cabo utilizando Affymetrix Genechip miRNA 2.0 microarray que incluye 1 100 miRNAs humanos cuya expresión se analizó en 5 líneas celulares de cáncer de mama: HCC1500, HCC1937, MDA-MB-231, MDA-MB-468 y HCC1806. Además, los estudios de metilación del ADN se realizaron con Illumina Human Methylation 850k Beadchip, analizando 21 muestras de mujeres jóvenes y 13 de mayores.

Se identificó un perfil de expresión de miRNAs diferente entre líneas celulares de cáncer de mama de mujeres mayores y jóvenes que se comparó con el perfil de miRNAs identificado en pacientes con cáncer de mama. Nuestro estudio nos ha permitido caracterizar las dos líneas celulares de cáncer de mama, HCC1937 y HCC1500, ambas de mujeres jóvenes, como buenos modelos para el estudio del cáncer de mama en pacientes menores de 35 años. Por otro lado, observamos un perfil de metilación diferente entre los grupos de edad, también en las regiones reguladoras de genes que codifican miRNAs. Específicamente, nuestros resultados mostraron una hipometilación global en tumores de mujeres jóvenes, así como una hipermetilación específica, centrada principalmente en las regiones “opensea” del genoma, en pacientes más jóvenes. También encontramos una aceleración en la edad epigenética de los tumores de mujeres jóvenes que podrían aumentar la inestabilidad del genoma. *HDAC5* estaba significativamente hipometilado y sobreexpresado en mujeres jóvenes con cáncer de mama. Los estudios funcionales con líneas celulares mostraron una reducción significativa en la proliferación y migración cuando el gen *HDAC5* fue inhibido por el inhibidor de HDACi, LMK-235. Esta inhibición fue evidente en los subtipos moleculares triples negativos de ambos grupos de edad. La hipometilación del miR-124-2 se asoció con una peor supervivencia en pacientes con cáncer de mama menores de 35 años en comparación con las mujeres mayores.

Las diferencias observadas en pacientes jóvenes con cáncer de mama en nuestro trabajo y otros, refuerzan la hipótesis de que el cáncer de mama que ocurre en mujeres muy jóvenes tiene características biológicas potencialmente únicas, más agresivas y complejas y se deben hacer más esfuerzos para lograr un tratamiento más personalizado.

INTRODUCTION

INTRODUCTION

1. Cancer

Cancer is a term for diseases in which abnormal cells divide without control and can invade nearby tissues. Cancer cells can also spread to other parts of the body through the blood and lymph systems. There are several main types of cancer and all of them are known as a malignant neoplasm. The word is from Ancient Greek νέος- neo "new" and πλάσμα plasma "formation, creation". Formerly the term tumor was used for any type of swelling. Nowadays in modern medicine, a tumor is called a neoplasm that, encapsulated or not, exceeds the limits of the tissue forming a protuberance. Not all neoplasms will end up bulging, nor being malignant. Neoplasms can be benign, premalignant (carcinoma *in situ*) or malignant (invasive carcinoma or cancer).

Malignant neoplasm is a general term for a heterogeneous group of several diseases, there are more than 200 types of cancer described, which have as common characteristics uncontrolled cell growth and the ability to invade adjacent or distant tissues, this process is called metastasis. The features of a cancerous process are well defined, being the most relevant the evasion of the cellular death, cellular immortalization, evasion of the growth suppression system, activation of the invasion capacity and metastasis, induction of angiogenesis and aberrant maintenance of proliferation signals.

Being a complex disease of variable penetrance, the risk of cancer depends on a large number of factors, both environmental and genetic, which may vary depending on the type of cancer.

1.1 Breast cancer

The ancient Egyptians were the first to describe breast tumours 3500 years ago in two different documents: The Edwin Smith Surgica Papyrus and The Ebers Papyrus. The latter is known as one of the most important medical documents of the ancient Valley of the Nile. The papyrus document was produced at the age of the pyramid between 3500 and 2500 B.C. that is around the Neolithic era (1). The papyrus of Ebers, was bought by the Egyptologist George Ebers in 1872 who published a facsimile with an English-Latin vocabulary of the document in 1875 (2).

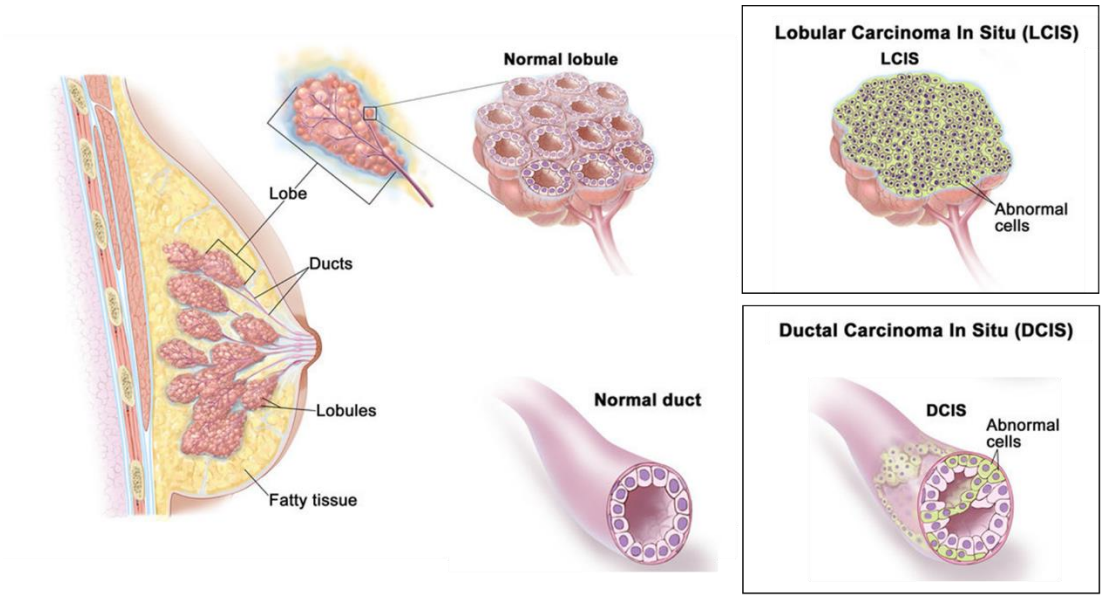
Breast cancer has been known to mankind since ancient times. It has been mentioned in almost every period of recorded history. Because of the visible symptoms especially at later stages the lumps that progress to tumours have been recorded by physicians from early times. This is more so because, unlike other internal cancers, breast lumps tend to manifest themselves as visible tumours.

Earlier, however, it was a matter of taboo and embarrassment that meant detection and diagnosis was rare. The mention of breast cancers in literature beyond medical journals and books was rare. Involvement of more women and actively bringing out the disease into the open is a recent phenomenon that is around three or four decades old. In the 1990's the symbol of breast cancer - the pink ribbon – brought out a revolution against this cancer.

2. Mammary gland

2.1 Anatomy

The mammary gland is an organ present in mammals whose function is to produce the milk that will feed the offspring. It is a specific type of apocrine gland, whose secretion is produced by excision of part of the cell along with secreted vesicles (Figure 1).



Modified from 2012 Teresa Winslow LLC

Figure 1. Anatomy of mammary gland. It represents a tumour process that originates either a lobular carcinoma *in situ* (LCIS) or a ductal carcinoma *in situ* (DCIS) according to the cells of origin.

The basic components of a mature mammary gland are the alveoli (hollow cavities, a few millimetres large) lined with milk-secreting cuboidal cells and surrounded by myoepithelial cells. These alveoli join to form groups known as lobules. Each lobule has a lactiferous duct that drains into openings in the nipple. The myoepithelial cells contract under the stimulation of oxytocin, excreting the milk secreted by alveolar units into the lobule lumen toward the nipple.

Maintaining the correct polarized morphology of the lactiferous duct tree requires another essential component – mammary epithelial cells extracellular matrix (ECM) which, together with adipocytes, fibroblast, inflammatory cells, and others, constitute mammary stroma (3). Mammary epithelial mainly contains myoepithelial basement membrane and the connective tissue. They not only help to support mammary basic structure, but also serve as a communicating bridge between mammary epithelia and their local and global environment throughout this organ's development.

2.2 Development

The mammary gland is the only organ whose development occurs in different stages, and reaches its maturity after embryonic development and birth (3, 4).

2.2.1 Embryonic and post-natal development

During embryonic development a rudimentary ductal system is formed, dependent on systemic and maternal hormones, as well as the paracrine communication between epithelial and mesenchymal cells by the parathyroid hormone-related protein (PTHrP). In the first weeks after birth there is a brief development of the milk ducts, but mostly this development occurs at puberty (5). Oestrogen promotes the differentiation of rudimentary ducts, and its growth is controlled by the epidermal growth factor EGF and extracellular matrix components (6).

2.2.2 Pregnancy and lactation

In the early stages of pregnancy, increases in oestrogen and progesterone promote the development of the tree's milk ducts, which reach their maximum during lactation. After

lactation, the involution of the mammary gland is triggered, and the majority of epithelial and glandular mammary tissue is lost drastically by programmed cell death or apoptosis. (4, 7, 8).

3. Epidemiology and risk factors of breast cancer

3.1 Incidence and mortality

Breast cancer is the leading cause of death from cancer among women worldwide (Figure 2). In 2012, 1.7 million of cases and 512 900 deaths from breast cancer were estimated (9). The highest registers for breast cancer are observed in countries of Eastern Europe and the United States, however, the lowest rates are registered in Africa and Asia (with the exception of Israel, where they present the highest rates). On the other hand, the highest mortality rates are found in the United States among black women, while the lowest are registered in Korean women (10).

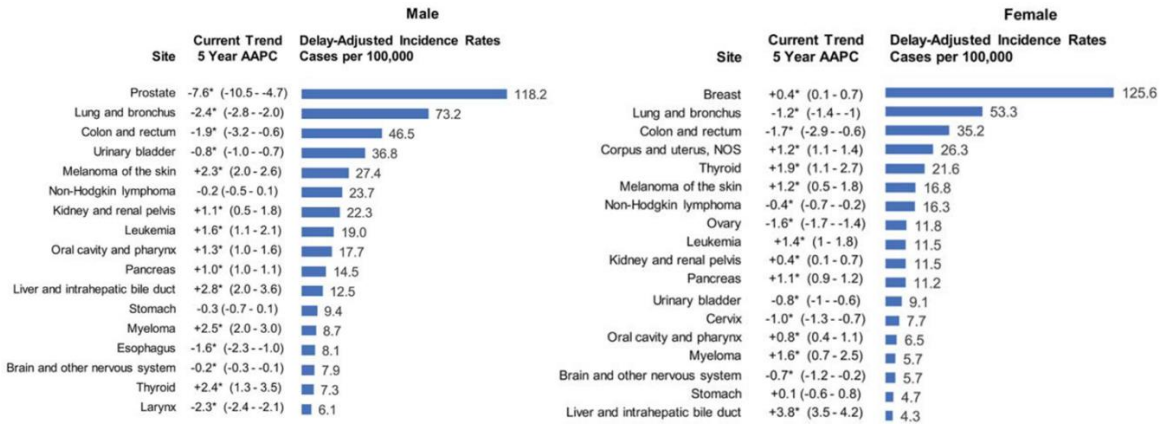


Figure 2. Age-standardized, delay-adjusted incidence rates and recent trends (2010-2014) are illustrated for the 17 most common cancers in men and the 18 most common cancers in women for all races/ethnicities combined and by sex. The 5- year average annual percent change (AAPC) is based on the join point trend from 1999 to 2014. An asterisk indicates that the AAPC is statistically significantly different from zero (2- sided t test or Z test; p-value<0.05). An update from Cronin et al 2018 (11) **.**

3.2 Breast cancer in Spain

In Spain, the same as at developed countries, cancer represents the main public health problem. Between 1989 and 2008, there were 115 080 deaths from breast cancer in women, which represents 18% of deaths in women related to cancer and 4% of total deaths in women. The areas with the highest cancer mortality rate are Catalonia and the Balearic Islands. Although this excess of mortality has been mitigated over time, it has increased in areas of western Andalusia in cities such as Huelva, Seville and Cádiz (12).

Breast cancer has a health, social and economic impact, because it affects mainly women between 45 and 65 years old and it has become in a chronic disease. In addition, breast cancer is the first cause of cancer in women. Currently about 16 000 new cases are diagnosed (13), and the standardized incidence rate has gone from being the lowest in Europe in the early 80s (54.8 cases per 100,000 women-year) to occupy an intermediate position among European countries in 2006 (93.6 cases per 100,000 women-year), according to International Agency for Research on Cancer (IARC) data.

3.3 Risk factors

3.3.1 Sex

Breast cancer is a rare pathology in men. The fact is that those men affected with breast cancer belong to a family with breast cancer antecedents. Therefore, just being a woman is the biggest risk factor for developing breast cancer (14).

3.3.2 External hormonal factors

Current users or those who have used hormone replacement therapy (HRT) have a higher risk of breast cancer. Before the association of breast cancer risk with HRT was known, many postmenopausal women carried out HRT for many years to counteract characteristic symptoms of menopause such as fatigue or hot flashes, as well as the reduction of bone mass. The evidence, from both epidemiological studies and clinical trials, indicates that the risk among the population of women receiving HRT depends on the duration of treatment, reducing by each year reaching the level of normal risk after 5 years of cessation in the treatment (15, 16).

The risk is also 15% higher in current users of oral contraceptives, however after 10 years without using them the risk equals to non-users, it is considered only a risk factor in young women, who have an incidence of basal breast cancer, so the use of oral contraceptives is not relevant as a risk factor for breast cancer in the general population (15, 17).

3.3.3 Internal hormonal factors

The time elapsed between the first period of menstruation and the first pregnancy is one of the main factors for breast cancer susceptibility (18). The early age of menarche (early puberty due to an increase in proteins and the percentage of fats in the diet composition) and the late age of the first pregnancy have been increased in the twenty-first century. These factors lead to an increase in hormonal stress resulting in an increase in the predisposition to the development of breast cancer throughout the life of these women (19).

Both, pregnancy and lactation, increase breast cancer risk in the short term, probably due to the increase in oestrogen levels in the first months of pregnancy. It has been proposed the existence of an association between the involution of the mammary gland after lactation and the

increased risk of developing breast cancer in the 5 years after pregnancy, this fact is due, according to some authors, to the rearrangement of the extracellular matrix and modification of the microenvironment, facilitating the proliferation of possible pre-malignant cells (20, 21).

However, there has been a negative correlation between the duration of breastfeeding and the incidence of breast cancer. A prolonged breastfeeding (over 12 months) has been linked to a reduction in the breast cancer risk showing a positive impact on both the mother's and the child's health (22). This beneficial effect of breastfeeding is due to the suppression of ovulation during the same, reducing the risk by 4.3% for each year of lactation (23).

3.3.4 Age

Age and sex, is considered the main risk factors for breast cancer (24). Before 30 years, the incidence of breast cancer is low, but from then until 50 years suffers a considerable increase, doubling every 10 years until the menopause is reached. From this moment the incidence increases in a slower way, perhaps in relation to the hormonal changes that occur, such as the decrease in circulating oestrogen levels, until around 70 years of age that the highest incidence rates are reached (24). The difference in the incidence of cancer in relation to age can be explained by the accumulation of factors throughout the life of the woman or by oncogenic stimuli that initiate the process with greater frequency in the elderly woman (24). However, in young women, breast cancer tends to be more aggressive, to express fewer steroid receptors [estrogen receptor (ER) and progesterone receptor (PR)] and more EGFR or human epidermal growth factor receptor, than in older women (25).

3.3.5 Environmental factors and life style

One of the main risk factors associated with breast cancer is the alcohol intake (26). Different studies show that alcohol intake both at early age (20 years old) or later in adulthood (over 60 years), are independently associated with the increased cancer risk, even when the intake is at low levels (19). Many studies, focusing mainly on postmenopausal women, suggest that diets rich in fruit, vegetables and fibre, as well as the reduction in animal products consumption and refined carbohydrates, can reduce the risk of cancer in this group (27-32). There are fewer studies in this direction for the group of premenopausal women but the results are similar (33-37). PREDIMED study (38) has been the first to show the benefits of the Mediterranean diet, characterized by the use of extra virgin olive oil and its benefit in breast cancer prevention.

On the other hand, overweight has been associated with an increased risk of breast cancer in postmenopausal women, finding a correlation between body mass index (BMI) and duration of overweight or obesity in postmenopausal women (39, 40). However, obesity in childhood and adolescence have been associated with a reduction in the breast cancer incidence in premenopausal women (41-45), but with an increase in the risk of suffering other cancer types such as endometrium (46), colon rectal (47) and pancreatic cancer (48).

Physical exercise has been associated with a certain protective effect against the development of breast cancer by regulating ovulatory cycles, the metabolism of endogenous hormones, BMI, the body fat composition and the immunity provided by this activity (49, 50). According to data obtained from the EPIC study, physical activity exerts a protective effect on the breast cancer risk, which decreases by 29% in premenopausal women and 19% in postmenopausal women (51).

Summarizing, alcohol consumption, overweight and sedentary lifestyle are related to 21% of global deaths from breast cancer (52).

3.3.6 Genetic predisposition

Women with close relatives who have been diagnosed with breast cancer have a higher risk of developing the disease. First-degree female relative (sister, mother, daughter) diagnosed with breast cancer doubled breast cancer risk. And this risk increases according to the number of first degree relatives affected, the age of diagnosis and the existence of bilateral breast cancer or breast cancer in men (53). About 10% of breast cancer cases have a family component, which can be attributed to genetic factors and genes of high, medium and low penetrance, mostly related to the metabolism and regulation of hormones, and DNA damage and repair.

In familial breast cancer, 20% corresponds to families with mutations in the high penetrance gens, *BRCA1* and *BRCA2* (*FANCD1*). Mutations in these genes increase the risk of breast cancer up to 80%, in addition to 15-40% of suffering from ovarian cancer and 65% of developing a second breast tumour. In these patients, breast cancer occurs at an early age and its evolution tends to have a poor prognosis. Both are tumour suppressor genes in the DNA repair pathway.

The frequency of mutations in the *BRCA1* and *BRCA2* genes varies depending on the population, but in general it remains low, with a prevalence of 1 in 860, and 1 in 740, respectively, explaining up to 25% of familial cases. There are other genes that can explain this familiar component, such as *TP53, PTEN, STK11* / *LKB, RAD51C*.

In addition to high-penetrance genes, mutations in many other genes cause an increased risk of breast cancer. This category includes mutations in *CHEK2, ATM, NBN, MRE11A, RAD50, BRIP1, PALB2* (54-56).

- *CHEK2* (Cell cycle checkpoint kinase 2) is a tumor suppressor gene. It codes for a kinase involved in the control of the cell cycle, regulating cell division and preventing uncontrolled growth.
- *ATM*, or ataxia telangiectasia mutated gene, codes for the phosphatidyl inositol kinase (*PI3K*), which is involved in the regulation of cell division control processes and DNA

repair. Together with the mTOR (rapamycin mammary route) protein, among others, they make up the group of class V phosphatidyl inositol-3-kinases. Mutations in this gene predispose to different types of cancer.

- *NBN* encodes the Nbs1 protein or nibrin, which complexes with MRE11A and RAD50 and binds to ATM. This complex plays an essential role in the detection of broken DNA chains and their repair, maintaining stability and regulating cell division. The accumulation of mutations in *NBN* causes an increased risk of developing cancer, and increase three times the breast cancer risk. It has been proposed that some inherited mutations in *RAD50* increase the risk of breast cancer, although the results are contradictory and some authors conclude that *RAD50* is not a risk factor.
- *BRIP1* and *PALB2*, also known as *FANCJ* and *FANCN*, respectively, take part of the gene group that make up the metabolic pathway associated with Fanconi anemia. Several genes of the same route have been associated with various cancers, including breast cancer.

Among all family cases diagnosed, the 65% of them do not present mutations in none of these genes, but they have a clear genetic or epigenetic component.

4. Heterogeneity and breast cancer classification

Breast cancer presents several classifications according to different aspects. It can be classified according to their histological origin, cell differentiation degree, stage, presence or absence of certain hormone receptors and molecular subtype. This is representative of the high heterogeneity of the disease, and the presence of multiple stratifications and groups, with

different prognosis and therapeutic approach. For this reason, is necessary to carry out an accurate classification and characterization of the tumour.

4.1 Histopathological classification

The tissue of the cells of the carcinoma and its histology are determined, also they can be *in situ*, if they have not crossed the tissue barriers. We can talk about invasive breast cancer, if they have infiltrated to others tissues.

In clinical practice, the most frequent types of breast cancer are:

- Invasive or infiltrating ductal carcinoma: Represents more than 80% of diagnosed cases, the cancer cells come from the cells that line the galactophore duct. Although this carcinoma can affect women of any age, it is more common in older ages. According to the American Cancer Society, approximately two-thirds of women who are diagnosed with invasive breast cancer are 55 years or older.
- Invasive or infiltrating lobular carcinoma: It is less common than ductal carcinoma, occurring in 10% of breast cancer patients. It originates in the cells that make up the lobes or glandular acini. Although, this type of carcinoma can appear at any age, invasive or infiltrating lobular carcinoma tend to appear around the age of 60.
- Medullary carcinoma: It represents around 3 to 5% of all breast cancer cases. It is called "medullary" carcinoma because the tumour is a soft and pulpy mass that resembles the medulla. It is more frequent in women from 45 to 55 years. Medullary carcinoma most often affects women who present a *BRCA1* mutation. The morphology of their cells is similar to aggressive and very anomalous cancer cells, but they do not act as such. Medullary carcinoma does not grow rapidly and usually does not spread outside the breast to the lymph nodes.

- Mucinous carcinoma: It represents around 2-3% of all cases of breast cancer. In this type of cancer, the tumour consists of abnormal cells embedded in accumulations of mucin. Mucinous carcinoma usually affects postmenopausal women, with an average age at diagnosis of 60 years or older. It is less likely to spread to the lymph nodes than other types of breast cancer and, moreover, is easier to treat.

4.2 Histological grade classification

Histological grade is an indicator of tumour aggressiveness. Low grade tumours (Grade I) are the most differentiated and with less mitotic activity, so they are the better prognosis group, compared to high grade tumours (Grade III), with greater anaplasia and proliferation, therefore, with worse prognosis. The intermediate group is represented by Grade II.

There is a statistically significant relationship between histological grade and disease-free survival (DFS) or overall survival (OS)(57). In 1988, Fischer observed a greater survival at 5 years among patients with a lower histological grade (58).

4.3 Stage classification

The stage is determined following the TNM staging system, a standardized method in which the cancer widespread is described and represents the primary tumour size (T), the lymph nodes affectation (N), and the appearance of distant metastasis (M). The stages are based on the 2010 staging manual of the AJCC (American Joint Committee on Cancer) (37).

4.4 Classification according to membrane receptors

Cells express membrane receptors for their interaction with the environment and to receive chemical stimuli that cause changes and consequences in the cell, such as hormones and growth factors. Regarding breast cancer cells, a characterization of three important receptors has been established to determine the prognosis and progression of breast cancer (59).

- Hormonal receptors: These are the hormone oestrogen receptors (ER) and progesterone receptor (PR). Abnormally high levels of oestrogen or its receptor promote tumour growth. A breast tumour is ER positive, when it expresses ER in its membrane, being detected by immunohistochemical tissue staining. Its survival rate is high, around 95% at 5 years when the cancer is early diagnosed.
- ERBB2 receptor: expressed by the homologous gene 2 of the avian viral erythrobatic leukemia oncogene (*ERBB2*), also called *CD340* (340 differentiation cluster), or *HER2/EGFR2* (Human epidermal growth factor receptor 2). This oncogene of the epidermal growth factor receptor family (*EGFR / ERBB*) is amplified, or overexpressed in approximately 10% to 30% of diagnosed breast cancers. Expression levels are low in normal breast epithelium, so tumours with *HER2* overexpression have a very proliferative and aggressive phenotype.

4.5 Molecular subtype classification

Perou and collaborators proposed the first molecular breast cancer classification, based on the analysis of the expression of certain genes through high throughput techniques (60). The classification involves five subtypes, with a characteristic expression profile representative of

each one and with a significant prognostic value and overall survival (Figure 3). These groups can separate themselves into ER+ and ER-, thus maintaining the previous molecular classifications (59, 61).

- Luminal subtypes: Subdivided into luminal A and luminal B. Both expressing groups of genes in a similar way to those normally expressed by the luminal epithelium of the breast, which is in contact with the lumen of the lactiferous duct. Both luminal A and luminal B subtypes express cytokines CK8 / 18, *GATA3*, and oestrogen receptors, so both would be part of the ER+ group. However, tumours classified or characterized as luminal A have a very high expression of ER, with low levels of proliferation genes, thus giving less aggressive phenotypes, low histological grade and very good prognosis. On the other hand, those denominated as luminal B express ER in smaller quantity and high levels of proliferation genes, resulting in tumours with higher histological grade and worse prognosis than luminal A. A significant proportion of tumours with *HER2* overexpression, but positive for endocrine receptors, it is grouped in this subtype, luminal B.

- HER2 subtype: This subtype is characterized by the presence of tumours with *HER2* overexpressed, apart from other genes in the HER pathway, located in the chromosomal band 17q12. They are completely negative for hormone receptors. Usually these tumours are aggressive, presenting a worse prognosis.

- Basal-like subtype: This subtype expresses mainly genes typical of basal or myoepithelial cells, such as *EGFR*, cytokines such as CK5, CK14, CK17, CD44, cadherins, caveolin 1 and 2 and nestin. They also give negative results for ER, PR and HER2, for that reason they

are also overlap with tumours defined as triple negative tumours, with a very aggressive and proliferative behaviour, with high histological grade, metaplastic areas, central necrosis, and nodular infiltration. Its histopathological characteristics resemble tumours typical of women with mutations in *BRCA1*.

- Normal-like subtype: It is not established as a representative subtype, and its clinical importance has not been determined, in addition, some authors have proposed that it is a false category, due to the quality of the tissue used, and may show a greater representation of fat cells. This category is grouped with samples of healthy breast tissue, expressing genes typical of adipocytes and other non-epithelial cells, including basal cells (59).

- Claudin-Low subtype: initially it was not defined as a subtype, it could be defined as a special basal-like subtype. They are ER, PR and HER2 negative. They are characterized by low expression levels of genes related to cell adhesion, such as claudins, ocluidins and E-cadherin. Although its expression resembles that of the basal-like subtype, it is inconsistent, showing low levels of luminal gene expression and high endothelial cell and lymphocyte cell markers. They are tumours with aggressive characteristics of the basal-like type, counting with a high number of carcinomas with histological grade II or III, which share features with the mesenchymal expression of squamous tumours.

Other breast cancer biomarkers

- Ki67: is the protein encoded by the *MKI67* gene, antigen identified by the monoclonal antibody Ki67, whose expression is evaluated as an indicator of tumour proliferation, acting as a prognostic and response marker, because the early suppression of *Ki67* after hormone therapy administration predicts the subsequent failure.

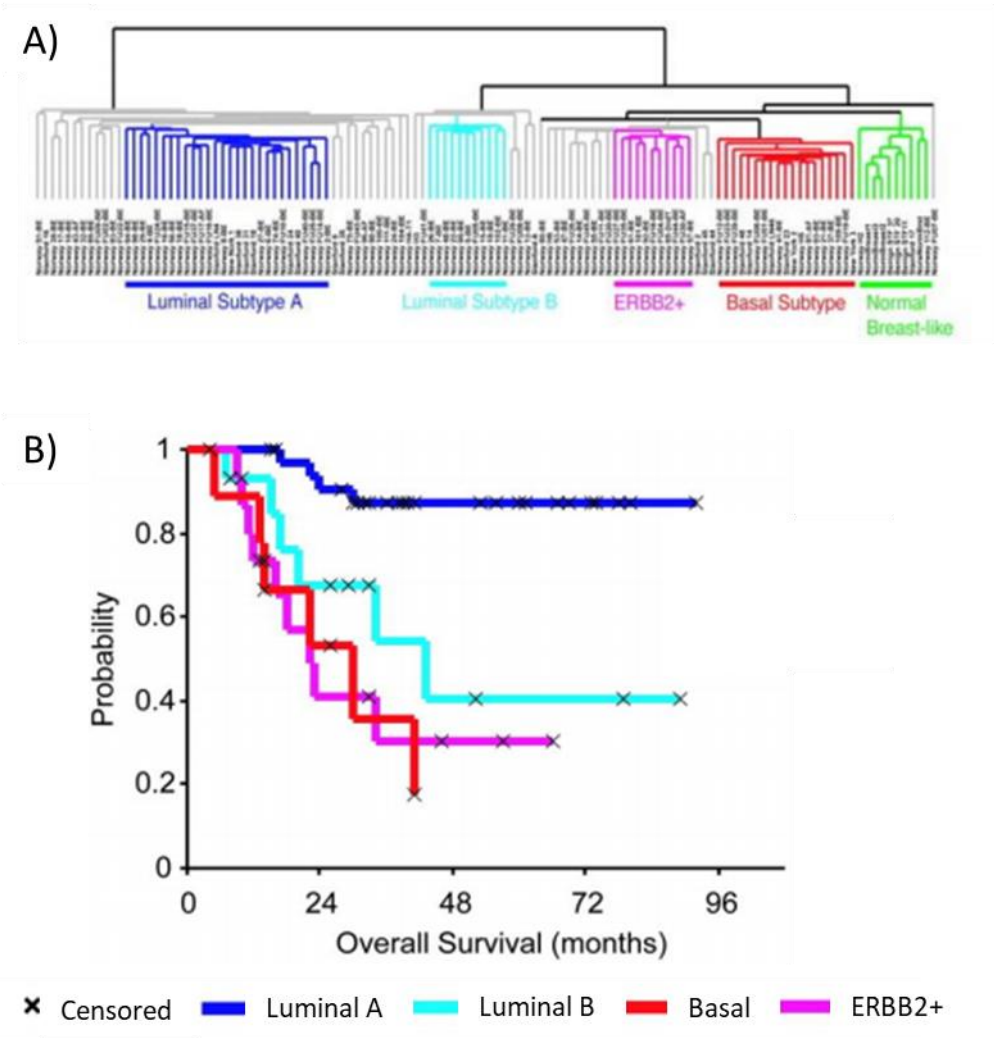


Figure 3. Molecular classification of breast cancer. Representation of the different groups in which breast cancer samples were classified: luminal A (blue), luminal B (cyan), ERBB2+ (pink), basal (red) and normal type (green) (**A**). Representation of the overall survival in months of the different subtypes described in A (**B**).

- P53: protein encoded by the *TP53* tumour suppressor gene. This gene is commonly mutated in cancer and its levels are studied in the clinical practice.
- EGFR: is a transmembrane receptor for epidermal growth factor, it has been found overexpressed in 10% of breast tumours, with a negative result for endocrine receptors, usually being basal-like.

4.6 Risk classification according to molecular profile

Gene expression studies by high throughput platforms have become a technique widely used in the study of diseases, including breast cancer. In addition, its value has been appreciated in the biomarkers identification with diagnostic and prognostic value and in an accurate patient classification.

In the clinical practice, gene panels are used whose expression classifies breast tumours according to their risk, facilitating the decision of treatment and an accurate prognosis. The most used are PAM50, which analyses the expression of 58 genes, classifying the tumour into one of the five subtypes and giving a measure of the risk of recurrence at 5 years; ONCOTYPE, similar to PAM50, analyses 21 genes in ER positive breast cancer cases, determining the likelihood of recurrence and the usefulness of chemotherapy to reduce that risk; Finally, MAMMAPRINT studies a panel of 70 genes and indicates the recurrence risk at 10 years (62-64).

There are multiple classifications according to molecular profiles, which may or may not be established as prognostic or therapeutic markers, but the heterogeneity and quantity of underlying processes that may be the object of a cancerous process or modified by this process is not clear (65).

5. Breast cancer treatment

The treatment of breast cancer depends partially on the state of the disease. According to the Breast Cancer Treatment there are five standard treatment types that are commonly used (66):

5.1 Surgery

Surgery is one of the methods used to eliminate breast cancer. The sentinel lymph node is the first one to receive lymphatic drainage where tumour formation will take place. A radioactive substance and / or blue marker is injected near the tumour, which flows through the lymphatic ducts, in this way, the first nodule to receive the injected substance is eliminated. Once the cells extracted by the biopsy have been analysed, the tumour is removed by surgery or mastectomy. If cancer cells have been observed in the sentinel node, it is desirable to remove more lymph nodes. There are different types of surgeries and we can distinguish between:

- Conservative surgery: it consists of eliminating cancerous tissue, as well as adjacent normal tissue, also known as partial mastectomy.
- Total mastectomy: complete elimination of the breast that presents the tumor. It is also known as a simple mastectomy.
- Modified radical mastectomy: consists of the complete elimination of the breast, as well as of the lymph nodes under the arm of the affected area, the area under the breast muscles, as well as part of the wall of the muscles of the breast.

5.2 Chemotherapy

Chemotherapy can be applied before surgery to remove the tumour in order to reduce the tumour size before the surgery. The treatment given before the surgery is known as preoperative or neoadjuvant therapy.

Even though cancer has been removed during surgery, many patients receive radiation therapy, chemotherapy or hormone therapy after surgery, with the goal of killing any remaining cancer cells. These treatments are called postoperative or adjuvant therapy.

Chemotherapy consists of the use of drugs to reduce the growth of tumour cells and can be applied systemically, affecting the entire body, or regional, acting on the area where the tumour is located.

5.3 Radiotherapy

Radiotherapy involves the use of high-energy x-rays or other types of radiation that kill or reduce the growth of cancer cells. There are two types of radiotherapy: external, applied outside the patient's body or internal, which consists in introducing the treatment directly to the tumour.

5.4 Hormonotherapy

Breast cancer is an extremely complex disease, but with the aim of make it simply, it can be separated into ER positive and ER negative, according to the oestrogen receptor expression (67). In the first case, the tumour uses oestrogen for its growth, establishing a clear target for its treatment that are focus on oestrogen production blocking its availability in order to reduce tumour growth. Endocrine therapies are directed to the ER positive group, which constitutes 64% of all tumours (68).

Oestrogen is one of the main female hormones regulating reproduction, it is produced by the ovaries until menopause, ceasing its activity. Post-menopausal women produce a limited amount of oestrogen, due to the adrenal glands produce androgens, a hormone that is metabolized into oestrogen by the aromatase enzyme.

- Oophorectomy: it is the simplest endocrine therapy, it consists in the surgical extirpation of the ovaries, reducing the production of oestrogen to a minimum. It is a drastic and invasive procedure, in which irreversible menopause is induced in a woman, depriving of her reproductive capacity and forcing, prematurely, all the symptoms associated with menopause.

- Selective oestrogen receptor modulators: they act by blocking oestrogen in the tissue, they are the most used method in pre-menopausal women, and the most important is tamoxifen.

- LHRH analogues: act at the hypothalamus-pituitary level blocking hormonal communication and inhibiting the synthesis of sex hormones, such as oestrogens and androgens. Luteinizing hormone-releasing hormone (LHRH) is secreted by neurons regulated by different neurotransmitters. Ultimately controls the levels of oestrogen and progesterone.

- Aromatase inhibitors (AIs): block the production of oestrogen by the aromatase enzyme, are considered the first line of therapy for metastatic breast cancer in post-menopausal women, ER positive.

5.5 Targeted therapy

Targeted therapy is a type of treatment that involves the use of drugs or other substances to specifically identify and attack cancer cells without normal cells damage. These include the therapies of monoclonal antibodies, tyrosine kinase inhibitors, PARP inhibitors and the mTOR

inhibitors. Among the monoclonal antibody therapies, the use of trastuzumab, an antibody that blocks the growth of the HER2 protein, stands out. A few years ago, HER2 positive tumours constituted breast cancer cases with lower survival. Implementation as a first line of treatment has considerably improved the prognosis, increasing 5-year survival to 77.1% to date (69, 70). On the other hand, lapatinib is an example of tyrosine kinase inhibitor treatment that acts by blocking the HER2 protein among others and is used for the treatment of cancer when it progresses after the action of trastuzumab.

6. Breast cancer in very young women

Breast cancer affects approximately 4-5% of women under 40 years of age every year. In Asia, these data are much higher, the Asian Breast Cancer Society reports that 13% of women diagnosed with breast cancer are under 40 years old and 5% under 35 years old. This increase in breast cancer affecting young women is very important because of their more aggressive behaviour compared to breast cancer in older women. This leads to a disproportionate number of deaths from breast cancer in young women per year (71, 72).

6.1 Determination of the “very young” threshold

What is considered young at the time of developing breast cancer has been a subject of deep discussion in the scientific community, considering young from 30, 35, 40 and even 45 years, complicating its study and reproducibility.

The European Society of Breast Cancer Specialists (EUSOMA) proposed the definition of young women, as a group of women with an age below 40 years. Although biology and endocrine status

are a continuum, the definition of age group is arbitrary (73). However, women under 40 years old have their own characteristics such as maintenance of fertility, pregnancy and lactation that are not maintained by women over 40 years old, it implies a different approach and establishing differences among older women, peri- and pre-menopausal (74).

A study in 1 703 patients from the same centre determined that the recurrence risk associated with age was a continuous linear parameter, showing a 4% decrease in recurrence and a 2% increase in the probability of death for each year of increase in age (75). However, these results have not been reproduced to date.

Han and collaborators in 2010, analysed 9 885 cases of breast cancer in women under 50 years of age. They were stratified into groups according to the age of the patients from 30 years (<30, 30-35, 35-40, 40-50), with a total of 1 443 patients below the age of 35 years (76). They observed that the group of women younger than 30 years showed worse survival than the group of 30-35 years, and this in turn, worse than the group of 35-40 years. However, the age groups 35-40 and 40-50 did not show differences, being able to be considered as a single group in terms of prognosis. These correlations were observed only in the group of ER positive patients and not in ER negative patients. The authors conclude that 35 years is a reasonable limit to define breast cancer in very young women, agreeing with other studies (77).

6.2 Breast cancer in very young women features

Young age of diagnosis for breast cancer has been established as an independent factor that has been associated with a high risk of relapse and death in a large number of studies (76-79). Biomarkers expression, including receptors, HER2 and proliferation markers are shown differentially in young women. Different studies recognize the presence of more aggressive tumour subtypes in young women (77, 80).

Several studies have suggested that breast cancer diagnosed at young ages are characterized by advanced stages at the time of diagnosis, more aggressive pathological features, higher rate of triple negative tumours and HER2 overexpression, higher rate of recurrence in any of the clinical stages, compared to older women, are the main causes of the more aggressive nature of these tumours in young women (81). So far, the increased risk of local recurrence in young women with breast cancer, the methods to preserve the fertility of these women, psychological interventions and the prospects of increasing survival rates are uncertain, unexplored and controversial (82).

Another disadvantage of having breast cancer at a young age is that endocrine receptors, HER2 and proliferation markers, appear differentially in young women. Many researchers believe that the effectiveness of hormone therapy is significantly lower in this group (83).

As mentioned above, there are new genomic methods such as Oncotype, Mammaprint, among others, that improve the prognosis and help decision making for treatment. These tests have been successfully integrated into clinical practice, but there is still some scepticism about whether they have the same prognostic value in young women with breast cancer because they were initially developed using postmenopausal women. For that reason, one of the future objectives is to clarify the role of these current tests in young women (84, 85).

7. Epigenetics

Between the 40s and the 60s, studies developed independently by researchers from different fields pointed to an alternative and reversible mechanism of transmission of information from generation to generation, different from the Mendelian inheritance patterns that were known. Specifically relevant were those of RA Brinck (studying the variation of pigmentation in corn), B McClintock (investigating retrotransposible elements in the corn genome) and H Crouse

(analysing translocations of the X chromosome in relation to sex determination)(86). The discovery of chromosomal imprinting by the latter - first discovered on the X chromosome, and next extended to other *loci* in the genome, suggested for the first time the existence of an alternative mechanism of gene expression regulation. The initial contributions in the matter came from scientific disciplines set aside at the time as biology of development and genetics - disciplines now inseparable. However, all these evidences converged in affirming that both the maintenance of the levels of gene expression during the life of an organism, as well as the information transmitted between generations, occurred without alteration of the DNA sequence. It was Conrad Waddington who coined the term epigenetics for the first time in 1952 (87).

The first articles that suggested an explanation for the turning on and off of gene activity were published in the 70s. Riggs' works on one hand (88), and Holliday & Pugh (89) on the other hand, suggested how methylation changes in DNA could have direct effects on the differential expression of genes throughout development. As they pointed out, the enzymes responsible for this process - later called DNA methyltransferases - would recognize specific sequences, or be recruited by other proteins linked to the target sequences; and they would also be able to recognize hemi-methylated substrates, so the changes would be transmissible through successive mitotic divisions. With the introduction of the DNA modification technique with sodium bisulphite in 1992, new doors were opened to the study of DNA methylation (90). This discovery accelerated research into epigenetics. By combining this technique with others such as PCR (91), genomic sequencing, or genomic analysis platforms, it has been possible to expand the study of epigenetics to a great variety of fields and contexts, having completed methylomas - sequencing of all the methylated nucleotides of a genome - of higher organisms (92, 93), and driven the sequencing of the human epigenome (94)

The concept initially described by C Waddington in 1952 marked the beginning, but the definition had to be expanded as the processes and their biological implications were specified. Epigenetics has been defined in a consensual manner as the "study area of those traits that are inherited in a stable manner, and arise from changes in chromosomes that do not result in alteration of the nucleotide sequence" (95).

There are four essential components with a key important role in gene regulation, development and carcinogenesis: DNA methylation, histone modifications, micro-RNAs and chromatin remodelling complexes (Figure 4).

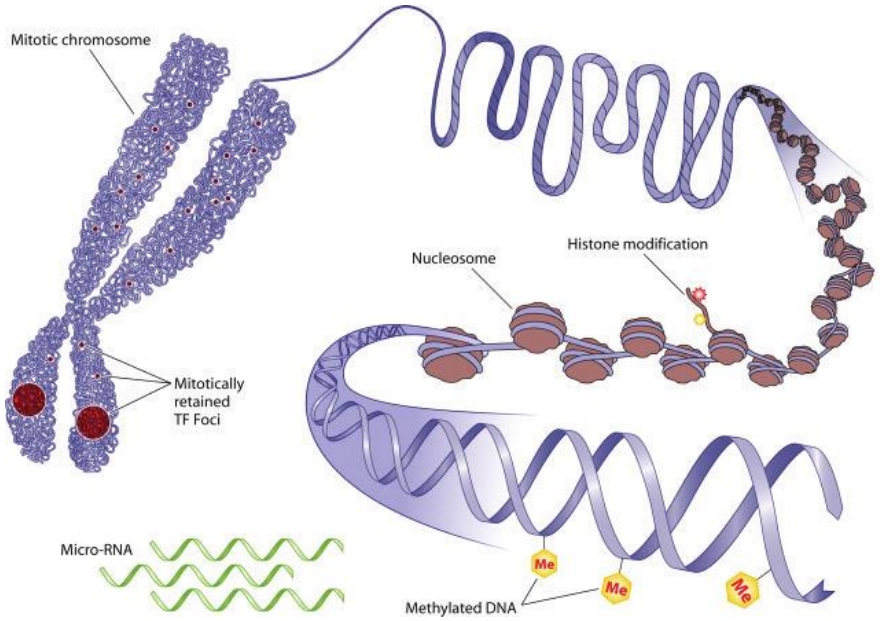


Figure 4. **Epigenetic processes, DNA methylation and histone modification.** Figure adapted from Derke 2015 (96).

7.1 DNA methylation

DNA methylation consists of the addition of a methyl group to the 5 'position of the pyrimidine ring of the cytosines which are followed by guanines called CpG nucleotides (Figure 5a). These CpG dinucleotides tend to concentrate in the promoter region of approximately 60% of the coding genes in humans, the CpG Islands forming regions between 200-2 000 bases with a CG ratio greater than 50% and generally not methylated in normal tissues, allowing gene activity when the corresponding transcription factors are available. DNA methylation is the epigenetic process studied in greater depth, given the easier availability of techniques to its study and its direct impact on gene activity (97). The enzymes responsible for DNA methylation are the DNA methyl transferases (DNMTs). These catalyse the transfer of a methyl group from S-adenosyl-L-methionine (SAM) to the 5 'position of the cytosine ring. The main components are the DNMTs 1, 3A and 3B with methylase activity and redundant functions in certain contexts, and the DNMT3L (DNMT3-like) that acts as a cofactor of the previous ones.

DNMT1 has been associated with maintenance methylation associated with DNA replication, so the same patterns are transmitted to the daughter cells. DNMT3A and 3B are involved in de novo methylation processes, which take place during embryonic development (98), or in pathological situations in which it occurs aberrantly. DNMT3L lacks methyltransferase activity but is directly related to epigenetic repression by recruiting histone deacetylase enzymes to methylated promoters (99, 100).

In eukaryotes, DNA methylation is necessary for the permanent silencing of certain *loci* (imprinted genes) and chromosomes (inactivation of the X chromosome in females). It is also necessary to coordinate transcriptional programs in development and differentiation processes (101). In the adult organism, DNA methylation determines tissue-specific transcription patterns that confer "cell identity": if the genetic code biologically individualises one organism against

the others, epigenetic processes confer identity, or phenotype, to each cellular type within the same organism. The distribution of 5-methyl cytosine in the genome follows a bimodal pattern, according to which the CpG dinucleotides included in CpG islands lack methylation, while those found outside are methylated (Figure 5c). This correlates with the transcriptional activity of the associated promoters: the CpG islands are associated with promoters of very active and evolutionarily conserved genes, whereas promoters not associated with CpG islands have lower transcription levels and present more evolutionary variability (102-104).

Although methylation in the promoter region of coding genes has been studied for decades, and its implication in transcriptional regulation has been characterized, the function of methylation present in the DNA of the body of genes has been little investigated. One of the proposed hypotheses was that intragenic methylation would be associated with the repression of transposons located in introns, or secondary gene promoters that would be silenced. Recent studies have observed a positive correlation between DNA methylation of the gene body and gene expression, both in plants and animals (105-107). While the hypermethylation of DNA in promoter regions represses the initiation of transcription, it seems to exert a synergistic effect on its elongation (108, 109), so the methylation changes in the body of the gene would modulate the gene expression degree. In this line, an association has been found between the hypermethylation of CpG islands outside promoters, and the tissue-specific activation of certain genes (110). In chronic myeloid leukaemia, changes in intragenic methylation have been described with positive and negative correlation with the expression of the affected genes, suggesting that depending on the area, this may have different effects (111). Nevertheless, the understanding of the role of methylation in intragenic regions, its relationship with the regulation mediated by the differential methylation of the promoter, and its effect on gene transcription, remain uncertain.

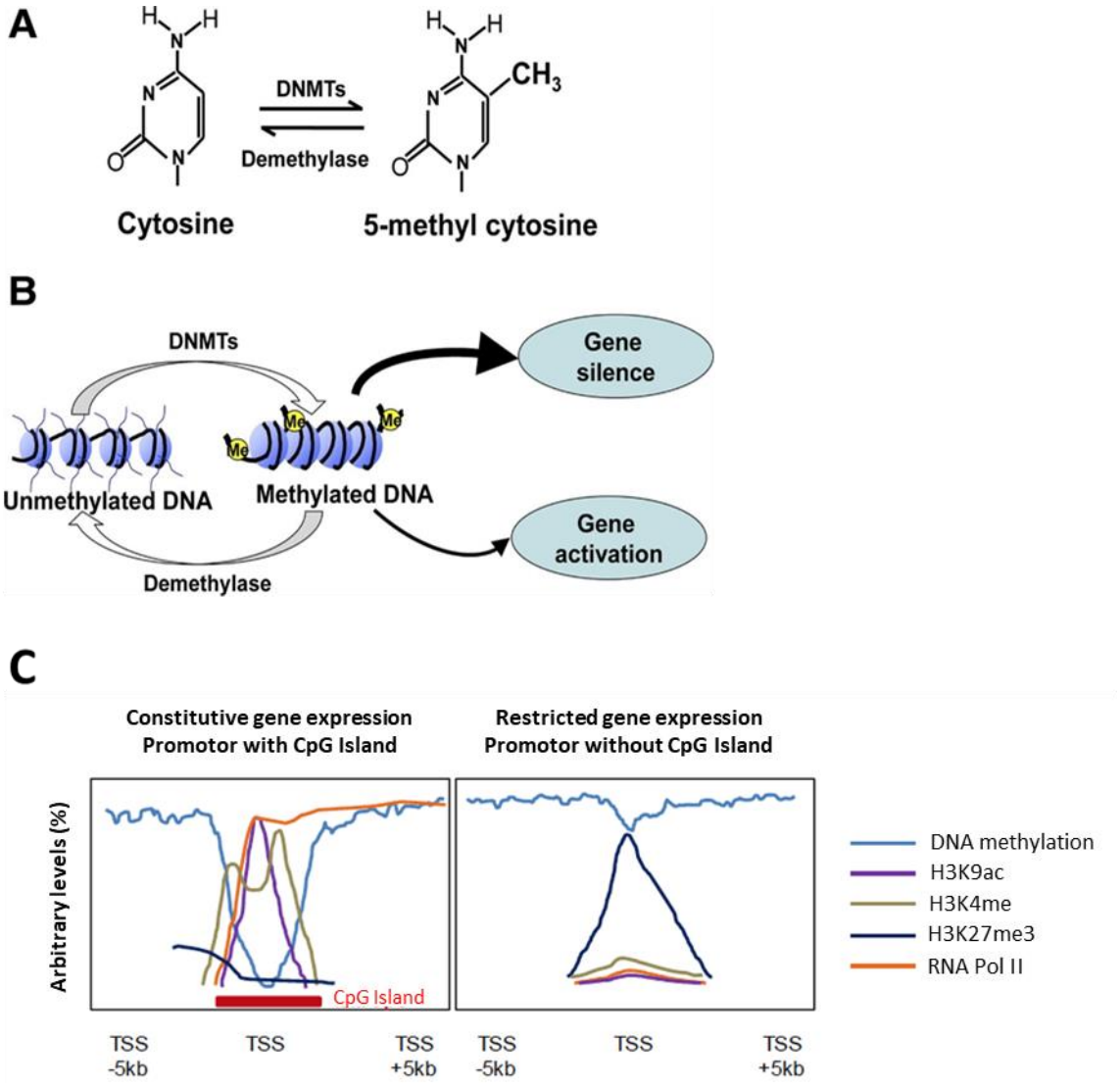


Figure 5. Methylation regulation and profiles. Regulation of DNA methylation of gene expression. DNA methylation is an addition of a methyl group at the C5 position of the cytosine ring (**A**). In most conditions, DNA methylation induced by DNMTs causes the recruitment of chromatin inactive proteins, leading to chromatin gene repression. Figure adapted from Yao and Li 2015 (112) (**B**). In somatic cells, constitutive expression genes usually have a CpG island in their promoter region in an unmethylated state and with the presence of histone modifications (H3K9ac, H3K4me) associated with transcriptional activation. Genes whose transcription is restricted to certain tissues or cell types are not usually associated with CpG islands, and DNA is methylated and associated with repressing histone tags (H3K9ac, acetylation of lysine 9 in histone 3, H3K4me, methylation of lysine 4 in histone 3; H3K27me3, trimethylation of lysine 27 in histone 3) restricting the access of the transcription machinery (**C**).

7.2 Non-coding RNA

The discovery of the micro-RNAs (miRNAs) shed light on the dark of non-coding genome, unknown to that moment, and that comprises 90% of the genome. The miRNAs were defined as molecules of about 22 bases that act as post-transcriptional regulators of gene expression in eukaryotes (113). MiRNAs are grouped into families according to their sequence homology, especially at the 5'-end of the mature miRNA (114). There are two types of miRNAs according to their genomic location: the intergenic miRNAs, located in non-coding regions, and the intragenic ones. The first ones are not associated to any gene and are transcribed by RNA Pol II from their own promoter. The intragenic miRNAs can be exonic or intronic, and their transcription depends on the gene in which they are inserted. Following transcription, the transcribed pre-miRNAs must be processed by the Drosha and Dicer enzymes, with RNase activity, and finally transported to the cytoplasm where they mainly perform their function.

MiRNAs have a blocking function on the expression of their target genes, which is based on the complementarity between the miRNA and the 3'-untranslated region (3'UTR) of the mRNA (messenger RNA), resulting in the inhibition of translation or the degradation of mRNA. Through this mechanism, miRNAs control hundreds of genes involved in a variety of cellular processes and signalling pathways (115). After the discovery of the miRNAs, many other classes of non-coding RNAs have been characterized (116-118). These vary in length: PIWI-interacting RNAs (piRNAs, 24-30 bases), small nucleolar RNAs (snoRNAs, 60-300 bases), and large intergenic non-coding RNAs (lincARNs,> 200 bases), among others. The functions they perform and their transcriptional regulation begins to be elucidated. Notably, lncRNAs have been implicated in epigenetic regulation processes by recruiting complex chromatin remodelling at specific loci. The classification of all the nc-RNAs that arise is complicated given the great structural diversity, the variety of processes in which they are involved, and the independent ways of processing and regulation (116-118).

8. Epigenetics and cancer

The first identification of the existence of aberrant DNA methylation in primary human tumours was three decades ago (119), currently the field of epigenetics has reached a position as relevant as genetics in cancer research. Numerous studies have identified specific epigenetic alterations of genes that occur in the development and progression of cancer. In the recent years, the introduction of high-throughput sequencing technologies and new arrays have accelerated and expanded our knowledge about the importance of a epigenetics in the processes of tumour formation (120).

8.1 Epigenetic alterations in breast cancer

In the breast cancer field, methylation plays an important role and methylation markers have been found useful for the diagnosis and monitoring of patients. An example is found in *ESR1*, whose methylation has been found to be a good survival predictor in patients treated with tamoxifen, while methylation in *ARHI* predicts survival only in patients not treated with tamoxifen (121). On the other hand, methylation of *CYP1B1* can also predict the response to tamoxifen. The methylation status of *ESR1* also has a prognostic value for the identification of luminal phenotypes with worse prognosis and greater resistance to hormonal treatment (122).

In addition, triple negative tumours characterized by the loss of expression of *ESR1*, *PR* and *HER2*, show alterations in DNA repair genes, mainly in *BRCA1* (123). The BRCA1 and BRCA2 proteins, as previously mentioned, are involved in hereditary breast carcinomas. However, in sporadic breast cancer, the silencing of the *BRCA1* gene occurs mainly through hypermethylation of its promoter, especially in the triple negative subgroup (124, 125). In this

way, Veeck et al (125) demonstrated that the hypermethylated *BRCA1* promoter predicts sensitivity to poly-(adenosine diphosphate)-ribose polymerase (PARP) inhibitors in non-hereditary triple negative tumours, suggesting its introduction as a potential biomarker of sensitivity to this compound.

Table 1. Epigenetic regulated genes in breast cancer*. Adapted from Davalos et al. (120)

Gene	Biological and clinical relevance	References
ESR1	Carcinogenesis, associated with ER- breast tumours and PR status; poor prognosis and poor response to anti-oestrogen agents	(121, 122, 126)
RASSF1A	Carcinogenesis, mainly in HR+ breast tumours; early detection of BC in cfDNA; poor prognosis in early stages of BC	(121, 124, 127-131)
TIMP3	Carcinogenesis, associated with ER- breast tumours	(121, 132)
BRCA1	Carcinogenesis in TNBC, sensitivity to PARPi and DNA damaging agents	(125, 130, 132, 133)
TWIST	Carcinogenesis	(129, 130, 134)
APC	Carcinogenesis	(130, 132, 135)
CDH13	Carcinogenesis, associated with TNBC phenotype	(130, 136, 137)
GSTP1	Carcinogenesis, associated with lymph nodes-positive BC	(128, 130, 132)
CCND2	Carcinogenesis, mainly in ER+ tumours	(128, 129, 134, 135)
HIN1	Carcinogenesis, associated with HR+ breast tumours	(128, 135, 137)
CST6	Carcinogenesis, poor prognosis in cfDNA of early stages of BC	(132, 138, 139)
***CDKN2A* (p16)**	Carcinogenesis, early detection of BC in cfDNA	(132, 140, 141)
PITX2	Carcinogenesis, poor prognosis in HR+ lymph node-negative BC	(127, 142-144)
CDH1	Carcinogenesis, associated with TNBC phenotype	(138, 145-147)

*In all the listed genes, aberrant CpG hypermethylation is the epigenetic event occurring in breast cancer. BC: Breast cancer; cfDNA: cell-free DNA; ER+: oestrogen receptor-positive; ER-: oestrogen receptor-negative; HR+: hormone receptor-positive; PARPi: PARP inhibitors; TNBC: triple-negative breast cancer.

Epigenetic regulation is a mechanism that is also controlling the epithelial-mesenchymal transition (EMT) in the progression of breast cancer. In this line, hypermethylation of E-cadherin (*CDH1*) is a common phenomenon in breast tumours and has been associated with the triple

negative phenotype (138, 147). The hypermethylation of *CDH1* is a factor of poor prognosis and has been postulated as a new target for the treatment of ER negative or HER2 negative tumours with aggressive characteristics (145, 146). In addition, there are a large number of epigenetically regulated genes that have been associated with breast cancer and are included in Table 1 (120).

8.2 Epigenetics today

The technological advances that have occurred in recent years have not only revolutionized the field of cancer but also the epigenetics. Methylation studies have increased our knowledge about the disease and have provided us with new techniques for cancer diagnosis, prognosis and therapy (120).

One of the key advances in the study of DNA methylation was the demonstration that the treatment of DNA with sodium bisulphite produced the deamination of unmethylated cytosines into uracils, while the methylated ones remained intact. The DNA bisulphite treatment followed by high-throughput technologies allows the generation of global DNA profiles. The sequencing technologies known as Next-generation technologies allow the sequencing of the complete genome. However, the costs associated with these techniques continue to be very high at present for their application to a large group of patients. Therefore, new techniques have been developed that consist in sequencing smaller and representative regions of the complete genome. One of the best known are the methylation arrays developed by Illumina. The Illumina GoldenGate Methylation Cancer Panel I array analysed 1 536 CpG dinucleotides (148), the *HumanMethylation27 BeadChip* array analysed 27 000 CpG sites (149), later the *HumanMethylation450 BeadChip* was developed and analysed up to 450 000 CpG sites (150, 151) until arriving at the current version, known as the Infinium MethylationEPIC BeadChip array that studies 850 000 CpGs (152), which is the most accurate platform for the global methylation

analysis. The 450k array interrogates more than 450 000 CpG sites, including not only the CpG islands, and sites around the transcription start sites of coding genes such as the known ones (shores and shelves), but also interrogates the body of the genes and the 3'UTR regions, as well as intergenic regions derived from the GWAS studies (151). The 450k array has been widely used in recent years for methylation studies, and was also selected as the platform for the studies of The Cancer Genome Atlas (TCGA) (153), due to its ability to analyse DNA methylation from paraffin embedded tissue samples (154). The recent version of the 850K platform incorporates new CpG sites located in enhancer regions identified by the ENCODE (the ENCyclopedia Of DNA Elements consortium) and FANTOM5 (Functional annotation of the mammalian genome) projects, and has become a powerful tool for new epigenetic studies (152).

HYPOTHESIS &

OBJECTIVES

Hypothesis

Breast cancer affecting women younger than 35 years old represents around 4-5 % of all breast cancers. Breast cancer in this group of age is typically more aggressive and presents potentially unique, aggressive and complex biological features. Unfortunately, they are not included in mammography screening programmes. Nowadays, there are genomic test available to improve prognosis and help at the decision-making process in the adjuvant setting. However, these test were developed using postmenopausal women at the first time and there is a great scepticism whether they have the same prognostic role in young women with breast cancer. All this statements make us to think that breast cancer in very young women may be a unique entity with a different molecular biology to breast cancer in older women and so, should be treated differently.

Epigenetic alterations such as DNA aberrant methylation and miRNA regulation has been recognised as characteristic events in malignant transformations and their study allow us to understand the molecular bases of cancer. However, the less in known about the molecular bases that distinguish breast cancer affecting young women from older counterparts. Understanding the role of the epigenome in younger women may lead to the development of novel epigenetic-based diagnostic strategies, taking the young age of these patients into account.

Objectives

The main objective of the present PhD project is to analyse the epigenetic mechanisms in terms of DNA methylation and miRNA expression, that occurs in breast cancer affecting women younger than 35 years old in comparison with older ones. To address these issues we propose the follow objectives:

1. To analyse the miRNA expression differences between two breast cancer cell lines derived from young women (HCC1500, HCC1937) and four from older patients (MCF-7, MDA-MB-231, HCC1806 and MDA-MB-468) and their comparison with the miRNA expression observed in tumour sample tissue analysed in previous group studies, in order to establish their utility as cellular models in breast cancer affecting young women studies.

2. To analyse the global DNA methylation profile in breast cancer samples from patients younger than 35 years old using high throughput platforms.

3. To evaluate the DNA methylation of gene encoding miRNAs in breast cancer samples from patients younger than 35 years old using high throughput platforms.

4. To identify potential targets in breast cancer in very young women and evaluate their clinical value in functional *in vitro* studies.

MATERIAL &

METHODS

MATERIAL

1. Cell lines samples

1.1 Cell lines and culture conditions

We included nine breast cancer cell lines in our study all from the American Type Culture Collection (ATCC, Rockville, MD). Four of them were obtained from young women: HCC1500, HCC1937, HBL100 and Hs 566(B). T. Additionally, 5 typical and widely used breast cancer cell lines from women older than 45 years old were used: MDA-MB-231, MDA-MB-468, MCF7, BT474 and HCC1806. The characteristics of cell lines and culture conditions are detailed in Table 2.

1.2 Cell doubling time

HCC1500, HCC1937, HBL100 and Hs566(B).T cell lines at the same passage 8 were used to plot the growth curve by using the 3-(a,5-dimethylthiazol-2-yl)-2,5-diphenyl tetrazolium bromide (MTT) method. The MTT method is described in the functional studies section from material and methods. The cells were plated into 96 well plates, 2 x 10^3 cells per well and 7 plates per cell line. One plate of each cell line was assayed every 24 hours. The absorbance was measured in a microtiter plate reader at 570 nm as the test wavelength and 690 nm as the reference wavelength. The number of cells was calculated by the result of MTT compared with the cell

number at the first day. The growth curves were plotted and the population doubling time of cell lines was calculated during the exponential growth phase of the cells.

Table 2. Cell lines characteristics and culture conditions.

Cell line	Subtype	Receptor	Tumour type	Age	Culture	Conditions	Supplements
HCC 1500	Basal	EGP2	IDC	32	RPMI	5%CO_2 37˚C	1% L-glu 10% FBS
HCC1937	Basal	EGP2	IDC	24	RPMI	5%CO_2 37˚C	1% L-glu 10% FBS
HBL 100	Basal	-	N	27	DMEM	5%CO_2 37˚C	1% L-glu 10% FBS
Hs 566(B).T	Unknown	-	Carcinoma	35	DMEM	5%CO_2 37˚C	1% L-glu 10% FBS
MDA-MB-231	Basal	EGFR, TGF-β	Carcinoma	51	RPMI	5%CO_2 37˚C	1% L-glu 10% FBS
MDA-MB-468	Basal	EGFR, TGF-α	Carcinoma	51	RPMI	5%CO_2 37˚C	1% L-glu 10% FBS
MCF7	Luminal	ER, IGFBP	IDC	69	RPMI	5%CO_2 37˚C	1% L-glu 10% FBS
BT474	Luminal	ER, PR,HER2	IDC	60	RPMI	5%CO_2 37˚C	1% L-glu 10% FBS
HCC1806	Basal	EGP2	Carcinoma	60	RPMI	5%CO_2 37˚C	1% L-glu 10% FBS

EGP2: Epithelial glycoprotein 2; EGFR: Epidermal growth factor receptor; TGF-β / α: transforming growth factor β / α; ER: oestrogen receptor; PR: progesterone receptor; HER2: hormonal oestrogen receptor 2; IGFBP: Insulin growth factor binding protein; RPMI: RPMI 1640 medium; FBS: fetal bovine serum; L-glu: L-glutamine; IDC: invasive ductal carcinoma; N: normal.

2. Commercial Material

Table 3. Table includes a list of all kits and probes used in the present study.

Product Name	Commercial Brand	Function
QIAamp DNA FFPE Tissue Kit	Qiagen	DNA extraction from FFPE tissue
RecoverAll ™ Total Nucleic Acid Isolation Kit	Ambion	RNA extraction from FFPE tissue
High Pure RNA Isolation Kit	Roche	RNA extraction from cell lines
Qubit® RNA HS Assay Kit	Molecular Probes® by Life Technologies	RNA quantification
Nuclear Extract Kit	Active Motif	Nuclear fraction protein extraction
Bio-Rad Protein Assay Kit	BioRad	Protein quantification by Bradford method
ECL Western Blotting Detection Reagents	GE Healthcare	Protein detection reagent
Anti-Histone H3 [1:3000] ab47915	Abcam	Primary antibody
beta Actin Antibody [1:1000] (PA1-183)	ThermoFisher Scientific	Primary antibody
Rabbit Immunoglobulins, Goat Anti-, Polyclonal, HRP, [1:5000] P0448	Agilent Dako	Secondary antibody
Affymetrix Genechip® miRNA 2.0 microarray	ThermoFisher Scientific	MiRNA expression analysis
Infinium MethylationEPIC BeadChip 850K	Illumina	Methylation study
Clariom™ D array	Applied Biosystems by Life Technologies	Gene expression study
High-Capacity cDNA Reverse Transcription kit	Applied Biosystems by Life Technologies	cDNA synthesis
MicroRNA Reverse Transcription Kit	Applied Biosystems by Life Technologies	cDNA synthesis for miRNAs

TaqMan® PreAmp Master Mix	Applied Biosystems by Life Technologies	cDNA Preamplification
TaqMan® MicroRNA Assays	Applied Biosystems by Life Technologies	Reagents for miRNA qRT-PCR
TaqMan® Gene Expression Assays	Applied Biosystems by Life Technologies	Reagents for mRNA qRT-PCR
LMK-235 inhibitor	Selleck Chemicals	HDACi inhibitor
MTT 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide	Sigma-Aldrich	MTT assay
hsa-miR-23a	Applied Biosystems by Life Technologies	MiRNA expression by qRT-PCR
hsa-miR-139-5p	Applied Biosystems by Life Technologies	MiRNA expression by qRT-PCR
hsa-miR-30c	Applied Biosystems by Life Technologies	MiRNA expression by qRT-PCR
hsa-miR-1207-5p	Applied Biosystems by Life Technologies	MiRNA expression by qRT-PCR
hsa-miR-1275	Applied Biosystems by Life Technologies	MiRNA expression by qRT-PCR
hsa-miR-3196	Applied Biosystems by Life Technologies	MiRNA expression by qRT-PCR
hsa-miR-92b	Applied Biosystems by Life Technologies	MiRNA expression by qRT-PCR
hsa-miR-9-1	Applied Biosystems by Life Technologies	MiRNA expression by qRT-PCR
hsa-miR-196a-1	Applied Biosystems by Life Technologies	MiRNA expression by qRT-PCR
hsa-miR-184	Applied Biosystems by Life Technologies	MiRNA expression by qRT-PCR
hsa-miR-551b	Applied Biosystems by Life Technologies	MiRNA expression by qRT-PCR
hsa-miR-124-2	Applied Biosystems by Life Technologies	MiRNA expression by qRT-PCR

RNU6B	Applied Biosystems by Life Technologies	MiRNA expression by qRT-PCR
RNU43	Applied Biosystems by Life Technologies	MiRNA expression by qRT-PCR
FOXI2 **Hs01653052_mH**	Applied Biosystems by Life Technologies	Gene expression by qRT-PCR
HOXD9 **Hs00610725_g1**	Applied Biosystems by Life Technologies	Gene expression by qRT-PCR
PCDH10 **Hs00252974_s1**	Applied Biosystems by Life Technologies	Gene expression by qRT-PCR
APIS3 **Hs00950999_m1**	Applied Biosystems by Life Technologies	Gene expression by qRT-PCR
APIB1 **Hs00153906_m1**	Applied Biosystems by Life Technologies	Gene expression by qRT-PCR
TIGIT **Hs00545087_m1**	Applied Biosystems by Life Technologies	Gene expression by qRT-PCR
CXCL17 **Hs01650998_m1**	Applied Biosystems by Life Technologies	Gene expression by qRT-PCR
HDAC5 **Hs00608351_m1**	Applied Biosystems by Life Technologies	Gene expression by qRT-PCR
HDAC9 **Hs01081558_m1**	Applied Biosystems by Life Technologies	Gene expression by qRT-PCR
NEIL3 **Hs00217387_m1**	Applied Biosystems by Life Technologies	Gene expression by qRT-PCR
EHF **Hs00171917_m1**	Applied Biosystems by Life Technologies	Gene expression by qRT-PCR
PAPSS2 **Hs00989928_m1**	Applied Biosystems by Life Technologies	Gene expression by qRT-PCR
PEX3 **Hs00920532_m1**	Applied Biosystems by Life Technologies	Gene expression by qRT-PCR
OTUD3 **Hs00418156_m1**	Applied Biosystems by Life Technologies	Gene expression by qRT-PCR

3. Tumour tissue samples

3.1 Sample data sets for DNA methylation assays

Illumina Infinium MethylationEPIC BeadChip samples from Hospital Clínico Universitario de Valencia (metEPICVal)

All the samples included in the Illumina Infinium MethylationEPIC BeadChip array (metEPICVal samples) were archived formalin fixed paraffin-embedded (FFPE) breast cancer tissues stored at the Hospital Clínico Universitario Valencia, Spain. We used 26 samples from breast cancer patients younger than 35 years (BCVY) and 15 samples from breast cancer patients older than 45 years (BCO). Normal tissue samples from very young women (NVY) (n = 2) and older women (NO) (n = 3) were also included in the methylation study and was from healthy donors. The clinical characteristics of the patients whose samples were included in the study are shown in Table 4. This study was approved by the Comité Ético de Investigación Clínica del Hospital Clínico Universitario de Valencia (CEIC-HCUV).

Infinium HumanMethylation450 BeadChip samples (HM450K samples)

We also used methylation data from a collaboration with Dr James Flanagan' s laboratory from Imperial College London that were previously published (155) and were analysed using Infinium HumanMethylation450 BeadChip (HM450K) (GSE72277). We retaining 11 samples from women aged younger than 35 years and 22 from women older than 45 years.

The Cancer Genome Atlas methylation data from breast cancer patients

We used methylation data from The Cancer Genome Atlas (TCGA). Methylation study was performed by HM450K array and includes data for 485 577 probes in 720 BCO and 27 BCVY samples.

3.2 Sample data sets for gene expression assays

The Cancer Genome Atlas gene expression data from breast cancer patients

We used gene expression data from TCGA. Gene expression includes data for 20 530 genes in 1 102 breast cancer (BC) samples (35 samples from BCVY and 924 from BCO) that were analysed by RNAseq.

Clariom D expression assay samples

In Clariom D array we analysed the expression of an independent data set that contain 42 breast cancer samples: 31 from BCVY and 11 from BCO. All samples included were archived formalin fixed paraffin-embedded BC tissues stored at the Department of Pathological Anatomy from the Hospital Clínico Universitario de Valencia, Spain. The clinical characteristics of the patients whose samples were included in the study are shown in Table 4. This study was approved by the Comité Ético de Investigación Clínica del Hospital Clínico Universitario de Valencia (CEIC-HCUV).

Table 4. Clinical tumour characteristics of samples included in the study.

	Methylation EPIC array N = 34		**qRT-PCR Validation set N = 40**		**TCGA Methylation N = 670**		**TCGA MiRNA expression N = 658**		**TCGA Gene expression N = 959**		**Clariom D Array N = 40**	
	BCVY (n = 21)	BCO (n = 13)	BCVY (n = 27)	BCO (n = 13)	BCVY (n = 27)	BCO (n = 643)	BCVY (n = 23)	BCO (n = 635)	BCVY (n = 35)	BCO (n = 924)	BCVY (n = 27)	BCO (n = 13)
Age mean (SD)	32.5 (2.7)	65.5 (8.5)	31.8 (3.2)	68.2 (7.3)	31.4 (2.9)	61.3 (11.0)	32.2 (2.7)	61.1 (11.3)	31.8 (2.9)	61.7 (11.1)	31.8 (3.2)	68.2 (7.3)
Histological subtype (%)												
Luminal A	2 (9.5)	2 (15.4)	3 (11.1)	4 (30.8)	12 (44.4)	232 (36.1)	8 (34.7)	212 (33.4)	14 (40.0)	356 (38.5)	3 (11.1)	4 (30.8)
Luminal B	8 (38.1)	4 (30.8)	11 (40.7)	5 (38.5)	4 (14.8)	105 (16.3)	4 (17.3)	93 (14.6)	6 (17.1)	166 (17.9)	11 (40.7)	5 (38.5)
TN	6 (28.8)	4 (30.8)	6 (22.2)	1 (7.7)	2 (7.4)	74 (11.5)	2 (8.7)	78 (12.2)	4 (11.4)	118 (12.7)	6 (22.2)	1 (7.7)
HER2*	5 (23.7)	3 (23)	7 (25.9)	3 (23.1)	0 (0)	29 (4.5)	1 (4.3)	38 (5.9)	2 (5.7)	58 (6.2)	7 (25.9)	3 (23.1)
ER status (%)												
ER positive	11 (52.4)	9 (69.2)	15 (55.5)	10 (76.9)	15 (55.6)	300 (46.6)	13 (56.5)	275 (43.3)	21 (60.0)	509 (55.1)	15 (55.5)	10 (76.9)
ER negative	10 (47.62)	4 (30.8)	10 (37.0)	3 (23.1)	4 (14.8)	90 (13.9)	2 (8.7)	94 (14.8)	6 (17.1)	145 (15.7)	10 (37.0)	3 (23.1)
Relapse (%)	4 (19.05)	1 (7.69)	7 (25.9)	0	3 (17.6)	49 (9.5)	2 (14.2)	49 (9.5)	3 (13.6)	65 (9.5)	7 (25.9)	0

SD: standard deviation; TN: triple negative subtype; ER: oestrogen receptor; BCVY: breast cancer in very young women; BCO: breast cancer in older women.
* A low proportion of BCVY and BCO patients from the methylation EPIC array and the qRT-PCR Validation set were classified as Luminal/Her2 but we considered them HER2 in the study.

3.3 Validation samples for real-time quantitative PCR

Gene and miRNA validation studies were performed in a different breast cancer sample set composed of 27 samples from BCVY and 13 from BCO. All samples analysed were archived formalin fixed paraffin-embedded BC tissues stored at the Department of Pathological Anatomy from the Hospital Clínico Universitario de Valencia, Spain. The clinical characteristics of the patients whose samples were included in the study are shown in Table 4. This study was approved by the Comité Ético de Investigación Clínica del Hospital Clínico Universitario de Valencia (CEIC-HCUV).

METHODS

1. Treatment of biopsied tissue samples for its preparation and conservation

1.1 Treatment of biopsied tissue for optimal freezing

The solid tissue samples were frozen in plastic cryo-tubes included in optimal cutting temperature, by rapid freezing by immersion in isopentane previously cooled to -110 ° C in a portable freezer (CoolSafe ™ 110, ScanVac, LaboGene ™ ApS, Lynge, Denmark). Finally, the samples were stored in mechanical freezers of -80° C for an indefinite time.

1.2 Paraffin embedded in biopsied tissue

After receiving the piece of tissue surgery, it is carved. This process consists in making sections of the sample for its best fixation. After this, the samples are fixed by immersion in a bath of 4% buffered formalin. Then, the samples undergo a progressive process of dehydration that allows to eliminate the water existing in the tissue. This operation is carried out using successive baths of 3 hours' duration, in alcohols of increasing graduation 50%, 70%, 80%, 95% and absolute alcohol. After dehydration the samples are subjected to a rinsing process, in which the dehydrating agent (ethanol in our case) is replaced by a miscible substance in the inclusion medium, xylene. To do this, the tissue samples are immersed in 3 successive xylene baths of 30 minutes each. Once this process is completed, the sample is infiltrated, which consists of completely infiltrating the histopathological sample in liquid paraffin at a temperature between 58° C and 65° C, depending on the melting point of the sample. To facilitate the complete

impregnation and total elimination of the xylol remains, two successive baths are made in different containers.

1.3 Histological control of samples

An additional cut of 3-5 µm was made of each samples, which was placed on slides pretreated with poly-L-lysine. Histological sections were stained with hematoxylin-eosin staining, as histological control of the cases to be studied. In this way, the percentage of tumour cells present in each sample analysed was determined.

2. DNA and RNA extraction

2.1 DNA extraction from paraffin tissue embedded

Paraffin tissue embedded DNA extraction

The DNA for the methylation study was extracted from tissue fixed in formaldehyde and embedded in paraffin, preserved at room temperature. Four sections of 20 µm were used, made in microtome at room temperature and collected in a 1.5 mL microtube. The DNA was extracted by the commercial kit QIAamp DNA FFPE Tissue Kit (Qiagen, Hiden, Germany).

Samples were deparaffinised adding 1000 µl of 100% xylol, shaking vigorously and leaving the mixture for 3 minutes at 56 ° C. The mixture was then centrifuged for 2 minutes at 10 000 x g and the xylol was removed. This step was repeated to achieve better deparaffination of the sample. Next, 2 washes of the sample were made with 100% ethanol, which allowed to eliminate the xylol residues and accelerate the drying of the sample, which was allowed to air dry for 30 minutes at room temperature. Next, 180 µl of buffer ATL and 16 µl of proteinase K, supplied in

the kit, were added, the mixture was incubated at 56 ° C overnight. Subsequently, samples were incubated at 90° C for 1h. At this time, 200 µl of the buffer AL kit solution and 200 µl of 100% ethanol were added to each sample. The mixture was passed through column, by centrifugation for a minute at 8.000 rpm. Next, a wash was carried out with 500 µl of the washing solution 1 and another with 500 µl washings with the washing solution 2 supplied in the kit. Finally, the total DNA was eluted with 60 µl (2 x 30 µl) of DNase-free water, and the extracted DNA was stored at -20 ° C for preservation after aliquoting 2 µl for the concentration measurement.

2.2 RNA extraction from paraffin tissue embedded and breast cancer cell lines

Paraffin tissue embedded RNA extraction

For the RNA extraction, the same material was used as for the DNA extraction mentioned in the previous section. In this case, 4 sections of 20 µm from breast cancer tumour paraffin embedded were used, also performed in microtome at room temperature and collected in a 1.5 mL microtube. The extraction of the RNA material was carried out using the commercial kit RecoverAll ™ Total Nucleic Acid Isolation Kit from ambion® (Applied Biosystems ™ by Life Technologies ™, Carlsbad, California, USA).

To each 1.5 mL tube containing the tissue sections 1000 µl of 100% xylol was added, shaking vigorously and leaving the mixture for 3 minutes at 50 ° C. The mixture was then centrifuged for 3 minutes at 10 000 x g and the xylol was removed. This step was repeated to achieve better deparaffination of the sample. Next, 2 washes of the sample were made with 100% ethanol, which allowed to eliminate the xylol residues and accelerate the drying of the sample, which was allowed to air dry for 30 minutes at room temperature. Next, 200 µl of digestion buffer and 4 µl of protease, supplied in the kit, were added, the mixture being incubated at 50 ° C for one hour and, subsequently, for 15 minutes at 80 ° C. At this time, 240 µl of the Isolation Additive kit

solution and 500 µl of 100% ethanol were added to each sample. The mixture was passed through column, by centrifugation for 30 seconds at 10 000 x g. Next, a wash was carried out with 700 µl of the washing solution 1 and another two 500 µl washings with the washing solution 2/3 supplied in the kit. After the washings, 60 µl of the DNase mixture was added to the centre of the column filter and left at room temperature for 30 minutes. Again, a wash was carried out with 700 µl of wash solution 1 and another two 500 µl washings with the 2/3 wash solution supplied in the kit. Finally, the total RNA (including miRNAs) was eluted with 70 µl (2 x 35 µl) of RNase-free water, and the extracted RNA was stored at -80 ° C for preservation after aliquoting 2 µl for the concentration measurement.

Breast cancer cell lines RNA extraction

RNA from breast cancer cell lines was extracted using the commercial High Pure RNA Isolation Kit from Roche. Breast cancer cell lines were cultivated by triplicate. RNA was extracted from cellular pellet from around 1 x 10^6 total cells.

Cells were resuspended in 200 µl of Phosphate Buffer Saline (PBS) and 400 µl of lysis/binding buffer was added. Mix was vortex and transferred to a column filter. After filtering, sample was incubated with DNase and DNase buffer for 15 min at room temperature. Next, a wash was carried out with 500 µl of the washing solution 1 and another 500 µl washing with the washing solution 2 supplied in the kit. Finally, the total RNA (including miRNAs) was eluted with 70 µl (2 x 35 µl) of RNase-free water, and the extracted RNA was stored at -80 ° C for preservation after aliquoting 2 µl for the concentration measurement.

All RNA experiments were done in RNase free environment.

3. Quantification methods and quality measurement of the extracted material

3.1 Quantification of DNA and total RNA by absorption spectrophotometry

Both the DNA and RNA extracted were quantified by UV-Vis absorption spectrophotometry using the Nanodrop 2000 equipment (Thermo Fisher Scientific Inc., Wilmington DE, USA). With only 1 μl of material, this equipment can determine the concentration and purity of the sample.

The concentration of nucleic acids is determined from the absorbance at 260 nm, following the Beer-Lambert Law:

$c = (A * \varepsilon)/b$

Where:

c = acid nucleic concentration in ng/μL

A= absorbance in AU

ε = the wavelength depending on the extinction coefficient in ng-cm /μL

b = distance in cm

The extinction coefficients generally accepted for nucleic acids are:

- Double-stranded DNA: 50 ng-cm/μL
- Single chain DNA: 33 ng-cm/μL
- RNA: 40 ng-cm/μL

The purity of the acid nucleic samples is analysed by the absorption ratio at 260 nm and 280 nm, with a ratio around 1.8 generally accepted as pure for DNA samples and a ratio of around 2.0 for pure RNA samples. If the ratio is lower in both cases, it may be indicating the presence of proteins, phenol or other contaminants that absorb at a wavelength of 280 nm.

On the other hand, the absorbance ratio between 260 nm and 230 nm is a second measure of the purity of acid nucleic. The values 260/230 for a pure acid nucleic sample are usually higher than the respective values for the ratio 260/280 and are commonly in the range between 1.8-2.2. If the ratio is appreciably lower, it may indicate the presence of contaminants copurified with the sample.

3.2 DNA quantification by PicoGreen

The Quant-iT PicoGreen dsDNA reagent is an ultrasensitive fluorescent acid nucleic used for the quantification of double-stranded DNA (dsDNA). This reagent is used in molecular biology for processes that require an exact quantification of the DNA material, such as sequencing studies, methylation, diagnostic applications, among others.

Although the most commonly used technique for the measurement of acid nucleic concentration is the determination of absorption at 260 nm, it has drawbacks. One of the main disadvantages of the absorption method is the inability to detect the presence of single-stranded nucleotides or proteins in the signal obtained, the interference caused by nucleic contaminants in the preparations of acid nucleic, the inability to distinguish between the DNA and RNA, and the low sensitivity of the assay.

The PicoGreen dsDNA quantification method can selectively detect a small quantity of 25 pg / mL of dsDNA in the presence of single stranded DNA, RNA and free acid nucleic. The assay is linear over three orders of magnitude and has little sequence dependency, which allows DNA from many sources to be accurately measured, including genomic, viral, minipreparation or PCR amplification products. On the contrary, it allows the quantification of very small amounts of up to 1 ng / mL of dsDNA.

In the quantification method by PicoGreen reagent, the DNA standard curve is first created. To this end, 2 g / mL of the stock solution of dsDNA in TE are prepared. The DNA concentration is determined based on the absorbance at 260 nm. For the construction of the standard curve, DNA of bacteriophage lambda or any purified DNA preparation is commonly used. The main characteristic of the PicoGreen Quant-iT reagent is that it remains linear in the presence of many components that commonly contaminate acid nucleic samples. From the dilution of pure dsDNA, the standard curve is constructed. We can create standard curves with five points ranging from 1 ng / mL to 1 μg / mL, although depending on the sample analysed we can create standard curves of lower or higher ranges. Once the standard dilutions are created, they are incubated for 2-5 min and the absorbance is read in a fluorescence plate reader. With the obtained values, the standard curve is constructed, which will be used to calculate the dsDNA concentrations of the samples. Then proceed to the preparation of samples that are diluted in TE to obtain a final volume of 1.0 mL. Next, 1.0 mL of the Quant-IT PicoGreen work solution is added to each sample. We incubate for 2-5 minutes at room temperature, protecting the mixture from light. Then, the fluorescence is measured using the parameters that we have previously used to generate our standard curve. From the values of the standard curve previously created, we measured the concentration values of the DNA of our samples.

3.3 RNA quantification by Qubit® 3.0

RNA extracted for Clariom D™ gene expression study was quantified using the Qubit® 3.0 (InvitrogenTM by Termo Fisher Scientific, Waltham, MA, USA) and the Qubit® RNA HS Assay Kit (Molecular Probes® by Life Technologies Carlsbad, California, USA). Qubit is a fluorometer that allows an accurately and sensitive measure of RNA quantity. The concentration or quality of the target molecule in the sample is reported by a fluorescent dye that emits a signal only when bound to the target, which minimizes the effects of contaminants on the result. Additionally, samples were prepared using the Qubit RNA HS Assay Kit, which provides an accurate and selective method for the quantitation of low-abundance RNA samples. The assay is highly selective for RNA and will not quantitate DNA, protein, or free nucleotides. Common contaminants, such as salts, free nucleotides, solvents, detergents, or protein, are well-tolerated in the assay. The kit provides concentrated assay reagent, dilution buffer, and pre-diluted RNA standards. Simply we diluted the reagent using the buffer provided, added the sample and read the concentration using the Qubit Fluorometer.

3.4 RNA total integrity analysis

Quality and concentration of total extracted RNA was determined with the RNA 6000 LabChip® and the 2100 Bioanalyzer (Agilent Technologies).

The integrity of total RNA was determined using Agilent 2100 Bioanalyzer equipment and Agilent RNA 6000 Nano Assay LabChip® (Agilent Technologies Inc., Santa Clara, CA, USA). This technology is based on capillary electrophoresis and allows the analysis of total RNA concentration between 25 and 500 ng / µl, using 1 µl of sample. The result of the analysis can be visualized as a conventional electrophoresis, in which RNA fragments separated by size are

observed, in addition to providing profiles or electropherograms for each sample. This technology applies a complex algorithm that automatically determines a value of RIN (RNA Integrity Number), with values between 1 and 10. The RIN considers the entire electrophoretic profile, including the presence and absence of degradation products, resulting in a reliable measure of RNA integrity. RIN value is very reproducible and allows us to make a numerical assessment of the integrity of the RNA and make comparisons between samples.

The LabChip® was prepared following the manufacturer's instructions. The reagents were left for 30 minutes at room temperature before use. The gel was prepared using a 550 µl column which was centrifuged at 1500 x g for 10 min for gel filtration.

We used 1 µl of the molecular weight marker (RNA 6000 ladder® Ambion) and 1 µl of RNA of each sample was denatured for 2 min at 70 ° C and introduced into the corresponding wells. Finally, the LabChip® was vortexed for 1 min at maximum power and introduced into the Agilent 2100 Bioanalyzer.

Agilent 2100 Bioanalyzer allowed visualizing the results in a computer by Expert Software v. B.01.02.SI136. The bioanalyzer uses fluorescence between 670 nm and 700 nm to determine the quality and integrity of total RNA. For each sample, the software calculated the quotient of the area of the ribosomal RNA peaks (28S / 18S), and the result was visualized by an electropherogram. The quotient should be around 2 since the molecular weight of 28S is double that of 18S.

3.5 Analysis of small RNAs

The integrity study of total RNA enriched with small RNAs was analysed using the Agilent 2100 Bioanalyzer and Agilent Small RNA Assay (Agilent Technologies Inc., Santa Clara, CA, USA). This technique evaluates the content of small RNA species using only 10 ng of total RNA. This technique is a faster and much more sensitive alternative to conventional agarose or polyacrylamide gels for the analysis and detection of small fragments of RNA in the range between 6 and 150 nucleotides. The profile or electropherogram of each sample allows easy visualization of the region corresponding to the miRNAs and, in addition, this assay determines the percentage and concentration of miRNAs present in the sample. The methodology followed is the same as that used in the analysis of the integrity of the total RNA detailed above.

4. Protein extraction and quantification

4.1 Protein extraction by cellular fractions

Nuclear protein extraction was preformed using the Nuclear Extraction Kit (Active Motif). Cells were collected in PBS, centrifuged and pellets were stored at -80 °C until their use. We added 200 µl of hypotonic buffer supplied by the kit (10x Buffer hypotonic, phosphatase inhibitor) to each cellular extract and the mixture were incubated 15 min in ice. Next, 10 µl of detergent included in the Kit was added and vortex 10 min. Samples were centrifuged 30 min at 14000 rpm and 4°C and the supernatant corresponded to the cytosolic fraction was stored at -20 °C.

Subsequently, we processed to the extraction of nuclear fraction. The pellet obtained in the previous step was resuspended in 30 µl of lysis buffer (lysis buffer, protease inhibitor and dithiothreitol 10M, all components were supplied by the kit). Mixture was vortex and incubated

30 min in ice and then all samples were subjected to sonication (4/5 pulses of 10 sec each at 38 of amplitude). At this point, the chromatin proteins were separated. Samples were centrifuged at 14000 rpm, 4 °C for 10 min and the supernatant were transferred in a new tub getting the protein nuclear fraction.

4.2 Protein quantification by Bradford

The protein concentration of the extracts was determined by a colorimetric assay using the Bio-Rad Protein Assay Kit (BioRad), based on the Bradford method (156), which measures the colour variation of the bright blue reagent of Coomassie G-250 (with a maximum absorbance of 465 nm to 595 nm). We mixed 25 µl of protein diluted 1:25 with 200 µl of the 1:5 diluted Coomassie bright blue reagent and the concentration was determined by measuring the absorbance at 595 nm. The curve of calibration was made with serial dilutions of BSA (bovine serum albumin), from 0.04 mg / mL to 0.32 mg / mL.

4.3 Western Blot assay

The western blot technique is used to separate out and detect specific proteins from a cell lysate or other solution containing different proteins mixed together. It is useful for quantifying the amount of a certain protein in a sample, its expression level, or for determining protein modification, such as phosphorylation. This is accomplished by incubating the proteins with sodium dodecyl sulfate (SDS), which denatures them and applies a negative charge, allowing them to be separated according to molecular weight by polyacrylamide gel electrophoresis (PAGE). The proteins can then be transferred from the gel to a nitrocellulose membrane that

can be probed for proteins of interest using specific primary antibodies coupled with a signal-inducing secondary antibody.

We prepared polyacrylamide gels. Acrylamide monomer is in a powder state before addition of water. Acrylamide is toxic to the human nervous system and is soluble in water and upon addition of water it polymerizes resulting in formation of polyacrylamide. A source of free radicals and a stabilizer, such as ammonium persulfate and tetrametiletilendiamina (TEMED) are added to initiate polymerization. The polymerization reaction creates a gel because of the added bisacrylamide, which can form cross-links between two acrylamide molecules. The acrylamide concentration of the gel can also be varied, generally in the range from 5% to 25%. Lower percentage gels are better for resolving very high molecular weight molecules, while much higher percentages of acrylamide are needed to resolve smaller proteins. The samples were subjected to polyacrylamide gels (4% stacking gel, 12% separating gel), for 2 hours at 120 V. Proteins were then transferred to a nitrocellulose membrane (Bio-Rad Laboratories, CA, USA) for 1 hour at 100 V using a traditional transfer system. Membranes were then blocked for non-specific interactions using BSA 5x [diluted with Tris-buffered saline with Tween20 (TBST) (10mM Tris, pH7.6, 137mM NaCl, 0.1% (v/v) Tween 20], followed by incubation with primary antibody diluted in BSA 1x overnight at 4°C (primary antibodies are indicated in material Table 3). After this, the membrane was washed for 5-10 minutes in BSA 1x three times. It was then incubated with horseradish peroxidase (HRP) conjugated secondary antibody for 1 hour at room temperature (secondary antibodies were diluted in BSA 1x and are indicated in material Table 3). Finally, membrane was washed with TBST for 5-10 minutes 3-4 times. Antibody-labelled proteins on the membrane were detected using ECL Western Blotting Detection Reagents (GE Healthcare, UK), which upon exposure to HRP luminesce. The light from this reaction was captured by Image Quant™ LAS4000 (GE Healthcare, Life Science, Oslo, Norway).

5. MiRNA expression analysis using Affymetrix Platform

RNA microarray is a solid surface to which a collection of RNA fragments binds. The surfaces used to fix the RNA are very variable and can be made of glass, plastic and even silicon. RNA microarrays are used to analyse the differential expression of genes, and the levels of thousands of them are monitored simultaneously. Basically, its operation consists of measuring the level of hybridization between the specific probe and the target molecule, which is usually indicated by fluorescence and through an image analysis, which indicates the level of expression of the gene by means of its messenger RNA. In the same way and under the same principle, the amount of expression of miRNAs can be analysed by microarrays with specific probes for them (157).

5.1 Affymetrix Platform

Sample RNA Hybridization using Affymetrix Genechip® miRNA 2.0

In this study, the Affymetrix Genechip® miRNA 2.0 microarray (Affymetrix Ltd., Santa Clara, CA) was used to measure the expression of non-coding RNA in three breast cancer cell lines (HCC1500, HCC1937 and HCC1806). To this end, total RNA from each sample was diluted to a homogeneous final concentration, 40 ng / µl in a volume of 10 µl. In the hybridization assay 8 µl was used, that means a total amount of 320ng. Samples were labelled with biotin and hybridized for 18 hours (according to the standard protocol established by the manufacturer), each sample on a GeneChip® miRNA 2.0 Array microarray (Affymetrix, Santa Clara, CA, USA), comprising a total of 4 553 small RNA sequences, including 1.105 from microRNAs.

RNA labelling with Biotin

For the RNA labelling, commercial Kit FlashTag ™ Biotin HSR RNA Labeling Kit (Affymetrix Ltd., Santa Clara, CA) was used. Briefly, to 8 µl of total RNA was added 2 µl of RNA Spike Control Oligos provided, to this mixture was added 5 µl of the Poly (A) Tailing Master Mix (containing 10X Reaction Buffer, 25mM Magnesium Chloride and 1 µl of ATP Mix diluted 1: 500 in 1 mM Tris buffer) and was mixed gently without vortexing, and incubated at 37 ° C for 15 minutes to allow attachment of the poly (A) tails. For ligation reaction, samples were centrifuged and 4 µl of FlashTag Biotin HSR Ligation Mix 5X and 2 µl of T4 DNA ligase were added separately to avoid self-ligations, mixed gently without vortex and centrifuged, after that, samples were incubated for 30 minutes at room temperature. The reaction was stopped by adding 2.5 µl of HSR Stop Solution, from the same kit. After mixing and centrifuging, 2.5 µl of sample were separated, to check the reaction efficiency by ELOSA QC Assay. Samples were stored at -20ºC until microarray hybridization, never exceeding two weeks.

RNA Hybridization with small RNA probes included in the microarray

For the hybridization reaction, the Affymetrix hybridization oven was previously prepared at 48 ° C, adjusting the agitation to 60rpm. The microarrays were left at room temperature for 15 minutes and marked appropriately. A pipette tip without filter was placed in the upper right section of the chip to allow a correct injection of the hybridization mixture and avoid bubbles.

The hybridization mixture was prepared with the following reagents provided, in this order: 50 µl of Hybridization Mix 2X, 15 µl of 27.5% Formamide, 10 µl of Dimethylsulfoxide (DMSO), 5 µl of the Eukaryotic Hybridization Controls 20X mixture, provided in GeneChip® Eukaryotic Hybridization Control Kit (Affymetrix Ltd., Santa Clara, CA), which was previously incubated at 65 ° C for 5 minutes, 1.7 µl of control oligos B2 at a concentration of 3nM (Control Oligo B2). The

amounts mentioned were per sample. The hybridization mixture was added to the biotin labelled RNA sample and the mixture was incubated at 99 ° C for 5 minutes followed by another 5 minutes at 45 ° C. Next, 100 µl of the prepared mixture was injected to each microarray, one by sample, after that, the previous pipette introduced was removed to facilitate the air exit and was covered with Tough-Spots® to minimize evaporation. Microarrays were placed in the hybridization oven, previously heated to 48 ° C, and were incubated under agitation at 60 rpm for 18 hours.

Washing and chip staining

Once the hybridization process finishes, microarrays were washed and detected. To this end, the remaining hybridization mixture was extracted on the microarray and replaced by the Array Holding buffer, to recover room temperature. After that, the Fluidics Station 450 washing station was prepared, which was used for the washing and detection of the microarrays following the protocol instructions. Briefly, consists of 10 washing cycles with washing buffer A (Wash Buffer A) at 30 ° C, followed by 6 cycles with washing buffer B (Wash Buffer B) at 50 ° C. For detection, probes were stained with Stain Cocktail 1 for 5 minutes at 35 ° C. After staining, samples were washed in 10 cycles with Wash Buffer A again (Wash Buffer A) at 30 ° C, and samples were re-incubated with Stain Cocktail 2 at 35 ° C for 5 minutes, followed by another 5 minutes with Stain Cocktail 1 at 35 ° C. Finally, the samples were washed in 15 cycles with washing buffer A (Wash Buffer A) at 35 ° C and microarrays were filled with Array Holding Buffer.

Microarray scanning and detection

After washing and staining the probes, microarrays were scanned using the GeneChip Scanner GCS3000 with self-loading (Affymetrix Ltd., Santa Clara, CA) that registered the images with provided software from Affymetrix® GeneChip® Command Console® 4.0 (Affymetrix Ltd., Santa

Clara, CA), which computed these images, saved in a file with extension type *.DAT, and transformed them into numerical and analysable intensity values, storing them in *.CEL files. Hybridization, washing, staining, and detection of Affymetrix Genechip® miRNA 2.0 microarrays were performed at Multigenic analysis service-UCIM.

5.2 Affymetrix array data preprocessing

To compare the miRNA expression profile from breast cancer cell lines from young women (CLVY) with older ones (CLO) we analysed the miRNA expression of 3 breast cancer cell lines from women older than 45 years (MDA-MB-231, MDA-MB-468 and MCF7) from previous group studies that were analysed using the same GeneChip® miRNA 2.0 Array (158).

Quality measurements and normalization of miRNA expression raw data (*.CEL) was analysed using the *affy* package from R Bioconductor. We performed a normalization of the intensity values using the RMA (Robust Multichip Average) function described in Irizarry et al (159). RMA is an algorithm used to create an expression matrix from Affymetrix data. The raw intensity values are background corrected, log2 transformed and then quantile normalized. Next a linear model is fit to the normalized data to obtain an expression measure for each probe set on each array. This function returns an expression measure log base 2 transformed.

Data intensity values normalized by RMA method were represented in a kernel density estimation plots for the different samples analysed. In statistics, kernel density estimation (KDE) is a non-parametric way to estimate the probability density function of a random variable. Kernel density estimation is a fundamental data smoothing problem where inferences about the population are made, based on a finite data sample. At this step, we could evaluate the

possibility that some sample/array presents an unacceptable level of noise and should be removed from the study. Raw and normalized microarray data will be available from Gene Expression Omnibus (GSE121396).

5.3 Statistical analysis of miRNA expression

In order to identified miRNA expression differences between breast cancer cell lines from young patients in comparison with older counterparts, we performed the statistical t-test study using *genefilter* package from R Bioconductor. Specifically, the *rowttest* implemented function allowed us to identified miRNA expression differences between age sample groups. P-values obtained were adjusted by Benjamini-Hochberg procedure to generate a false discovery rate and FDR p-values < 0.05 were accepted as differentially expressed between groups. MiRNA expression differences between breast cancer cell lines and tumour sample together from young and older patients were analysed following the same procedure.

5.4 Comparative miRNA expression study between cell lines and FFPE breast cancer samples

In order to compare the results from breast cancer cell lines and tumour samples from patients, we re-analyze the hybridation chips from breast cancer FFPE tumor samples from previous group studies (160) and cell lines in a combining study. BC samples from patients account with 12 samples from women older than 45 years old and 21 from women younger than 35 years. The combining study was analyzed and normalized following the same steps as breast cancer cell lines study previously explained. We selected miRNAs that were differentially expressed

between young and old women from both sample types, cell lines and patients, with a FDR p-value < 0.05. Significant miRNAs were represented in an unsupervised hierarchical clustering.

6. DNA methylation analysis by Infinium Human Methylation EPIC BeadChip

Infinium MethylationEPIC BeadChip (EPICarray) allows researchers to interrogate over 850,000 methylation sites quantitatively across the genome at single-nucleotide resolution. Multiple samples, including FFPE, can be analyzed in parallel to deliver high-throughput power while minimizing the cost per sample.

Infinium HD array technology enables a pan-enhancer and coding region view of the methylome that can be used for epigenome-wide association studies on a variety of human tissues. The EPICarray was introduced in 2016 in order to replace the Infinium Human Methylation 450K BeadChip. The Illumina Infinium HumanMethylation450 BeadChip platform (450K; Illumina Inc., CA, USA) for DNA methylation studies (151) has reached a predominant place in the market and the scientific arena, being not only the platform selected for The Cancer Genome Atlas (TCGA) studies (161), but also for numerous studies interrogating tumours outside these consortiums (162, 163), the aging process (164) or interindividual variability (165). The versatility of the described microarray has also been shown by its capacity to determine 5-methylcytosine DNA patterns from formalin-fixed paraffin-embedded samples (154) and for the recently identified 5-hydroxymethylcytosine mark (166, 167). However, the first array developed miss most of the information located in some sequences that has been seen that have a high impact in the transcription although they are distant from the transcription start site. This is particularly true for enhancers regions. The existence of chromosomal territories in the interphase nucleus is

now widely accepted, as is the looping and contact of DNA elements interspersed at great genomic distance. Such loops are of great relevance for transcription given that they afford the proximity and communication between gene promoters and their enhancers (168) or super enhancers (169, 170). Given that methylation status can affect the binding of cognate transcription factors (171), it is highly probable that DNA methylation differences in enhancer sequences exert a major role in cell and tissue functionality and that altered profiles could contribute to human diseases. The novelty of the new methylation EPICarray is the inclusion of 853 307 CpG probes, among them 333 265 CpG sites are located in enhancer regions identified by ENCODE (172, 173) and FANTOM5 (174) projects. Additionally, there are major number of probes located in regions outside CpG islands, non-CpG methylated sites identified in human stem cells, DNase hypersensitive sites, miRNA promoter regions and, moreover, it includes more than 90% of the content contained on the Illumina Human Methylation 450k BeadChip.

6.1 DNA methylation BeadChip Array study

Illumina Infinium BeadChips are sophisticated silicon-based array devices. An IntelliHyb seal separates the sample sections of the slide so that you can run multiple samples simultaneously.

Each individual sample section holds oligonucleotide probe sequences that are attached to beads assembled into the microwells of the BeadChip substrate. Because the microwells outnumber the distinct bead types, multiple copies of each bead type are present in the array. This built-in redundancy improves robustness and measurement precision. The BeadChip manufacturing process includes hybridization-based quality controls of each array feature, allowing consistent production of high-quality, reproducible arrays.

On Illumina infinium methylation platforms, probes are classified according to their position to CpG Islands, and the closest annotated gene. According to its relative position to CpG islands, loci are classified into three categories: (1) loci located inside a CpG island, called as Island; (2) loci located at <2,000 bp of a CpG island, named Shores (North or South) and (3) OpenSea regions that refers to regions away from CpG Islands. According to their relative position to the annotated genes, probes are classified in: (1) within the promoter (Transcription start site TSS1500 / TSS200), when they are located in a region between 1,500 bp upstream and the start of transcription point; (2) in the 5'UTR region (Untranslated region); (3) in gene, or body gene region; (4) at 3'UTR region; (5) or in intergenic regions.

Quality check of FFPE DNA samples

Bisulfite conversion

This process uses the Zymo EZ DNA Methylation Kit to convert unmethylated cytosines (C) in genomic DNA to uracil (U), though leaving methylated cytosines (C) unchanged for methylation analysis.

Methylation detection in bisulfite-converted DNA is based on the different sensitivity of cytosine and 5-methylcytosine to deamination by bisulfite. Under acidic conditions, cytosine undergoes conversion to uracil, though methylated cytosine remains unreactive. An effective bisulfite conversion protocol is a prerequisite for a successful Infinium HD Assay for Methylation. Incomplete conversion of cytosine to uracil can result in false- positive methylation signals, and can reduce the overall quality of the assay data.

Three hundred nanograms of FFPE DNA were distributed on a 96-well plate, and processed using the EZ-96 DNA Methylation kit (Zymo Research Corp., CA, USA). For bisulfite conversion DNA was incubated with conversion reagent in a thermal cycler that involves 30 seconds at 95 °C and

1h at 50 °C during 16 cycles. The denaturation process is necessary for bisulfite conversion because the conversion reagent only works on single-stranded DNA. The DNA was hold at 4°C for 10 min in the thermal cycler.

Samples were cleaning using the provided spin columns to wash off the remaining conversion reagent. Columns were desulphonated with desulphonation buffer and incubated at room temperature for 15 min, then they were washed twice to remove the desulphonation buffer. Finally, elution buffer was added and centrifugated and bisulfite-corverted DNA samples were transfer into a new plate and stored at -25°C until next step.

FFPE restoration

The Infinium HD FFPE Restore protocol restores degraded FFPE DNA to a state that is amplifiable by the Infinium HD FFPE methylation whole genome amplification protocol. Bisulfite-converted DNA (bs-DNA) from FFPE samples was processed as previously described. The DNA was denatured with NaOH for 10 min at room temperature. An 1h reaction at 37°C was then performed with PPR (Primer Pre Restore) reagent and AMR (Amplification Mix Restore reagent) reagents supplied by the kit manufacturer, in which DNA repair is accomplished. DNA was cleaned with a ZR-96 DNA Clean & Concentrator-5 kit (Zymo Research Corp.) and eluted in 13 µl of ERB. Cleaned DNA was then denatured for 2 min at 95°C, followed by ligation incubation at 37°C for 1 h with RST and CMM reagents. The resulting material was cleaned with ZR-96 DNA Clean & Concentrator-5 kit (Zymo Research Corp.) and eluted in 10 µl of DiH20.

Array hybridization

Eight microliter of restored FFPE bisulfite converted DNA were used for hybridization on Infinium Human MethylationEPIC 850K BeadChip, following the Illumina Infinium HD Methylation protocol. This consisted of a whole genome amplification step followed by enzymatic end-point fragmentation, precipitation and resuspension. The resuspended samples were hybridized on Human MethylationEPIC 850K BeadChips at 48°C for 16 h. Then unhybridized and non-specifically hybridized DNA were washed away, followed by a single nucleotide extension using the hybridized bisulfite-treated DNA as a template. The nucleotides incorporated were labeled with biotin (ddCTP and ddGTP) and 2,4-dinitrophenol (DNP) (ddATP and ddTTP). After the single-base extension, repeated rounds of staining were performed with a combination of antibodies that differentiated DNP and biotin by fixing them different fluorophores. Finally, the BeadChip was washed and protected in order to scan it.

Scanning BeadChips

The Illumina HiScan SQ scanner is a two-color laser (532 nm/660 nm) fluorescent scanner with a 0.375 µm spatial resolution capable of exciting the fluorophores generated during the staining step of the protocol. The intensities of the images were extracted using GenomeStudio (2010.3) Methylation module (1.8.5) software. Bisulfite conversion, FFPE restoration, array hibridization and scanning were performed at the *Unidad de Genotipado y Diagnóstico Genético (UGDG)-Unidad Central de Investigación en Medicina de la Universidad (UCIM).*

Control probes included in the Array

Sample Independent Controls

Sample-independent controls evaluate the performance of specific steps in the process flow.

- Staining Controls

Staining controls are used to examine the efficiency of the staining step in both the red and green channels. Staining controls have dinitrophenyl or biotin attached to the beads. These controls are independent of the hybridization and extension step. Both red and green channels can be evaluated using the Staining Controls.

- Extension Controls

Extension controls test the extension efficiency of A, T, C, and G nucleotides from a hairpin probe, and are therefore sample-independent. Both red (A, T) and green (C, G) channels are monitored.

- Hybridization Controls

The hybridization controls test the overall performance of the entire assay using synthetic targets instead of amplified DNA. These synthetic targets complement the sequence on the array perfectly, allowing the probe to extend on the synthetic target as template.

The synthetic targets are present in the Hybridization Buffer at 3 levels, monitoring the response from high-concentration (5 pM), medium-concentration (1 pM), and low- concentration (0.2 pM) targets. All bead type IDs should result in signal with various intensities, corresponding to the concentrations of the initial synthetic targets. Performance of hybridization controls should only be monitored in the green channel.

- Target Removal Controls

Target removal controls test the efficiency of the stripping step after the extension reaction. In contrast to allele-specific extension, the control oligos are extended using the probe sequence as template. This process generates labelled targets. The probe sequences are designed such that extension from the probe does not occur.

All target removal controls should result in low signal compared to the hybridization controls, indicating that the targets were removed efficiently after extension. The target removal controls are present in the Hybridization Buffer. Performance of target removal controls should only be monitored in the green channel.

Sample-Dependent Controls

The sample-dependent controls can be used to evaluate performance across samples. These control oligos are designed for bisulfite-converted human genomic DNA sequences. Because target sequences do not contain CpG dinucleotides, the performance of the control oligos does not depend on the methylation status of the template DNA.

- Bisulfite-Conversion Controls

These controls assess the efficiency of bisulfite conversion of the genomic DNA. The Infinium Methylation probes query a (C/T) polymorphism created by bisulfite conversion of non-CpG cytosines in the genome.

Bisulfite-Conversion I: These controls use the Infinium I probe design and allele-specific single base extension to monitor efficiency of bisulfite conversion. If the bisulfite conversion reaction

was successful, the "C" (Converted) probes match the converted sequence and are extended. If the sample has unconverted DNA, the "U" (Unconverted) probes are extended. There are no underlying C bases in the primer landing sites, except for the query site itself. Performance of bisulfite conversion controls C1, C2, and C3 should be monitored in the green channel, and controls C4, C5, and C6 should be monitored in red channel.

Bisulfite-Conversion II: These controls use Infinium II probe design and single base extension to monitor the efficiency of bisulfite conversion. If the bisulfite conversion reaction was successful, the "A" base is incorporated and the probe has intensity in the red channel. If the sample has unconverted DNA, the "G" base is incorporated across the unconverted cytosine, and the probe has elevated signal in the green channel.

- Specificity Controls

Specificity controls are designed to monitor potential nonspecific primer extension for Infinium I and Infinium II assay probes. Specificity controls are designed against nonpolymorphic T sites.

Specificity I: These controls are designed to monitor allele-specific extension for Infinium I probes. The methylation status of a particular cytosine is carried out following bisulfite treatment of DNA by using query probes for unmethylated and methylated state of each CpG locus. In assay oligo design, the A/T match corresponds to the unmethylated status of the interrogated C, and the G/C match corresponds to the methylated status of C. G/T mismatch controls check for nonspecific detection of methylation signal over unmethylated background. PM controls correspond to A/T perfect match and should give high signal. MM controls correspond to G/T mismatch and should give low signal. Performance of GT Mismatch controls

should be monitored in both green and red channels. The Controls dashboard table lists expected outcome for controls probes.

Specificity II: These controls are designed to monitor extension specificity for Infinium II probes and check for potential nonspecific detection of methylation signal over unmethylated background. Specificity II probes should incorporate the "A" base across the nonpolymorphic T and have intensity in the red channel. If there is nonspecific incorporation of the "G" base, the probe has elevated signal in the green channel.

- Negative Controls

Negative control probes are randomly permutated sequences that should not hybridize to the DNA template. Negative controls are particularly important for methylation studies because of a decrease in sequence complexity after bisulfite treatment. The mean signal of these probes defines the system background. It is a comprehensive measurement of background, including signal resulting from cross-hybridization and nonspecific extension and imaging system background. We can use the Average signal and standard deviation of 600 negative controls to establish detection limits for the methylation probes. Performance of negative controls should be monitored in both green and red channels.

- Nonpolymorphic Controls

Nonpolymorphic controls test the overall performance of the assay, from amplification to detection, by querying a particular base in a nonpolymorphic region of the bisulfite genome. They let us compare assay performance across different samples. One nonpolymorphic control has been designed to query each of the 4 nucleotides (A, T, C, and G).

Methylation BeadChip Array Analysis

Data pre-processing

In methylation section is described the process of hybridization and scanning of the arrays. The intensities of the images were extracted using the GenomeStudio (v.2011.1) Methylation module (1.9.0) software. Raw microarray data is available from Gene Expression Omnibus (GEO) (GSE100850).

DNA methylation was measured at 853 307 CpG sites. Before array processing, we perform a background subtraction and colour correction for the dye bias seen in Infinium II probes, as well as removal of bad quality samples using *minfi* package implemented in R Bioconductor (175). For that, we start reading the *. IDAT files obtained. Using *minfi* package we constructed the methylated and unmethylated signals that finally were transformed into β values. The methylation score for each CpG was represented as a β value according to the fluorescent intensity ratio representing any value between 0 (unmethylted) and 1 (completely methylated).

Quality control step

Minfi package includes efficient and reliable functions for quality control step. One of them consist of a simple quality control plot that uses the log median intensity in both the methylated and unmethylated channels. When plotting these two medians against each other, it has been observed that good samples cluster together, while failed samples tend to separate and have lower median intensities. The line separating “bad” from “good” samples represent a useful cutoff, which may have to be adapted to a specific dataset. Another application for further

explore the quality of the samples is the β value densities of the samples that can be represented and make a judgement.

As we explain in the previous section, the EPICarray contains several internal control probes which were implemented to check technical aspects of the array such as staining, hybridization, target removal, extension, bisulphite conversion, specificity, non-polymorphism, and negative controls The values of these control probes can be extracted using *minfi* package in order to be analysed. We removed samples that were >3 standard deviation outside the normal distribution of each probe set across all samples. Additionally, probes that were not detected in more than 20% of the samples were excluded from the analyses and we selected samples with >98% of probes detected.

Single Nucleotide Polymorphism (SNP)

Methylation studies strongly recommend to drop the CpG probes that contain either SNP at the CpG interrogation or at the single nucleotide extension. Because the presence of SNPs inside the probe body or at the nucleotide extension can have important consequences on the downstream analysis, *minfi* offers the possibility to remove such probes and they were removed. Finally, a total of 30 271 SNPs loci were excluded from subsequent analysis.

Cross-reactive probes

It has been previously reported than about 6% of the probes on the 450K array co-hybridize to alternate genomic sequences, therefore potentially generating spurious signals. However, for the latest EPICarray a list of probes that are located at sides affected by polymorphisms and/or

have a potential to cross-hybridise has been published (176). Consequently, cross reactive probes were removed in our analysis.

Pre-processing and normalization

Different normalization procedures have been described. In our analysis we performed the preprocessFunnorm function. This normalization process implements the functional normalization algorithm developed in (177) that uses the internal control probes present on the array to infer between-array technical variation. This is particularity recommended for studies comparing conditions with kwon large-scale differences.

Methylation statistical analysis

The methylation differences were assessed using Wilcoxon rank sum test and generalized linear models (GLM), p-values were adjusted by Benjamini-Hochberg procedure to generate a false discovery rate (FDR). We considered significant CpG sites those with both p-value < 0.05 and a minimum change of ±0.1 in β-values. Flow-diagram of the analysis is shown in Figure 6. First, we analysed methylation differences between normal tissue samples (NVY and NO) and both BC data sets (BCO and BCVY) respectively. The aim of this first analysis was to remove methylation probes that were differentially methylated between groups just for the cancer condition. Thus, from significant CpG sites we removed those probes common to both cancer sets. Finally, we analysed only probes differentially methylated in BCVY and compared the methylation status in BCO. In this case an α value of 0.01 was set.

Additionally, generalized liner models were assessed to evaluate if the methylation differences observed were influenced by different aspects such as oestrogen receptor status or molecular cancer subtypes.

Probes were classified into different categories according to their description in the Illumina manifest file, depending on the function (promoter, gene body, TSS or UTRs), relation to the CpG context (North/South (N/S) shore, N/S shelf and island), enhancer regions provided by ENCODE and FANTOM5 projects and ENCODE regulatory elements (CpG sites in transcription factor biding sites [TFBS], open chromatin regions and DNase hypersensitive regions). Methylation differences by categories in the significant probe set were assessed by Welch's t-test and p-value < 0.001 were considered to be statistically significant.

6.2 Validation of DNA methylation with combined and TCGA data

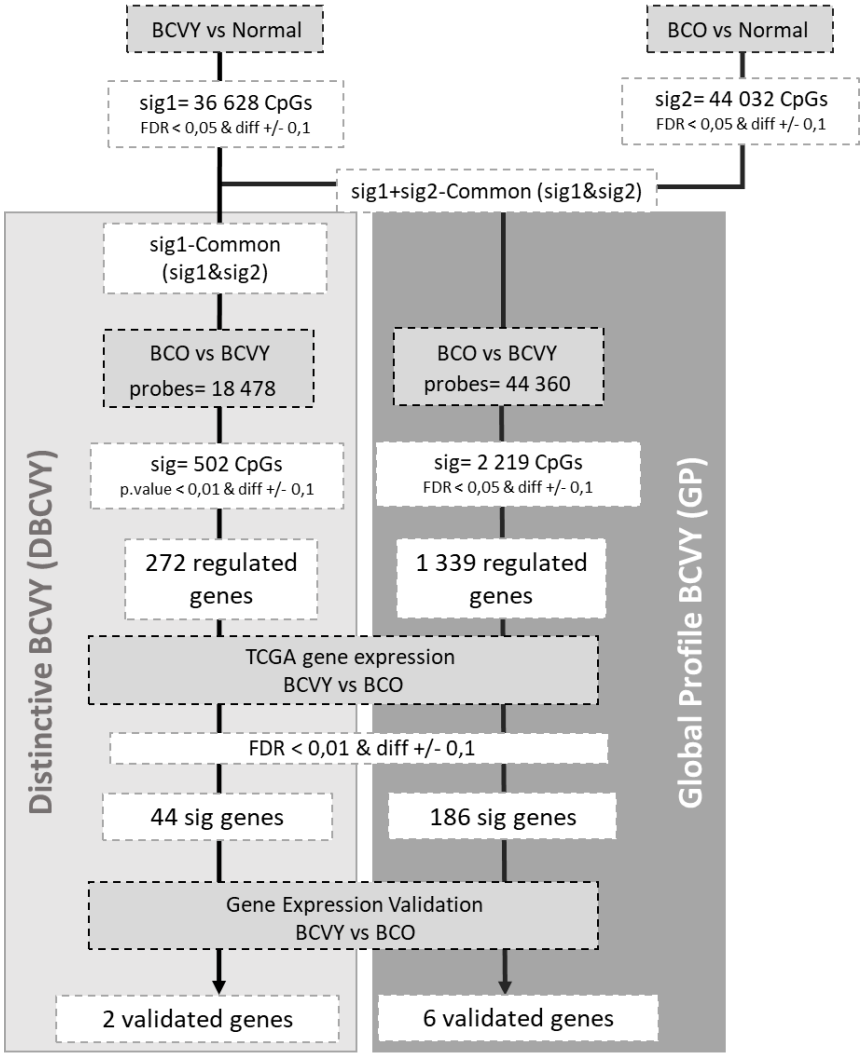
We performed a combined study using methylation data from Flower et al (155) analysed by HM450K array and data from metEPICVal study previously described. First, we analysed HM450K array samples according to the methods described in the data pre-processing section. Thus, we retained methylation levels for 405 068 probes present in both metEPICVal and Flower et al.'s (155) HM450K study in a total of 69 samples (32 BCO samples, 32 from BCVY, 3 from NO and 2 from NVY women). TCGA methylation data includes β-values for 485 577 probes in 720 BCO and 27 BCVY samples. We used both data sets to analyse methylation differences found between BCVY and BCO.

6.3 Gene expression study of genes regulated by differentially methylated probes

Expression of the genes regulated by the CpGs sites that were differentially methylated in BCVY was analysed using gene expression data from TCGA for BC. Gene expression study was done with 50 permutations, and samples were randomly selected and balanced by subtype. We used *metap* R package to combine p-values for each gene in each study; p-values were adjusted by Benjamini-Hochberg FDR procedure; genes with FDR < 0.01 and gene expression differences ±0.1 were considered statistically significant (Figure 6).

6.1 Estimation of DNA methylation age

A vast literature characterizes genes or genomic regions that either get hypermethylated or hypomethylated with age (177-186). The epigenomic landscape varies markedly across tissue types (187-189) and many age-related changes depend on tissue type (184, 190). But several recent studies have shown that age-dependent CpG signatures can be defined independently of sex, tissue type, disease state and array platform (178, 179, 182, 191-193). According to this statements, Horvath 2013 used a collection of publicly available DNA methylation data sets for defining and evaluating an age predictor. Specifically, the age estimator was developed using 8 000 sample from 82 Illumina DNA methylation array datasets, encompassing 51 healthy tissues and cell types. The approach consists in a DNA methylation age clock based on the methylation values of 353 CpG regions, which are transformed to DNAm (DNA methylation) age using a calibration function that was named epigenetic clock. The calibration function reveals that the epigenetic clock has a high ticking rate until adulthood, after which it slows to a constant ticking rate. Using the training data sets, Horvath (194, 195) used a penalized regression model (Elastic net regularization) to regress a calibrated version of chronological age on 21 369 CpG probes



	EPIC study		
	Methylation	**TCGA**	**Gene exp Validation**
BCVY	N=21	N=39	N=27
BCO	N=13	N=50	N=13
Normal	N=5	-	-
Total Probes	793 483CpGs	1 339 (GP) 272 genes (DBCVY)	14 genes

Figure 6. **Diagram of analysis procedure for DNA methylation study.** Diagram include work flow for the statistical analysis performed for metEPICVal samples. Sig and Sig1/2 = significant probes; N = sample number; Probes = total number of probes for each analysis; BCVY: breast cancer in very young women; BCO: breast cancer in old women; Diff = methylation differences between BCVY and BCO; FDR = false discovery rate; TCGA: The Cancer Genome Atlas.

that were present both on the Illumina 450K and 27K platform and had fewer than 10 missing values. DNAm age is defined as estimated ("predicted") age. The elastic net predictor automatically selected 353 CpGs. 193 of the 353 CpGs correlate positively with age while the remaining 160 CpGs correlate negatively with age. R software and a freely available web-based tool can be found at the following website (https://labs.genetics.ucla.edu/horvath/htdocs/dnamage/).

Estimated age, also referred to as DNA methylation age, has the following properties: first, it is close to zero for embryonic and induced pluripotent stem cells; second, it correlates with cell passage number; third, it gives rise to a highly heritable measure of age acceleration; and, fourth, it is applicable to chimpanzee tissues (which are used as human analogs for biological testing purposes). Salient features of Horvath's epigenetic clock include its high accuracy and its applicability to a broad spectrum of tissues and cell types. Since it allows one to contrast the ages of different tissues from the same subject, it can be used to identify tissues that show evidence of accelerated age due to disease.

By contrasting DNA methylation age (estimated age) with chronological age, one can define measures of age acceleration. Age acceleration can be defined as the difference between DNA methylation age and chronological age. Alternatively, it can be defined as the residual that results from regressing DNAm age on chronological age. The latter measure is attractive because it does not correlate with chronological age. A positive/negative value of epigenetic age acceleration suggests that the underlying tissue ages faster/slower than expected.

DNA methylation age was analysed according to Horvath epigenetic clock model using the R software. Taking into account that epigenetic clock was developed using methylation data from 450k and 27k methylation arrays, we found 18 CpG values that were missing in the new EPICarray and were replaced by the mean of their values from the Flower et al., (155) data set analysed by the Human Methylation 450K array without age distinction. Differences in age acceleration between breast cancer groups were analysed by Wilcoxon rank sum test.

6.2 Copy Number Aberration study (CNA)

Conumee R package

Copy number aberrations (CNA) were obtained from the EPICarray study by implementing the *conumee* package in R Bioconductor (196). Although the primary purpose of these arrays is the detection of genome-wide DNA methylation levels, it can additionally be used to extract useful information about copy-number alterations, e.g. in clinical cancer samples. The approach was initially described in Sturm et al., 2012 (197).

CNA analysis was performed using a two-step approach. First, the combined intensity values of both 'methylated' and 'unmethylated' channel of each CpG were normalized using a set of normal controls that present a flat genome not showing any copy-number alterations. This step is required for correcting for probe and sample bias caused by GC-content, type I/II differences or technical variability. Secondly, neighbouring probes were combined in a hybrid approach, resulting in bins of a minimum size and a minimum number of probes. This step is required to reduce remaining technical variability and enable meaningful segmentation results.

For the *conumee* package it is recommended that methylation array data was loaded using the *minfi* package previously used. Segmentation is performed using the circular binary segmentation (CBS) algorithm implemented in the *DNAcopy* package (198). The package requires the raw *. IDAT files that were loaded using the *minfi* package. CNA analysis is separated into four parts:

- Fist, a normalization of a single query sample to a set of control samples by multiple linear regression is performed. For best results control samples of matched normal tissues that are profiled within the same experiment are used (which are likely to have the same technical bias). Essentially this regression analysis yields the linear combination of control samples that most closely fits the intensities of the query sample. Subsequently, the log2-ratio of probe intensities of the query sample versus the combination of control samples are calculated and used for further analysis. At this step, the five normal control samples included in our analysis were used.

- Secondly, a combination of probes within predefined genomic bins are used. Bins are previously generated. Intensity values are shifted to minimize the median absolute deviation from all bins to zero to determine the copy-number neutral state.

- Thirdly, an optional step could be performed to include detail interest regions that have been previously created.
- Finally, a segmentation of the genome into regions of the same copy-number state is performed.

Conumee package was used to perform complete genome plots of samples that were classified into different groups according to their CNA profile

ChAMP R package

ChAMP R package is another developed package for methylation studies from data obtained from HumanMethylation arrays (199). However, one function implemented in the package allow us to identify copy number alterations. The function uses the intensity values for each probe to count copy number and determine if copy number alterations are present.

Basically, the CNA analysis is similar to previous R package described, and is based in the comparison of the copy number between case samples and control samples. Otherwise, *ChAMP* package already included some blood control samples itself.

There is an interesting function in this package, that was not in the previous one, and is based in the representation of CNA according to the groups analysed. At this step, we used the present package to analyse CNA according to breast cancer groups and compare between them.

DNAcopy R package

DNAcopy is another R package for CNA studies and as the previous mentioned, it used the intensity values detected to determine the copy number alterations (198). The package includes different plot options such as the complete genome CNA representation in a single plot. However, the most relevant representation that we used from the package is the CNA mean values representation from individual chromosomes.

7. DNA methylation of genes encoding miRNAs

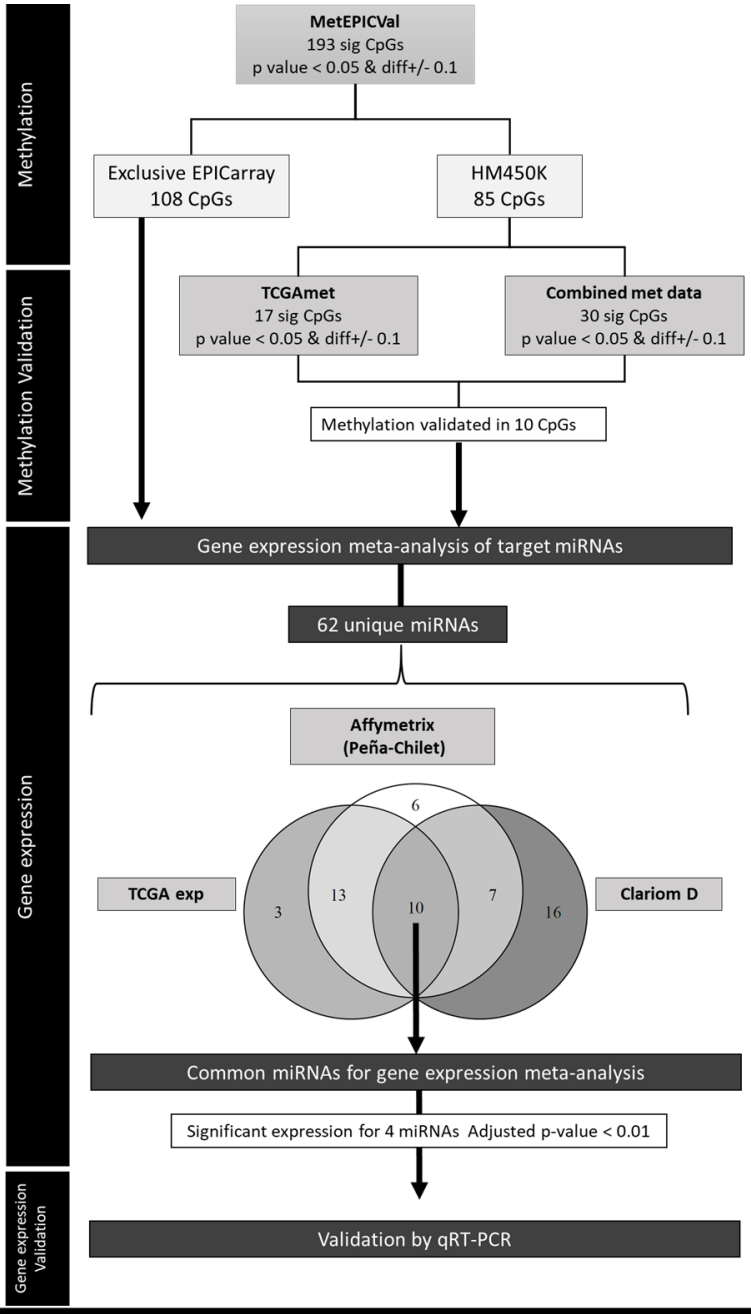
This part of the study was focused on probes associated to genes encoding for miRNAs. According to the information available in the Illumina annotation file, EPICarray platform includes 9 961 probes that are linked to miRNAs. After pre-processing described in Methylation Array BeadChip analysis section, a set of 6 567 CpG probes has been obtained which were associated with 1 264 unique miRNAs and were used in the present study.

Analysis of probes distribution by genes coding miRNAs revealed that there were probes that were associated to one miRNA and probes associated to multiple miRNAs, the maximum number was for probes that were regulating 7 miRNAs. Additionally, miRNAs may be regulated by unique or multiple probes (1-293).

7.1 Statistical analysis for genes encoding miRNAs

We considered all CpG probes regulating gene-encoding miRNAs to identify miRNAs differentially methylated between breast cancer in young and older patients. We analysed methylation differences using the Wilcoxon Rank Sum test. CpG sites with both p-value < 0.05 and a minimum change of ±0.1 in β-values between BCVY and BCO were considered significant. Additionally, we performed a GLM analysis to assess whether specific methylation differences observed between BCVY and BCO were independent of molecular subtype and ER status. The study procedure diagram is included in Figure 7.

MiRNA probes were classified into different categories according to their description in the Illumina manifest file, depending on function (promoter, gene body, TSS or UTRs), relation to the CpG context (N/S shore, N/S shelf and island), enhancer regions provided by ENCODE and FANTOM5 projects, and ENCODE regulatory elements (CpG sites in transcription factor binding



	Methylation			miRNA expression			
	MetEPIC Val	TCGA met	Combined met data	TCGA exp	Affymetrix	Clariom D	qRT PCR samples
BCVY	N=21	N=39	N=32	N=39	N=21	N=27	N=27
BCO	N=13	N=50	N=32	N=50	N=12	N=13	N=13
Total Probes	6,567 miRNA CpGs	2,771 miRNA CpGs	2,931 miRNA CpGs	1,046 miRNA genes	768 miRNA precursors	1,217 miRNA genes	5 miRNAs

Figure 7. Workflow diagram of the procedure for DNA methylation of genes encoding miRNAs study. The diagram shows the sample set analysed and the significant probes obtained in each step. The table includes the sample and probe size for each sample set. *diff: methylation differences between BCVY minus BCO; sig: significant; BCVY: breast cancer in very young women; BCO: breast cancer in old women; TCGA: The Genome Cancer Atlas; qRT-PCR: quantitative real time PCR; N: sample number.

sites [TFBS], open chromatin regions and DNase hypersensitive regions). Methylation differences by categories in the significant probe set were assessed by Welch's t-test and p-value < 0.001 were considered to be statistically significant.

Univariate Cox analysis was used to analyse the effect of clinical variables and miRNA methylation on patients' relapse-free survival (RFS) and overall survival (OS). Multivariate Cox method was performed to adjust for clinical variables such as oestrogen receptor status and molecular subtype that could be influencing in relapse and/or survival.

7.2 Validation of miRNA methylation with combined and TCGA data

We performed a combined study using methylation data from Flower et al (155) analysed by Infinium HumanMethylation450 BeadChip (HM450K) and data from metEPICVal study previously described. First, we analysed HM450K array samples according to the methods described in the data pre-processing section. Thus, we retained methylation levels for 405 068 probes present in both metEPICVal and Flower et al.'s (155) HM450K study in a total of 64 samples (32 BCO samples, 32 from BCVY) that were used in the validation study.

Additionally, we used methylation data from an independent breast tissue samples available from The Cancer Genome Atlas, that includes data for 485 577 probes in 720 BCO and 27 BCVY samples.

Both studies were performed using the Illumina HM450K array, which only includes half of the probes present in the EPICarray. To validate our results from the BCVY-BCO study we selected miRNA differentially methylated probes that were included in HM450K and a Wilcoxon Rank Sum test was performed between BCVY and BCO tumour samples.

7.3 Meta-analysis of gene encoding miRNAs expression regulated by differentially methylated probes

Expression of miRNAs regulated by differentially methylated probes were analysed using data from Clariom™ D samples (explained in Clariom D gene expression analysis section). This study includes expression for 1,217 miRNA precursors. Additionally, in the meta-analysis we included previous miRNA expression published data from Peña-Chilet et al. (160), which were analysed by GeneChip® miRNA 2.0 Array (Affymetrix, Santa Clara,CA, USA) with GEO accession number GSE48088. This data set includes 21 BCVY and 12 BCO samples. Furthermore, we downloaded gene expression from TCGA and used data from 39 BCVY and 50 BCO samples, randomly selected and balanced by subtype.

We performed a meta-analysis study to combine p-values from three studies by Fisher method using the *metap* R package; p-values were adjusted by Benjamini-Hochberg FDR procedure; genes with FDR < 0.01 and gene expression differences ±0.1 were considered statistically significant.

8. Relapse free survival (RFS) and overall survival (OS) studies

Relapse free survival is defined as the time from random assignment until first recurrence (loco-regional or distant metastasis) or death due to any cause, whichever was observed first. Overall survival was the time from random assignment until death due to any cause. Our study collets the follow-up data from metEPICVal samples and we used the follow-up from TCGA patients in the miRNA methylation study. For the miRNA expression study we used clinical data from TCGA and METABRIC that analyses 1 109 samples older than 45 years old and 37 from women younger than 35 years old with breast cancer.

For RFS and OS studies we performed two test:

- Log-rank (based on Mantel-Cox) test

The log-rank test statistic compares estimates of the hazard functions of the two groups at each observed event time. It is constructed by computing the observed and expected number of events in one of the groups at each observed event time and then adding these to obtain an overall summary across all-time points where there is an event.

- Gehan-Breslow-Wilcoxon tests

The log-rank statistic can be used when observations are censored. If censored observations are not present in the data then the Wilcoxon rank sum test is appropriate.

After analysis, percentages of RFS and OS by days were represented. The graphs plotted represents RFS and OS probabilities on Y axis and time past after entry into the study on X axis. This kinds of curves are known as relapse/survival curves and in them the proportion surviving or not relapsing remains unchanged between the events, even if there are some intermediate censored observations.

9. Clariom D gene expression analysis

We used Clariom™ D array (Applied Biosystems by Life Technologies, Carlsbad, California, USA) for gene expression study of 42 breast cancer patients (13 from BCO and 27 from BCVY). Clariom™ D provides a highly detailed overview of the transcriptome identifying genes, exons

and alternative splicing events that give rise to coding and long noncoding RNA isoforms. Specifically, array identifies more than 134,700 genes, 542,500 transcripts and detects around 484,900 exon-exon splice junctions. Total RNA was isolated using RecoverAll Total Nucleic Acid Isolation Kit (Applied Biosystems by Life Technologies, Carlsbad, California, USA) following the manufacturer's protocol. RNA concentration was measured by Qubit® 3.0 (InvitrogenTM by Termo Fisher Scientific, Waltham, MA, USA) using the Qubit® RNA HS Assay Kit (Molecular Probes® by Life Technologies Carlsbad, California, USA) previously mentioned. In order to assess RNA quality, we performed a real-time quantitative PCR (qRT-PCR) and observed good amplification plots for 25 ng of RNA. Finally, hybridization in the Clariom™ D array was performed with 25 ng of total RNA. Raw files (*.CEL) were normalized by RMA method using the R Bioconductor *affy* package.

10. Gene expression validation technics (RNA and miRNAs) by real time quantitative PCR

The evaluation of expression of genes that were significantly different expressed between groups was carried out by real time quantitative PCR (qRT-PCR). It is a conventional polymerase chain reaction (PCR), in which the quantity of product amplified by each cycle of the PCR is measured at real time, thus obtaining an amplification curve to quantify the initial amount of product.

The PCR technique was developed for the first time by Mullis et al., 1986, and since then it has been one of the basic techniques in laboratories around the world (200). Its main aim is the specific amplification of nucleotides from a DNA template. This technique is based on two concepts: denaturation and renaturation of temperature-dependent DNA, and the resistance to high temperatures of the polymerase from *Termus aquaticus* (Taq). Taq polymerase is

responsible of the synthesis of nucleotide sequences using a template strand, currently there are various polymerases resistant to high temperatures. The PCR consists of increasing and decreasing the temperature alternately in cycles, following these steps:

- Denaturalization: The temperature rises above the level of DNA denaturation. Both strands of DNA separate.
- Hybridization: The temperature decreases to a level where it allows the binding of the Taq polymerase and the primers to denatured DNA strands, which are used as a template.
- Elongation: The temperature decreases allowing the synthesis of double-stranded DNA by Taq polymerase. Once synthesized, Taq polymerase is released from the strand.

After the first temperature cycle is finished, another one starts following the same steps, this time, in addition to the complete DNA strand as a template, it has short fragments of the specific nucleotide sequence as a substrate, whose quantity will increase exponentially.

The sequence specificity is achieved using primers complementary to the nucleotide sequence problem, which are joined in step 2 of the temperature cycle and allow the initiation of DNA polymerization by the Taq. We obtain the DNA fragment of a specific sequence amplified several orders of magnitude as final product, allowing us to detect the presence of small amounts of a specific sequence.

The qRT-PCR is a variation of this technique, first described by Higuchi et al. 1993, in which the amount of amplified product (amplicon) is detected at real time in each cycle (201). This was achieved by incorporating a fluorescent intercalating agent (EtBr) into the reaction, which allowed the detection of the product synthesized at time T, by accumulating fluorescence. The kinetics of the accumulation of fluorescence with each cycle depends directly on the initial amount of DNA, and can thus extrapolate that amount.

At present, double-stranded DNA binding agents (such as SYBR Green) and fluorescent probes linked to the primers are used, which only emit fluorescence when the elongation takes place in each cycle of the PCR, thus increasing fluorescence intensity by increasing the amount of amplified product that is synthesized. Detecting only the amplified sequence is a much more specific and precise method.

In the analysis of expression of both mRNAs and miRNAs we have used TaqMan® probes. This method is characterized by a fluorochrome bound to a specific probe with a sequence complementary to the fragment that we want to amplify, as well as, a fluorescence extinguishing element, when the elongation phase of the PCR is produced, the fluorochrome know as reporter is excised from the probe and separated of the quencher, thus emitting its fluorescence (Figure 8a). To carry out the PCR and detect the fluorescence, instruments adapted for this purpose are required. For the quantification of RNA by this technique, an additional step is required, the synthesis through the retrotranscription of complementary DNA (cDNA) using RNA as a template, by a reverse transcriptase.

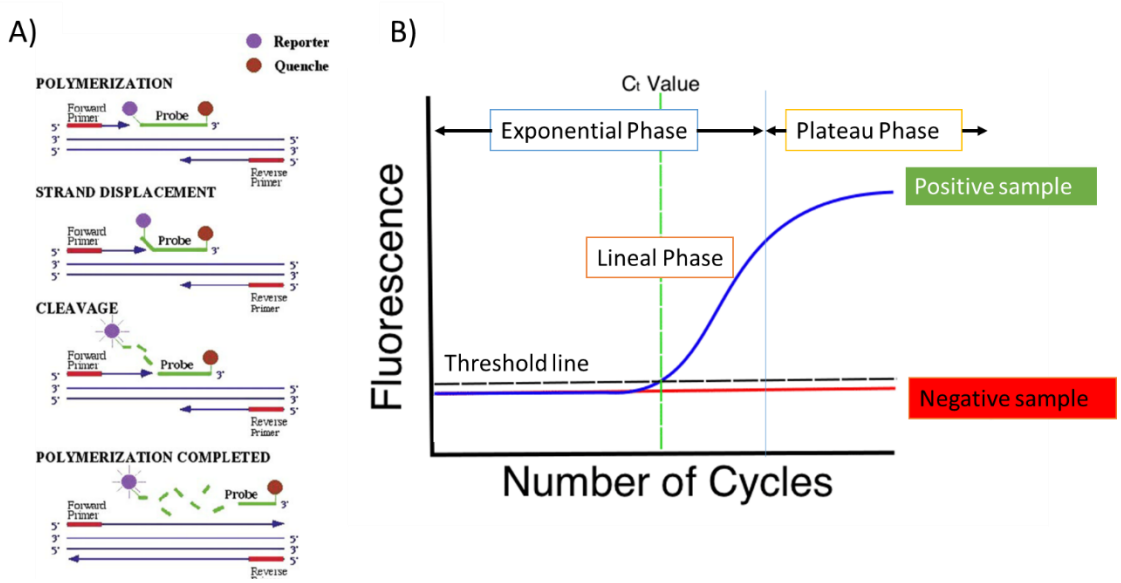


Figure 8. Real time quantitative PCR. Method used by TaqMan probes to report fluorescence (**A**). Representation of the typical curve obtained in a qRT-PCR. The X axis shows the cycle of the PCR reaction and the Y axis represents the fluorescence intensity emitted by the fluorochrome that is directly related to the quantity of the amplified product. The curve in red shows the pattern of a negative sample and would correspond with a basal expression, while the blue curve shows the typical pattern of an amplified sample (**B**).

The kinetics of the reaction consist of a sigmoid curve typical of a logistic function, so we obtain for each sample a curve similar to that shown in Figure 8b. The curve consists of 3 phases, a basal phase, an exponential phase, in which the quantity of product increases exponentially, until the enzyme reaches its maximum yield and goes to the linear phase, where the kinetics of the enzyme follows a linear behaviour, so that its slope can be calculated. The plateau or stationary phase is the last phase, in which the amount of product reaches the level of saturation of the enzyme and is characterized by a slow increase. The threshold cycle or Ct is that cycle in which the curve intersects with an intensity level called the threshold, which marks a limit above which a sample is considered significantly positive.

Reverse transcription

The synthesis of the complementary DNA (cDNA) was carried out from identical amounts of total RNA (200ng), starting from a total RNA dilution of 50 ng / µl.

Reverse transcription of mRNA

For the analysis of gene expression, cDNA synthesis was performed using the High-Capacity cDNA Reverse Transcription kit (Applied Biosystems ™ by Life Technologies ™, Carlsbad, California, USA). In this case the synthesis of cDNA was based on the use of random primers. For each 10 µl reaction, 200 ng of total RNA were incubated with 1 µl of 10X RT buffer, 0.4 µl of 100mM dNTPs, 1 µl of 10X random primers, 0.5 µl of MultiScribe ™ Reverse Transcriptase 50 U / µl , 0.5 µl of RNase inhibitor and nuclease-free water for 10 minutes at 25 ° C, followed by 120 minutes at 37 ° C and 5 minutes at 85 ° C in the thermocycler (Eppendorf, Hamburg, Germany) using 96-well plates .

Reverse transcription of miRNA

For the analysis of small RNAs, the commercial kit TaqMan® MicroRNA Reverse Transcription Kit (Applied Biosystems ™ by Life Technologies ™, Carlsbad, California, USA) was used for the analysis of small RNAs. In this case, the synthesis of cDNA is specific for each mature miRNA or snoRNA, therefore a mixture of primers was used to perform a single cDNA synthesis per sample, according to the protocol for Creating Custom RT and Preamplification Pools using TaqMan® MicroRNA Assays (Applied Biosystems ™ by Life Technologies ™, Carlsbad, California, USA).

A mixture of the miRNA / snoRNA specific primers to be studied was prepared at a concentration of 0.5X each. For each 15 µl reaction, 200 ng of total RNA were incubated with 6 µl of the primer

mixture, 0.3 µl of 100 mM dNTPs, 3 µl of MultiScribe ™ Reverse Transcriptase 50 U / µl, 1.5 µl of buffer RT 10X and 0.2 µl of RNase inhibitor 20 U / µl for 30 minutes at 16 ° C, followed by 30 minutes at 42 ° C and 5 minutes at 85 ° C in the thermocycler (Eppendorf, Hamburg, Germany) using 96-well plates.

Pre-amplification

The pre-amplification consists of a first amplification phase before carrying out the qRT-PCR, in this way the amount of the initial substrate is increased and its subsequent detection is facilitated in cases where it is limited.

It was carried out in the cases that a low material detection was systematically obtained.

For this, a mixture of the probes that were to be used in the qRT-PCR was prepared at 0.01 concentration, 3 µl of that mixture and 6 µl of TaqMan® PreAmp Master Mix were added to 2.5 µl of the cDNA sample obtained from the reverse transcription reaction in a 96-well plate. The preparations were incubated in thermal cycler 2 minutes at 50 ° C, 10 minutes at 95 ° C and 50 cycles of denaturation at 95 ° C for 15 seconds and amplification at 60 ° C for 1 minute, and repeated 14 cycles. Subsequently, the preamplificated product was diluted in a 1:10 ratio prior to carrying out the qRT-PCR in order not to saturate the amplification reaction.

Real-time quantitative PCR (qRT-PCR)

The recovery analyses of small size RNAs (miRNA and snoRNA) and mRNAs were carried out by qRT-PCR using the QuantStudio™ 5 System, the TaqMan® MicroRNA Assays and TaqMan® Gene Expression Assays kits (Applied Biosystems ™ by Life Technologies ™, Carlsbad, California, USA), respectively.

For each 10 µl reaction, 1 µl of the synthesized cDNA was mixed with 0.5 µl of the corresponding TaqMan® Assay 20X, 5 µl of the corresponding TaqMan® Universal Master Mix 2X and 3.5 µl of nuclease-free water. The amplification conditions were as follows: 2 minutes at 50 ° C, 10 minutes at 95 ° C and 50 cycles of denaturation at 95 ° C for 15 seconds and amplification at 60 ° C for 1 minute. PCR triplicate was performed for each sample using 384-well plates. Each plate was designed to include the same number of trials together with the appropriate endogenous control in each case, as well as, an equitable representation of case and control samples to avoid possible methodology biases.

Relative expression values using the $2^{-\Delta\Delta Ct}$ method

This method assumes certain assumptions and a specific experimental design, previously described in Livak et al (202). The initial quantity of molecules normalized by a reference gene or endogenous and in relation to a control sample is given by:

$$2^{-\Delta\Delta Ct}$$

As explained above, the amplification reaction follows a sigmoidal kinetics, the equation that describes the amplification at its exponential base by PCR is given by:

$$X_n = X_0\,(1 + E_X)^n$$

Where:

X_n= Number of amplicon molecules at n reaction cycle.

X_0 = Initial number of sample molecules.

E_X = Amplification efficiency in the interest segment of the curve.

n = Number of cycles.

The threshold cycle (Ct) indicates the number of cycles required for the quantity of amplified product to reach a fixed threshold following this equation:

$$X_T = X_0\,(1 + E_X)^{C_{T,X}} = K_X$$

Where:

X_T = Number of molecules at threshold cycle.

X_0 = Initial number of sample molecules.

E_X = Amplification efficiency in the interest segment of the curve.

$C_{T,X}$ = Threshold cycle for a given amplification.

K_X = Constant at the exponential phase of the amplification.

We would have a similar equation for the mRNA amplification of the reference gene or normalization in each of the samples:

$$R_T = R_0\,(1 + E_R)^{C_{T,R}} = K_R$$

R_T= Number of mRNA molecules of the reference gene in the threshold cycle.

R_0 = Initial number of mRNA molecules of the reference gene.

E_R= Efficiency of the amplification of the reference gene.

$C_{T,R}$ = Threshold cycle for the reference gene amplification.

K_R = Constant in the exponential phase of the reference gene.

Dividing X_T by R_T, we get the expression:

$$\frac{X_T}{R_T} = \frac{X_0\,(1 + E_X)^{C_{T,X}}}{R_0\,(1 + E_R)^{C_{T,R}}} = \frac{K_X}{K_R} = K$$

Assuming that the efficacy of the sample amplification and the reference gene is the same, we have to:

$$E_X = E_R = E$$

$$X_N(1+E)^{\Delta C_T} = K$$

Where:

X_N= amount of problem molecule normalized per sample

ΔC_T = Difference between the threshold cycle of the problem gene and the threshold cycle for the reference gene.

Modifying this in the expression, we obtain:

$$X_N = K(1+E_R)^{-\Delta C_T}$$

Final step consists of dividing X_N of a sample (q) by X_N of the calibration sample (cb, sample control):

$$\frac{X_{N,q}}{X_{N,cb}} = \frac{K(1+E)^{-\Delta C_{T,q}}}{K(1+E)^{-\Delta C_{T,cb}}} = (1+E)^{-\Delta\Delta C_T}$$

$$\Delta\Delta C_T = \Delta C_{T,q} - \Delta C_{T,cb}$$

We calculate the relative expression, replacing in the formula:

$$2^{-\Delta\Delta C_T}$$

First, we calculate the arithmetic mean of the C_Ts:

$$\bar{M}_{C_T} = \frac{C_{T,1} + C_{T,2} + \cdots + C_{T,n}}{n} = \frac{\sum_{i=1}^{n} C_{Ti}}{n}$$

Where n is the number of replicates and $C_{T,i}$ the C_T values replicated per sample. It is also useful to calculate the standard deviation of the threshold cycles (SD_{Ct}), in order to guarantee a reliable mean and a minimum variation between replicates, the SD_{Ct} should be less than 0.1, this value corresponds to a coefficient of variation (C_V) of the percentage.

$$SD_{Ct} = \sqrt{\frac{1}{n-1}\sum_{i=1}^{n}\left(C_{T,i} - \bar{M}_{C_T}\right)^2}$$

Where n indicates the number of replicates by sample.

$$C_V = \frac{SD_{Ct}}{\bar{M}_{C_T}}\ 100$$

Once the mean values of C_T have been obtained, to normalize the values, we calculate the difference by sample between the own C_T and the C_T of the reference gene corresponding to the same sample problem.

$$\Delta C_T = C_{T,X} - C_{T,R}$$

Once the normalized C_T values are obtained per sample (q), the difference of each sample is obtained with the normalized C_T value of the calibrator (the calibrator consists of a sample with a null condition, that is, a control, which we use to establish a baseline reference limit).

$$\Delta\Delta C_T = \Delta C_{T,q} - \Delta C_{T,cb}$$

This value will be substitute in the equation derived from the relative expression, thus obtaining the expression values of the problem gene:

$$2^{-\Delta\Delta C_T}$$

By this method, the relative expression of all samples in the trials was estimated, taking into account the cases in which the SD of triplicate values exceeded 0.1 and, furthermore, one of the triplicate values was clearly deviated from the others two (attributable to an error in the qRT-PCR), this was eliminated leaving the other two as valid. With these values the pertinent statistical analyses were carried out. The QuanStudio™ Design and Analysis Software v1.4 (Applied Biosystems ™ by Life Technologies ™, Carlsbad, California, USA) was used to extract Ct data and results were processed by R Bioconductor.

11. Pathway enrichment analysis

The aim of the enrichment analysis is to elucidate in which pathways are involved the miRNAs or genes differentially methylated or expressed.

The enrichment analysis of routes in this work was carried out using enrichment online tools: Enrichr and DIANA-miRPath v3.0, of DianaTOOLS, already described (203-205).

MiRNAs enrichment analysis by Diana miRPath

For the enrichment analysis by Diana miRPath, only the significantly different expressed miRNAs for the Affymetrix study and miRNAs regulated by differentially methylated genes encoding miRNAs were analysed. Rosetta conversion online tool (http://www.miriadne.org/rosetta/convert) was used to convert miRNAs names from different platforms to miRBase v21.0 and a list of miRNAs for each study was generated. We considered only the experimentally validated miRNA-gene interaction included in TarBase. After this, the enrichment analysis of the target genes was carried out in the KEGG route database (version 67).

The algorithm used is Fisher's exact one-tailed test, because the null hypothesis is to detect routes only enriched in specific targets. In addition, the EASE value was obtained, a conservative adjustment method for Fisher's statistic (206). Both F-values of Benjamini and Hochberg were applied to both p-values as a correction for multiple comparisons. Two strategies for the analysis of enrichment were approached, the first one server combined the target genes of the selected miRNAs in a common supergroup (union of genes), this supergroup was incorporated in the enrichment analysis; On the other hand, the significance of the miRNAs is calculated individually among all possible pairs of miRNA-route, performing the enrichment analysis repeatedly; once this is done, the server combines the calculated individual p-values and provides a combined p-value using the Fisher's probability combination method, using a meta-analysis algorithm to join the p-values of two independent tests that are based on the same hypothesis.

Gene enrichment analysis by EnrichR

Enrichment analysis for genes that were differentially methylated and expressed in the DNA methylation study was performed in Enrichr online tool (http://amp.pharm.mssm.edu/Enrichr/enrich). This software performs an enrichment analysis of multiple genes using different gene-set libraries such as Reactome, Gene Ontology or KEGG. Pathways obtained with a p-value < 0.05 were considered significantly enriched.

12. Functional studies *in vitro*

12.1 Treatment with LMK-235 inhibitor

Breast cancer cell lines were treated with LMK-235 inhibitor (Selleck Chemicals) with a variable dose between 0.2-20 µM for 72 hours. Media were replaced (with drugs added) every day. LMK235 is a HDAC inhibitor, previously shown to be a novel and specific inhibitor of human HDAC4 and 5. Structure of HDACi LMK-235 inhibitor is represented in Figure 9.

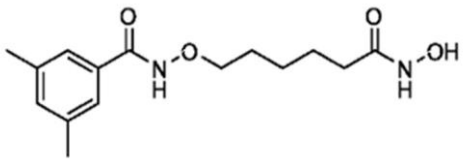


Figure 9. **LMK-235 inhibitor**. Structure of the class IIa HDACi LMK-235.

12.2 Cell proliferation assay

This assay reports the tumorigenic capacity of cells. The protocol used is based on a colorimetric MTT assay. We seeded and cultured 2.000 cells in 96-well plates. Next, cells were treated with a specific drug dose and the MTT assay was performed after 24, 48 and 72 hours after treatment. The MTT assay is a colorimetric assay for assessing cell metabolic activity. NAD(P)H-dependent cellular oxidoreductase enzymes may, under defined conditions, reflect the number of viable cells present. These enzymes are capable of reducing the tetrazolium dye MTT 3-(4,5-dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide to its insoluble formazan, which has a purple colour. MTT, a yellow tetrazole, is reduced to purple formazan in living cells. A solubilisation solution, specifically, dimethyl sulfoxide, is added to dissolve the insoluble purple formazan product into a coloured solution. The absorbance of this coloured solution was quantified by measuring at 590 by a spectrophotometer. Absorbance values that are lower than the control cells indicate a reduction in the rate of cell proliferation. Conversely a higher absorbance rate indicates an increase in cell proliferation. Each experiment was performed in triplicate and repeated at least twice. Average values for triplicates were calculated. Absorbance observed at different drug concentrations was compared with their respective controls not treated. Cell viability was calculated with the following equation:

$$\% \text{ Cell Viability} = (100 \times \text{Condition Absorbance}) / \text{Control Absorbance}$$

12.3 Scratch Wound-Healing assay

The wound-healing assay is simple, inexpensive, and one of the earliest developed methods to study directional cell migration *in vitro*. This method mimics cell migration during wound healing

in vivo. The basic steps involve creating a "wound" in a cell monolayer, capturing the images at the beginning and at regular intervals during cell migration to close the wound, and comparing the images to quantify the migration rate of the cells. In the present study, cells were seeded in 6-well plates at density of 5 x 10^5 cells/well and incubated overnight until they reached 70% confluence. A pipette tip was used to generate a wound in the cell layer. Cells were then treated with LMK-235 (20 µM). Each experiment was performed in triplicate and repeated at least twice. Images were obtained at 0, 24 and 48 hours at the same position and percentage of cell migration was evaluated using ImageJ, a programme for digital image processing.

RESULTS

RESULTS

1. Characterization of breast cancer cell lines from very young women

In order to identified a good cellular model for breast cancer studies affecting very young women, we characterized four breast cancer cell lines from women younger than 35 years that were commercially available and are indicated in Table 2. In terms of morphology all breast cancer cell lines grew forming adherent monolayers and showed epithelial morphology (Figure 10).

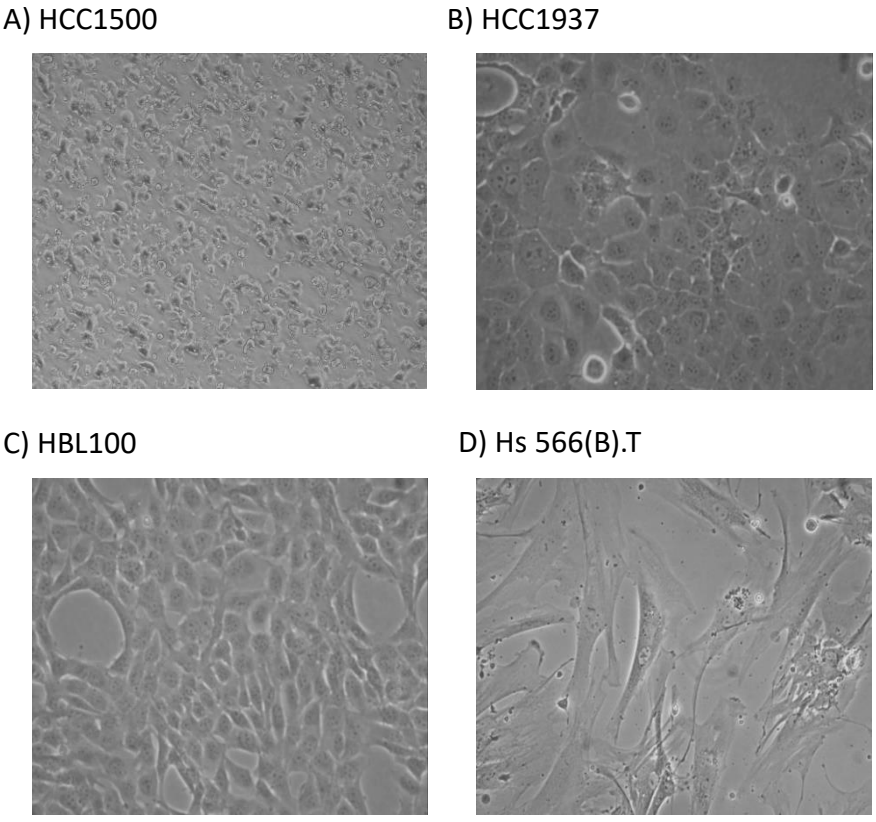


Figure 10. Morphology of breast cancer cell lines from very young women. Images were taken at 40x and all breast cancer cell lines were near 95% of confluence. (**A**) HCC1500; (**B**) HCC1937; (**C**) HBL100 and (**D**) Hs 566(B).T.

The grown kinetics of HCC 1500, HCC 1937 and HBL100 cell lines were determined as described in material and methods section at passage 8. The growth curves of these cell lines were shown in Figure 11. The population doubling time of HCC1500, HCC1937 and HBL100 were 4.56, 3.78 and 4.58 days respectively.

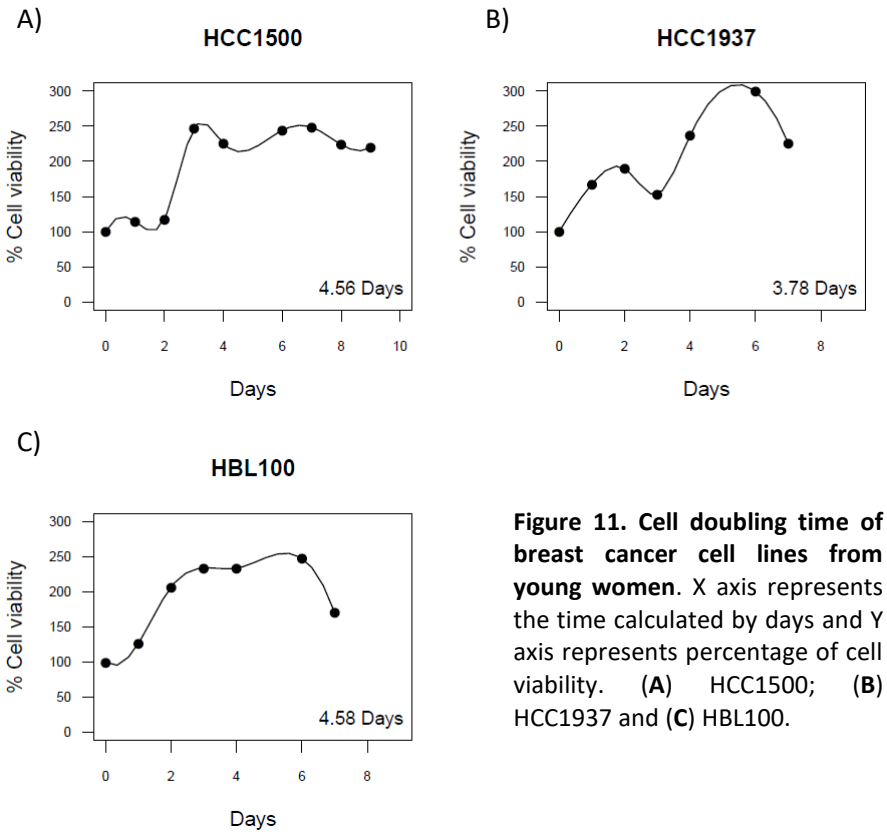


Figure 11. Cell doubling time of breast cancer cell lines from young women. X axis represents the time calculated by days and Y axis represents percentage of cell viability. (**A**) HCC1500; (**B**) HCC1937 and (**C**) HBL100.

Hs566(B).T cell line is less frequently used in breast cancer studies and few information is published in the bibliography about it. Moreover, this breast cancer cell line presented a slow grown kinetics and was excluded in the following studies.

On the other hand, HBL100 cell line was originally described as a breast transformed but non-tumorigenic cell line established *in vitro* from young women apparently healthy, but recently it has been found to contain Y chromosome (ATCC website http://www.lgcstandards-

atcc.org/Products/Cells_and_Microorganisms/Cell_Lines/Misidentified_Cell_Lines.aspx?geo_country=es) and has been discontinued by ATCC. For that reason, we excluded it in the further studies.

Finally, only HCC1500 and HCC1937 cell lines were used in the breast cancer studies that affect very young women.

2. MiRNA expression analysis by Affymetrix array

2.1 MiRNA expression in breast cancer cell lines from BCVY and BCO

Analysis of array intensity and quality control

After array normalization by RMA (Robust Multiarray Average) we represented the Kernel density estimator in order to evaluate the variability among the different arrays composing our cell samples. At this step, we are evaluating the possibility that some sample/array presents an unacceptable level of noise and should be removed from the study. The Kernel density estimator showed high right symmetry for our cell samples (Figure 12a). Analogously, we appreciated a symmetry in the expression variability within array and among arrays in the box diagrams representation (Figure 12b). In both plots, the mean Log Intensity values were around 1. The symmetry among arrays guarantee that the results from arrays could be comparable to each other.

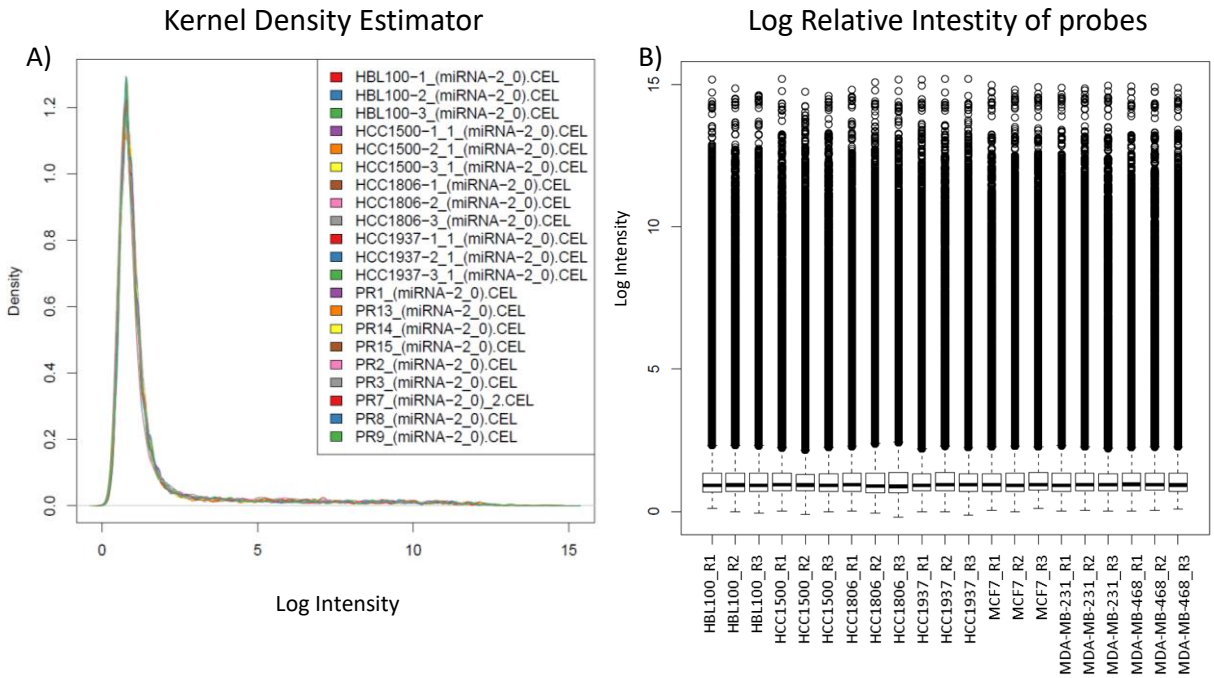


Figure 12. Logarithmic intensity distribution of probes included in each array. Kernel estimator of the intensities for the different microarrays for breast cancer cell lines (**A**). Box plot representations of the probe level expression for breast cancer cell lines analysed (**B**).

MiRNA expression in breast cancer cell lines

After initial pre-processing, array maintained 2 275 probes for human miRNAs. Among them, 1 121 were pre-miRNA transcript probes and were removed in the study. Finally, we analysed the differential expression for 1 154 miRNAs. The statistical analysis showed that 302 miRNAs were differentially expressed between breast cancer cell lines from young women (HCC1500 and HCC1937) and older counterparts (MCF-7, MDA-MB-231, MDA-MB-468 and HCC1806) (p-value < 0.05) and after adjusting for FDR (FDR < 0.05) 155 miRNAs remained significant. Values obtained are included in Annex I.

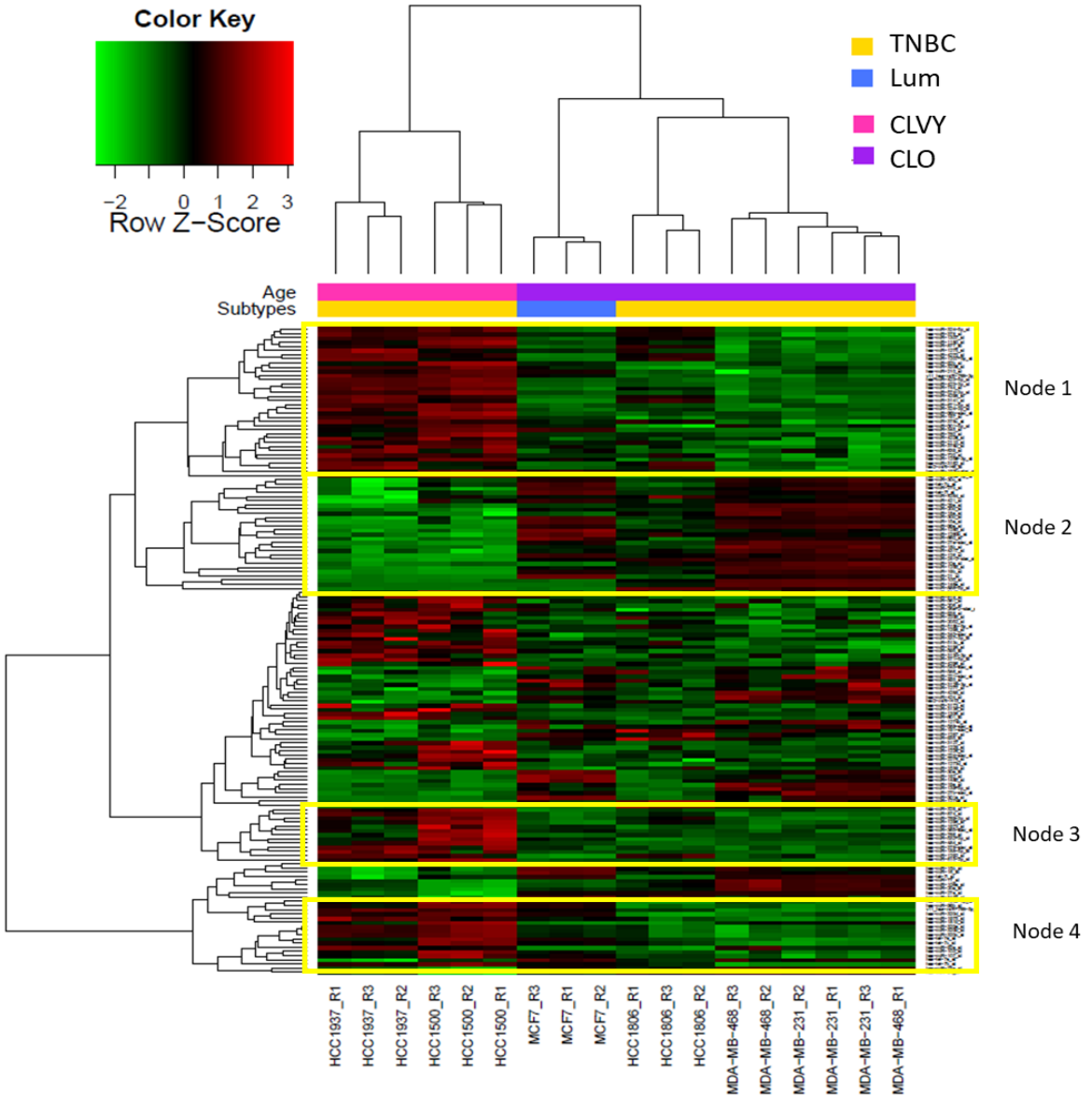


Figure 13. Hierarchical clustering centered on the median and visualization of the expression of significantly deregulated miRNAs in cell lines from young women. Heatmap representing the expression of 155 miRNAs that were significantly deregulated in breast cancer cell lines from young women. The expression is shown in different intensities, green when it is below the median, red above and in black unchanged with respect to the median. Cell lines are represented in pink colour for younger than 35 years old and older are indicated in purple. Molecular subtypes are indicated as: triple negative (yellow) and luminal (blue). TNBC: triple negative breast cancer; Lum: luminal; CLO: breast cancer cell lines from women older than 45 years old; CLVY: breast cancer cell lines from women younger than 35 years old. Replicates are represented as a R.

Average linkage hierarchical clustering with data from the 155 significant miRNAs was performed to obtain clusters of data set. Heatmap representation shows two major subgroups,

separating CLVY from CLO (Figure 13). Biological triplicates for each breast cancer cell lines presented a similar miRNA profile and clustered together in the Heatmap representation, with the exception of one triplicate of MDA-MB-468 cell line. These results show a good reproducibility among biological triplicates. Additionally, we found a global higher miRNA expression in younger cell lines, specifically 60% of significant miRNAs were overexpressed CLVY (red colour in Heatmap representation) and the rest 40% were repressed in this group.

We highlighted four sub-nodes that presented homogeneous expression between age groups. Node 1,3 and 4 shown a group of miRNA overexpressed in CLVY that were selected for further pathway enrichment analysis. Interestingly, we identified a group of 24 significant miRNAs that presented a miRNA expression pattern similar for HCC1806 cell line and CLVY and they are indicated in node 2. MiRNAs for each sub-node are included in Annex II.

Pathway enrichment analysis for breast cancer cell lines

Enrichment study show 22 pathways that were significantly deregulated (FDR adjusted p-value < 0.05) by the significant miRNAs. We found pathways related with extracellular matrix (ECM) receptor interaction (hsa04512) or proteoglycans (hsa05205). Interestingly, we identified pathways related with nervous system such as axon guidance (hsa04360), hippo signalling pathways (hsa04390), GABAergic synapse (hsa04727) or neurotrophic signalling (hsa04722). Additionally, we identified pathways related with ErbB (hsa04012), Wnt (hsa04310), Ras (hsa04014), PI3K-Akt (hsa04151) or MAPK signalling (hsa04010). Focal adhesion (hsa04510) and adherent junctions (hsa04520) were, also, deregulated by significant miRNA found. List of deregulated pathways identified are represented in Figure 14a.

MiRNA expression study shown a group of 24 miRNAs (included in Node 2) that presents a similar expression between CLVY and HCC1806. We sought to analyse the pathways in which these miRNAs were involve. MiRNA enrichment analysis shown about 15 deregulated pathways with a FDR adjusted p-value < 0.001 (Figure 14b). The most significant pathway, in which 19 miRNAs (among the 24 studied), were taking part was related with extracellular matrix receptor interaction (hsa04512). Additionally, cancer related pathways such as proteoglycans (hsa05205) were observed. Interestingly, we identified relevant signalling pathways in cancer such as mTOR (hsa04150), PI3K-Akt (hsa04151), ErbB (hsa04012), Ras (hsa04014), Wnt (hsa04310) and MAPK (hsa04010). Focal adhesion pathways (hsa04510) were deregulated by the miRNA studied. Furthermore, we have to highlight the implication of 23 miRNAs in signalling pathways regulating pluripotency of stem cells (hsa04550).

When studying miRNAs from the selected sub-nodes indicated in Figure 13, KEGG pathway enrichment analysis revealed several pathways overrepresented with a FDR p-value < 0.05. Most of the pathways observed were deregulated by miRNA from different sub-nodes. However, we identified particular cluster altered pathways. MiRNAs from node 1 were taking part in cancer metabolisms related pathways such as central carbon (hsa05230) or choline (hsa05231). Signalling pathways related with calcium (hsa04020), insulin (hsa04910), MAPK (hsa04010), oxytocin (hsa04921), TFG-beta (hsa04350) were also deregulated by miRNA from node 1. MiRNAs from node 3 were taking part in cell adhesion (hsa04514), N-Glycan biosynthesis (hsa00510) and nervous system related pathways such as hippo signalling (hsa04390). Node 4 shown biological proliferation or differentiation-related pathways relevant in cancer development or progression as AMPK (hsa04152), mTOR (hsa04150), PI3K-Akt (hsa04151), oestrogen (hsa04915) and Wnt (hsa04310) signalling pathways among others. More information

about node miRNA's enrichment and FDR p-values associated with each pathway are shown in Annex III.

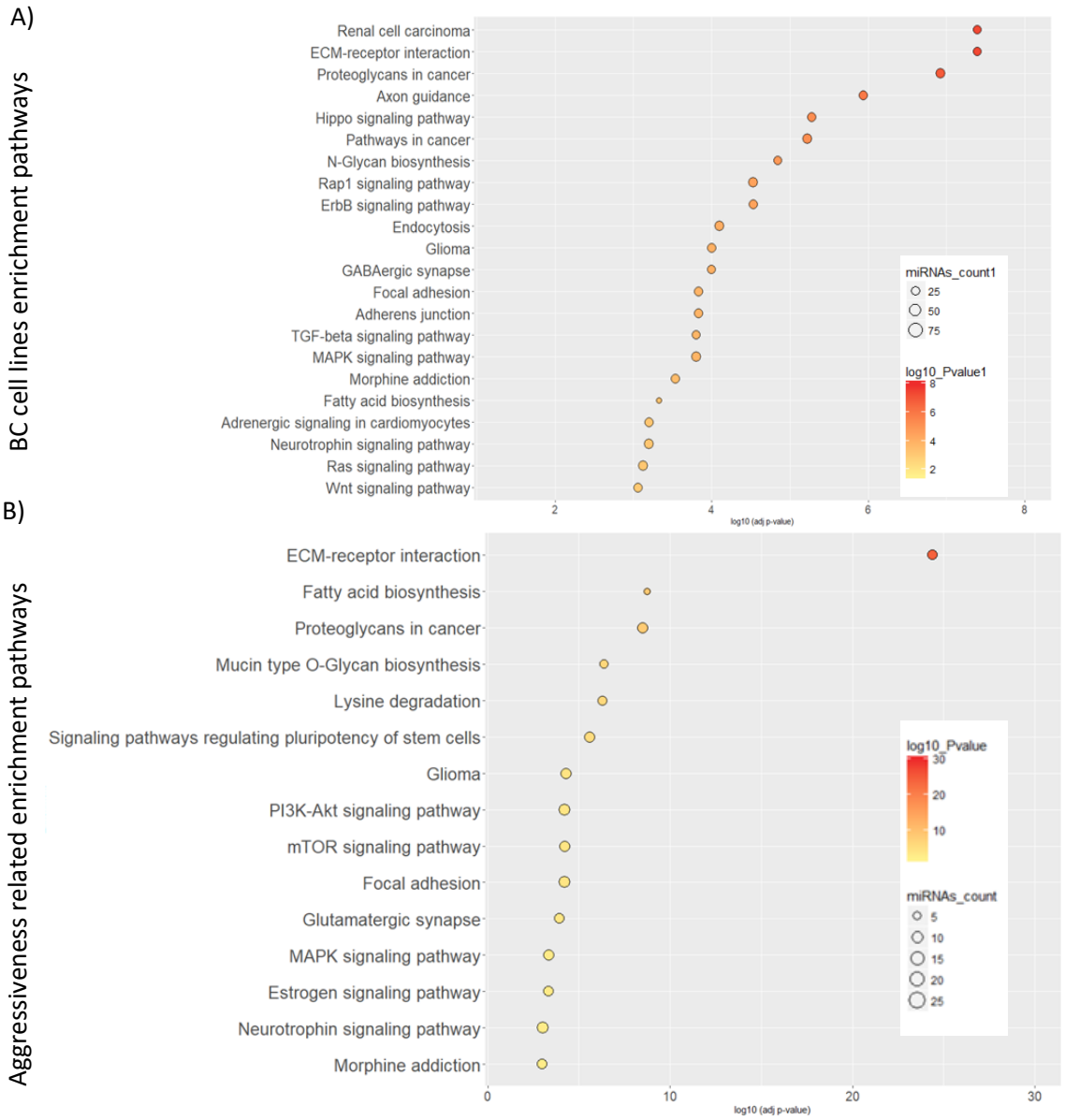


Figure 14. Pathways significantly deregulated by miRNA differentially expressed in breast cancer cell lines from young women. Enrichment study was performed by DIANA mirPath and pathways obtained presented a FDR adjusted p-value < 0.001. (**A**) Pathways altered by the 155 significant miRNAs deregulated in breast cancer cell lines from young women; (**B**) Pathways in which the 24 miRNA aggressiveness-related were involved. Dot colour indicates a logarithmic transformation of BH adjusted p-values (orange for most significant pathways and yellow for lower p-values) and miRNA count indicates the number of miRNAs involved in the represented pathways.

Validation by qRT-PCR of miRNA expression in breast cancer cell lines

The nature of the data and the methods used, can produce false positives results due to the technique used or to the stochastic processes. Therefore, we must evaluate the robustness of the data obtained and confirm that the selected miRNAs are deregulated in young women. To this end, a study of the expression of the miRNAs selected according to the previous detailed process was carried out, using a different technique, more direct, such as the analysis of the expression by qRT-PCR. We analysed three repressed miRNAs (miR-23a, miR-139-5p, miR-30c) and 4 overexpressed (miR-1207-5p, miR-1275 and miR-3196 and miR-92b). We performed the validation step in a different set of cellular extracts using new biological triplicates of breast cancer cell lines included in the array.

After miRNA amplification by PCR, we analysed the mean Ct for replicates, the standard deviation and calculated the ΔΔCt, as previously described in material and methods section. We remove outlier values and performed the normality test Shapiro-Wilk. All samples tested presented a p-value < 0.05, then there were evidences that the data were not normally distributed. After that, we performed the Wilcoxon rank sum test for non-parametric samples to compare relative expression values for each miRNA between CLVY vs. CLO.

Results shown significant differences between cell lines from young and old women for five of 7 selected miRNAs: miR-1275 (FDR p-value = 2.6×10^{-3}), miR-92b (FDR p-value = 2.6×10^{-3}), miR-30c (FDR p-value = 7.5×10^{-3}), miR-23a (FDR p-value= 1.6×10^{-2}) and miR-1207-5p (FDR p-value = 4.4×10^{-2}) (Figure 15). MiR-23a and miR-30c expression values agree with downregulation

observed in younger cell lines, while the rest significant miRNAs shown contrary results between qRT-PCR validation and previous array expression data.

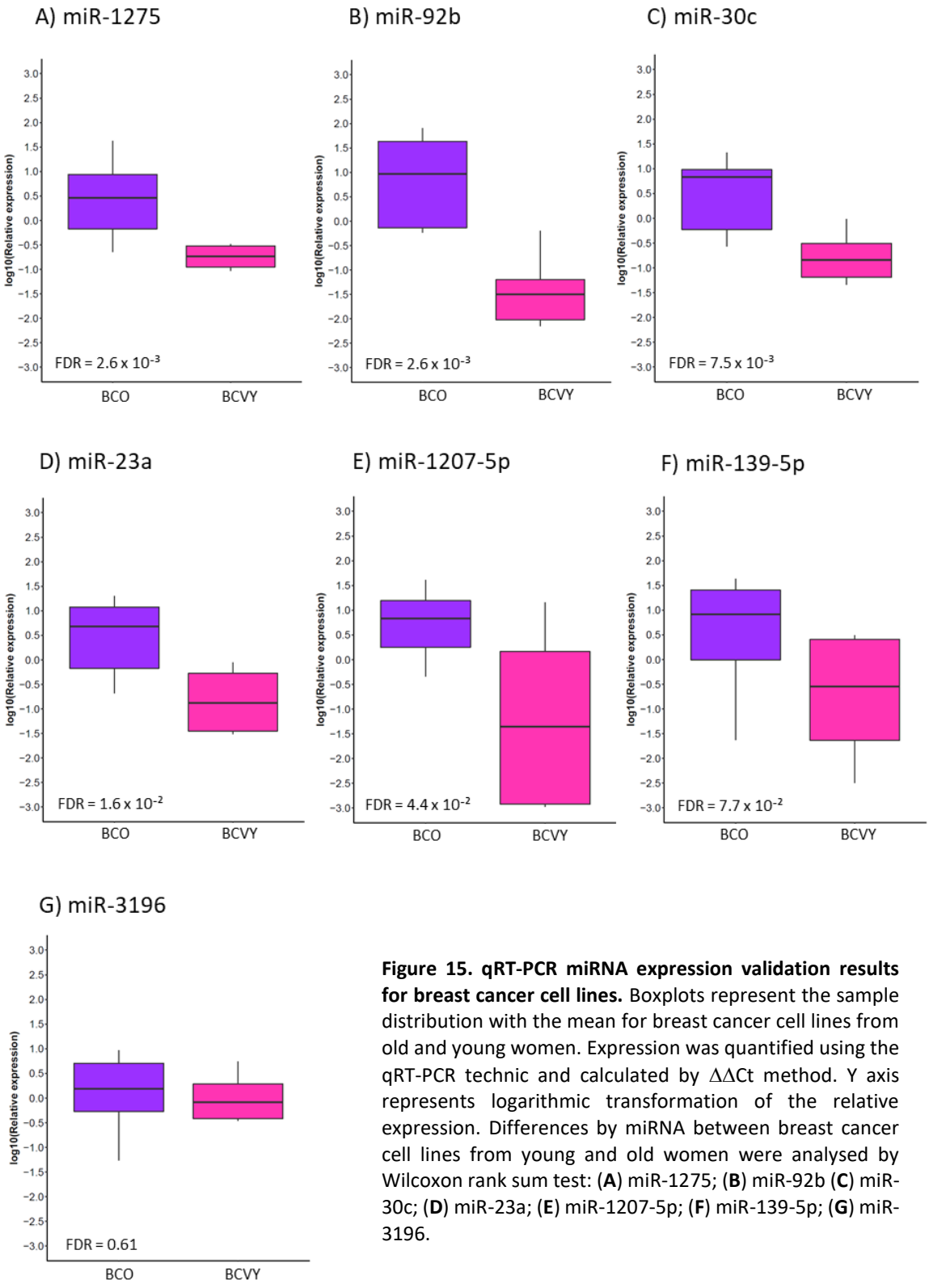


Figure 15. qRT-PCR miRNA expression validation results for breast cancer cell lines. Boxplots represent the sample distribution with the mean for breast cancer cell lines from old and young women. Expression was quantified using the qRT-PCR technic and calculated by ΔΔCt method. Y axis represents logarithmic transformation of the relative expression. Differences by miRNA between breast cancer cell lines from young and old women were analysed by Wilcoxon rank sum test: (**A**) miR-1275; (**B**) miR-92b (**C**) miR-30c; (**D**) miR-23a; (**E**) miR-1207-5p; (**F**) miR-139-5p; (**G**) miR-3196.

2.2 Comparison miRNA expression between breast cancer cell line and FFPE breast cancer samples

In order to evaluate the reproducibility of the miRNA profile observed in cell lines, we performed a comparative study between miRNA expression in breast cancer cell lines and breast cancer FFPE tumour samples obtained in previous group studies (160).

Analysis of array intensity and quality control

We re-analysed the hybridation chips from breast cancer FFPE tumour samples and cell lines in a combined study. The combined study was normalized following the same steps as breast cancer cell lines study, previously described. We performed an array normalization by RMA (Robust Multiarray Average) of all samples together. After normalization, we represented the Kernel density estimator in order to evaluate the variability among the different arrays from cells and patients. We found higher similarity among cell lines. Patients samples presented major dispersion in intensity values. However, the mean Log Intensity values were around 1 for all samples (Figure 16). The symmetry among arrays guarantee that the results from cell and patients arrays could be comparable to each other and could be analysed in the same study.

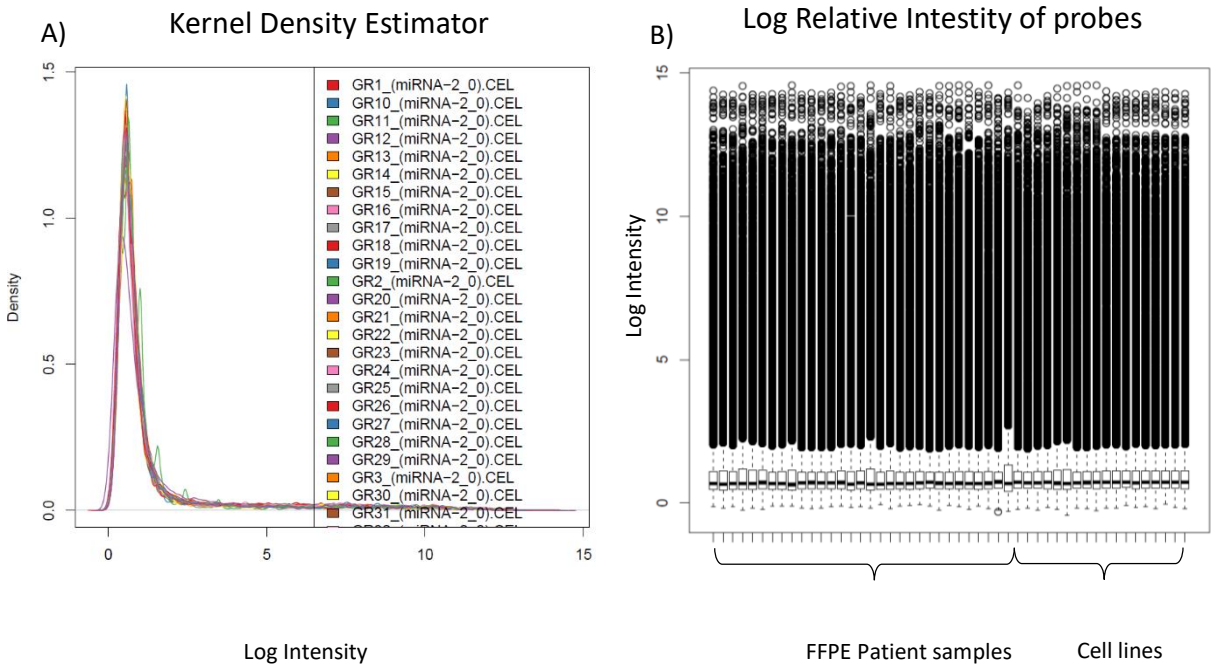


Figure 16. Logarithmic intensity distribution of probes included in each array. Kernel estimator of the intensities for the different microarrays for breast cancer cell lines (**A**). Box plot representations of the probe level expression for breast cancer cell lines analysed (**B**).

MiRNA expression in breast cancer cell lines and patients

The differential expression between young and old samples (cell lines and patients) of 1 154 human miRNA included in the array were analysed by t-test. Results shown 305 miRNAs differentially expressed in both, breast cancer samples from very young women patients (BCVY) and CLVY, comparing with breast cancer older patients (BCO) and CLO, with p-value < 0.05 and 132 of them were adjusted by FDR < 0.05. The 54% of significant miRNAs were upregulated in younger samples and the remaining 46% were repressed on them. List of significant miRNAs are summarized in Annex IV.

We performed an average linkage hierarchical clustering with data from the 132 significant miRNAs to obtain clusters of data set (Figure 17). With the exception of HCC1806 cell line and

few samples from BCO patients, heatmap representation shows two major subgroups, separating BCVY and CLVY samples from older counterparts. In agreement with previous observations in breast cancer cell lines, HCC1806, presents a miRNA profile more similar to CLVY and BCVY. Furthermore, in this combined study, the expression of HCC1806 cell line, is similar to younger samples for all miRNAs studied. Two sub-nodes were selected for further enrichment pathway analysis and miRNAs included are summarized in Annex V.

Next, we studied whether miRNA expression differences observed between BC age groups were related with molecular subtypes. GLM analysis showed that miRNA expression differences that distinguish young from old samples were specific of age and not related with any molecular subtype (p-value = 0.66). Supporting that, no molecular subtype clusters were observed in heatmap representation.

Pathway enrichment analysis for breast cancer cell lines and patients

Enrichment pathway analysis with the significant miRNAs that were differentially expressed in breast cancer cell lines and patients shown 23 deregulated pathways and most of them were previously observed in the miRNAs expression study in BC cell lines (Figure 18). Again, the most significant pathway obtained was related with ECM receptor interaction with an FDR adjusted p-value = 3.6×10^{-13} in which 74 miRNAs of the 132 studied were implicated. RAS, RAP1, ERBBB, PI3K-AKT, mTOR and WNT signalling pathways among other were significantly deregulated in younger patients. We could identify, also, nervous system related pathways such as axon guidance, hippo signalling pathways or glutamatergic synapse, as previously observed in cell lines. Pathways related with adherent junctions and focal adhesion were significantly altered by miRNAs deregulated.

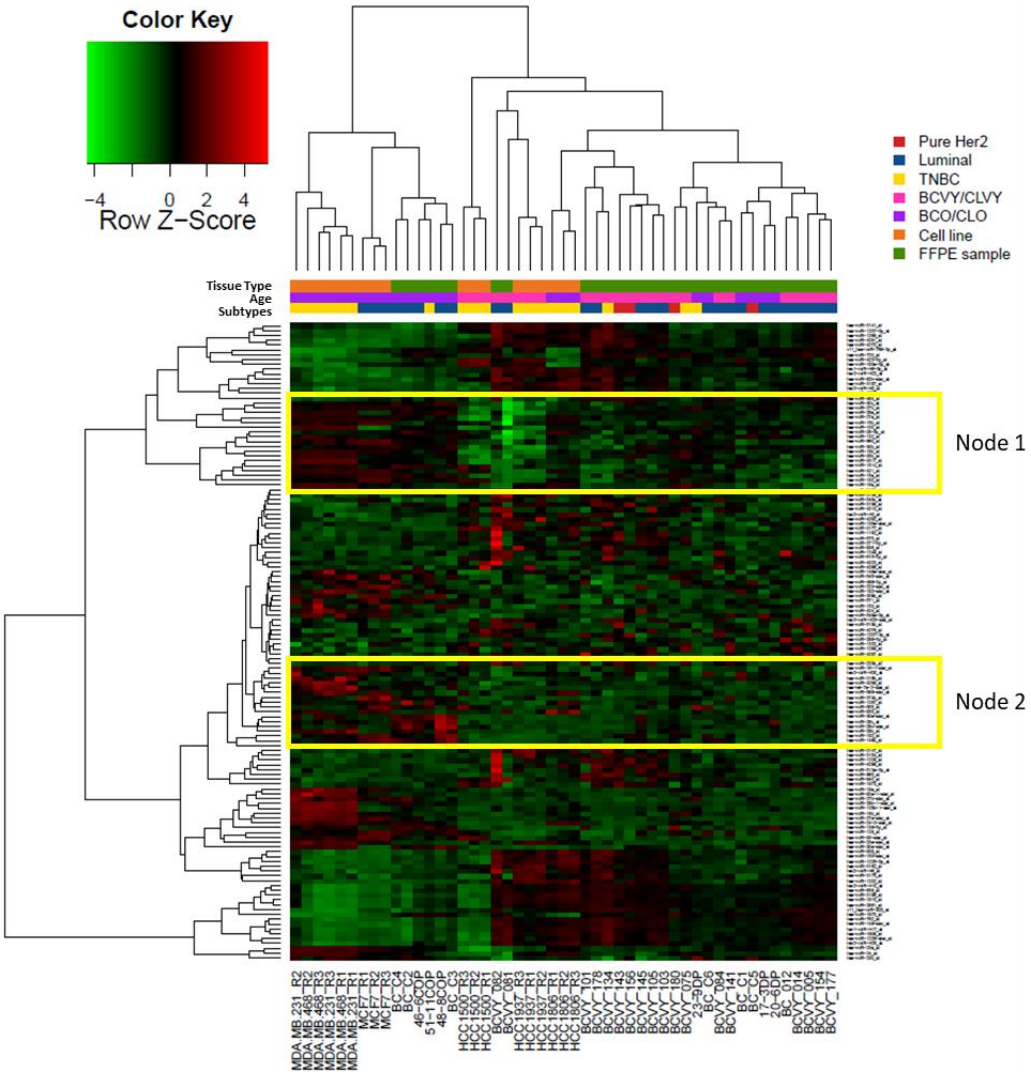


Figure 17. Hierarchical clustering centered on the median representation of the expression of significantly deregulated miRNAs between young and old samples (patients and cell lines). Heatmap representing the expression of 132 miRNAs that were significantly deregulated between young and old samples from BC patient tissue and cell lines. The expression is shown in different intensities, green when it is below the median, red above and in black unchanged with respect to the median. MiRNA sub-nodes 1 and 2 are highlighted in yellow. Young samples are represented in pink colour and older are indicated in purple. Molecular subtypes are indicated as: Pure Her2 (red), triple negative (yellow) and luminal (blue). Sample precedence is indicated as: orange for breast cancer cell lines and green for FFPE patient tissue samples. TNBC: triple negative breast cancer; Lum: luminal; BCO: breast cancer tumours from women older than 35 years old; BCVY: breast cancer tumours from women younger than 35 years old; CLO: BC cell lines from older patients; CLVY: BC cell lines from very young patients; FFPE: Formalin-fixed, paraffin-embedded.

We analysed the enrichment pathways by the two selected sub-nodes. Node 1 shown multiple pathways altered and involved in different functions. We identified deregulation in AMPK, ERBB, mTOR or prolactin signalling pathways, among others. Most of them were previously observed in the global enrichment study for BC patients and cell lines. Interestingly, node 2 miRNAs were taking part in circadian rhythm and cardiomyopathies. List of deregulated pathways by sub-nodes are summarized in Annex VI.

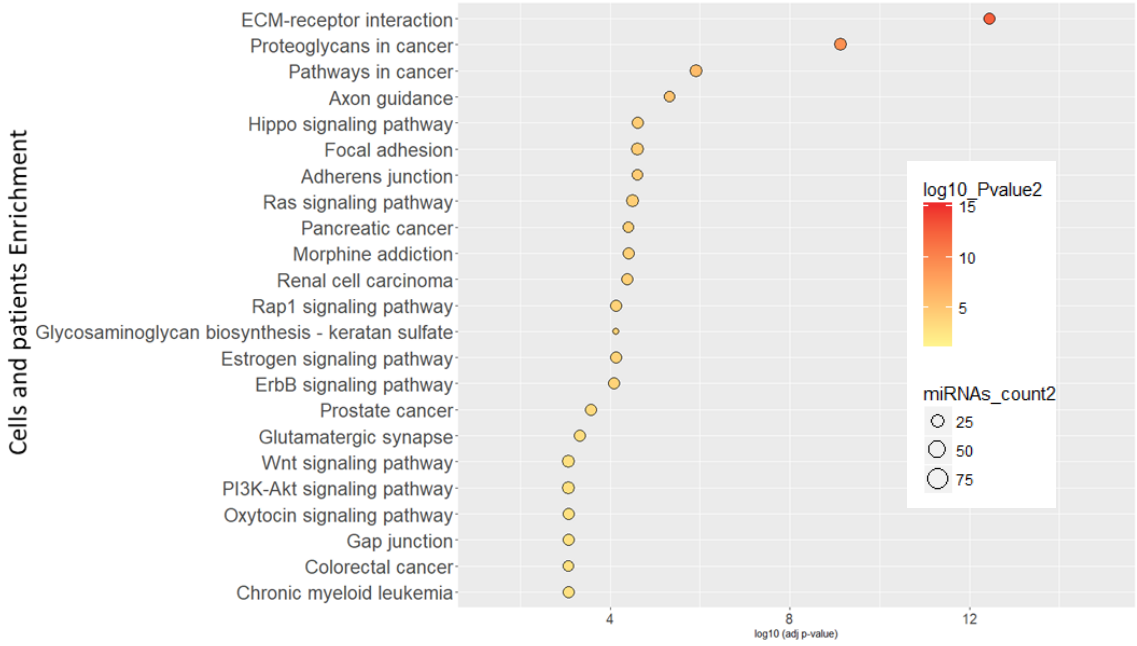


Figure 18. Pathways significantly deregulated by the 132 miRNA differentially expressed in the combined study with breast cancer cell lines and FFPE samples. Enrichment study was performed by DIANA mirPath and pathways obtained presented a FDR adjusted p-value < 0.001. Dot colour indicates a logarithmic transformation of FDR adjusted p-values (orange for most significant pathways and yellow for lower p-values) and miRNA count indicates the number of miRNAs involved in the represented pathways.

2.3 Common significant miRNA from breast cancer cell lines and tumour patient samples

We compared the results obtained in the present thesis (miRNA data in cell lines) with miRNA data from combined study (using patient tissue and cell lines) and data previously described by

our group (160) and combined study with patients and cell lines. Significant miRNA obtained from different analysis are represented in Venn diagram (Figure 19). Results shown 10 miRNAs that were significantly deregulated in younger samples in the three studies (miR-1207-5p, miR-1275, miR-1973, miR-220b, miR-23a, miR-27b*, miR-3141, miR-3162, miR-4270 and miR-4317). MiR-23a downregulation was validated by qRT-PCR in both, young cell lines and BC patient samples in Peña-Chilet (160). This result supports the miR-23a role in BCVY women and their potential biomarker function in this group of age.

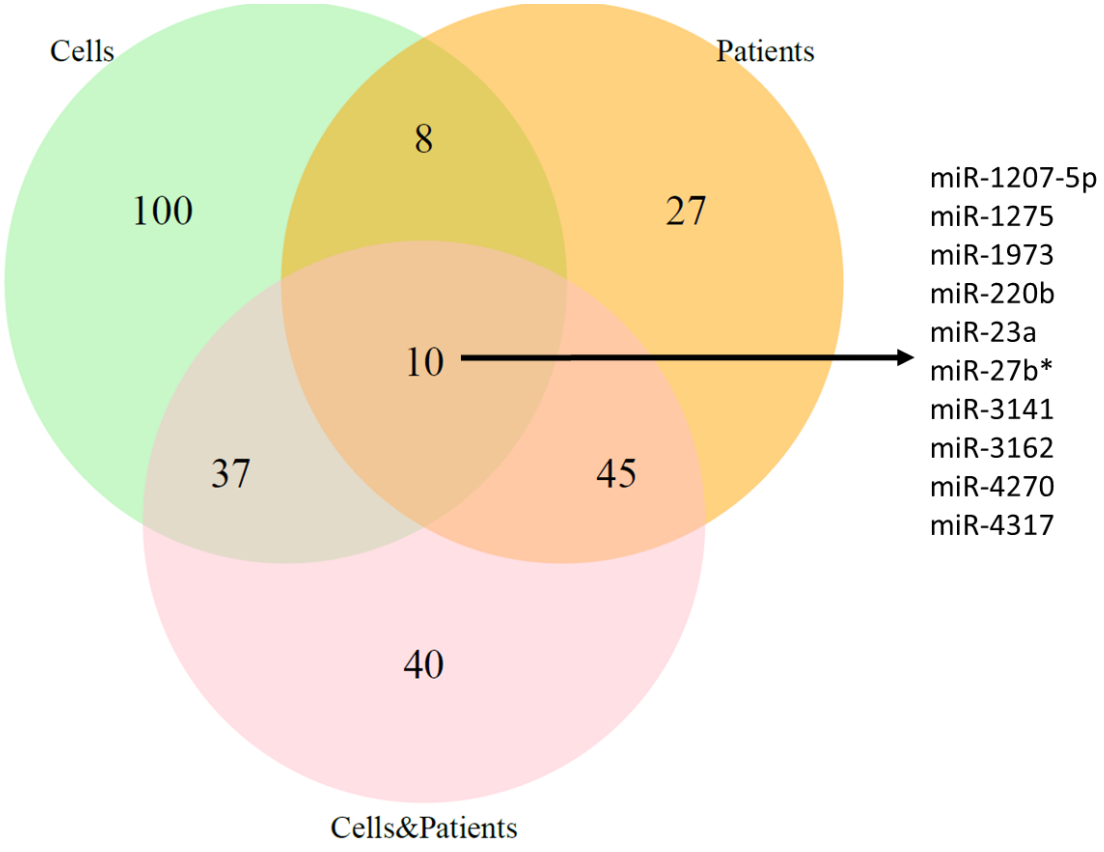


Figure 19. Venn diagram representation of significant miRNAs obtained in the three statistical analysis. Diagram show 10 miRNAs that were significantly different expressed in young samples, compared with older counterparts, in the three studies with adjusted FDR p-value < 0.05.

2.4 Clinical importance of miR-23a expression as an independent prognostic factor for relapse-free survival and overall survival in BCVY

To test the hypothesis of the potential association of miR-23a repression in BCVY with RFS and/or OS, we analysed the follow-up data of TCGA and METABRIC samples with miRNA expression data. Specifically, TCGA miRNA expression data includes 635 women older than 45 years old and 23 BCVY samples. METABRIC analyses 1.109 samples older than 45 years old and 37 from women younger than 35 years old. We next performed a univariate cox regression study to determine whether repression or overexpression of miR-23a were correlated with underlying clinical conditions in BCVY and BCO. Survival studies were performed using survival R package.

RFS study did not show significant correlation between miR-23a expression status and relapse neither in BCO (p-value = 0.09) nor in BCVY women (p-value = 0.36) (Figure 20a). Additionally, nine samples from BCVY patients were removed in the RFS study because they did not present follow-up data. Although, no significant results were achieved, BCVY patients with miR-23a overexpression presented higher relapse incidence. However, no conclusions can be draw because of the limited sample size for BCVY patients.

We could not obtain significant association between OS and miR-23a expression neither in BCO (p-value= 0.06) nor in BCVY (p-value = 0.52) samples (Figure 20b). However, we observed poor survival for BCVY patients with miR-23a repression comparing with BCO who presented worse prognosis when miR-23a was overexpressed.

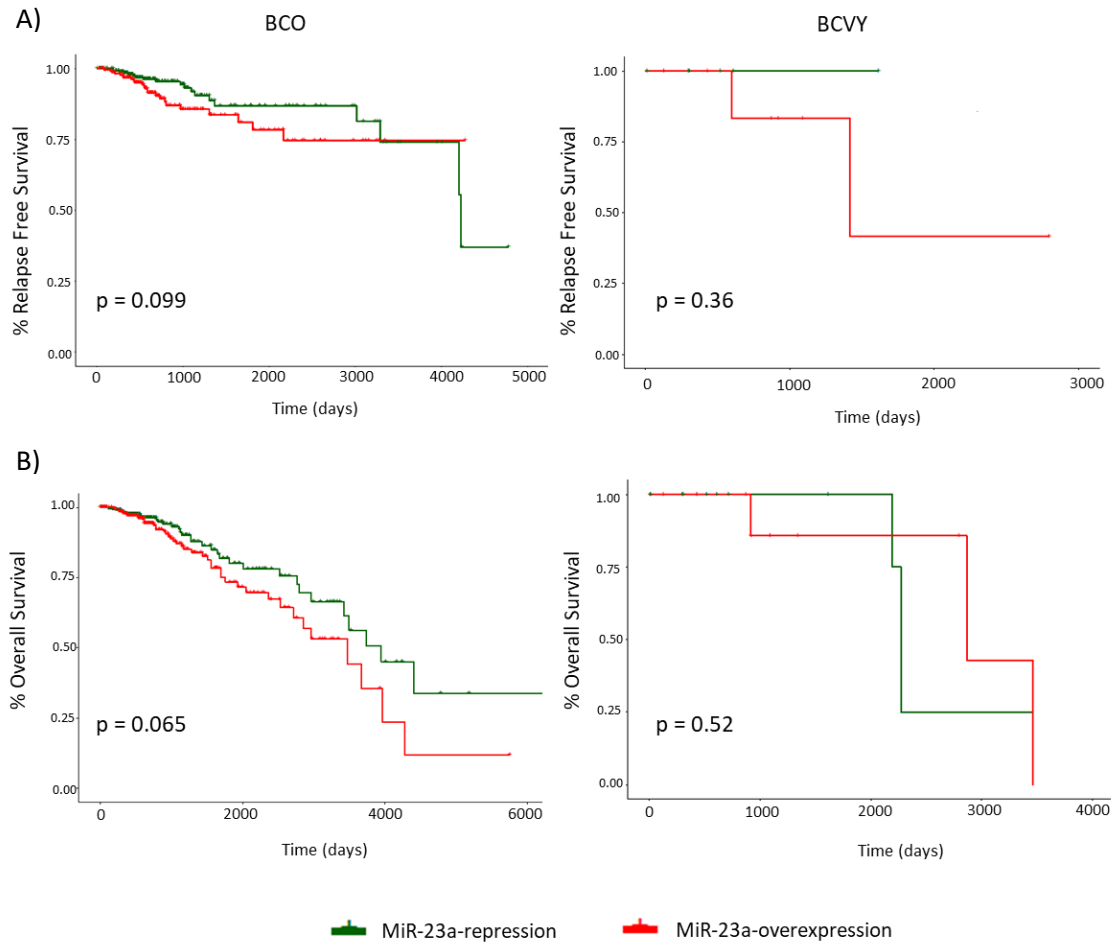


Figure 20. Representation of relapse free survival and overall survival curves for miR-23a expression analysed in breast cancer samples from TCGA and METABRIC. Relapse free survival in BCO (left) and BCVY (right) (**A**); Overall survival in BCO (left) and BCVY (right) (**B**). X axis represents the follow-up time by days and Y axis indicates the percentage of relapse free survival or overall survival. Green curves represent miR-23a repression and red colour line represents miR-23a overexpression. P-values were obtained by univariate cox analysis. BCVY: breast cancer in very young women; BCO: breast cancer in old women.

3. DNA methylation analysis by Infinium Human Methylation EPIC array

3.1 DNA methylation BeadChip Array study

Array pre-processing

Raw methylation data (*. IDAT files) were imported using illuminaio tool implemented in *minfi* package. Then a number of quality control metrics are examined to determine the success of

the bisulphite conversion and subsequently array hybridation. The EPICarray contains several internal control probes, previously mentioned, that have been used to assess the quality control of different preparation steps (bisulfite conversion, hybrydation, etc). The values of these control probes for each samples were extracted and analysed. We excluded for further methylation studies samples that presented a value, for some of the quality controls, lower than two standard deviations from the mean threshold. Specifically, six samples were excluded because of they fail at the bisulfite conversion step.

Additionally, we used the quality control plots from *minfi* package for interactive visualization (Figure 21). All samples clustered together in the plot with the exception of one which was separated and presented lower median intensity, so this sample was excluded for further analysis.

We remove probes that failed in 20% of samples. Furthermore, we check whether existed samples in which the 5% of probes failed, however, after quality control process, any sample met the criteria. After pre-processing, we analysed 809.822 CpG sites in 21 samples from BCVY, 13 from BCO, 2 NVY and 3 from NO in our metEPICVal study. Table 5 summarized probe number and sample number keep after every quality control step.

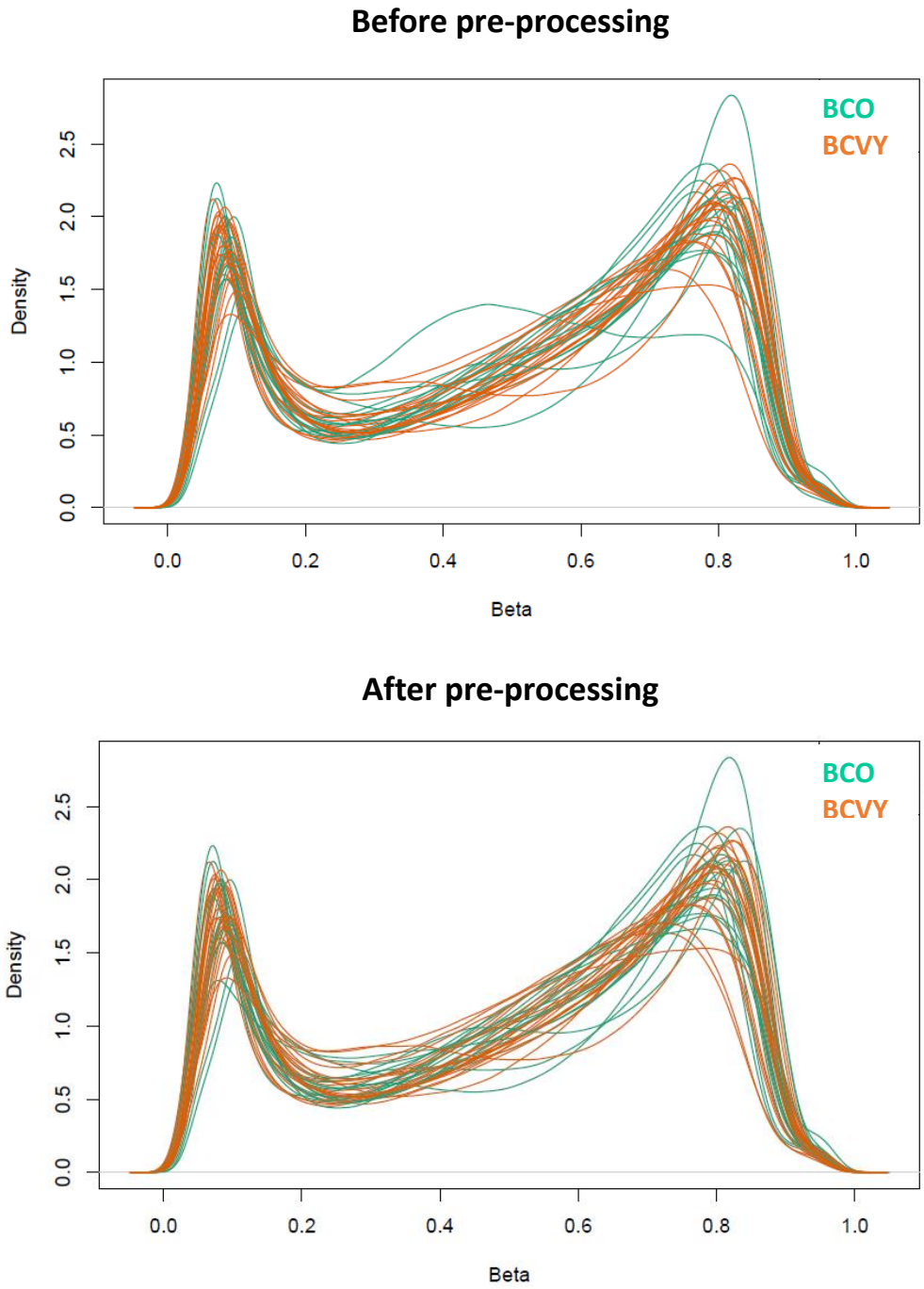


Figure 21. Density plots of Beta values for BCO and BCVY samples. Beta value densities before pre-processing (**A**). Beta values after pre-processing (**B**). BCVY: breast cancer in very young women; BCO: breast cancer in old women.

Table 5. Summary of CpG probes and samples remaining in the MetEPICVal and KFdata after each quality control step.

	MetEPICVal		KFdata	
Filter	**Probes**	**Samples**	**Probes**	**Samples**
Raw data	866.836	46	485.512	52
Control probes	866.836	39	485.512	49
Probes success in 80% of samples	865.223	39	480.543	49
Samples with 98% of successful probes	865.223	39	480.543	47
SNPs	835.387	39	463.312	47
Cross Reactive Probes	809.822	39	435.939	47
Filtered data	**809.822**	**39**	**435.939**	**47**

Global hypomethylation in BCVY

Statistical analysis of the metEPICVal study samples revealed 44 032 CpG sites differentially-methylated between BCO vs normal tissue samples and 36 628 CpG for BCVY vs. normal tissues. After removing 18 150 different methylated CpGs that were common to both the control and experimental groups, we analysed the differences in methylation between BCVY and BCO; we identified 2 219 CpG sites significantly differentially methylated in BC depending on patient age. The diagram of the analysis procedure is included in Figure 6. There was a general hypomethylation profile in BCVY in which 69% of significant CpG sites were hypomethylated and the remaining 31% were hypermethylated compared to BCO (Figure 22).

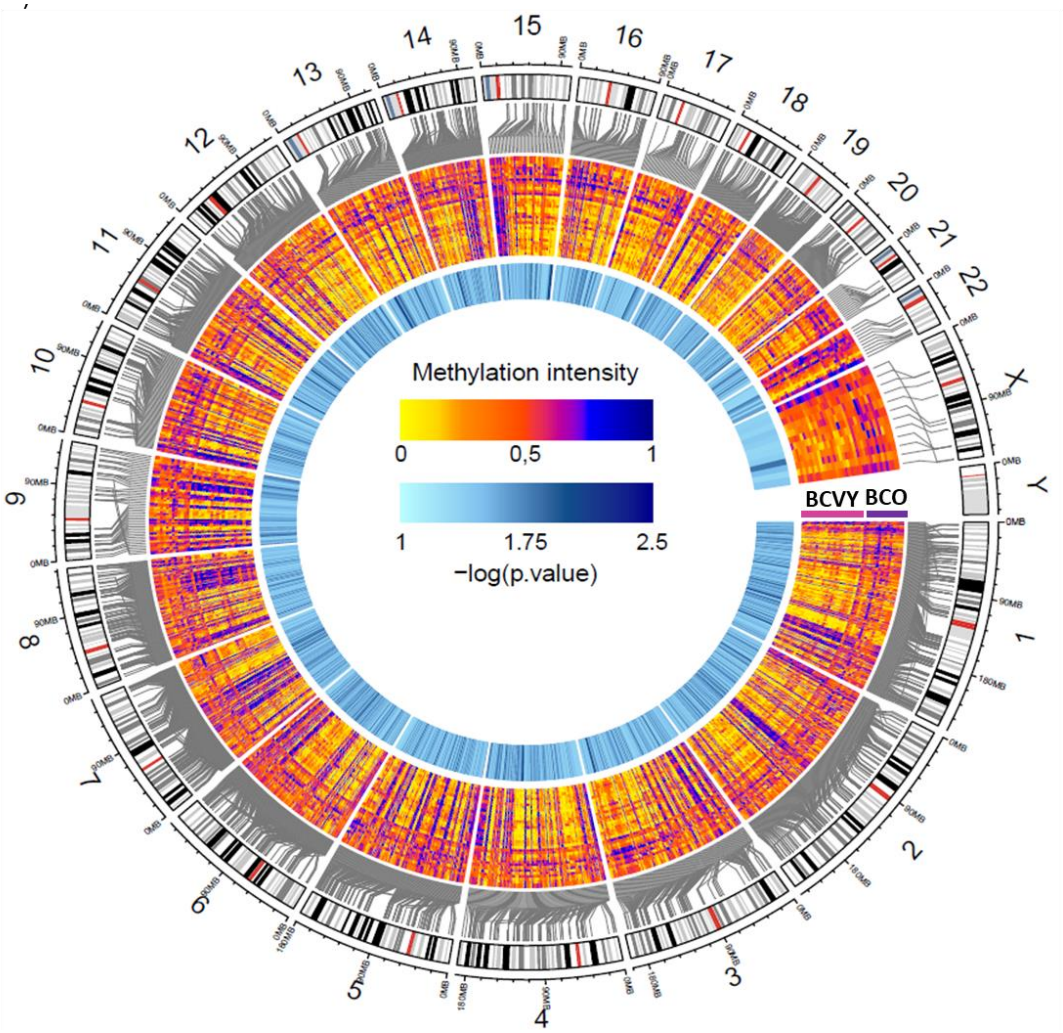


Figure 22. Differential methylation study in breast cancer in very young women (BCVY) vs. their older (BCO) counterparts from metEPICVal samples. Circular plot represents DNA methylation levels for significantly differentially CpG-methylated probes in BCVY and BCO after removing the methylation differences from normal tissue samples. The average methylation levels in 2219 CpG probes are expressed as β-values (0-1) and are graded from yellow (unmethylated) to blue (methylated). Probabilities are represented as the negative logarithm of the p-value graded from light to dark blue (less to more significant).

Distinctive Hypermethylation Profile BCVY

We re-analysed methylation differences between BCVY and BCO in the 18 478 BCVY-specific probes. This identified 502 CpG sites that were significantly differentially-methylated in BCVY vs. BCO in normal tissue. The diagram of the analysis procedure is included in Figure 6. Of those, 462 CpG sites were also significantly associated with age by GLM analysis (Annex VII). Only 16%

of distinctive BCVY CpG sites were hypomethylated and 84% were hypermethylated. While hierarchical clustering showed two principal sample groups: one consisting of BCVY samples and other including BCO and both normal tissues (NO and NVY), no molecular subtype clusters were found (Figure 23). Thus, although BCVY samples are globally hypomethylated, we identified a group of CpG sites with higher methylation, in contrast to the BCO, NVY, and NO samples. We had not an overrepresentation of any subtype among any cancer group (BC p-value = 0.72, BCVY p-value = 0.098 and BCO p-value = 0.472). GLM analysis showed that methylation differences that distinguish BCVY from BCO were specific of age and not related with ER status (p-value = 0.21) or molecular subtypes (p-values > 0.05) except for Luminal/Her2 which were associated with BCVY methylation differences and this result is in agreement with the major similarity in the methylation profile for BCO-Luminal/Her2 samples with BCVY patients observed in Figure 23. However, luminal/her2 sample size is quite limited and no significant conclusions can be drawn. Thus methylation differences observed indicate a BCVY specific methylation signature independent of molecular subtypes and ER status.

Genomic and functional context of significant CpG sites

In terms of CpG context, the global methylation differences between BCVY and BCO were significant for CpG islands and the adjacent regions (N/S Shore) and for non-CpG islands (open-sea). However, their specific profiles were different: while CpG islands and near regions (N/S shore) were hypomethylated, open-sea regions were hypermethylated in BCVY samples (Figure 24a). Functionally, the most significant differences were found in proximal promoters (CpG sites located within 200 bp, or 1 500 bp upstream of transcription start sites and in 5'UTR) enhancers regions and DNase hypersensitivity sites (p-value < 2.2×10^{-16}). Additionally, we found

methylation differences in gene body regions and Open Chromatin. All functional regions for the significant CpG sites analysed were hypomethylated in BCVY (Figure 24b).

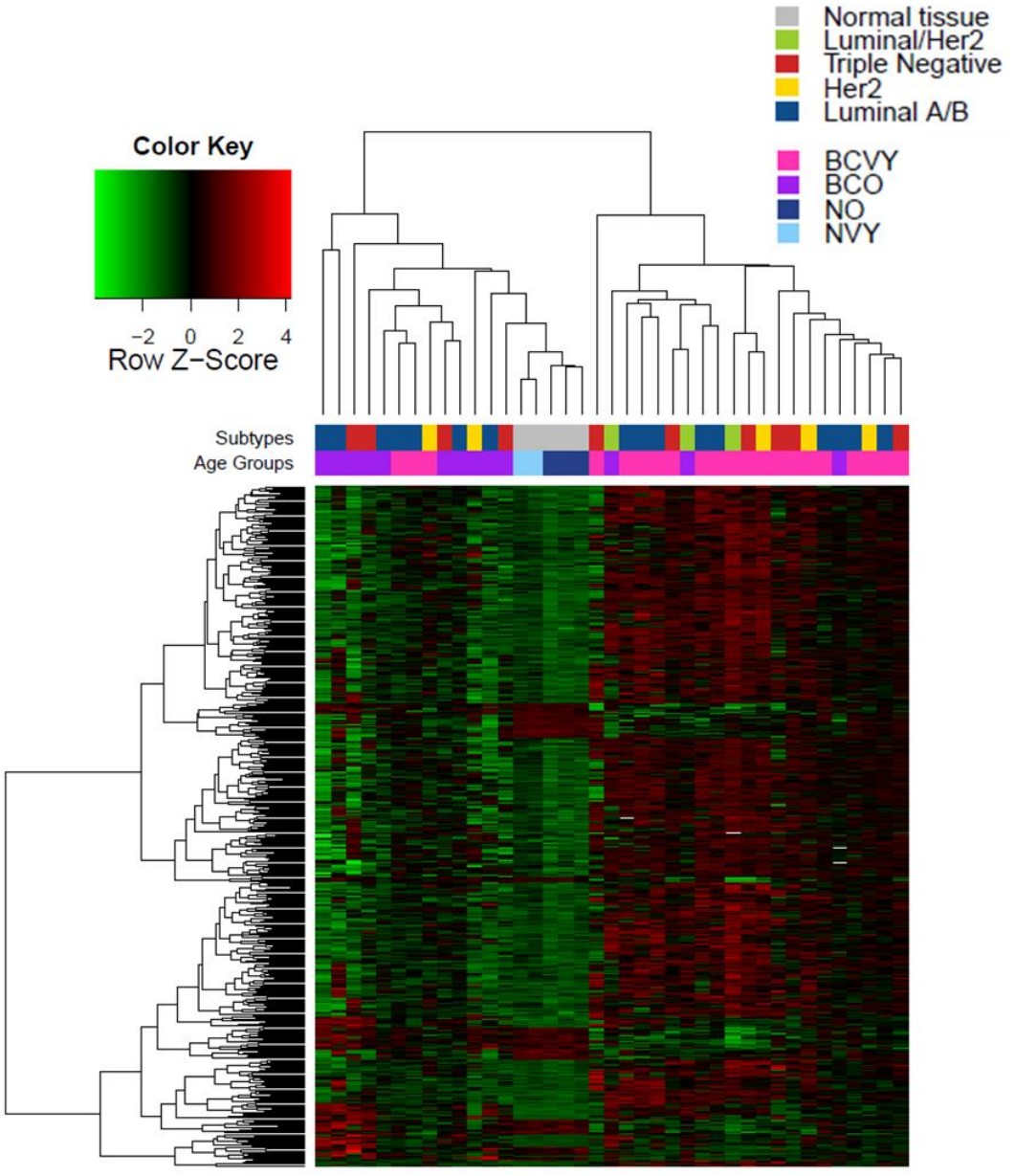


Figure 23. Heatmap representing a supervised cluster performed according to a hierarchical clustering methods centred on the median of the methylation levels at the 502 CpG sites distinctive in BCVY compared to BCO and normal tissue samples. Hypermethylated CpG probes (red) and hypomethylated probes (green). Samples studied are represented as BCVY (light pink), BCO samples (purple), normal tissue samples from women younger than 35 years (light blue), and normal samples from women older than 45 years (dark blue). Moreover, subtype bar indicates the molecular subtype for each sample, indicating luminal/Her2 (green), triple negative (red), her2 (yellow) and luminal A/B subtype (dark blue) and normal tissue is represented in grey colour. BCVY: breast cancer in very young women; BCO: breast cancer in old women; NO: normal old samples; NVY: normal young samples.

Regarding hypermethylation of BCVY was localized mainly in open-sea and regions distant from islands (Figure 24c). Significant CpG probes overlapped in all functional categories included and we were able to find significant methylation differences between BCVY and BCO for all functional categories with the exception of 3'UTR regions that were poorly-represented among significant CpG sites. Interestingly, we found methylation differences in important regulatory genome regions such as enhancers and TFBS (Figure 24d).

Methylation in Illumina Infinium EPIC BeadChip-specific probes

To measure the contribution of EPIC850K to metEPICVal study, we split significant probes into those that were included in both EPIC850K and HM450K arrays from those exclusive for EPIC850K, analysing them according to their functional and genomic distribution. From the 2 219 significant probes, 783 were new in EPIC850K array and 1 432 were common to both arrays. Interestingly, most of the new and significant EPIC850K probes were hypermethylated in BCVY (65.5%). However, the vast majority of significant hypomethylated probes had already been included in the HM450K. The major contribution of new EPIC850K probes was observed in open-sea and TFBS regions (68% and 57.3%, respectively). About gene context, 45% of significant probes targeting gene bodies were exclusively in EPIC850K array (Figure 24e and 24f).

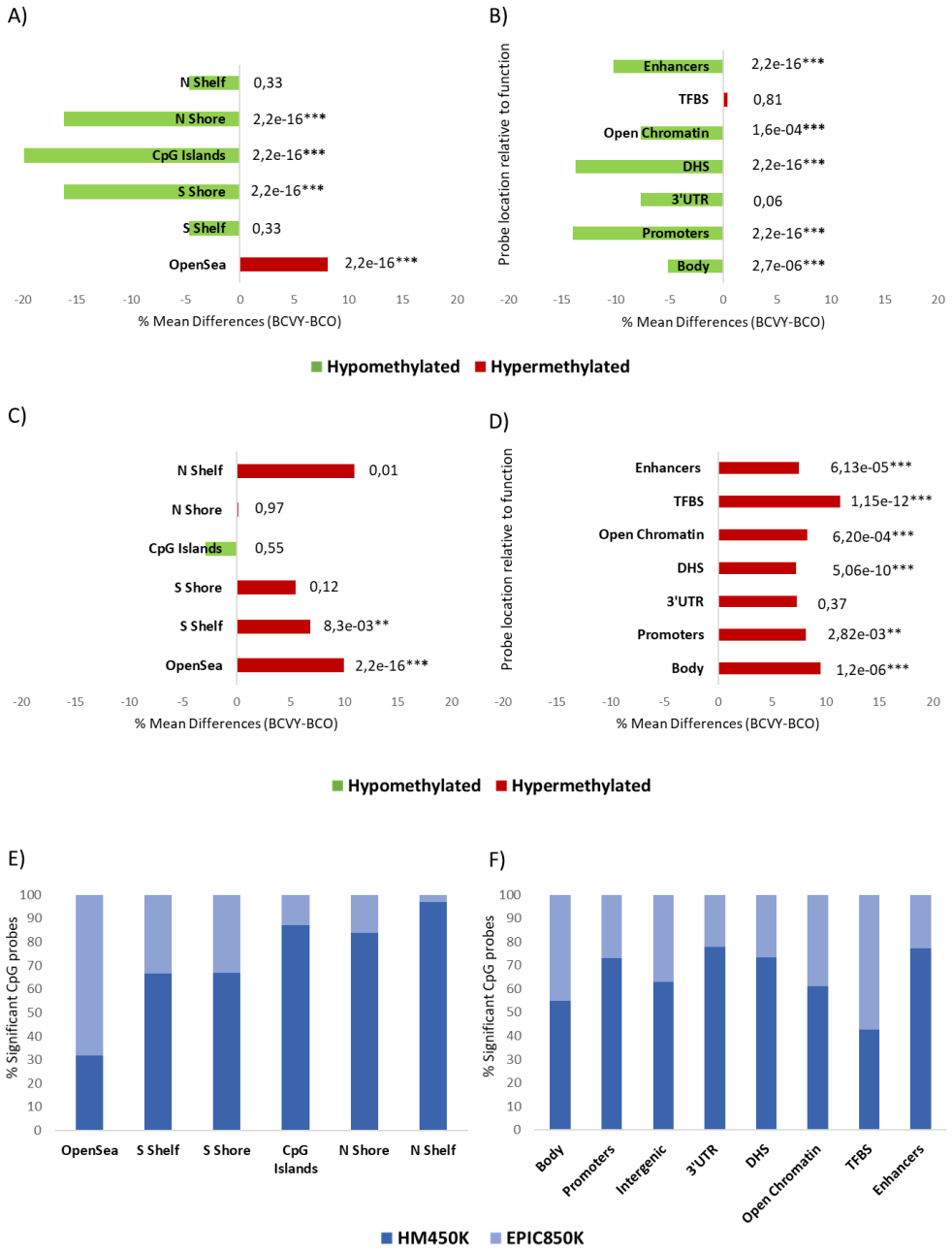


Figure 24. Genomic and functional context of significant CpG sites which are differentially methylated in BCVY and BCO from metEPICVal samples analysed by Infinium MethylationEPIC BeadChip, and comparison with HM450K array. Percentage of methylation differences for statistically significant CpG sites from BCVY-BCO comparison according to location of the CpG relative to the island (open-sea, island, N/S shore and N/S shelf) (**A**) and to the UCSC gene region feature category and regulatory elements (**B**); Percentage of methylation differences for distinctive CpG sites in BCVY samples according to the location of the CpG relative to the CpG island (**C**) and to the regulatory elements (**D**). Red bars represent hypermethylation in BCVY and green bars represents hypomethylation; Percentage of significant metEPICVal probes that are in methylation EPIC BeadChip array (EPIC850K) and probes that are common to the EPIC850K and HM450K array (HM450K) according to CpG location (**E**) and functional classification (**F**). Light blue colour represents percentage of significant probes which are in methylation EPIC array and dark blue colour are probes that are as well as in HM450K array. *P ≤ 0.05, **P ≤ 0.01, ***P ≤ 0.001 statistically significant. N/S: north/south; upstream or downstream to the CpG island. BCVY: breast cancer in very young women; BCO: breast cancer in old women; UTR: Untranslated region; TFBS: transcription factor binding site; DHS: DNasa I hypersensitive sites.

3.2 Validation of DNA methylation with combined and TCGA data

We next moved to validate our DNAm findings in BCVY and BCO metEPICVal samples in two 450K data sets (combined data and TCGA). The combined study includes methylation values for 69 samples and 845 samples were included in TCGA data. Considering significant metEPICVal probes present in HM450K (1 432 probes) for the global hypomethylation profile in BCVY, 657 were also significant in the combined study (45.8%) (Figure 25a). Similarly, we observed consistent result in the TCGA population (116 out of 1 254 CpG sites β-values available, 9.25%) (Figure 25b). In both validation cohorts, the vast majority of significant probes were hypomethylated in BCVY (99% in combined data and 95,7% in TCGA data).

Additionally, 35.06 % and 25.29 % of the total significant probes in the distinctive BCVY methylation profile were also included in HM450K array in combined and TCGA data, respectively. After performing the Wilcoxon rank sum test restricted to significant CpG sites from distinctive hypermethylation signature in BCVY included in HM450K validation sets we observed 15 significant CpG sites of 176 for the combined data and 10 significant CpG sites of 127 for TCGA.

Based on localization and function of the probes, all categories presented significant hypomethylation for BCVY (p-value < $2{,}2 \times 10^{-16}$). Nevertheless, validation studies were missing for most of the probes located in enhancer regions, TFBS, Open Chromatin and open-sea regions, exclusive to the EPIC850K array.

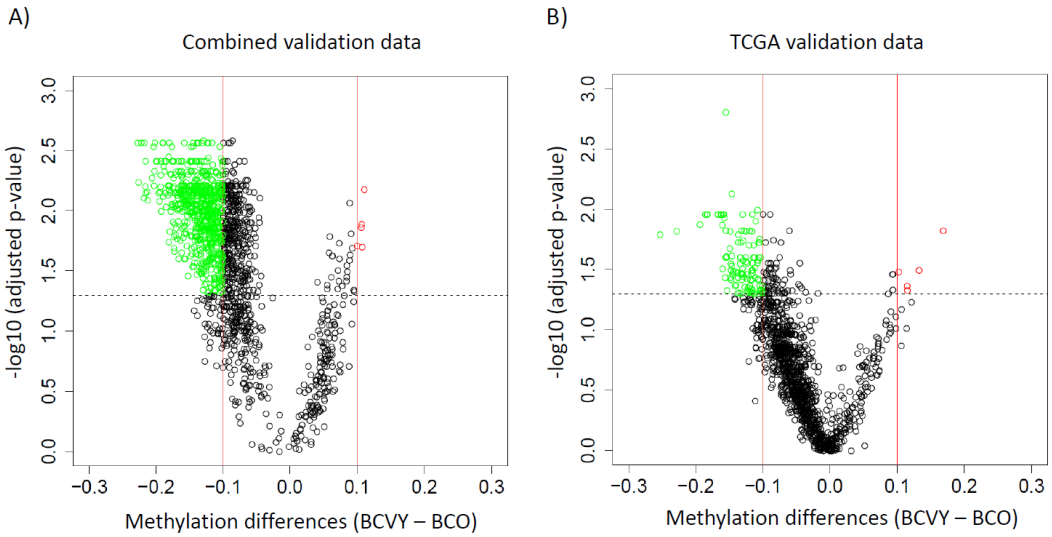


Figure 25. Global hypomethylation BCVY validation in combined and TCGA data sets. Significant probes from global hypomethylation profile in BCVY obtained from metEPICVal samples were selected and validated in combined data (**A**) and TCGA data (**B**). Green dots represent hypomethylation in BCVY and red dots hypermethylation. Dotted line indicates minus logarithm of adjusted p-value threshold applied and red lines delimit methylation differences between BCVY and BCO ±0,1. BCVY: breast cancer in very young women; BCO: breast cancer in old women.

3.3 Gene expression TCGA

We analysed gene expression of target genes for 2 219 significant CpG probes from the BCVY vs BCO comparison. We subsequently performed a meta-analysis to combine p-values for each gene in each permutation study, which identified 573 genes that were both significantly differentially expressed and regulated by CpG sites different methylated in BCVY and BCO women. Next, we filtered these significant genes according to their correlation between methylation and gene expression, retaining BCVY-sample overexpressed genes which were regulated by hypomethylated probes as well as repressed genes regulated by hypermethylated probes, to obtain a total of 186 genes with an expression–methylation correlation. Gene expression for genes regulating CpGs sites from BCVY distinctive profile were analysed following

the same procedure. We could identify a set of 44 genes differentially expressed in BCVY and regulated by regions differentially methylated.

Global hypomethylation was detected in different functional category regions, however, important regulatory elements such as promoter and enhancer sites were highly enriched in the hypomethylation pattern and gene expression for genes affected was evaluated using TCGA data (Figure 26a and 26b respectively). Methylation list with probes for promoters and enhancers are included in Annex VIII and Annex IX respectively. TFBS and enhancers regions were significantly represented in the hypermethylated distinctive profile found in BCVY and we analysed expression of genes regulated by them using TCGA data (Figure 26c and 26d respectively). Methylation list with probes for enhancers and TFBS are included in Annex X and Annex XI, respectively.

3.4 Pathway enrichment analysis for methylation data

Significant genes obtained from TCGA data that were regulating significant CpGs from global hypomethylation profile in BCVY were involved in neuronal system pathways related with transmission across synapsis, GABA receptor activation and different neurotransmitter cycles among others. These results were previously observed in the miRNA expression study. Additionally, 22% of pathways were related with extracellular matrix organization including sulfate/heparin metabolism, proteoglycans metabolism or fibroblasts activation.

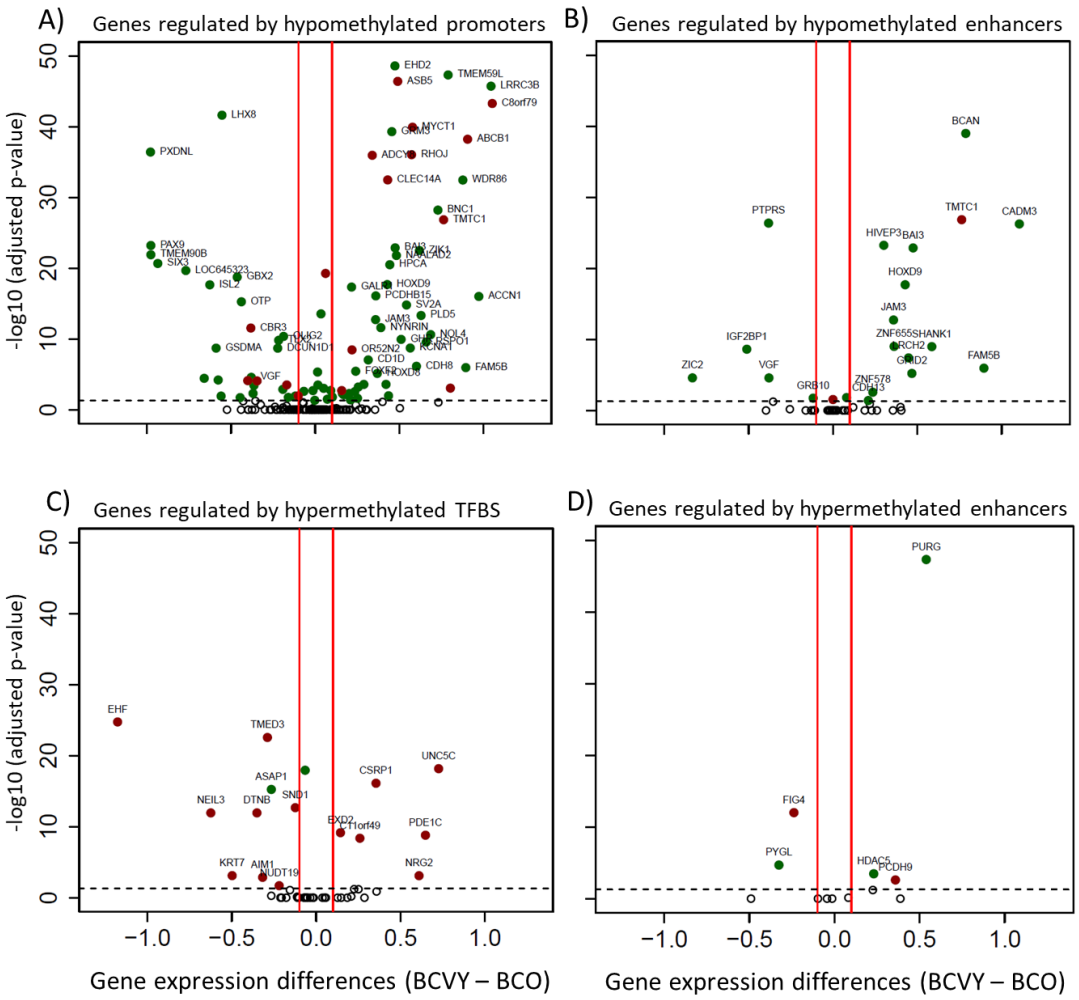


Figure 26. TCGA expression for genes regulated by significantly different methylated CpG sites. Genes regulated by differentially methylated CpG sites from global BCVY hypomethylation profile found in metEPICVal samples localized in promoters (**A**) and enhancers (**B**). Genes regulated by differentially methylated CpG sites from distinctive BCVY hypermethylation signature found in metEPICVal samples localized to TFBS (**C**) and enhancers (**D**); Green dots represent genes regulated by hypomethylated CpG regulatory sites and red dots those regulated by hypermethylated CpG regulatory sites. BCVY: breast cancer in very young women; BCO: breast cancer in old women; TFBS: Transcription factor binding sites.

Pathways related with cell adhesion represented 16% of the significant pathways, and some genes were involved in insulin secretion regulation (Figure 27a). Complete pathway enrichment is included in Table 6. The enrichment pathway analysis for significant genes that characterize exclusively BCVY showed interesting pathways deregulated (Table 7). One of the most noteworthy was related with Notch and Notch1 signalling (18%). Immune system pathways

represented 11% of total in which Major Histocompatibility Complex I and II were important. Pathways related with DNA repair and vesicular traffic and biogenesis were highly represented including lysosome, Golgi-associated, clathrin-derived or coat complex protein I vesicles (Figure 27b).

Significant genes which promoters and enhancers were differentially-methylated in the global hypomethylation profile for BCVY were involved in cell adhesion related pathways. However, we found an enrichment of insulin metabolism pathways for genes regulated by enhancers differentially methylated in the distinctive hypermethylation profile for BCVY and repair pathways were characteristics for TFBS distinctively hypermethylated in BCVY. According to previous results, there were an overrepresentation in all regulatory categories for both signatures (hypo and hypermethylation) of genes involved in neuronal pathways such as glutamatergic synapse, axon guidance and related pathways, although they were not significant.

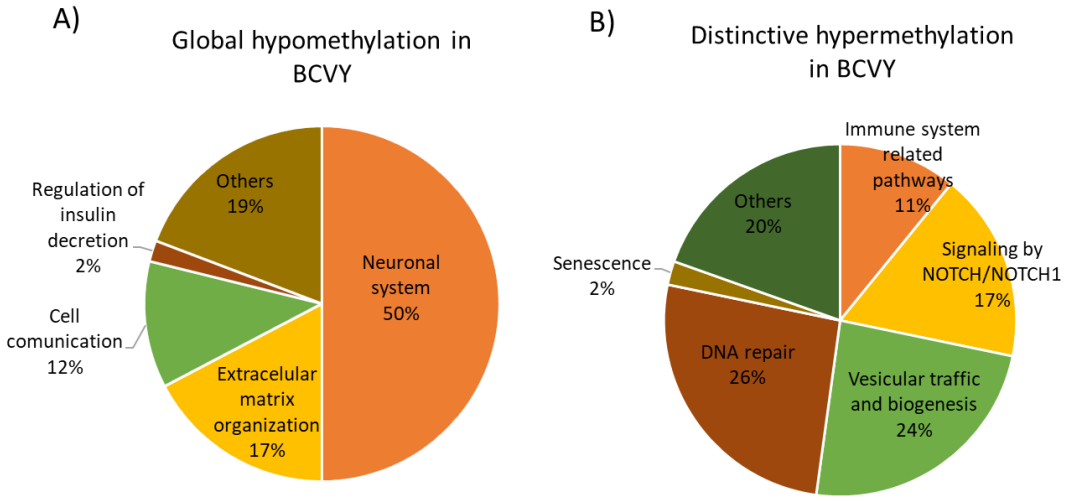


Figure 27. Pathway enrichment study. Percentage of significant enriched signalling pathways associated to genes differentially expressed and regulated by global hypomethylated CpG sites in BCVY (**A**) and regulated by distinctive CpG probes generally hypermethylated in BCVY (**B**). BCVY: breast cancer in very young women.

Table 6. **Pathway enrichment results for genes regulated by the CpG probes globally hypomethylated in BCVY.**

Reactome Term	p-value	Category	Genes
Neuronal System_Homo sapiens_R-HSA-112316	1.53E-07	Nervious System	*GABRB2;SNAP25;GABBR; KCNC2;CACNA2D1; GABRA4;KCNA1;KCNH8; KCNA6;ADCY8;CPLX1; PTPRD;LRFN5;GNB5; GRIA4*
Transmission across Chemical Synapses_Homo sapiens_R-HSA-112315	1.65E-04	Nervious System	*GABRB2;SNAP25;GABBR1 ;CACNA2D1;GABRA4; GNB5;ADCY8;CPLX1; GRIA4*
Voltage gated Potassium channels_Homo sapiens_R-HSA-1296072	6.73E-04	Nervious System	*KCNC2;KCNH8;KCNA1; KCNA6*
G alpha (i) signalling events_Homo sapiens_R-HSA-418594	1.87E-03	Nervious System	*RGS4;GRM3;GABBR1; GRM7;GALR1;RGS12; GNB5;ADCY8*
Neurotransmitter Receptor Binding And Downstream Transmission In The Postsynaptic Cell_Homo sapiens_R-HSA-112314	2.15E-03	Nervious System	*GABRB2;GABBR1; GABRA4;GNB5;ADCY8; GRIA4*
SALM protein interactions at the synapse_Homo sapiens_R-HSA-8849932	1.07E-03	Nervious System	*PTPRD;LRFN5;GRIA4*
GABA receptor activation_Homo sapiens_R-HSA-977443	1.71E-03	Nervious System	*GABRB2;GABBR1; GABRA4;ADCY8*
Cell-cell junction organization_Homo sapiens_R-HSA-421270	2.50E-03	Cell adhesion	*CADM3;PVRL3;F11R; CDH8*
Uptake and actions of bacterial toxins_Homo sapiens_R-HSA-5339562	1.96E-03	Bacterial toxins	*SNAP25;SV2A;TXNRD1*
Potassium Channels_Homo sapiens_R-HSA-1296071	2.33E-03	Nervious System	*GABBR1;KCNC2;KCNH8; KCNA1;KCNA6*
Nectin/Necl trans heterodimerization_Homo sapiens_R-HSA-420597	1.75E-03	Cell adhesion	*CADM3;PVRL3*
Adherens junctions interactions_Homo sapiens_R-HSA-418990	2.94E-03	Cell adhesion	*CADM3;PVRL3;CDH8*
Neurotoxicity of clostridium toxins_Homo sapiens_R-HSA-168799	3.69E-03	Nervious System	*SNAP25;SV2A*
Glucagon-like Peptide-1 (GLP1) regulates insulin secretion_Homo sapiens_R-HSA-381676	6.96E-03	Regulation of insulin secretion	*KCNC2;GNB5;ADCY8*

Regulation of insulin secretion_Homo sapiens_R-HSA-422356	6.61E-03	Regulation of insulin secretion	*SNAP25;KCNC2;GNB5; ADCY8*
Class C/3 (Metabotropic glutamate/pheromone receptors)_Homo sapiens_R-HSA-420499	6.07E-03	G-protein coupled receptor (GPCRs)	*GRM3;GABBR1;GRM7*
GABA A receptor activation_Homo sapiens_R-HSA-977441	6.27E-03	Nervious System	*GABRB2;GABRA4*
GPCR ligand binding_Homo sapiens_R-HSA-500792	9.23E-03	G-protein coupled receptor (GPCRs)	*GRM3;GABBR1;EDNRB;GRM7;GALR1;AGTR1;GNB; TAC1;QRFPR;DRD5*
G alpha (z) signalling events_Homo sapiens_R-HSA-418597	8.96E-03	Nervious System	*RGS4;GNB5;ADCY8*
Cell junction organization_Homo sapiens_R-HSA-446728	8.51E-03	Cell adhesion	*CADM3;PVRL3;F11R; CDH8*
Phase 0 - rapid depolarisation_Homo sapiens_R-HSA-5576892	7.92E-03	Nervious System	*CACNG7;CACNA2D1; FGF12*
GABA synthesis. release. reuptake and degradation_Homo sapiens_R-HSA-888590	1.33E-02	Nervious System	*SNAP25;CPLX1*
Serotonin Neurotransmitter Release Cycle_Homo sapiens_R-HSA-181429	1.19E-02	Nervious System	*SNAP25;CPLX1*
Defective B3GALT6 causes EDSP2 and SEMDJL1_Homo sapiens_R-HSA-4420332	1.33E-02	Extracelular Matrix Organization	*BCAN;GPC5*
Norepinephrine Neurotransmitter Release Cycle_Homo sapiens_R-HSA-181430	1.19E-02	Nervious System	*SNAP25;CPLX1*
Acetylcholine Neurotransmitter Release Cycle_Homo sapiens_R-HSA-264642	1.07E-02	Nervious System	*SNAP25;CPLX1*
Defective B3GAT3 causes JDSSDHD_Homo sapiens_R-HSA-3560801	1.33E-02	Extracelular Matrix Organization	*BCAN;GPC5*
Defective B4GALT7 causes EDS. progeroid type_Homo sapiens_R-HSA-3560783	1.33E-02	Extracelular Matrix Organization	*BCAN;GPC5*
Heparan sulfate/heparin (HS-GAG) metabolism_Homo sapiens_R-HSA-1638091	1.39E-02	Extracelular Matrix Organization	*BCAN;GPC5;HS3ST4*
ECM proteoglycans_Homo sapiens_R-HSA-3000178	1.46E-02	Extracelular Matrix Organization	*BCAN;LAMA1;HAPLN1*
Integration of energy metabolism_Homo sapiens_R-HSA-163685	1.95E-02	Metabolism	*SNAP25;KCNC2;GNB5; ADCY8*

Dopamine Neurotransmitter Release Cycle_Homo sapiens_R-HSA-212676	1.91E-02	Nervious System	*SNAP25;CPLX1*
Glutamate Neurotransmitter Release Cycle_Homo sapiens_R-HSA-210500	2.08E-02	Nervious System	*SNAP25;CPLX1*
Ligand-gated ion channel transport_Homo sapiens_R-HSA-975298	2.42E-02	Nervious System	*GABRB2;GABRA4*
Diseases associated with glycosaminoglycan metabolism_Homo sapiens_R-HSA-3560782	2.42E-02	Extracelular Matrix Organization	*BCAN;GPC5*
Glycosaminoglycan metabolism_Homo sapiens_R-HSA-1630316	2.65E-02	Extracelular Matrix Organization	*BCAN;GPC5;HS3ST4; PAPSS2*
Phase 1 - inactivation of fast Na+ channels_Homo sapiens_R-HSA-5576894	2.59E-02	Nervious System	*CACNG7;CACNA2D1*
G alpha (q) signalling events_Homo sapiens_R-HSA-416476	3.32E-02	Nervious System	*EDNRB;AGTR1;GNB5; TAC1;QRFPR*
Cell-Cell communication_Homo sapiens_R-HSA-1500931	3.41E-02	Cell adhesion	*CADM3;PVRL3;F11R;CDH 8*
Peptide ligand-binding receptors_Homo sapiens_R-HSA-375276	3.45E-02	Cell adhesion	*EDNRB;GALR1;AGTR1; TAC1;QRFPR*
Amino acid transport across the plasma membrane_Homo sapiens_R-HSA-352230	3.35E-02	Nervious System	*SLC6A15;SLC7A3*
HS-GAG biosynthesis_Homo sapiens_R-HSA-2022928	3.35E-02	Extracelular Matrix Organization	*GPC5;HS3ST4*
Extracellular matrix organization_Homo sapiens_R-HSA-1474244	4.93E-02	Extracelular Matrix Organization	*BCAN;LAMA1;ADAMTS1; F11R;HAPLN1;JAM3*
Diseases of glycosylation_Homo sapiens_R-HSA-3781865	4.89E-02	Extracelular Matrix Organization	*BCAN;ADAMTS18;GPC5*
Tachykinin receptors bind tachykinins_Homo sapiens_R-HSA-380095	4.56E-02	Cell adhesion	*TAC1*

Table 7. **Pathway enrichment results for genes regulated by CpG probes distinctively hypermethylated in BCVY.**

Reactome Term	p-value	Category	Genes
Nef mediated downregulation of MHC class I complex cell surface expression_Homo sapiens_R-HSA-164940	2.10E-04	Immune system related pathways	*AP1B1;AP1S3*
Nef-mediates down modulation of cell surface receptors by recruiting them to clathrin adapters_Homo sapiens_R-HSA-164938	9.67E-04	Immune system related pathways	*AP1B1;AP1S3*
The role of Nef in HIV-1 replication and disease pathogenesis_Homo sapiens_R-HSA-164952	1.72E-03	Immune system related pathways	*AP1B1;AP1S3*
Lysosome Vesicle Biogenesis_Homo sapiens_R-HSA-432720	2.69E-03	Vesicular traffic and biogenesis	*AP1B1;AP1S3*
NOTCH1 Intracellular Domain Regulates Transcription_Homo sapiens_R-HSA-2122947	4.80E-03	Signaling by NOTCH/NOTCH1	*HDAC5;HDAC9*
Golgi Associated Vesicle Biogenesis_Homo sapiens_R-HSA-432722	6.29E-03	Vesicular traffic and biogenesis	*AP1B1;AP1S3*
Signaling by NOTCH1 HD+PEST Domain Mutants in Cancer_Homo sapiens_R-HSA-2894858	6.99E-03	Signaling by NOTCH/NOTCH1	*HDAC5;HDAC9*
Constitutive Signaling by NOTCH1 HD+PEST Domain Mutants_Homo sapiens_R-HSA-2894862	6.99E-03	Signaling by NOTCH/NOTCH1	*HDAC5;HDAC9*
Signaling by NOTCH1 in Cancer_Homo sapiens_R-HSA-2644603	6.99E-03	Signaling by NOTCH/NOTCH1	*HDAC5;HDAC9*
Constitutive Signaling by NOTCH1 PEST Domain Mutants_Homo sapiens_R-HSA-2644606	6.99E-03	Signaling by NOTCH/NOTCH1	*HDAC5;HDAC9*
Signaling by NOTCH1 PEST Domain Mutants in Cancer_Homo sapiens_R-HSA-2644602	6.99E-03	Signaling by NOTCH/NOTCH1	*HDAC5;HDAC9*
Clathrin derived vesicle budding_Homo sapiens_R-HSA-421837	1.04E-02	Vesicular traffic and biogenesis	*AP1B1;AP1S3*
trans-Golgi Network Vesicle Budding_Homo sapiens_R-HSA-199992	1.04E-02	Vesicular traffic and biogenesis	*AP1B1;AP1S3*
Vitamin B2 (riboflavin) metabolism_Homo sapiens_R-HSA-196843	1.10E-02	Vitamine B2 metabolism	*ENPP1*
Signaling by NOTCH1_Homo sapiens_R-HSA-1980143	1.10E-02	Signaling by NOTCH/NOTCH1	*HDAC5;HDAC9*
Transport and synthesis of PAPS_Homo sapiens_R-HSA-174362	1.31E-02	PAPS synthesis	*PAPSS2*

GDP-fucose biosynthesis_Homo sapiens_R-HSA-6787639	1.31E-02	Sugar biosyntesis	*GMDS*
Membrane Trafficking_Homo sapiens_R-HSA-199991	1.34E-02	Vesicular traffic and biogenesis	*KIF26B;AP1B1; TMED3;AP1S3*
COPI-dependent Golgi-to-ER retrograde traffic_Homo sapiens_R-HSA-6811434	1.37E-02	Vesicular traffic and biogenesis	*KIF26B;TMED3*
MHC class II antigen presentation_Homo sapiens_R-HSA-2132295	2.16E-02	Immune system related pathways	*AP1B1;AP1S3*
Synthesis of PIPs at the late endosome membrane_Homo sapiens_R-HSA-1660517	2.18E-02	PI metabolism and related pathways	*FIG4*
Removal of the Flap Intermediate from the C-strand_Homo sapiens_R-HSA-174437	2.18E-02	DNA repair	*RPA3*
Vesicle-mediated transport_Homo sapiens_R-HSA-5653656	2.26E-02	Vesicular traffic and biogenesis	*KIF26B;AP1B1; TMED3;AP1S3*
Processive synthesis on the C-strand of the telomere_Homo sapiens_R-HSA-174414	2.39E-02	DNA repair	*RPA3*
Golgi-to-ER retrograde transport_Homo sapiens_R-HSA-8856688	2.44E-02	Vesicular traffic and biogenesis	*KIF26B;TMED3*
Signaling by NOTCH_Homo sapiens_R-HSA-157118	2.96E-02	Signaling by NOTCH/NOTCH1	*HDAC5;HDAC9*
Mismatch repair (MMR) directed by MSH2:MSH6 (MutSalpha)_Homo sapiens_R-HSA-5358565	3.04E-02	DNA repair	*RPA3*
Mismatch repair (MMR) directed by MSH2:MSH3 (MutSbeta)_Homo sapiens_R-HSA-5358606	3.04E-02	DNA repair	*RPA3*
Removal of the Flap Intermediate_Homo sapiens_R-HSA-69166	3.04E-02	DNA repair	*RPA3*
Synthesis of PIPs at the early endosome membrane_Homo sapiens_R-HSA-1660516	3.04E-02	PI metabolism and related pathways	*FIG4*
Mismatch Repair_Homo sapiens_R-HSA-5358508	3.25E-02	DNA repair	*RPA3*
Processive synthesis on the lagging strand_Homo sapiens_R-HSA-69183	3.25E-02	DNA repair	*RPA3*
Host Interactions of HIV factors_Homo sapiens_R-HSA-162909	3.27E-02	Immune system related pathways	*AP1B1;AP1S3*
Translesion synthesis by REV1_Homo sapiens_R-HSA-110312	3.46E-02	DNA repair	*RPA3*
Translesion synthesis by POLI_Homo sapiens_R-HSA-5656121	3.68E-02	DNA repair	*RPA3*
Translesion synthesis by POLK_Homo sapiens_R-HSA-5655862	3.68E-02	DNA repair	*RPA3*

ABC transporters in lipid homeostasis_Homo sapiens_R-HSA-1369062	3.68E-02	Lipid transport	*PEX3*
Formation of Senescence-Associated Heterochromatin Foci (SAHF)_Homo sapiens_R-HSA-2559584	3.68E-02	Senescence	*UBN1*
Synthesis of PIPs at the Golgi membrane_Homo sapiens_R-HSA-1660514	3.89E-02	PI metabolism and related pathways	*FIG4*
Translesion Synthesis by POLH_Homo sapiens_R-HSA-110320	4.10E-02	DNA repair	*RPA3*
PCNA-Dependent Long Patch Base Excision Repair_Homo sapiens_R-HSA-5651801	4.10E-02	DNA repair	*RPA3*
Lagging Strand Synthesis_Homo sapiens_R-HSA-69186	4.31E-02	DNA repair	*RPA3*
Telomere C-strand (Lagging Strand) Synthesis_Homo sapiens_R-HSA-174417	4.73E-02	DNA repair	*RPA3*
Resolution of AP sites via the multiple-nucleotide patch replacement pathway_Homo sapiens_R-HSA-110373	4.73E-02	DNA repair	*RPA3*
Gap-filling DNA repair synthesis and ligation in GG-NER_Homo sapiens_R-HSA-5696397	4.94E-02	DNA repair	*RPA3*

3.5 Validation step by qRT-PCR of genes deregulated that were regulated by CpG probes differentially methylated in BCVY.

In order to validate gene expression obtained for the global hypomethylation, we selected 3 hypomethylated genes (*FOXI2, HOXD9* and *PCDH10*). Among them, *HOXD9* (p-value = $6x10^{-3}$) and *PCDH10* (p-value = $4.7x10^{-2}$) were significantly upregulated in BCVY previously observed in TCGA gene expression data (Figure 28a and 28b). Gene expression validation was performed using qRT-PCR according to the material and methods description.

In addition, for the distinctive hypermethylation profile in BCVY, we selected *APIS3, APIB1, TIGIT* and *CXCL17* for their involvement in pathways related with immune system, *HDAC5* and *HDAC9* because they were implicated in Notch/Notch1 signalling pathways, *NEIL3* for its involvement in

Table 8. Results of qRT-PCR validation analysis.

Gene	BCVY mean	BCO mean	Mean differences	p-value
NEIL3	1.91	0.58	1.33	$1x10^{-4}$*****
PAPSS2	12.37	3.43	8.94	0.001*****
HOXD9	6.47	1.27	5.20	0.005*****
HDAC5	19.4	4.46	14.94	0.010*****
HDAC9	7.76	3.08	4.68	0.037*****
AP1B1	5.64	2.89	2.75	0.037*****
PCDH10	10.82	2.14	8.68	0.047*****
TIGIT	20.08	4.61	15.47	0.081
EHF	0.96	1.76	-0.80	0.248
CXCL17	0.88	0.99	-0.11	0.502
AP1S3	1.93	3.12	-1.19	0.74
OTUD3	1.27	1.2	0.07	0.859
PEX3	1.79	1.47	0.32	0.896

* Significant p-values < 0.05
BCVY: breast cancer in very young women; BCO: breast cancer in old women.

DNA repair, and *EHF, PAPSS2, PEX3* and *OTUD3* because of their association with carcinogenesis. Six of them were confirmed to have a differential expression across groups: *APIB1* (p-value = $3.7x10^{-2}$), *TIGIT* (p-value = $8.1x10^{-2}$), *HDAC5* (p-value = $1x10^{-2}$), *HDAC9* (p-value = $3.7x10^{-2}$), *NEIL3* (p-value < $1x10^{-4}$) and *PAPSS2* (p-value = $1x10^{-3}$). Despite significant differences in gene expression, only *HDAC5* expression was consistently upregulated in both the qRT-PCR validation and TCGA data (Figure 28c). As expected, *EHF*, *CXCL17* and *AP1S3* were downregulated in BCVY, however the difference did not reach statistical significance. The mean expression values for each gene obtained in the validation step are included in Table 8. The putative functional implications of the genes whose expression was consistent with our TCGA study are summarised in Table 9.

Table 9. Published information about validated genes and implication in cancer. Table includes statistically significant genes that were validated by qRT-PCR which expression was correlated with gene expression from TCGA analysis. Moreover, we included non-significant genes whose expression were consistent with TCGA study results.

Gene	p-value	Function	Published papers
HDAC5	0.01**	Histone deacetylases (HDACs) constitute a family of enzymes that play important roles in the epigenetic regulation of gene expression and contribute to the growth, differentiation and apoptosis of cancer cells. *HDAC5* was extensively expressed in human BC tissues, and high *HDAC5* expression was associated with an inferior prognosis. *HDAC5* enhances Notch 1 expression at both the mRNA and the protein level in glioma cell lines. *HDAC5* promotes glioma cell proliferation, suggesting that this effect involves the upregulation of Notch 1.	(207, 208)
HOXD9	0.005**	Hox gene expression is regulated by epigenetic control. CpG islands in Hox promotor genes are commonly methylated. Gene expression alteration could be important in oncogenesis and tumour suppression. *HOXD9* activation is a negative prognostic marker in glioblastoma, hepatocellular carcinoma, and cervical cancer.	(209)
PCDH10	0.048*	Protocadherin 10 is a tumour suppressor gene, and is frequently inactivated epigenetically in multiple carcinomas: pancreatic cancer, gastric cancer, prostate cancer, breast cancer, and other carcinomas.	(210)
EHF	0.248	*EHF* overexpression has been related with activation of HER family and poor survival in gastric cancer. The *EHF* gene has been characterised as a breast, prostate, and lung tumour suppressor gene. Loss of *ESE3/EHF* induces epithelial mesenchymal transition, stem cell-like features, and tumour-initiating and metastatic properties in prostate epithelial cells.	(211, 212)
CXCL17	0.502	Chemokine ligand 17 levels are upregulated in the early intraepithelial stages of human pancreatic carcinogenesis and are involve in increasing anti-tumour immune response. For hepatocellular carcinoma *CXCL17* is mainly produced by tumour-infiltrating neutrophils and can be used as an independent indicator of poor prognosis.	(213)
AP1S3	0.740	Adaptor protein complex 1 (*AP-1*) is an evolutionary conserved heterotetramer that promotes vesicular trafficking between the trans-Golgi network and endosomes. *AP1S3* silencing disrupts the endosomal translocation of the innate pattern-recognition receptor *TLR-3* (Toll-like receptor 3) and results in marked inhibition of downstream signalling.	(214)

*Refers to significant p-values analyzed by Wilcoxon rank sum test: *P ≤ 0.05, **P ≤ 0.01, ***P ≤ 0.001.

3.6 Accelerated epigenetic aging in BCVY

We estimated the DNAmAge for metEPICVal samples (Annex XII). Mean DNAmAge for BCVY was 51.44 years and for BCO was 64.98 years. We observed in metEPICVal sample set a significant correlation between chronological age and DNAmAge for BCO tumours (r = 0.52, p-value = 0.039). Although, no significant age correlation was found for BCVY (r = 0.14, p-value = 0.52), results shown higher age-acceleration significantly different to BCO (p-value = 5.5×10^{-4}) (Figure 29a-b-c). We have reproduced the same analyses to TCGA methylation data detecting also a significant age-acceleration for BCVY (p-value = 6.5×10^{-3}) (Figure 29d-e-f). We also examined DNAmAge vs. chronological age in normal samples from TCGA data, identifying an extremely good correlation decreasing in women with advanced age (r = 0.88, p-value = 8.4×10^{-33}) (Figure 29g-h). Despite homogeneously low age-accelerated values for healthy tissues compared with a more dispersed data in cancer samples, non-significance was reached (Figure 29i).

Strikingly, we observed an increased age-acceleration in BCVY-ER+ samples for metEPICVal set (p-value = 0.001), with a similar trend in the BCVY-TCGA cohort, the latter reached statistical significance. Nevertheless, we found an enrichment of ER+ tumours for BC samples with increased DNAmAge acceleration and we corroborated this association with TCGA (p-value < 2.2×10^{-16}) (Figure 29j-k-l). Age-acceleration was not associated with relapse or clinical subtypes. Nonetheless, there was an enrichment of luminal/her2 subtype in BCVY samples with increased age-acceleration values for metEPICVal samples that could not be reproduced in TCGA data. Similarly, all BCVY samples with relapse presented increased age-acceleration compared with BCO samples that suffered a relapse (results not shown). However, there is a limitation in the BCVY sample size in TCGA data and no conclusions can be drawn.

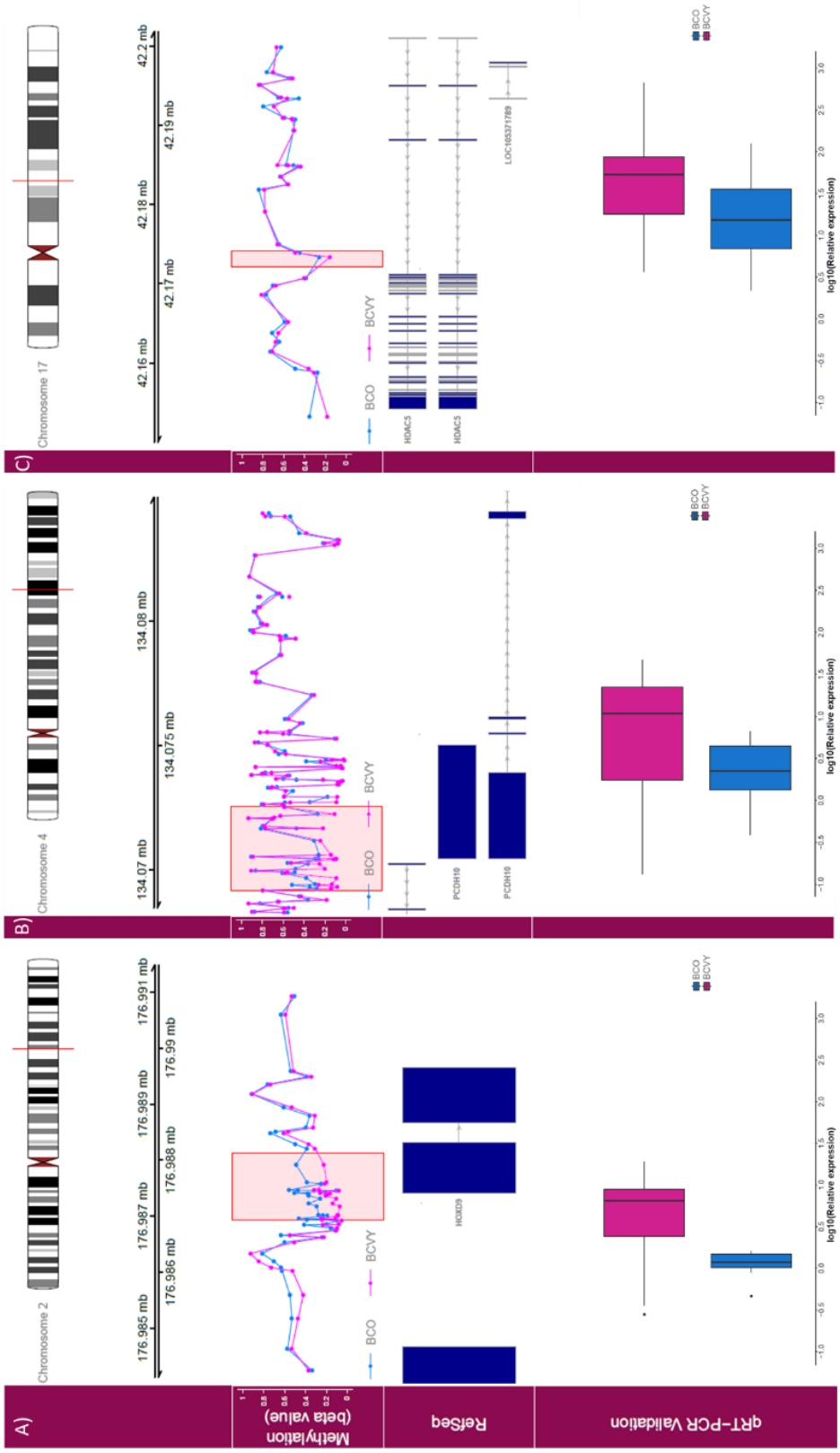


Figure 28. Representation of methylation in the regulatory regions for significantly differentially -expressed and validated genes. First track represents the chromosome position of the different methylated region. Methylation track shows the β-values for BCVY (pink) and BCO (blue) methylation. RefSeq track represents genome position of the significant CpG regions obtained. Validation results from qRT-PCR plotted in the qRT-PCR Validation track show gene expression for BCVY (pink) and BCO (blue). (**A**) HOXD9 methylation profile and validation analysis; (**B**) PCDH10 methylation profile and validation analysis; (**C**) HDAC5 methylation profile and validation analysis. Measures of the expression were quantified using a qRT-PCR technique and calculated using the ΔΔCt method, X axis represents logarithmic transformation of the relative expression. Boxplots represent the sample distribution with the mean for BCO and BCVY patients. All genes were significant (*P ≤ 0.05) according to the Wilcoxon rank sum test. BCVY: breast cancer in very young women; BCO: breast cancer in old women.

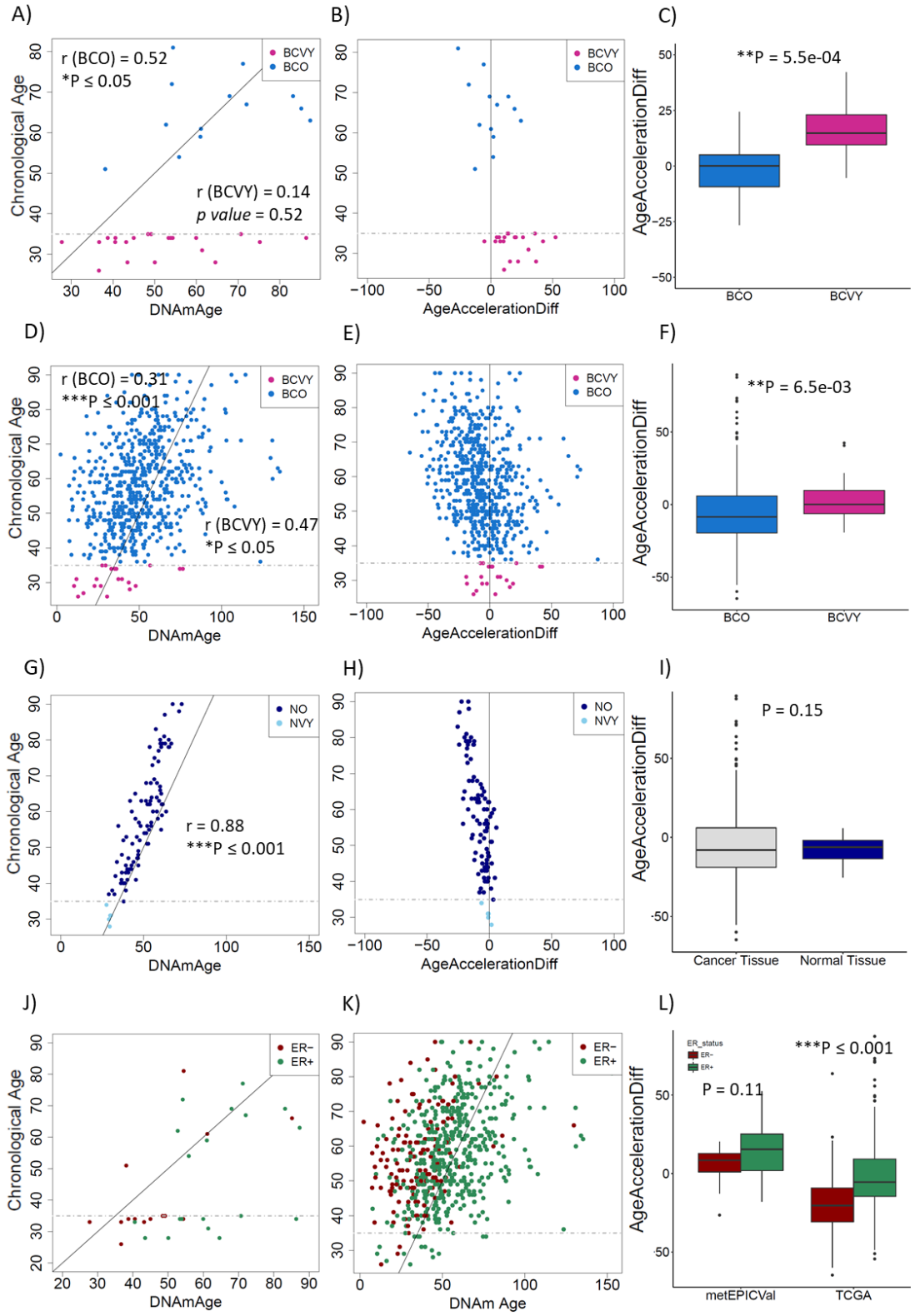


Figure 29. DNA methylation age studies. Results for metEPICVal samples according to: Chronological age versus DNAmAge by age groups (**A**); Age-acceleration differences versus chronological age (**B**); Boxplot for age-acceleration differences between BCO and BCVY (**C**). Analogous results for TCGA samples according to: Chronological age versus DNAmAge by age groups (**D**); Age-acceleration differences versus

chronological age (**E**); Boxplot for age-acceleration difference values between BCO and BCVY (**F**). Normal breast tissue results: Chronological age versus DNAmAge by age groups (**G**); Age-acceleration differences versus chronological age (**H**); Comparison of age-acceleration difference values between cancer and normal tissues (**I**). Representation of chronological age versus DNAmAge by oestrogen receptor positive and negative for metEPICVal (**J**) and TCGA samples (**K**); Age-acceleration differences by oestrogen receptor status for metEPICVal and TCGA data sets (**L**). Correlation values (r) for chronological age and DNAmAge by age group are included in plot representation. *P ≤ 0.05, **P ≤ 0.01, ***P ≤ 0.001. BCVY: breast cancer in very young women; BCO: breast cancer in old women; NO: normal old samples; NVY: normal young samples.

3.7 Copy Number Aberrations analysis

CNV analysis suggested high heterogeneity among samples. Despite that, we have been able to identified two clear groups based on CNVs, one very stable with few CNV which we classified into two subgroups: low-CNV, presenting a CNV profile similar to healthy tissue samples, and medium-CNV, with very few CNVs. The second group high-CNV; presented a more varied landscape with numerous alterations and multiple gains and losses in several regions across all chromosomes. We further divided the high-CNV group into three clusters with characteristic CNV profiles: CNV-chr8, CNV-chr1 and complex-CNV (Figure 30a). The first subgroup exhibited chromosome 8q amplification and the second 1q amplification; no specific CNV regions were detected for the third subgroup which showed high-genomic instability.

We observed no significant correlations between CNV and BC patient age, although we did find potential differences in the CNV-chr1 subgroup, which presented a chromosome 3p deletion and 3q gain, and a less evident amplification at chr16p and chr16q deletion, in BCVY but not BCO samples. In the CNV-chr8 subgroup, there was an evident chr8q amplification and chr16p gain and chr6q loss in BCO but not in BCVY samples (Figure 30b). However, larger datasets will be required to further analyse these trends.

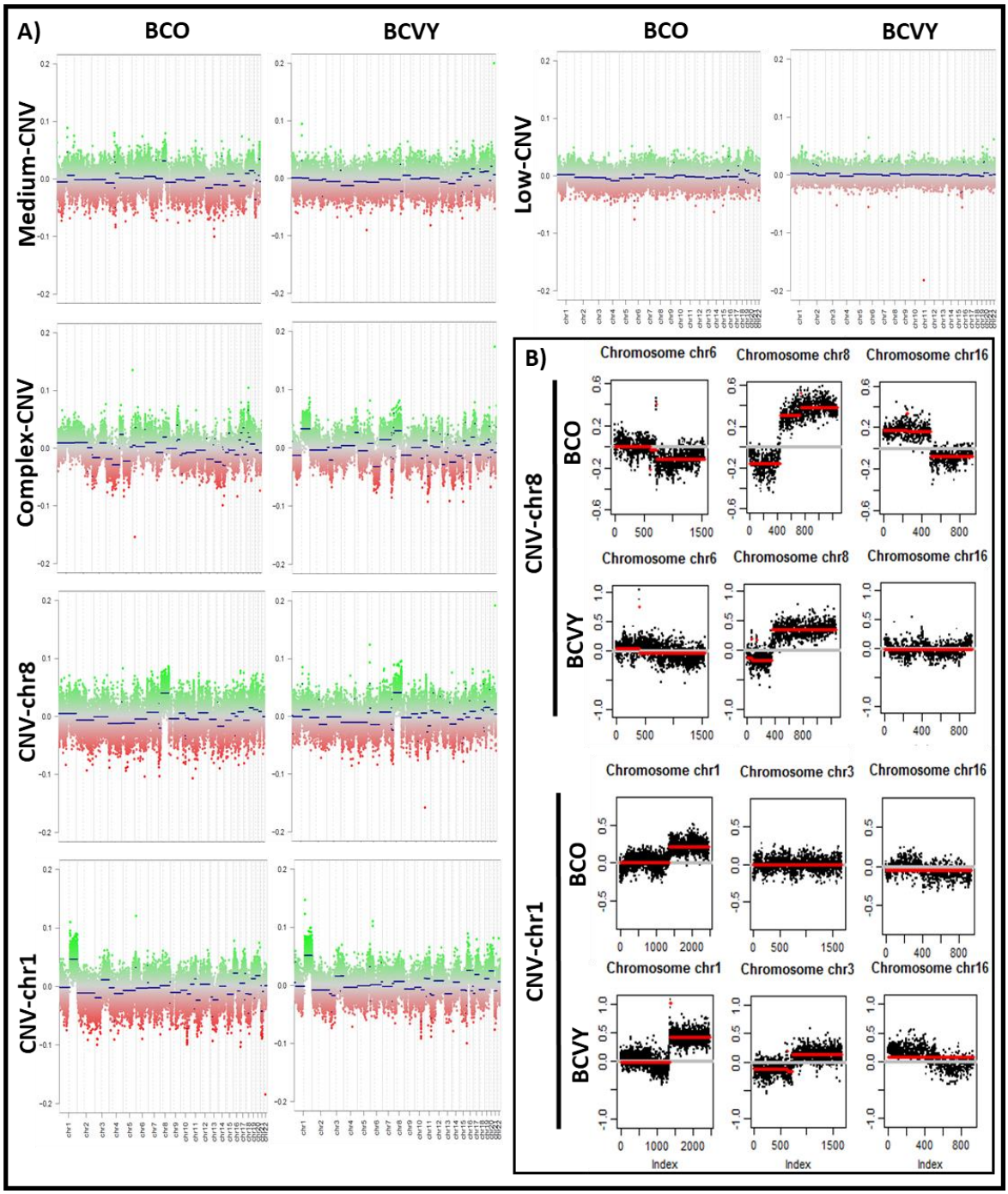


Figure 30. Copy number Variation profiles by subgroups for BCVY and BCO. Log_2 ratio of probe intensities are represented in y axis, x axis represents chromosomes segments. (**A**) Copy number subgroup profiles; (**B**) Zoomed image for CNV-chr8 and CNV-chr1 subgroups. BCVY: breast cancer in very young women; BCO: breast cancer in old women.

4. Functional studies with breast cancer cell lines and LMK-235 inhibition

Epigenetic modification of gene expression explicitly operates through chromatin remodelling, such as nucleosome structural modification. Acetylation and deacetylation controlled by histone acetyltransferases and histone deacetylases (HDACs) are the common methods of post-translational histone modification (215). HDACs are not only regulators of mitosis, cell differentiation, apoptosis, and chromatin organization but are also involved in the regulation of metabolism, learning, memory, and immune response (216). As clinically validated cancer targets, their inhibition has been proven to be a successful strategy for the development of novel anticancer drugs (217). Numerous pre-clinical *in vivo* and *in vitro* studies support the potential of HDAC inhibitors and are included in Table 10. The HDAC inhibitor-induced cell death is mediated through several pathways. Cell growth arrest, differentiation, induction of apoptosis, antiangiogenic effects, and effects on DNA repair and on mitosis have been observed in cancer cells after HDAC inhibition (218).

Methylation studies showed loss of methylation in CpG probes regulating *HDAC5* in BCVY patients and this hypomethylation correlates with a reduced *HDAC5* expression in this group of age, that was validated by qRT-PCR. After the importance of HDACs in cancer and the potential inhibitors commercially available, we decided to analyse its inhibition effect using the HDACi LMK-235. LMK-235 is a class IIa-selective HDAC inhibitor with high potency against *HDAC4* and *HDAC5*. We compared the inhibition effect using two breast cancer cell lines from young women (HCC1500 and HCC1937) and four from older women with breast cancer (MDA-MB-231, MCF-7, HCC1806 and BT-474).

Table 10. HDAC inhibitors studied in clinical trials (Addapted from (219)).

Name	Chemical structure	Specificity
Romidepsin (FK228)	Cyclic peptide	Class I HDACs
Vorinostat (SAHA)	Hydroxamic acids	Class I, II, IV HDACs
Panobinostat (LBH589)	Hydroxamic acids	Class I, II, IV HDACs
Quisinostat (JNJ26481585)	Hydroxamic acids	Class I, II, IV HDACs
Givinostat (ITF2357)	Hydroxamic acids	Class I, II HDACs
CKD-581	Hydroxamic acids	Class I HDACs
Belinostat (PXD101)	Hydroxamic acids	Class I, II, IV HDACs
Abexinostat (PCI-24781)	Hydroxamic acids	Class I, II HDACs
Fimepinostat (CUDC-907)	Hydroxamic acids	Class I, II HDACs+PI3K
Tinostamustine (EDO-S101)	Hydroxamic acids	Bendamustine–vorinostat fusion
Ricolinostat (ACY-1215)	Hydroxamic acids	HDAC6
Citarinostat (ACY-241)	Hydroxamic acids	HDAC6
Entinostat (MS-275)	Benzamide	Class I HDACs
Tacedinaline (CI-994)	Benzamide	Class I HDACs
AR-42	Benzamide	Class I, II, IV HDACs
4SC-202	Benzamide	Class I HDACs
CXD101	Benzamide	Class I HDACs

4.1 MTT cell proliferation assay using LMK-235

The previous mentioned breast cancer cell lines were exposed to increasing concentrations of LMK-235 (0, 0.625, 1.25, 2.5, 5, 10 and 20 µM) for 24, 48 and 72 hours. We observed a relative proliferation decreased in a dose- and time- dependent manner for some breast cancer cell lines and results could be observed in Figure 31. After 48 hours of treatments, cell viability of MDA-MB-231 and HCC1806 cell lines was severely compromised with low doses of LMK-235. Additionally, HCC1937 cell line experimented a notably viability decreased after 72 hours of

treatment showing similar results to MDA-MB-231 and HCC1806 cell lines. Despite no significant results were obtained for HCC1500 cell line after 48 hours of treatment, the young cell line shown a viability reduction of 50% after 72h of treatment with low doses of LMK-235 inhibitor. Breast cancer cell lines that experience viability reduction presented a triple negative subtype. Contrary, the breast cancer cell lines MCF-7 and BT-474, both from luminal subtypes, did not presented a significant viability reduction in comparison with the previous ones.

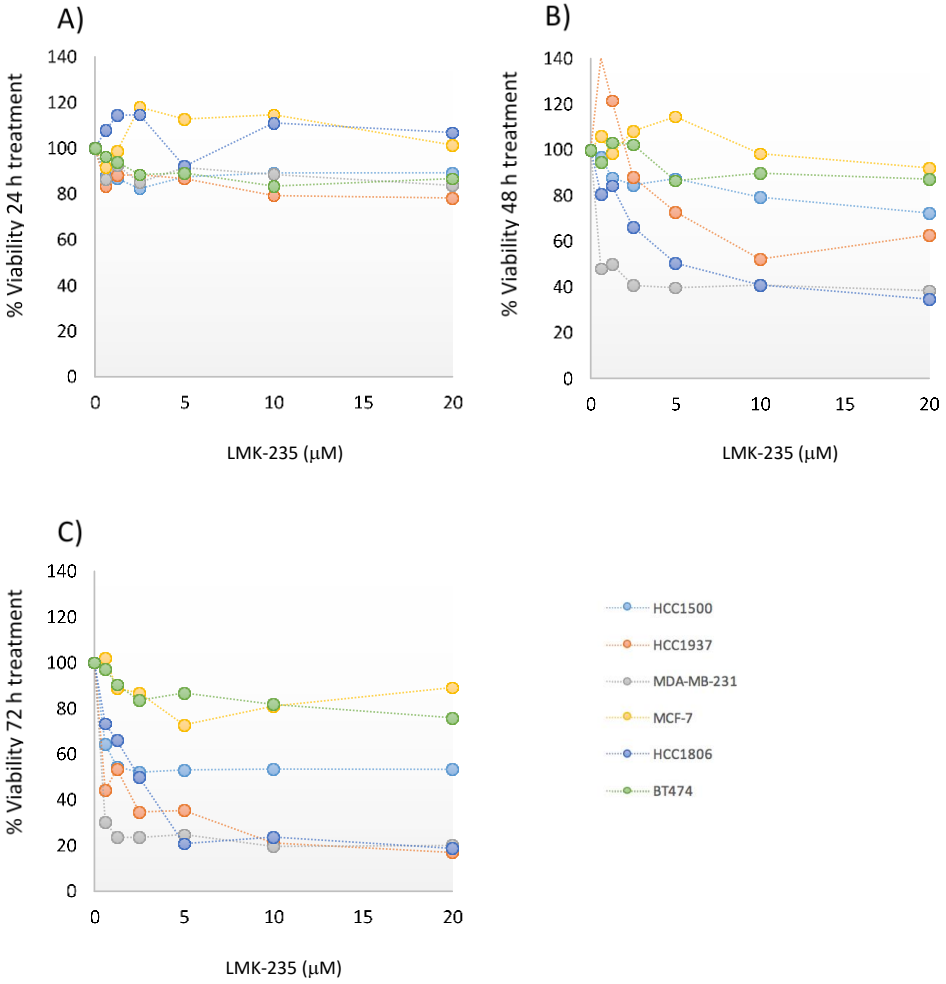


Figure 31. LMK-235 inhibits BC cell proliferation. HCC1500, HCC1937, MDA-MB-231, MCF-7, HCC1806 and BT-474 cell lines were treated with LMK-235 (0 to 20 µM) for 24 hours (**A**); 48 hours (**B**) and 72 hours (**C**). Cell proliferation was determined by MTT assay. Points indicate the mean of at least three independent experiments.

4.2 Migration reduction of breast cancer cell lines treated with LMK-235 measured by wound-healing assay

Breast cancer cell lines were treated with 20 µM of LMK-235 iHDAC inhibitor. A scratch was performed at 0 hours and the percentage of wound closure was measured using images at 0-48 hours. Images of wound-healing by cell line are included in Figure 32.

Wound-healing assays demonstrated that the LMK-235 inhibitor reduce significantly the migration of HCC1937, HCC1806 and MDA-MB-231 breast cancer cell lines. Specifically, we observed a ~23% of cell migration reduction in the HCC1937 cell line, ~ 14.7% for MDA-MB-231 and HCC1806 cell line shown not only important cell migration reduction but also cell apoptosis. HCC1500 CLVY treated with LMK-235 presented a cell migration reduction of ~5.3%. However, this last CLVY shows reduced cell doubling time and their grown rates are slower compared with the other cell lines and although no significant results were obtained in comparison with its control, HCC1500 treated with LMK-235 presented considerably cell apoptosis at 48 hours when the cells were treated. Contrary, breast cancer cell lines MCF-7 and BT474, both from luminal subtypes, did not shown significant cell migration reduction when cells were treated with LMK-235 (Figure 33).

These results showed a higher cell migration reduction in cell lines from triple negative subtypes and among them two were from CLVY (HCC1937 and HCC1500). Results were in agreement with previous observations by MTT assays.

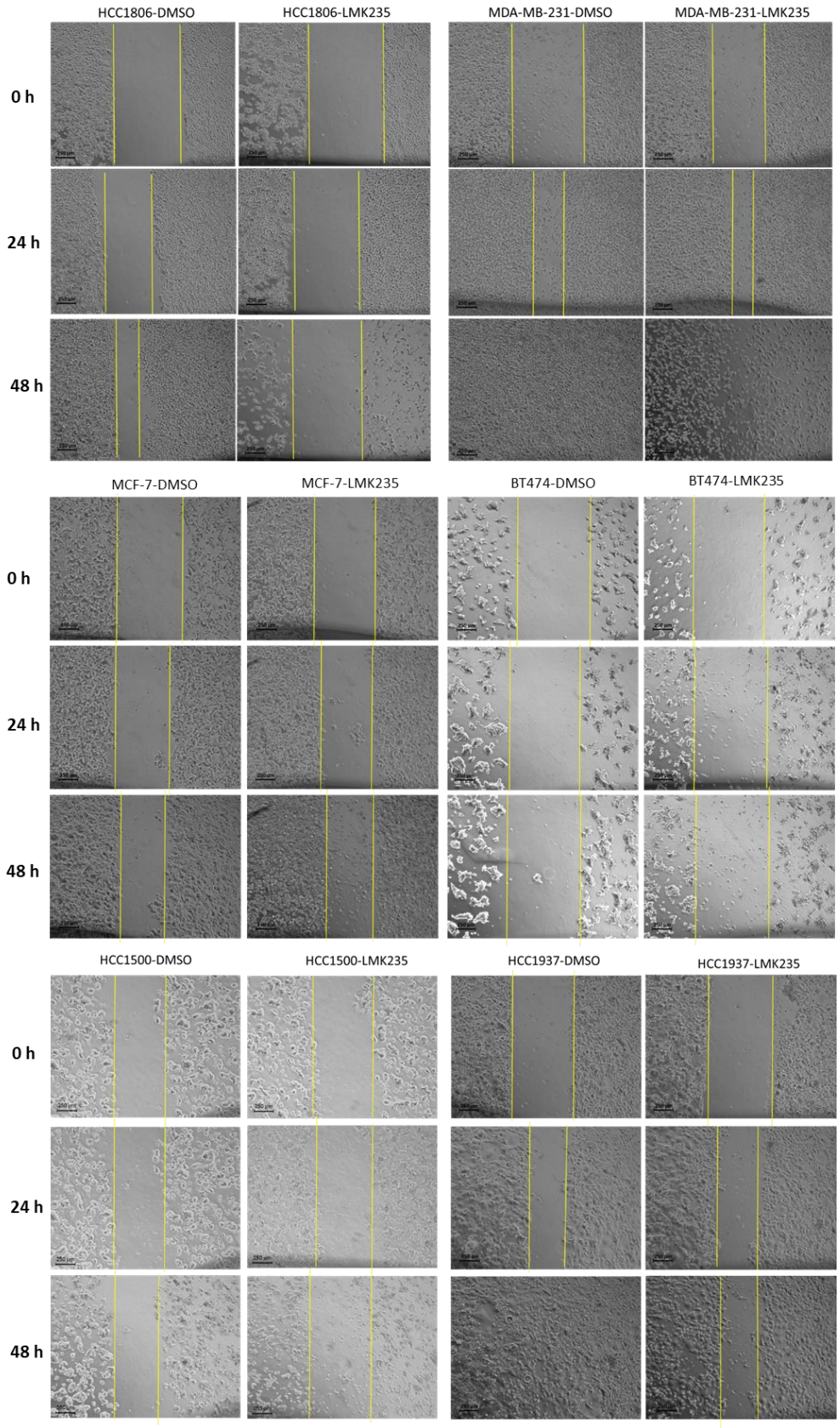


Figure 32. Wound-healing cell migration assay. HCC1806, MDA-MB-231, MCF-7, BT474, HCC1500 and HCC1937 cell migration was assessed by scratch wound-healing assay after treatment with LMK-235 (20 µM) for 48 hours. Three separate experiments were conducted and representative results are shown. Magnification, 40x.

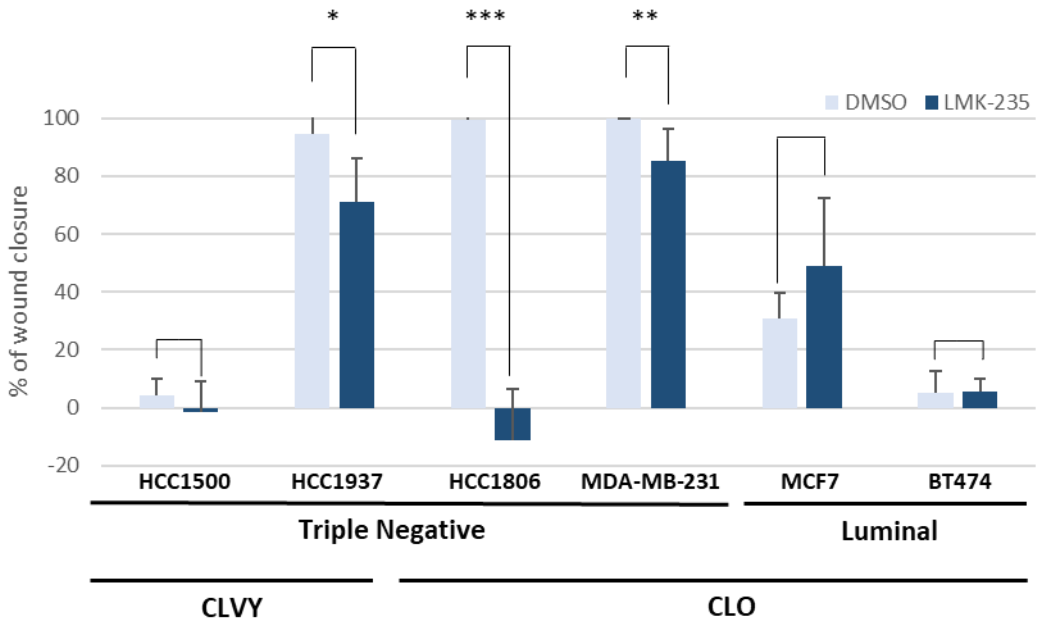


Figure 33. Wound-healing cell migration assay analysis of wound closure. HCC1806, MDA-MB-231, MCF-7, BT-474, HCC1500 and HCC1937 cell migration was assessed by scratch wound-healing assay after treatment with LMK-235 (20 µM) for 48 hours. Three separate experiments were conducted and representative results are shown. Columns, mean ±SD of percentage of wound closure in the three independent experiments. Light blue bars represent control DMSO cells and dark blue bars represent cells treated with LMK-235. *P ≤ 0.05, **P ≤ 0.01, ***P ≤ 0.001 statistically significant. CLO: BC cell lines from older patients; CLVY: BC cell lines from very young patients.

4.3 Accumulation of acetyl-H3 after LMK-235 treatment in breast cancer cell lines

The effects of *HDAC5* inhibition were examined using the HDACi LMK-235. The HDAC specificity of LMK-235 was validated by measuring its effect on the level of acetyl-histone H3. Western-blot studies showed accumulation of acetyl-histone H3 in breast cancer cell lines after 48 hours of LMK-235 (20 µM) treatment (Figure 34). We observed acetyl-histone H3 accumulation both in luminal breast cancer cell lines, such as BT474 and triple negative breast cancer cell lines (HCC1937). These results demonstrate the specific inhibition of HDACs which catalyse the removal of acetyl groups from N-acetyllysine residues of histones therefore, their inhibition produces an accumulation of acetyl-histones and the acetyl-H3 is one of the most extensively modified.

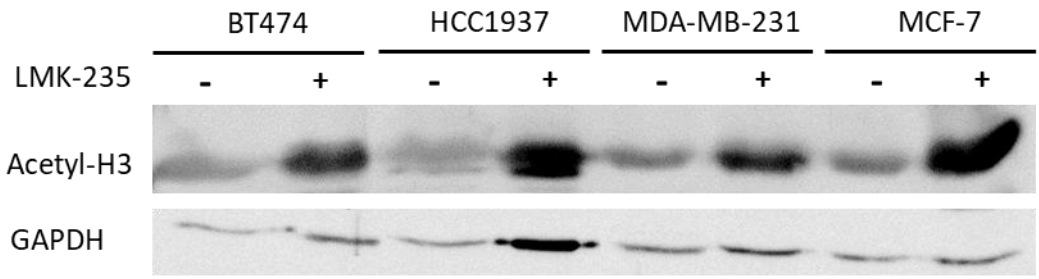


Figure 34. Accumulation of acetyl-histone H3 after 48 hours of LMK-235 (20 µM) treatment in breast cancer cell lines examined by western-blot. The levels of acetyl-histone H3 were examined by western blot. GAPDH was used as a loading control.

5. DNA methylation of genes encoding miRNAs

5.1 Methylation differences in probes regulating genes encoding miRNAs

In order to evaluate the methylation differences in genes encoding miRNAs between BCVY and BCO patients we analysed data from 6 567 CpG sites regulators of genes encoding miRNAs present in the EPICarray. Wilcoxon rank sum test shown 193 CpG probes that were significantly differentially methylated in BCVY (p-value < 0.05 and methylation differences ± 0.1) and that regulated 83 unique miRNA encoding genes. Among them, 90 were hypomethylated and 103 hypermethylated in BCVY with (Figure 35 and Figure 36). Significant differentially methylated CpG sites obtained are included in Annex XIII. However, hierarchical clustering showed two principal sample groups: one consisting of BCVY samples and other including BCO, four BCO samples clustered with BCVY patients. GLM analysis showed that methylation differences distinguishing BCVY from BCO were not related to ER status (p-value = 0.06) or molecular subtypes (p-values = 0.65) and no molecular subtype clusters were identified in heatmap representation (Figure 35).

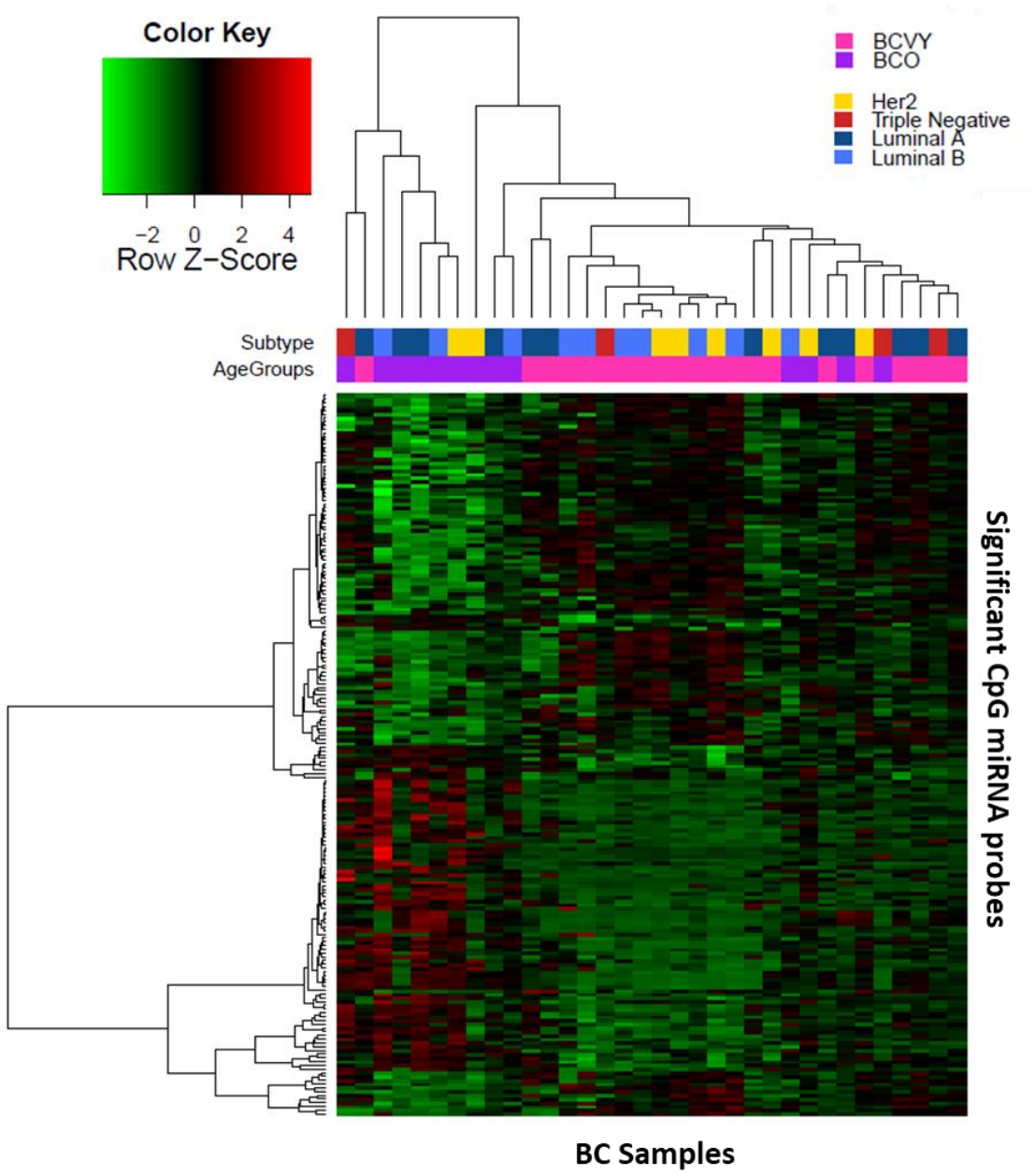


Figure 35. Differential miRNA methylation study in BCVY vs. BCO from metEPICVal samples. Heatmap representing a supervised cluster centred on the median of the methylation levels at the 193 CpG sites that regulated miRNA genes distinctive in BCVY. Hypermethylated CpG probes in BCVY (red) and hypomethylated probes (green). Samples represented as BCVY (light pink) and BCO samples (purple). Molecular subtypes are indicated as: Her2 (yellow), triple negative (red), luminal A (dark blue) and luminal B (light blue). BC: breast cancer; BCVY: breast cancer in very young women; BCO: breast cancer in old women.

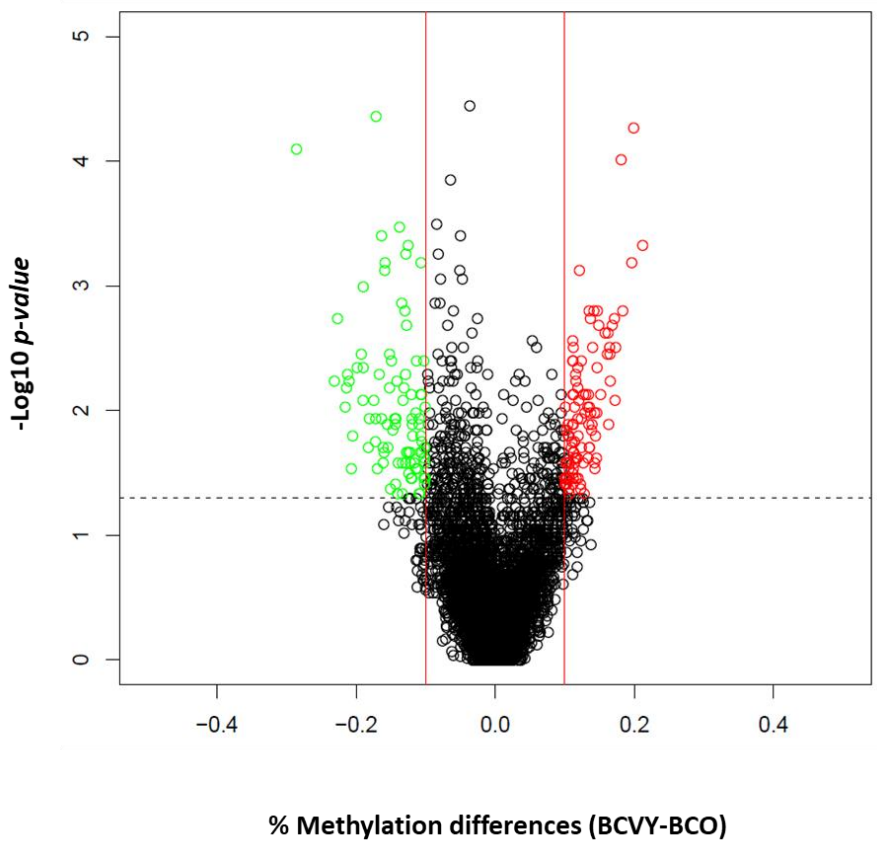


Figure 36. Volcano-plot representation of methylation for significant CpG regulators of miRNA genes. Hypomethylated probes in BCVY (green) and hypermethylated (red). Red lines delimit ±0.1 methylation differences between BCVY vs. BCO and the dotted line represents a p-value threshold of 0.05. BCVY: breast cancer in very young women; BCO: breast cancer in old women.

Genomic and functional context of significant CpG sites

In terms of CpG context, significant methylation differences for miRNA CpG probes were localized in islands (p-value= 5.29 x 10^{-7}) and regions away from them called open sea (p-value = 4.73 x 10^{-5}). Again, and as we had previously observed in the methylation study, significant CpG probes localized in island regions were mainly hypomethylated in BCVY and those localized in open sea sites were hypermethylated in this age group (Figure 37a). We analysed the functional localization of differentially methylated miRNA probes and the most important differences were identified in DNasa hypersensitive sites (p-value = 4 x10^{-3}), which were generally hypermethylated in BCVY. Although not significant, transcription factor binding sites, promoters

and gene body regions presented methylation differences of more than 4% between BCVY and BCO (Figure 37b).

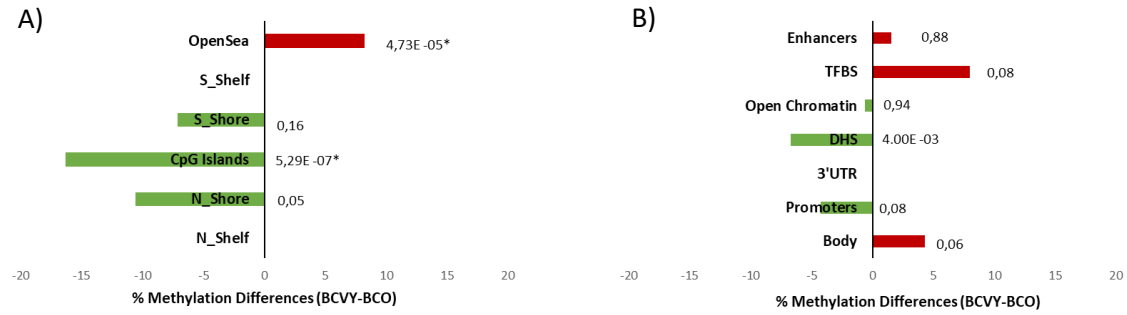


Figure 37. Genomic and functional context of significant CpG site regulators of genes encoding miRNAs which are differentially methylated in BCVY-metEPICVal samples. Percentage of methylation differences for statistically significant CpG sites from BCVY-BCO comparison according to location of the CpG relative to the island (**A**) and to the UCSC gene region feature category and regulatory elements (**B**). Red bars represent hypermethylation in BCVY and green bars represents hypomethylation. *statistically significant p-values ($p<0.001$). N/S: north/south; upstream or downstream to the CpG island. BCVY: breast cancer in very young women; BCO: breast cancer in old women; UTR: Untranslated region; TFBS: transcription factor binding site; DHS: DNasa I hypersensitive sites.

5.2 MiRNA methylation validation in two populations

Both data validation sample sets were analysed using HM450K array, which includes around 50% of probes included in the EPICarray. Of 193 significant differentially methylated miRNA probes identified in the metEPICVal sample set, only 85 were shared with the HM450K array and could be analysed in the validation study, whereas the 108 remaining probes were exclusively for the EPICarray and could not be included in the validation study. We verified the methylation differences for 17 CpG sites between BCVY and BCO in TCGA data (p-value < 0.05 and methylation differences ± 0.1). Combined data showed 30 differentially methylated miRNA CpG probes among the 85 analysed in the validation study. Finally, we were able to validate a total of 10 methylation probes regulating miRNAs that were significant in both validation data sets, and all of them were hypomethylated in BCVY compared to older patients (Table 11).

Table 11. Table of CpG probe regulators of miRNA genes that were significantly differentially methylated in BCVY compared with BCO samples in the validation study using TCGA and the combined study data sets. This table includes a list of 10 significant CpG probes and the corresponding miRNA that were regulated by them.

		TCGA		Combined study	
Validated CpG probes	miRNA gene	Methylation differences (BCVY - BCO)	p-value	Methylation differences (BCVY - BCO)	p-value
cg04735310	MIR196A1	-0,11	4,13E-50	-0,15	5,57E-04
cg07234865	MIR9-1	-0,13	1,80E-45	-0,14	6,72E-05
cg08737296	MIR124-3	-0,13	7,30E-23	-0,16	3,84E-03
cg07792478	MIR124-2	-0,14	3,92E-21	-0,10	1,03E-02
cg05474726	MIR124-2	-0,12	2,66E-20	-0,14	4,21E-03
cg22333214	MIR137	-0,15	2,74E-20	-0,12	1,41E-02
cg04947764	MIR184	-0,11	5,67E-18	-0,18	4,73E-04
cg25147193	MIR181C	-0,10	5,61E-17	-0,14	8,03E-03
cg16407471	MIR129-2	-0,10	5,60E-12	-0,13	2,84E-02
cg00210994	MIR548G	-0,10	2,11E-07	-0,15	2,29E-02

BCVY: breast cancer in very young women; BCO: breast cancer in old women.

5.3 Meta-analysis expression of miRNAs regulated by significant CpG probes

To elucidate whether differentially methylated CpG probes affected miRNA gene expression, we analysed expression of affected miRNA in a meta-analysis using three gene expression data sets (TCGA, Clariom™ D and Peña-Chilet et al. (160)). We analysed expression in miRNAs regulated by the 10 previously validated CpG sites and we included miRNAs regulated by 108 CpG probes that were differentially methylated in BCVY-BCO and exclusive of the EPICarray and therefore excluded from the validation study in the HM450K data sets. Finally, we had a total of 118 CpG probes that regulated 62 unique miRNAs. Gene expression for unique miRNAs was evaluated by t-test in the three data sets separately. Only 10 miRNAs were included in the three studies and

could therefore be evaluated in the meta-analysis (Table 12), and 4 of them were significantly de-regulated with adjusted p-value < 0.05 (miR-9-1, miR-184, miR-551b and miR-196a-1).

Table 12. MiRNA expression results from meta-analysis. Table shows miRNA expression values from the meta-analysis of the 10 miRNAs common to the three studies: TCGA, Clariom ™D and Peña-Chilet et al study (160).

miRNAs	p-value	Adjusted p value	Gene expression differences
hsa-mir-196a-1	3,92E-47*	3,92E-46	-0,28
hsa-mir-184	1,31E-34*	6,55E-34	0,26
hsa-mir-9-1	6,12E-03*	2,04E-02	-0,13
hsa-mir-551b	1,45E-02*	3,62E-02	-0,38
hsa-mir-548n	1,46E-01	2,92E-01	-0,06
hsa-mir-383	3,49E-01	5,82E-01	0,28
hsa-mir-548f-1	5,74E-01	8,21E-01	-0,28
hsa-mir-490	8,05E-01	9,65E-01	0,14
hsa-mir-181c	9,46E-01	9,65E-01	0,11
hsa-mir-662	9,65E-01	9,65E-01	-0,05

*Refers to significant p-values

We also analysed whether we observed different expression in miRNAs regulated by differentially methylated regions. For this purpose, we selected unique miRNAs for each significant region obtained (CpG Island, open sea and DNase hypersensitive sites) and their expression was analysed in a meta-analysis using the three previously mentioned data sets. We identified 14 differentially methylated miRNAs regulated by CpG islands regions and 3 of them were also differentially expressed between BCVY and BCO (miR-181c, miR-196a-1 y miR-212). Significant open sea regions regulated 55 miRNA genes and miR-184, miR-211 and miR-383 were significantly deregulated in the meta-analysis. Finally, DNase hypersensitive regions significant in methylation analysis regulated 38 miRNAs and 5 of them presented different expression (miR-196a-1, miR-184, miR-345, miR-212 and miR-211). Results are shown in Table 13. Interestingly,

we found some miRNAs that were regulated by more than one significant CpG categories and were deregulated in BCVY compared with older patients.

5.4 Pathway enrichment analysis of genes encoding miRNAs significantly deregulated

We performed a pathway enrichment analysis of significant miRNA genes obtained in the meta-analysis that were regulated by different methylated regions, using DIANA miRPath tool. We considered only the experimentally validated miRNA-gene interaction included in TarBase. KEGG pathways enriched are reported in Figure 38. The miRNAs miR-9-1 and miR-196a-1 were involved in pathways related to adherent junctions (hsa04520), proteoglycans involved in cancer (hsa05205) or transcriptional misregulation in cancer (hsa05202), among others. MiR-196a-1 took part in multiple pathways, the vast majority of them implicated in other cancer types.

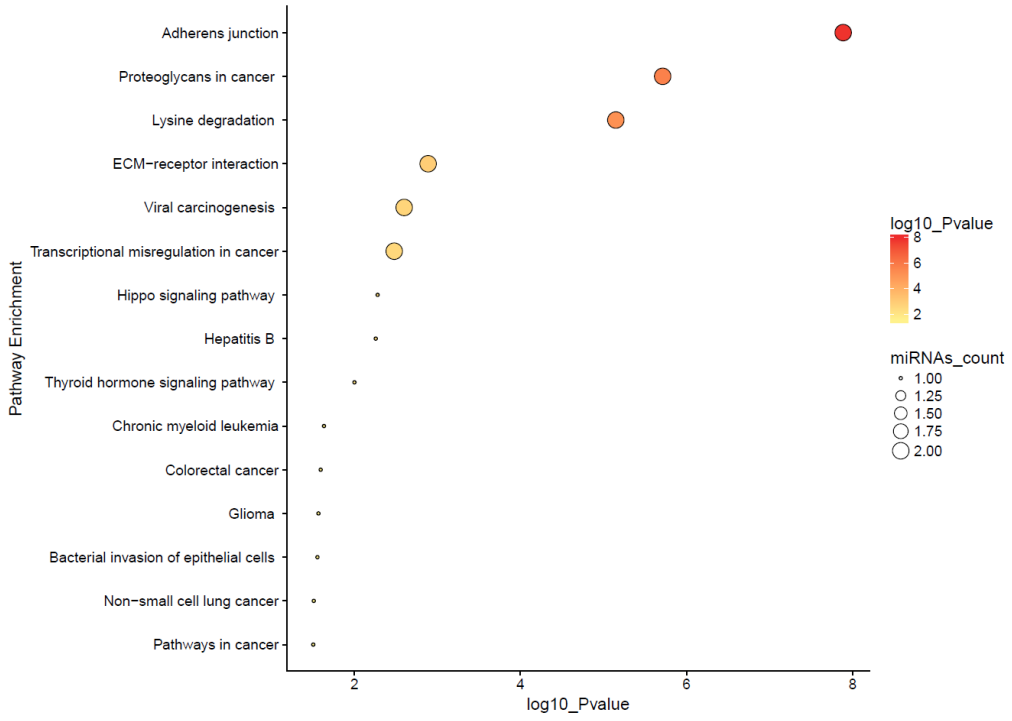


Figure 38. Pathway enrichment analysis results obtained by DIANA mirpath. Plot represents the main pathways in which miRNA regulated by significant CpG probes in BCVY were involved. Dot colour indicates p-values and miRNA count indicates the number of miRNAs involved in the represented pathways.

Table 13. MiRNA expression results from meta-analysis by category regions. Expression of miRNAs that were regulated by CpG probes localized at the most different methylated regions between BCVY and BCO were analysed separately in the meta-analysis. The table shows miRNAs that were differentially methylated and differentially expressed in BCVY vs. BCO taking into account category regions analysed. P-value and adjusted p-value by Benjamini & Hochberg were calculated and differences in gene expression and methylation between BCVY and BCO are indicated in the table.

Island probes				
miRNAs	p-value	p-adjusted	gene expression differences (BCVY - BCO)	mean methylation differences (BCVY - BCO)
hsa-mir-181c	$1,29x10^{-41}$	$5,15 x10^{-41}$	-0,27	-0,12
hsa-mir-196a-1	$1,39 x10^{-05}$	$2,79 x10^{-05}$	-0,06	-0,11
hsa-mir-212	$2,19 x10^{-02}$	$2,92 x10^{-02}$	-0,13	-0,17
OpenSea probes				
miRNAs	p-value	p-adjusted	gene expression differences (BCVY - BCO)	mean methylation differences (BCVY - BCO)
hsa-mir-184	$9,82 x10^{-41}$	$9,82 x10^{-40}$	0,25	-0,18
hsa-mir-211	$1,28 x10^{-03}$	$6,41 x10^{-03}$	-0,2	0,17
hsa-mir-383	$3,39 x10^{-03}$	$1,13 x10^{-02}$	-0,38	0,12
DNase hypersensitive regions				
miRNAs	p-value	p-adjusted	gene expression differences (BCVY - BCO)	mean methylation differences (BCVY - BCO)
hsa-mir-196a-1	$3,79 x10^{-40}$	$4,17 x10^{-39}$	-0,27	-0,12
hsa-mir-184	$2,10 x10^{-36}$	$1,16 x10^{-35}$	0,26	-0,18
hsa-mir-345	$2,52 x10^{-21}$	$9,25 x10^{-21}$	0,04	0,11
hsa-mir-212	$5,84 x10^{-05}$	$1,61 x10^{-04}$	-0,07	-0,11
hsa-mir-211	$2,34 x10^{-04}$	$5,14 x10^{-04}$	-0,2	0,17

BCVY: breast cancer in very young women; BCO: breast cancer in old women.

Additionally, we performed a pathway enrichment analysis taking into account the miRNAs deregulated by localization of significant categories obtained. MiRNAs regulated by CpG probes localized in island regions were involved in multiple pathways related to cancer (transcriptional misregulation in cancer (hsa05202), prostate (hsa05215), lung (hsa05223) and endometrial cancer pathway (hsa05213), among others, adherent junctions (hsa04520), oestrogen signalling pathways (hsa04915), among others (Figure 39a). For deregulated miRNAs in BCVY whose methylation was regulated by open sea regions, we identified pathways related to cancer as previously observed for island miRNAs and, moreover, they were implicated in TNF signalling pathways (hsa04668), ECM-receptor interaction (hsa04512) and prolactin signalling pathways (hsa04917) (Figure 39b). Finally, for DNase hypersensitive sites we obtained, again, most of the pathways previously observed in island regions (Figure 39c).

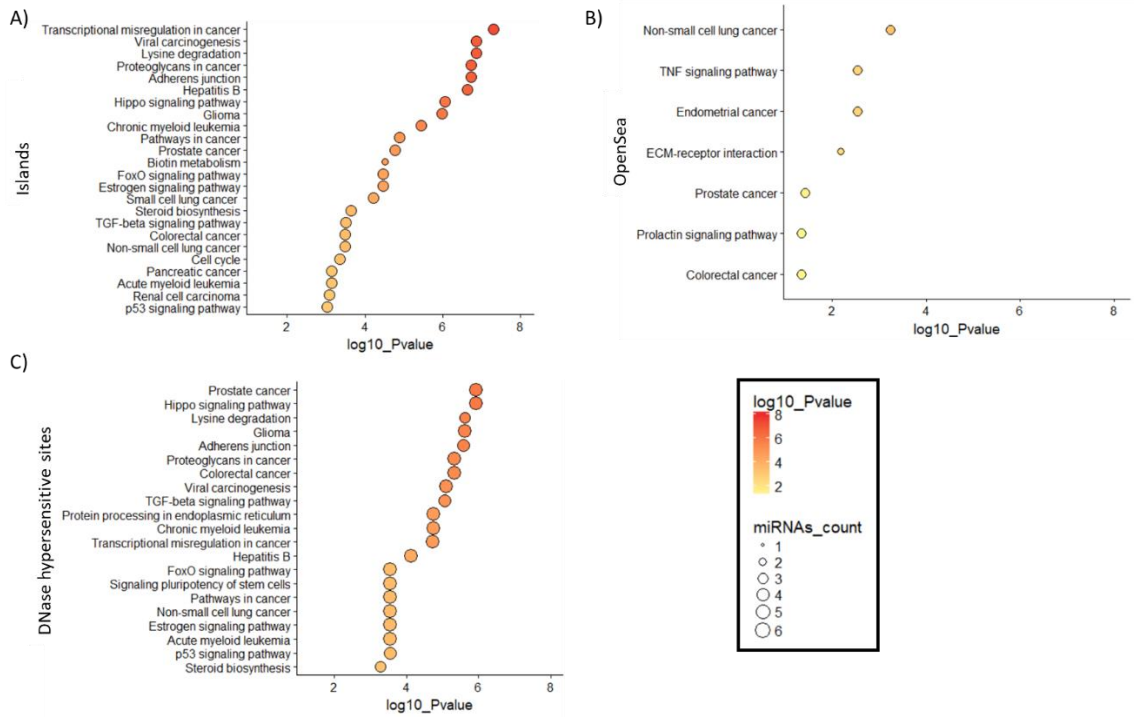


Figure 39. Pathway enrichment analysis results for miRNAs deregulated by localization of significant categories obtained: (**A**) Islands; (**B**) Open Sea; (**C**) DNase hypersensitive sites. Dot colour indicates p-

values (orange for most significant pathways and yellow for lower p-values) and miRNA count indicates the number of miRNAs involved in the represented pathways.

5.5 qRT-PCR validation of miRNA expression

To validate the differential expression of significant miRNAs obtained in the meta-analysis that were regulated by differentially methylated CpG sites in BCVY we performed a validation study by quantitative real-time PCR (qRT-PCR). Validation was performed on an independent set of samples with similar characteristics to those used in the metEPICVal study. We analysed the expression of the four significant miRNAs from the meta-analysis (miR-9-1, miR-184, miR-551b and miR-196a-1) and additionally, we analysed the expression of the precursor miR-124-2 that was hypomethylated in both methylation validation data sets but could not be evaluated in the expression study. MiR-124-2 was significantly overexpressed in BCVY samples (p-value = $4x10^{-3}$) (Figure 40a) compared with older patients and this result correlated with the significant hypomethylation detected in BCVY.

MiR-9-1 (p-value = 0.07), miR-196a-1 (p-value = 0.36) and miR-184 (p-value = 0.7) were overexpressed in BCVY, which are in agreement with the hypomethylation detected and miR-551b was repressed in BCVY (p-value = 0.07) that was consistent with the hypermethylation found. However, the differences did not reach statistical significance (Figure 40b-e).

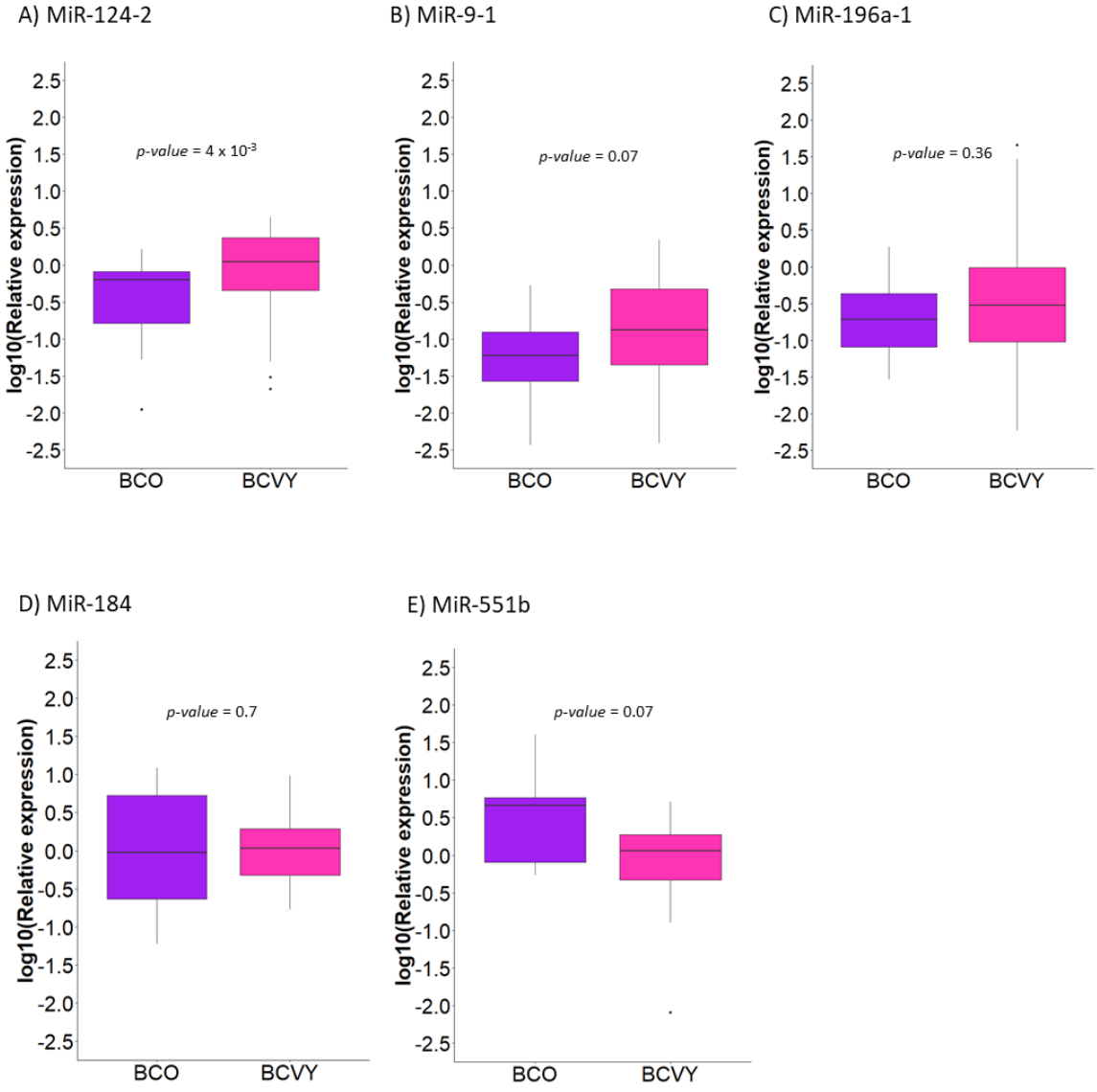


Figure 40. qRT-PCR expression validation results. Boxplots represent the sample distribution with the mean for BCO and BCVY patients. Expression was quantified using the qRT-PCR technic and calculated by ΔΔCt method. Y axis represents logarithmic transformation of the relative expression. Differences by miRNA between BCVY and BCO were analysed by Wilcoxon rank sum test: (**A**) miR-124-2; (**B**) miR-9-1; (**C**) miR-196a-1; (**D**) miR-184; (**E**) miR-551b. BCVY: breast cancer in very young women; BCO: breast cancer in old women; qRT-PCR: quantitative real time PCR.

5.6 Clinical importance of miRNA methylation as an independent prognostic factor for relapse-free survival and overall survival in BCVY

Although no significant expression differences were identified for most of the miRNAs, we wondered whether the 5 significantly differentially methylated miRNAs in BCVY were associated with relapse-free survival (RFS) and overall survival (OS). To test this hypothesis, we investigated the metEPICVal samples and methylation samples from TCGA that present follow-up data. We next performed a univariate cox regression study to determine whether hypo- or hypermethylation of significant miRNAs were correlated with underlying clinical conditions in BCVY and BCO. Whereas no significance was reached in the RFS analysis, miRNA hypomethylation was related with higher relapse risk in both BC groups (BCO and BCVY) for all miRNAs analysed with the exception of miR-184, in which hypomethylation reduces the risk of relapse in BCVY, contrary to the results for BCO (Figure 41). For OS studies, univariate cox analysis showed that miR-124-2 hypermethylation was significantly related with reducing survival in BCO (p-value = 9×10^{-4}) (Figure 42). Multivariate analysis shown a significant relation with miR-124-2-hypermethylation in BCO and poor survival (p-value 1.02×10^{-3}, hazard ratio HR = 3.23). However, oestrogen receptor status and molecular subtype were not contributing to survival in BCO patients. Although no significant results were obtained in univariate cox study among BCVY survival and miR-124-2 methylation, we observed a contrary effect on them, in which poorer survival was related with miR-124-2 hypomethylation. BCVY sample size is limited compared with older patients and further studies with larger dataset should be addressed to evaluate mir-124-2-hypomethylation and prognosis in this group. Additionally, OS analysis demonstrated a good prognostic factor for miR-184 hypomethylation in both BCVY and BCO groups, and a reduced survival rate for miR-196a-1-hypomethylation in BCVY: however, these results did not reach statistical significance (Figure 42).

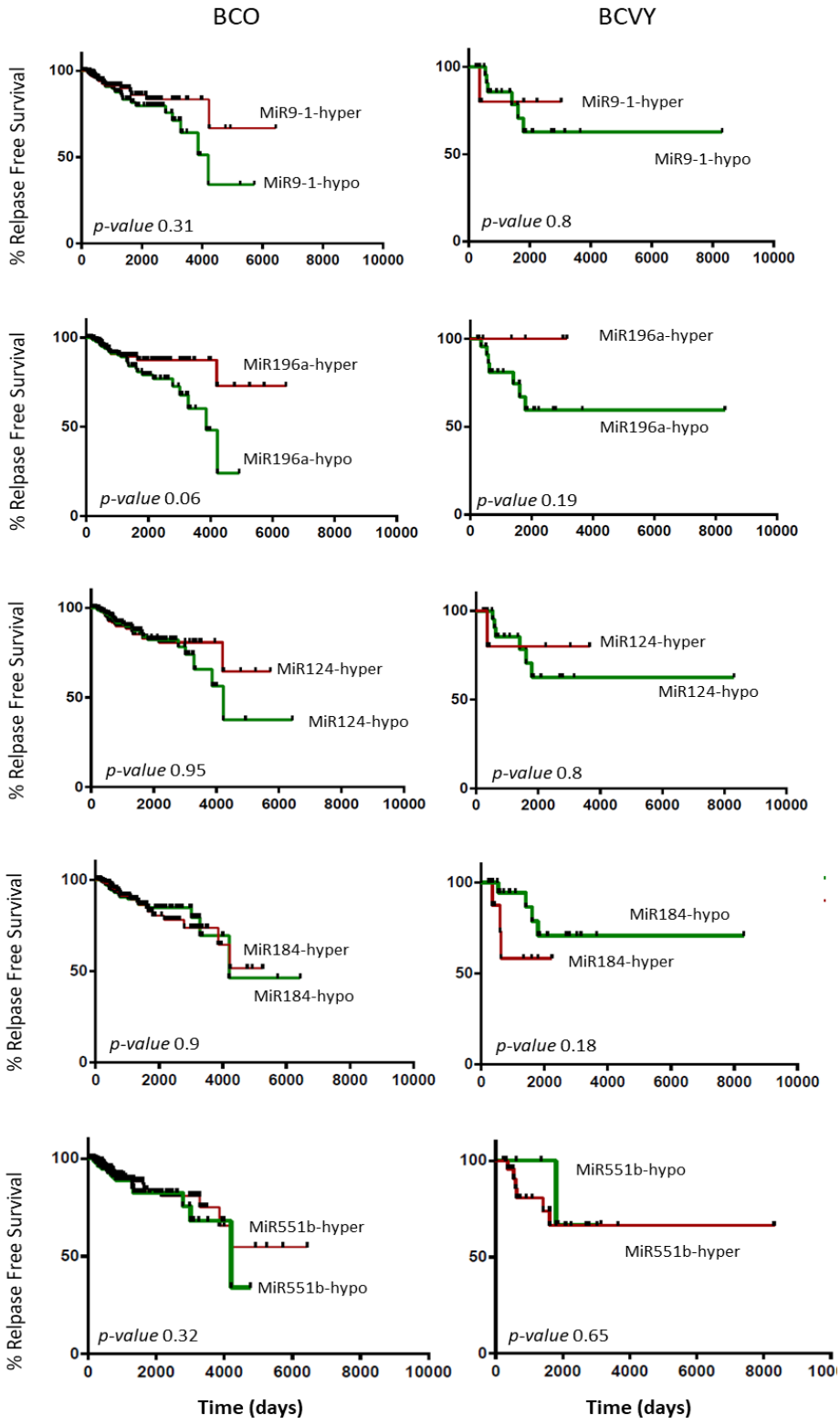


Figure 41. Representation of relapse-free survival curves for significantly differentially methylated miRNAs according to their methylation status. The first column curves represent RFS in BCO patients and the second RFS in BCVY. X axis represents the follow-up time by days and Y axis indicates the percentage of relapse-free survival. Green curves represent miRNA hypomethylation and red line hypermethylation. P-values obtained by univariate cox analysis. BCVY: breast cancer in very young women; BCO: breast cancer in old women.

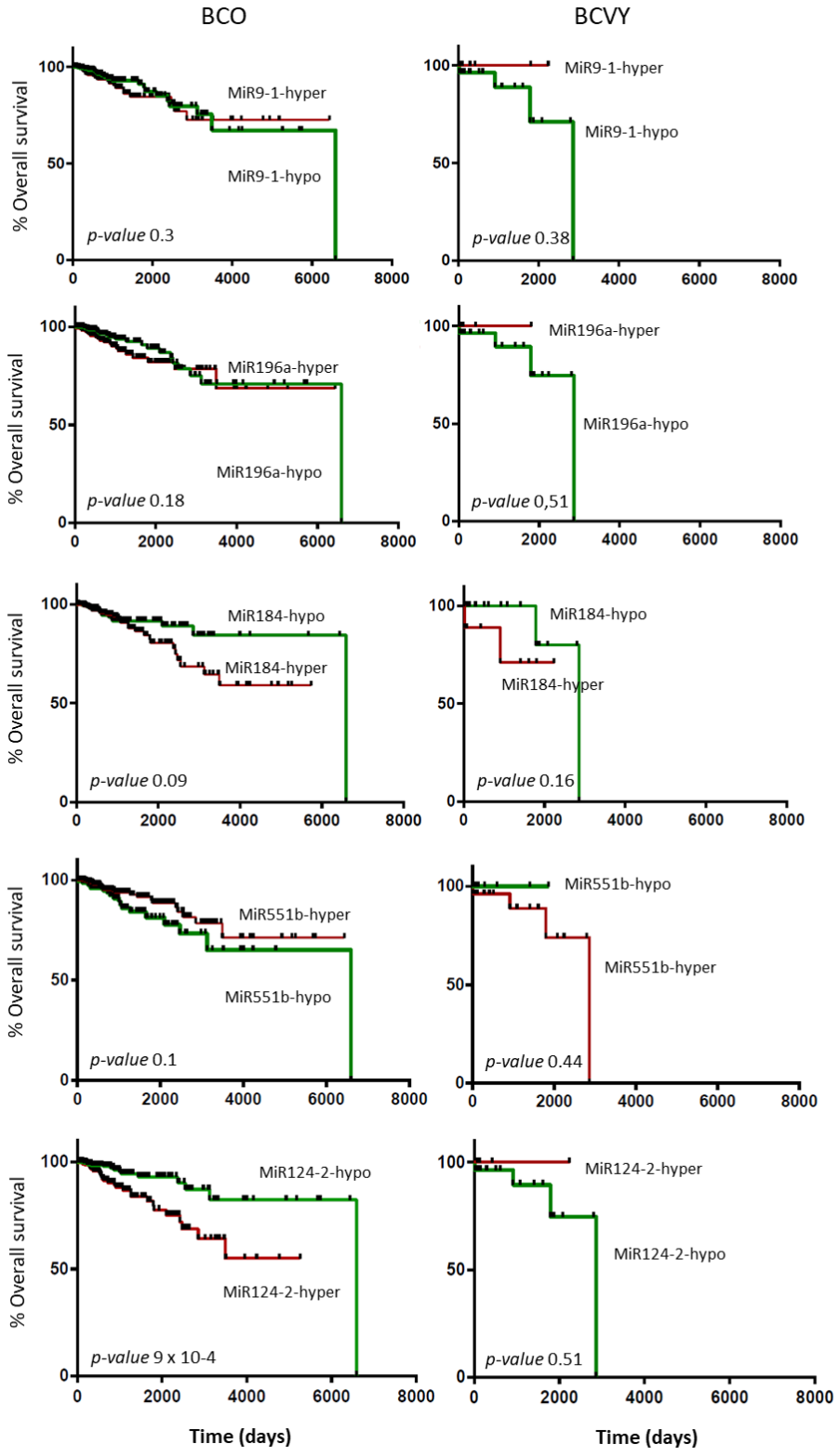


Figure 42. Representation of overall survival curves for significantly differentially methylated miRNAs according to their methylation status. The first column curves represent OS in BCO patients and the second OS in BCVY. X axis represents the follow-up time by days and Y axis indicates the percentage of overall survival. Green curves represent miRNA hypomethylation and red colour line hypermethylation. P-values obtained by univariate cox analysis. BCVY: breast cancer in very young women; BCO: breast cancer in old women.

DISCUSSION

DISCUSSION

1. MiRNA profile in breast cancer cell lines from young women and their validation as a cellular model for breast cancer research in very young patients

MiRNAs, a novel class of non-coding RNAs, have been proposed by numerous studies as promising diagnostic and therapeutic tools for human cancers because of their critical roles in various biological processes during tumor development, such as cell proliferation, differentiation, apoptosis and migration, and subsequent tumorigenesis (220, 221).

1.1 MiRNA profile characterization in breast cancer cell lines from very young women

The differential miRNA expression study between the two CLVY (HCC1500 and HCC1937) and CLO (HCC1806, MCF-7, MDA-MB-231 and MDA-MB-468) revealed a group of 155 miRNAs that were differentially expressed between cell lines from different age groups. The correspondent cell line replicates cluster together, indicating good reproducibility of miRNA expression profile among biologic replicates. With the exception of the luminal MCF-7, all breast cancer cell lines in our study were triple negative (TNBC), so we could not extract any conclusions about miRNA expression differences among molecular subtypes.

We identified a set of 24 miRNAs that presented similar expression patter between HCC1806 and breast cancer cell lines from young women. HCC1806 was established from a tumour of a 60 years old patient with TNBC (222) and was subsequently confirmed to lack ER/PR/HER2 (222)

as well as p53 (223), breast cancer markers of aggressiveness. Although, biologically and genetically, TNBC is a heterogeneous group of tumours (224), the majority (∼80–90%) falls into the classification of basal-like subtype (225, 226). Basaloid TNBC is characterized by expression of cytokeratins 5, 6, 14, and 17 (227-230), epidermal growth factor receptor (230-232), c-Kit (230), mutated BRCA1 (230, 231, 233), and mutated or deleted p53 (234, 235). Patients with basaloid TNBC have higher tumour mitotic index (236) and a worse prognosis than patients with triple-negative tumours that do not express basal markers (225, 230, 237). Basaloid TNBC subgroup has a tendency to generate larger tumours (227) with frequent lymphovascular invasion (238, 239) and metastasis to multiple sites, whereas non-basaloid tumours typically metastasize only to one site (225). This features make HCC1806 breast cancer cell line a more aggressive TNBC group. Breast cancer affecting young women are characterized with a more aggressive, more proliferative tumours and higher size than older patients. We hypothesise that, the group of 24 miRNAs with similar expression profile between HCC1806 and CLVY may be related with aggressiveness that is a similar feature among breast cancer in young patients and the previous description of HCC1806 cell line. With these common features between them in mind, enrichment pathway study shown that most of pathways deregulated by these miRNAs were previously obtained in the breast cancer cell lines study. However, a remarkable deregulation was observed for pluripotency of stem cells related pathways. A growing body of evidence has demonstrated a direct link between epithelial-to-mesenchymal transition progression and the gain of cancer stem cells-like properties, including the ability to self-renew and tumorigenicity. Additionally, failure of conventional treatments has commonly been associated with cancer stem cells survival (240-243). The major aggressive features attributed to HCC1806 cell line and its similarity to younger cell lines observed in our study, make us to associate an aggressiveness role to the set of 24 miRNAs identified. Further studies should be addressed to deeply in the potential role of these miRNAs in cancer progression and more aggressive tumour features.

We were able to validate some miRNAs that were differentially expressed in breast cancer cell lines from young women. The miR-23a and miR-30c validation results agree with downregulation in younger cell lines obtained in the miRNA array.

1.2 Validation of the promising use of HCC1500 and HCC1937 cell lines as a cellular model for breast cancer studies in very young women

The combined study analyzing miRNA expression data from breast cancer cell lines and patients shown good correlation between young women samples, both from cell lines and tumours, and a different miRNA profile between them and samples from older women (cells and patients). Specifically, we identified 132 miRNAs differentially expressed between young and older samples and 47 of them were previously observed in the breast cancer cell line study. In the same way as breast cancer cell line study, HCC1806 cell line presented a miRNA expression profile more similar to young samples. These results agree with previous observations in cell line's study, where HCC1806 presented a similar pattern to CLVY for the miRNAs that we suggest that are related with the aggressive features of breast cancer affecting very young women. The description of HCC1806 as an aggressive TNBC cell line, and their higher correlation with breast cancer affecting very young women in terms of miRNA expression, supports the idea that breast cancer affecting young women presents more aggressive tumours characteristics with poor outcome than older counterparts.

Multiple pathways obtained in the enrichment analysis were observed both in the breast cancer cell line study as well as in the combined study with cell lines and tumour samples. Most of them were related with signaling pathways important in cancer such as mTOR, PI3K-AKT, ERBB, RAS, WNT and MAPK. Additionally, we identified deregulated pathways related with nervous system that has been previously observed in other works in breast cancer in young women. In

agreement with this, previous articles have highlighted the importance of neural elements in the cancer microenvironment in which promotes cell growth. However, the mechanisms mediating neuronal influences on cancer growth and progression are likely incompletely understood (244). Cell lines have been used as models for breast cancer research. However, there are few breast cancer cell lines available from young women (<35 years old) and some of them have been eliminated from breast cancer studies due to their uncertain origin such as MDA-MB-435 cell line (31 years old) which was described as a melanoma cell line 14 years ago. The reproducibility and similarity between the tumour sample study and the present breast cancer cell line analysis, supports the use of HCC1500 and HCC1937 as a suitable model for the study of breast cancer affecting very young women.

1.3 Poor survival for BCVY patients with miR-23a downregulation

The combined study shown 10 potential miRNA deregulated in younger breast cancer that were significant in the three analysis (cell lines, tumour samples and their combination). Among them, the miR-23a downregulation was validated in our breast cancer cell lines study as well as in the tumour patient samples from Peña-Chilet (160). Although we could not obtain significant results, we identified an association between miR-23a-dowregulation and poor survival for young patients in comparison to older ones, in which results shown a contrary association. MiRNA-23a is a critical regulator in carcinogenesis and aberrant miR-23a expression has been detected in many cancers. Analysis of breast cancer tissue-based miRNA panel shows that miR-23a possibly functions as an oncogenic governor and promotes breast cancer progression (245). However, miR-23a also acts as a negative regulator of oncogene in several types of cancer, and an example is their down-regulation in nephroblastomas (246) and pancreatic cancer (247). Despite of using data from two huge data sets (METABRIC and TCGA) BC samples from women younger than 35

years old is quite limited compared with older counterparts. Considering that, further studies should be addressed to detail the potential role of miR-23a-downregulation as a poor prognostic biomarker in breast cancer affecting young women.

The similarity in the miRNA expression between breast cancer cell lines and tumour samples from BC patients, demonstrates the suitability of HCC1500 and HCC1937 as cellular models for future studies of breast cancer affecting very young. Additionally, the miR-23a-dowregulation was detected and validated in breast cancer cell lines and tumour samples from young women. Although, no significant results were reached, we observed an association between miR-23a-downregulation and poor survival rates in very young women with breast cancer. We encourage the future study of miR-23a-downregulation in large cohort of breast cancer in very young women in order to validate their potential role as a poor prognosis biomarker in this group of age.

2. DNA methylation characterization in breast cancer affecting very young women

2.1 Identification of a global hypomethylation profile and a distinctive hypermethylation pattern in BCVY

In the second part of the study, we analysed methylation patterns using EPIC850K methylation array. The overall pattern shows a global hypomethylation signature in BCVY which is irrelevant of probe location. Significant hypomethylation was detected in CpG Islands and adjacent regions located mainly in promoters and enhancer regulatory regions. Several epidemiological case-control studies have reported global hypomethylation in peripheral blood of cancer patients, suggesting systematic effect of hypomethylation on disease predisposition (248, 249). However,

our results reported that global hypomethylation is presumably more important in BCVY. In addition, we have identified a distinctive signature of hypermethylation in 502 CpG sites characteristic of BCVY tumours, localized mainly in regions around CpG islands and open-sea, as well as, overlapping in important regulatory expression sites such as enhancers and TFBS. Previous works revealed distinctive methylation patterns among BC subtypes, with hypermethylation for luminal B and hypomethylation for TNBC (250, 251). TNBC subtype is more frequent in young women (252, 253) and although in our sample set we have several TNBC, these are not significantly overrepresented in BCVY samples and GLM analysis demonstrate that methylation differences observed are characteristic of BCVY and independent of molecular subtypes and ER status.

According to the literature, with few exceptions, mammalian aging is more commonly associated with CpG hypomethylation patterns, especially at repetitive DNA sequences (254). Age-related DNAm changes are also more prominent in CpG islands (255). Thus, while general and localised DNA hypomethylation is detected during aging, hypermethylation would simultaneously occur at specific CpG sites with the latter presumably repressing the expression of particular genes (194, 195, 256). Therefore, we would expect higher hypomethylation in CpG island regions for older women. On the contrary, our results show stronger CpG island hypomethylation in BCVY samples. Several known CpG sites that are predictive of age have been identified in different studies (194, 195, 255) and therefore, these probes have been excluded from our analysis to avoid that the results were confounded by age related methylation changes.

The global hypomethylation identified in the BCVY group, leads us to infer a higher gene deregulation in addition to a greater genomic instability, that could justify the aggressiveness of these tumours (254).

2.2 Increased DNA methylation age acceleration in breast cancer tumours from very young women

Additionally, we have detected noticeable discrepancies between chronological and DNAmAge in BCVY tumours, which results in a significant increased age-acceleration compared with BCO in both datasets analysed.

Previous studies suggest that epigenetic age-acceleration might also be an indicator for a more aggressive course of tumour disease (257, 258). This age-acceleration has been seen in a previous study in breast tissue from healthy females respect its paired blood sample (259). However, normal breast tissue from TCGA shown a strong correlation between chronological age and DNAmAge. This correlation is better than the one obtained for cancer tissues, suggesting that acceleration observed in BCVY is disease specific and not tissue related.

Furthermore, we observed a higher age-acceleration for ER+ tumours in both data sets. These results support previous observations from Horvath and colleagues (194, 195). Although we should expect higher acceleration in older women, because their major proportion of ER+ tumours and negative acceleration values for BCVY because aggregates major proportion of TN subtypes (194, 195), our results strongly indicate a higher age-acceleration in BCVY. Additionally, we found that BCVY metEPICVal samples that were ER+ presented a significant higher epigenetic age, a trend seen also in the BCVY samples from TCGA dataset.

BC risk is linked with endogenous hormone exposure which is reduced in the post-menopausal status. Thus, we hypothesized that the acceleration found in BCVY may be partially influenced by ER status, but another factors such as hormonal levels could be involved in the increased DNAmAge and contributing with their poor outcome.

2.3 Validation of methylation patterns identified in BCVY

We could reproduce global hypomethylation found in BCVY, both in combined study and TCGA data sets with 99% and 95,7% significant probes respectively. We could not detect the distinctive hypermethylated profile for BCVY in any of the two validation populations, due to the low proportion of open-sea probes in the HM450K array. Despite that, CpG shores, shelves and open-sea regions has been associated with different carcinomas and, reprogramming-specific differentially methylated regions, with mechanistic relevance for the expression of associated genes (260, 261). Our study suggests that open-sea regions may play an important role in gene regulation in different cancer types and specifically in BCVY, but because these regions are underrepresented in HM450K studies, further research will be required to test this hypothesis.

2.4 Importance of genes differentially methylated and expressed in BCVY

Furthermore, we showed that global hypomethylation in BCVY was significantly overrepresented in genes involved in neuronal-system processes, as well as, in extracellular matrix organisation and cell communication. In agreement with these results, previous articles have highlighted the importance of neural elements in the cancer microenvironment in which promotes cell growth. However, the mechanisms mediating neuronal influences on cancer growth and progression are likely incompletely understood (244).

Gene expression analysis of TCGA data identified 186 genes whose expression correlated with the global hypomethylation signature found in BCVY. As expected, two of three genes included in our validation analysis in a different sample set, *HOXD9* and *PCDH10*, were also overexpressed in BCVY. Both, promoter and enhancer CpG sites of *HOXD9* were hypomethylated. Previous works showed that silencing *HOXD9* in liver cancer inhibited epithelial–mesenchymal transition,

migration, and invasion in vitro and decreased tumorigenic and metastatic capacities in vivo (262). This phenomenon is essential for tumour cells dissemination and metastasis and fits with the major invasive ability of BCVY tumours.

Protocadherin 10 (*PCDH10*) downregulation is involved in several types of cancer acting as a potential novel tumour suppressor gene (263). Our results showed hypomethylation in a *PCDH10* gene-regulation region and correlates with higher expression of this gene in BCVY, as also validated by qRT-PCR**.** While *PCDH10* expression has been related to inhibition of invasion and metastasis, it has been suggested that it does so in combination with other proteins such as p53, which control apoptosis and cell growth and Shi and colleagues (264) showed that the overexpression of *PCDH10 per se* does not have a significant impact on cell growth.

Further, we found that hypermethylation in BCVY modulates characteristic pathways related with immune system, DNA repair, vesicular trafficking and biogenesis as well as with Notch/Notch1 signalling, and probably has a role in the aggressive pathophysiology of the disease. Of the 11 genes selected from the distinctive BCVY for validation step, only *HDAC5* showed results coherent with TCGA gene expression. Despite the hypermethylated profile, *HDAC5* CpG regulatory regions, and specially enhancer sites, were hypomethylated and was coincident with overexpression validated in BCVY.

Our observations suggest that the methylation pattern seen in BCVY differs to that observed in BCO. BCVY presents a global hypomethylation profile that can be validated with TCGA data, and a specific hypermethylation signature seen in metEPICVal samples, mainly present in open-sea regions, that could play an important role in the biological complexity of BCVY tumours. Hypomethylation signature found in BCVY may be contributing to the increased DNAm age-acceleration detected. We propose that BCVY aggressiveness could be potentiated by genome instability triggered by global hypomethylation found. Moreover, hormonal factors affecting young women could be responsible of the accelerated DNAmAge and may be associated with

the poor prognosis of BCVY, in which hormone exposure are high and more important than in BCO. Therefore, the differences observed in BCVY in our work and others, strengthen the hypothesis that BC arising in young women has a potentially unique, aggressive and complex biological features and more efforts should be done to reach a more personalized treatment.

2.5 *HDAC5* as a potential therapeutic target in breast cancer

Histones are acetylated on the side chains of lysine residues, which neutralizes lysine's positive charge, leading to open chromatin structure by reducing interaction between histone and negatively charged DNA (265, 266). Thus, histone acetylation increases the accessibility of proteins, such as transcription factors, to promoters and enhancers, thereby mediating active gene expression. Acetylated histones also function as binding sites for numerous proteins with bromodomains, which often activates gene transcription (265). In contrast, deacetylation of histone is associated with chromatin condensation and transcriptional repression (265). Analogous to histone methylation, histone acetylation is reversibly controlled by two large families of enzymes: histone lysine acetyltransferases (KATs) and HDACs.

As for *HDAC5*, high expression has been correlated with poorer prognosis in patients with BC or pancreatic neuroendocrine tumours and has been attributed with oncogenic effects (267, 268). Our study showed higher *HDAC5* expression in BCVY compared to BCO samples. HDAC inhibitors are the most extensively investigated epigenetic drugs in clinical studies (219). *In vitro* studies suggest that the HDAC5 inhibitor, LMK-235, could be a novel therapeutic strategy for treating BC (207). We evaluate the inhibition effect of LMK-235 in breast cancer cell lines from young women and compared it with breast cancer cell lines from older patients. Our results demonstrated important reduction in the progression, not only in breast cancer cell lines from young women but also in all triple negative subtypes, that were age independent, and that has

been previously observed at different studies (269-272). There is a limitation in the number of breast cancer cell lines from women younger than 35 years' old that are commercially available. Our study only includes two breast cancer cell lines from young women and both were from triple negative subtype, impeding the possibility of analyse their role in different molecular subtypes from BCVY patients. We propose the study of HDAC5 inhibitors, like LMK-235, in a larger cohort of breast cancer cell lines from young women and from different molecular subtypes in order to evaluate their inhibition effect. Due to the higher expression of *HDAC5* in BCVY patients, we should expect higher reduction in the proliferation in this age group in which the HDACi inhibitors could have a potential therapeutic role.

3. DNA methylation of gene encoding miRNA in breast cancer affecting young women

Several miRNAs have been involved in cancer pathogenesis (273). Accordingly, altered miRNA expression profiles have been found in every type of human cancer including colon, brain, lung and breast cancer (274, 275) suggesting miRNAs as possible biomarkers for early cancer detection. Furthermore, miRNAs not only function as part of the epigenetic machinery, but are also epigenetically modified by DNA methylation themselves. DNA methylation plays a key role in silencing numerous cancer-related genes affecting several processes, and there is considerable evidence supporting the idea that DNA methylation is actively involved in the dysregulation of miRNAs in cancer (276).

Breast cancer affecting young women has been previously associated with singular molecular biology that could be responsible for the poor prognosis that characterizes these tumours. The previous results in the study of DNA methylation in BCVY have shown differing methylation signatures for breast cancer in young women than in older patients. Although the mechanism

underlying miRNA dysregulation in cancer is not yet fully understood, recent studies have shown that epigenetic mechanisms play an important role in regulating miRNA expression (277).

3.1 Differences in DNA methylation of miRNA encoding genes between breast cancer samples in young and old women

In this section, we analysed the differences in DNA methylation of miRNA encoding genes between breast cancer samples in young and old women. Results from the methylation study in metEPICVal samples performed in the Illumina Infinium MethylationEPIC BeadChip showed 193 CpG sites that were significantly differentially methylated between groups. Most highlighted methylation differences were localized in open sea and island CpG regions. Specifically, open sea sites were hypermethylated and islands and regions near them were hypomethylated in BCVY. These results are in agreement with our previous global methylation observation in BCVY patients, that found hypomethylation localized in CpG island regions and hypermethylation in open sea and regions away from the islands. Relative to that, it has been seen that several gene-encoding miRNAs are more frequently hypermethylated and consequently repressed in regions away from the islands than the CpG island itself (276), agreeing with our observations.

In terms of regulatory localization, most significantly hypomethylated probes in BCVY were located on DNasa I hypersensitive sites (DHSs). These regions of chromatin are sensitive to cleavage by DNase I enzyme, and chromatin has lost its condensed structure, making the DNA more accessible. This remodelled state is essential to increased transcriptional activity. DHSs therefore tended to be enriched on highly expressed genes throughout whole gene regions while not showing significant changes for low and silently expressed genes. Also, DHSs are enriched in regions away from CpG islands, suggesting preference to act within active chromatin domains that present low density CpG islands (278). In contrast to the previously mentioned

hypermethylation observed in regions away from islands, DHS regions present hypomethylation in BCVY, contributing to the more accessible DNA and consequently upregulating their expression.

3.2 Methylation validation of gene encoding miRNAs

In order to validate the results found in the metEPICVal sample, we used data from two previous methylation analysis (TCGA methylation and combined data) performed with HM450K array, with the limitation that we could only analyse 85 CpG probes that were shared between EPICarray and HM450K. The methylation validation study identified 10 CpG significant probes, all of them hypomethylated in young women with breast cancer. These results are in agreement with HM450K array content, given that the number of CpG islands probes is higher than open sea probes, the latter poorly presented in the HM450K array. Based on this we were able to validate some of the hypomethylation probes for BCVY localized in islands regions or near them, using the HM450K validation data sets. However, hypomethylation was concentrated mainly in open sea regions, which were underrepresented in HM450K data sets and could not be validated.

Advances in microarray and sequencing technologies have enabled comprehensive analysis of the epigenome and miRNA expression in cancer cells, which has led to the identification of miRNA which are frequent targets of aberrant DNA methylation in cancer (279). In the present study we were able to validate significant methylation differences in a set of differentially methylated miRNAs in cancer such as: miR-181c, miR-129-2, miR-196a-1, miR-137, miR-9-1 and miR-124, of which methylation differences were observed in miR-124-2 and miR-124-3.

3.3 MiRNAs differentially methylated and expressed in BCVY and their roles in cancer

We explored the expression of miRNAs regulated by differentially methylated probes in BCVY taking into account their genomic/functional category. Some significant miRNAs were regulated by probes situated in more than one genomic/functional category. Differentially deregulated MiR-196a-1 and miR-212 in BCVY were in turn regulated by differentially methylated regions located in islands and DHS sites. Another similar case is miR-211 and miR-184, which were regulated by probes situated in open sea and DHS sites and were deregulated in BCVY. These results suggest an important regulatory role for DNA methylation in the miRNA-encoding genes which are significantly deregulated in BCVY vs. BCO in CpG from different regions.

Next, meta-analysis of expression in miRNAs regulated by validated regulatory CpG probes and new significant probes included in the EPICarray, revealed 4 miRNAs (miR-9-1, miR-184, miR-551b and miR-196a-1) that were differentially expressed and methylated in BCVY. Although miR-124-2 expression could not be evaluated in the meta-analysis expression study, its hypomethylation was validated in both methylation data sets and qRT-PCR revealed significant miR-124-2 upregulation in BCVY. However, none of the four miRNAs identified (miR-9-1, miR-184, miR-551b and miR-196a-1) could be validated by qRT-PCR.

Enrichment analysis showed that the significant differentially methylated miRNAs have functions linked to cancer. Most of them were implicated in pathways related with adherent junctions, important in cancer initiation and progression. Furthermore, some miRNAs were involved in extracellular matrix organization (ECM) through proteoglycans (hsa05205) and ECM-receptor interaction (hsa04512).

3.4 Clinical relevance of miR-124-2 in BCVY patients

MiRNA-methylation has been extensively investigated as a prognostic factor, and survival analysis performed with our data and TCGA revealed that hypermethylation of miR-124-2 was significantly associated with poorer OS in BCO, in contrast to BCVY, in which miR-124-2-hypomethylation presented the lowest survival rates. These results suggest that hypomethylation of miR-124-2 may be a potential prognostic risk factor specific to BCVY. Although no significant results were reached for the rest miRNAs, our study suggested that miR-184-hypomethylation could be a good prognostic factor in breast cancer patients, in which methylation loss increases survival and reduces relapse rates. Conversely, miR-9-1 and miR-196a-1 hypomethylation were associated with reduced survival and higher relapse rates.

Within the human genome, three independent loci (miR-124-1, miR-124-2 and miR-124-3) encode the identical mature miR-124, and all are associated with CpG islands, which have been described as targets of hypermethylation in colon, stomach, liver, leukaemia and cervix cancers (280). MiR-124 exerts a tumour suppressor effect by targeting cyclin-dependent kinase 6 (CDK6), and epigenetic silencing of miR-124 leads to CDK6 activation and Rb phosphorylation (280, 281). Interestingly, our results show a hypomethylation of miR-124-2 gene in BCVY compared with older patients and a higher miRNA expression. Additionally, the miR-124-2 has been found to be the most abundant miRNA expressed in neuronal cells and their differentiation. Neuronal elements have been described that are related to cancer microenvironment promoting cell growth, although the mechanisms mediating neuronal influences on cancer growth and progression are likely incompletely understood (244). Results of the present thesis have shown that most of the pathways deregulated in BCVY were related to neuronal-system processes However, more work is needed to gain insight into the miRNA-124-2-hypomethylation role in

breast cancer affecting young women and its potential role as a biomarker of poor survival in BCVY.

Several limitations of this section need to be acknowledged. In the case of miRNA expression meta-analysis, the three datasets employed for the study do not include all significantly differentially methylated miRNAs. Thus, the miRNAs that were significantly deregulated in some but not all of the three studies were not included in the meta-analysis and information about expression status could not be evaluated; despite obtaining the methylation status of significant CpG regions for BCVY, we could not translate them into differences in gene expression.

Results of methylation from genes encoding miRNAs are already published (282) and the article is included in Annex XIV.

CONCLUSIONS

CONCLUSIONS

1. The similarity in the miRNA expression between breast cancer cell lines and tumour samples from BC patients demonstrates the suitability of HCC1500 and HCC1937 as cellular models for future studies of breast cancer affecting very young women. Furthermore, the miR-23a-dowregulation has been detected and validated in breast cancer cell lines and tumour samples from young women and we observe an association between miR-23a-downregulation and poor survival rates in very young women with breast cancer.

2. Our observations show that the methylation pattern seen in breast cancer of young women differs to that observed in older patients. Breast cancer of young women presents a global hypomethylation profile, and a specific hypermethylation signature, mainly present in open-sea regions.

3. We detect a significant increased DNA methylation Age acceleration in breast cancer of young women tumours. This acceleration is disease specific and not tissue related. Additionally, breast cancer in very young women that are oestrogen receptor positive presents a significant higher epigenetic age. This effect suggests that the acceleration found in breast cancer affecting young women may be partially influenced by oestrogen receptor status, but other factors such as hormonal levels could be involved in the increased DNA methylation age and contributing with their poor outcome.

4. Gene enrichment analysis points up that the global hypomethylation profile found in breast cancer affecting young women is significantly overrepresented in genes involved in neuronal-system processes, as well as, in extracellular matrix organisation and cell communication. Otherwise, hypermethylation signature observed in breast cancer in very young women modulates characteristic pathways related with immune system, DNA repair, vesicular trafficking and biogenesis as well as with Notch/Notch1 signalling.

5. Global hypomethylation in breast cancer of young women has been validated both in the combined study and TCGA data set. However, the distinctive hypermethylated profile for very young women could not be validated due to the low proportion of open-sea probes in the Human Methylation 450K array. We conclude that the validation study detects significant expression differences between breast cancer age groups for *HDAC5*, *HOXD9* and *PCDH10* genes.

6. *HDAC5* gene is significantly hypomethylated and therefore overexpressed in breast cancer tumours from young patients. The LMK-235 iHDAC inhibitor shows a significant proliferation reduction not only in breast cancer cell lines from young patients but also in triple negative subtypes. Additional molecular subtypes from young patients could help to elucidate the role of HDAC inhibitors in breast cancer affecting young women.

7. We identify a different methylation pattern for genes encoding miRNAs in breast cancer of young women tumours in comparison to older women. Specifically, open sea sites are hypermethylated and islands and regions near them are hypomethylated in breast cancer affecting very young women.

8. Our results show a significant hypomethylation of CpG sites regulating miR-124-2 expression in breast cancer affecting very young women. We observe a relationship between miR-124-2-hypomethylation with significantly better survival rates for older breast cancer patients as opposed to the worse prognosis observed in young breast cancer women, identifying it as a potential specific survival biomarker in this last group of age.

REFERENCES

REFERENCES

1. Helgason CM. Commentary on the significance for modern neurology of the 17th century B.C. Surgical Papyrus. The Canadian journal of neurological sciences Le journal canadien des sciences neurologiques. 1987;14(4):560-3.
2. Lukong KE. Understanding breast cancer - The long and winding road. BBA clinical. 2017;7:64-77.
3. Watson CJ, Khaled WT. Mammary development in the embryo and adult: a journey of morphogenesis and commitment. Development. 2008;135(6):995-1003.
4. Radisky DC, Hartmann LC. Mammary involution and breast cancer risk: transgenic models and clinical studies. Journal of mammary gland biology and neoplasia. 2009;14(2):181-91.
5. Sternlicht MD. Key stages in mammary gland development: the cues that regulate ductal branching morphogenesis. Breast cancer research : BCR. 2006;8(1):201.
6. Schedin P, O'Brien J, Rudolph M, Stein T, Borges V. Microenvironment of the involuting mammary gland mediates mammary cancer progression. Journal of mammary gland biology and neoplasia. 2007;12(1):71-82.
7. Baxter FO, Neoh K, Tevendale MC. The beginning of the end: death signaling in early involution. Journal of mammary gland biology and neoplasia. 2007;12(1):3-13.
8. Lund VE, Makinen JI. Ectopic pregnancy and seasonality. Human reproduction. 1996;11(3):683-4.
9. Ferlay J SI, Ervik M, Dikshit R, Eser S, Mathers C. GLOBOCAN 2012 v1.0, Cancer Incidence and Mortality Worldwide: IARC CancerBase No. 11 2015. Available from: http://globocan.iarc.fr.
10. Torre LA, Siegel RL, Ward EM, Jemal A. Global Cancer Incidence and Mortality Rates and Trends--An Update. Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology. 2016;25(1):16-27.
11. Cronin KA, Lake AJ, Scott S, Sherman RL, Noone AM, Howlader N, et al. Annual Report to the Nation on the Status of Cancer, part I: National cancer statistics. Cancer. 2018;124(13):2785-800.
12. Lopez-Abente G, Aragones N, Perez-Gomez B, Pollan M, Garcia-Perez J, Ramis R, et al. Time trends in municipal distribution patterns of cancer mortality in Spain. BMC cancer. 2014;14:535.
13. Pollan M, Ramis R, Aragones N, Perez-Gomez B, Gomez D, Lope V, et al. Municipal distribution of breast cancer mortality among women in Spain. BMC cancer. 2007;7:78.
14. Jemal A, Tiwari RC, Murray T, Ghafoor A, Samuels A, Ward E, et al. Cancer statistics, 2004. CA: a cancer journal for clinicians. 2004;54(1):8-29.
15. La Vecchia C. Oral contraceptives, menopause hormone replacement therapy, and risk of stroke. Maturitas. 2004;47(4):265-8.
16. Stefanick ML, Anderson GL, Margolis KL, Hendrix SL, Rodabough RJ, Paskett ED, et al. Effects of conjugated equine estrogens on breast cancer and mammography screening in postmenopausal women with hysterectomy. Jama. 2006;295(14):1647-57.
17. La Vecchia C, Altieri A, Franceschi S, Tavani A. Oral contraceptives and cancer: an update. Drug safety. 2001;24(10):741-54.
18. Colditz GA, Frazier AL. Models of breast cancer show that risk is set by events of early life: prevention efforts must shift focus. Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology. 1995;4(5):567-71.

19. Liu Y, Colditz GA, Rosner B, Berkey CS, Collins LC, Schnitt SJ, et al. Alcohol intake between menarche and first pregnancy: a prospective study of breast cancer risk. Journal of the National Cancer Institute. 2013;105(20):1571-8.
20. McDaniel SM, Rumer KK, Biroc SL, Metz RP, Singh M, Porter W, et al. Remodeling of the mammary microenvironment after lactation promotes breast tumor cell metastasis. The American journal of pathology. 2006;168(2):608-20.
21. Schedin P. Pregnancy-associated breast cancer and metastasis. Nature reviews Cancer. 2006;6(4):281-91.
22. Collaborative Group on Hormonal Factors in Breast C. Breast cancer and breastfeeding: collaborative reanalysis of individual data from 47 epidemiological studies in 30 countries, including 50302 women with breast cancer and 96973 women without the disease. Lancet. 2002;360(9328):187-95.
23. Dzhandzhalia MT. [Risk of the development of breast cancer during gynecological pathology]. Georgian medical news. 2006(137):32-4.
24. Vogel VG. Epidemiology, genetics, and risk evaluation of postmenopausal women at risk of breast cancer. Menopause. 2008;15(4 Suppl):782-9.
25. Pollan M, Pastor-Barriuso R, Ardanaz E, Arguelles M, Martos C, Galceran J, et al. Recent changes in breast cancer incidence in Spain, 1980-2004. Journal of the National Cancer Institute. 2009;101(22):1584-91.
26. Chen WY, Rosner B, Hankinson SE, Colditz GA, Willett WC. Moderate alcohol consumption during adult life, drinking patterns, and breast cancer risk. Jama. 2011;306(17):1884-90.
27. Buckland G, Travier N, Cottet V, Gonzalez CA, Lujan-Barroso L, Agudo A, et al. Adherence to the mediterranean diet and risk of breast cancer in the European prospective investigation into cancer and nutrition cohort study. International journal of cancer. 2013;132(12):2918-27.
28. Engeset D, Dyachenko A, Ciampi A, Lund E. Dietary patterns and risk of cancer of various sites in the Norwegian European Prospective Investigation into Cancer and Nutrition cohort: the Norwegian Women and Cancer study. European journal of cancer prevention : the official journal of the European Cancer Prevention Organisation. 2009;18(1):69-75.
29. Fung TT, Hu FB, Hankinson SE, Willett WC, Holmes MD. Low-carbohydrate diets, dietary approaches to stop hypertension-style diets, and the risk of postmenopausal breast cancer. American journal of epidemiology. 2011;174(6):652-60.
30. Key TJ, Appleby PN, Spencer EA, Travis RC, Roddam AW, Allen NE. Cancer incidence in vegetarians: results from the European Prospective Investigation into Cancer and Nutrition (EPIC-Oxford). The American journal of clinical nutrition. 2009;89(5):1620S-6S.
31. Mai V, Kant AK, Flood A, Lacey JV, Jr., Schairer C, Schatzkin A. Diet quality and subsequent cancer incidence and mortality in a prospective cohort of women. International journal of epidemiology. 2005;34(1):54-60.
32. Potter J, Brown L, Williams RL, Byles J, Collins CE. Diet Quality and Cancer Outcomes in Adults: A Systematic Review of Epidemiological Studies. International journal of molecular sciences. 2016;17(7).
33. Bakker MF, Peeters PH, Klaasen VM, Bueno-de-Mesquita HB, Jansen EH, Ros MM, et al. Plasma carotenoids, vitamin C, tocopherols, and retinol and the risk of breast cancer in the European Prospective Investigation into Cancer and Nutrition cohort. The American journal of clinical nutrition. 2016;103(2):454-64.
34. Eliassen AH, Hendrickson SJ, Brinton LA, Buring JE, Campos H, Dai Q, et al. Circulating carotenoids and risk of breast cancer: pooled analysis of eight prospective studies. Journal of the National Cancer Institute. 2012;104(24):1905-16.
35. Eliassen AH, Liao X, Rosner B, Tamimi RM, Tworoger SS, Hankinson SE. Plasma carotenoids and risk of breast cancer over 20 y of follow-up. The American journal of clinical nutrition. 2015;101(6):1197-205.

36. Emaus MJ, Peeters PH, Bakker MF, Overvad K, Tjonneland A, Olsen A, et al. Vegetable and fruit consumption and the risk of hormone receptor-defined breast cancer in the EPIC cohort. The American journal of clinical nutrition. 2016;103(1):168-77.
37. Jung S, Spiegelman D, Baglietto L, Bernstein L, Boggs DA, van den Brandt PA, et al. Fruit and vegetable intake and risk of breast cancer by hormone receptor status. Journal of the National Cancer Institute. 2013;105(3):219-36.
38. Toledo E, Salas-Salvado J, Donat-Vargas C, Buil-Cosiales P, Estruch R, Ros E, et al. Mediterranean Diet and Invasive Breast Cancer Risk Among Women at High Cardiovascular Risk in the PREDIMED Trial: A Randomized Clinical Trial. JAMA internal medicine. 2015;175(11):1752-60.
39. Arnold M, Jiang L, Stefanick ML, Johnson KC, Lane DS, LeBlanc ES, et al. Duration of Adulthood Overweight, Obesity, and Cancer Risk in the Women's Health Initiative: A Longitudinal Study from the United States. PLoS medicine. 2016;13(8):e1002081.
40. Neuhouser ML, Aragaki AK, Prentice RL, Manson JE, Chlebowski R, Carty CL, et al. Overweight, Obesity, and Postmenopausal Invasive Breast Cancer Risk: A Secondary Analysis of the Women's Health Initiative Randomized Clinical Trials. JAMA oncology. 2015;1(5):611-21.
41. Berkey CS, Gardner JD, Frazier AL, Colditz GA. Relation of childhood diet and body size to menarche and adolescent growth in girls. American journal of epidemiology. 2000;152(5):446-52.
42. Berkey CS, Willett WC, Frazier AL, Rosner B, Tamimi RM, Rockett HR, et al. Prospective study of adolescent alcohol consumption and risk of benign breast disease in young women. Pediatrics. 2010;125(5):e1081-7.
43. Colditz GA, Bohlke K, Berkey CS. Breast cancer risk accumulation starts early: prevention must also. Breast cancer research and treatment. 2014;145(3):567-79.
44. Gunther AL, Karaolis-Danckert N, Kroke A, Remer T, Buyken AE. Dietary protein intake throughout childhood is associated with the timing of puberty. The Journal of nutrition. 2010;140(3):565-71.
45. Rogers IS, Northstone K, Dunger DB, Cooper AR, Ness AR, Emmett PM. Diet throughout childhood and age at menarche in a contemporary cohort of British girls. Public health nutrition. 2010;13(12):2052-63.
46. Dougan MM, Hankinson SE, Vivo ID, Tworoger SS, Glynn RJ, Michels KB. Prospective study of body size throughout the life-course and the incidence of endometrial cancer among premenopausal and postmenopausal women. International journal of cancer. 2015;137(3):625-37.
47. Kantor ED, Udumyan R, Signorello LB, Giovannucci EL, Montgomery S, Fall K. Adolescent body mass index and erythrocyte sedimentation rate in relation to colorectal cancer risk. Gut. 2016;65(8):1289-95.
48. Genkinger JM, Kitahara CM, Bernstein L, Berrington de Gonzalez A, Brotzman M, Elena JW, et al. Central adiposity, obesity during early adulthood, and pancreatic cancer mortality in a pooled analysis of cohort studies. Annals of oncology : official journal of the European Society for Medical Oncology. 2015;26(11):2257-66.
49. Bianchini F, Kaaks R, Vainio H. Weight control and physical activity in cancer prevention. Obesity reviews : an official journal of the International Association for the Study of Obesity. 2002;3(1):5-8.
50. Friedenreich CM, Orenstein MR. Physical activity and cancer prevention: etiologic evidence and biological mechanisms. The Journal of nutrition. 2002;132(11 Suppl):3456S-64S.
51. Lahmann PH, Friedenreich C, Schuit AJ, Salvini S, Allen NE, Key TJ, et al. Physical activity and breast cancer risk: the European Prospective Investigation into Cancer and Nutrition. Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology. 2007;16(1):36-42.

52. Danaei G, Vander Hoorn S, Lopez AD, Murray CJ, Ezzati M, Comparative Risk Assessment collaborating g. Causes of cancer in the world: comparative risk assessment of nine behavioural and environmental risk factors. Lancet. 2005;366(9499):1784-93.
53. BL PB. World cancer report. Cancer IAfRo. International Agency for Research on Cancer (IARC). 2008.
54. Damiola F, Pertesi M, Oliver J, Le Calvez-Kelm F, Voegele C, Young EL, et al. Rare key functional domain missense substitutions in MRE11A, RAD50, and NBN contribute to breast cancer susceptibility: results from a Breast Cancer Family Registry case-control mutation-screening study. Breast cancer research : BCR. 2014;16(3):R58.
55. Heikkinen K, Rapakko K, Karppinen SM, Erkko H, Knuutila S, Lundan T, et al. RAD50 and NBS1 are breast cancer susceptibility genes associated with genomic instability. Carcinogenesis. 2006;27(8):1593-9.
56. Litman R, Peng M, Jin Z, Zhang F, Zhang J, Powell S, et al. BACH1 is critical for homologous recombination and appears to be the Fanconi anemia gene product FANCJ. Cancer cell. 2005;8(3):255-65.
57. Rauschecker HF, Sauerbrei W, Gatzemeier W, Sauer R, Schauer A, Schmoor C, et al. Eight-year results of a prospective non-randomised study on therapy of small breast cancer. The German Breast Cancer Study Group (GBSG). European journal of cancer. 1998;34(3):315-23.
58. Fisher B, Redmond C, Fisher ER, Caplan R. Relative worth of estrogen or progesterone receptor and pathologic characteristics of differentiation as indicators of prognosis in node negative breast cancer patients: findings from National Surgical Adjuvant Breast and Bowel Project Protocol B-06. Journal of clinical oncology : official journal of the American Society of Clinical Oncology. 1988;6(7):1076-87.
59. Cornejo KM, Kandil D, Khan A, Cosar EF. Theranostic and molecular classification of breast cancer. Archives of pathology & laboratory medicine. 2014;138(1):44-56.
60. Perou CM, Sorlie T, Eisen MB, van de Rijn M, Jeffrey SS, Rees CA, et al. Molecular portraits of human breast tumours. Nature. 2000;406(6797):747-52.
61. Vargo-Gogola T, Rosen JM. Modelling breast cancer: one size does not fit all. Nature reviews Cancer. 2007;7(9):659-72.
62. Galanina N, Bossuyt V, Harris LN. Molecular predictors of response to therapy for breast cancer. Cancer journal. 2011;17(2):96-103.
63. Mehta R, Jain RK, Badve S. Personalized medicine: the road ahead. Clinical breast cancer. 2011;11(1):20-6.
64. Nielsen TO, Parker JS, Leung S, Voduc D, Ebbert M, Vickery T, et al. A comparison of PAM50 intrinsic subtyping with immunohistochemistry and clinical prognostic factors in tamoxifen-treated estrogen receptor-positive breast cancer. Clinical cancer research : an official journal of the American Association for Cancer Research. 2010;16(21):5222-32.
65. Lehmann BD, Bauer JA, Chen X, Sanders ME, Chakravarthy AB, Shyr Y, et al. Identification of human triple-negative breast cancer subtypes and preclinical models for selection of targeted therapies. The Journal of clinical investigation. 2011;121(7):2750-67.
66. Breast Cancer Treatment (PDQ(R)): Patient Version. PDQ Cancer Information Summaries. Bethesda (MD)2002.
67. Jatoi I. Breast cancer screening: past, present and future. Future oncology. 2015.
68. Zelnak AB, O'Regan RM. Optimizing Endocrine Therapy for Breast Cancer. Journal of the National Comprehensive Cancer Network : JNCCN. 2015;13(8):e56-64.
69. Baselga J, Mendelsohn J. Type I receptor tyrosine kinases as targets for therapy in breast cancer. Journal of mammary gland biology and neoplasia. 1997;2(2):165-74.
70. Baselga J, Seidman AD, Rosen PP, Norton L. HER2 overexpression and paclitaxel sensitivity in breast cancer: therapeutic implications. Oncology. 1997;11(3 Suppl 2):43-8.
71. Gabriel CA, Domchek SM. Breast cancer in young women. Breast cancer research : BCR. 2010;12(5):212.

72. Lee HB, Han W. Unique features of young age breast cancer and its management. Journal of breast cancer. 2014;17(4):301-7.
73. Cardoso F, Loibl S, Pagani O, Graziottin A, Panizza P, Martincich L, et al. The European Society of Breast Cancer Specialists recommendations for the management of young women with breast cancer. European journal of cancer. 2012;48(18):3355-77.
74. Anderson WF, Jatoi I, Sherman ME. Qualitative age interactions in breast cancer studies: mind the gap. Journal of clinical oncology : official journal of the American Society of Clinical Oncology. 2009;27(32):5308-11.
75. de la Rochefordiere A, Asselain B, Campana F, Scholl SM, Fenton J, Vilcoq JR, et al. Age as prognostic factor in premenopausal breast carcinoma. Lancet. 1993;341(8852):1039-43.
76. Han W, Kang SY, Korean Breast Cancer S. Relationship between age at diagnosis and outcome of premenopausal breast cancer: age less than 35 years is a reasonable cut-off for defining young age-onset breast cancer. Breast cancer research and treatment. 2010;119(1):193-200.
77. Cancello G, Maisonneuve P, Rotmensz N, Viale G, Mastropasqua MG, Pruneri G, et al. Prognosis and adjuvant treatment effects in selected breast cancer subtypes of very young women (<35 years) with operable breast cancer. Annals of oncology : official journal of the European Society for Medical Oncology. 2010;21(10):1974-81.
78. Fredholm H, Eaker S, Frisell J, Holmberg L, Fredriksson I, Lindman H. Breast cancer in young women: poor survival despite intensive treatment. PloS one. 2009;4(11):e7695.
79. Gnerlich JL, Deshpande AD, Jeffe DB, Sweet A, White N, Margenthaler JA. Elevated breast cancer mortality in women younger than age 40 years compared with older women is attributed to poorer survival in early-stage disease. Journal of the American College of Surgeons. 2009;208(3):341-7.
80. Azim HA, Jr., Michiels S, Bedard PL, Singhal SK, Criscitiello C, Ignatiadis M, et al. Elucidating prognosis and biology of breast cancer arising in young women using gene expression profiling. Clinical cancer research : an official journal of the American Association for Cancer Research. 2012;18(5):1341-51.
81. Anders CK, Hsu DS, Broadwater G, Acharya CR, Foekens JA, Zhang Y, et al. Young age at diagnosis correlates with worse prognosis and defines a subset of breast cancers with shared patterns of gene expression. Journal of clinical oncology : official journal of the American Society of Clinical Oncology. 2008;26(20):3324-30.
82. Jankowska M. Sexual functioning in young women in the context of breast cancer treatment. Reports of practical oncology and radiotherapy : journal of Greatpoland Cancer Center in Poznan and Polish Society of Radiation Oncology. 2013;18(4):193-200.
83. Kroman N, Jensen MB, Wohlfahrt J, Mouridsen HT, Andersen PK, Melbye M. Factors influencing the effect of age on prognosis in breast cancer: population based study. Bmj. 2000;320(7233):474-8.
84. Ademuyiwa FO, Cyr A, Ivanovich J, Thomas MA. Managing breast cancer in younger women: challenges and solutions. Breast cancer. 2016;8:1-12.
85. Ribnikar D, Ribeiro JM, Pinto D, Sousa B, Pinto AC, Gomes E, et al. Breast cancer under age 40: a different approach. Current treatment options in oncology. 2015;16(4):16.
86. Crouse HV. Translocations in Sciara: their bearing on chromosome behavior and sex determination: Univ. Missouri Res. Bull; 1943.
87. Waddington CH. Selection of the genetic basis for an acquired character. Nature. 1952;169(4302):625-6.
88. Riggs AD. X inactivation, differentiation, and DNA methylation. Cytogenetics and cell genetics. 1975;14(1):9-25.
89. Holliday R, Pugh JE. DNA modification mechanisms and gene activity during development. Science. 1975;187(4173):226-32.

90. Frommer M, McDonald LE, Millar DS, Collis CM, Watt F, Grigg GW, et al. A genomic sequencing protocol that yields a positive display of 5-methylcytosine residues in individual DNA strands. Proceedings of the National Academy of Sciences of the United States of America. 1992;89(5):1827-31.
91. Herman JG, Graff JR, Myohanen S, Nelkin BD, Baylin SB. Methylation-specific PCR: a novel PCR assay for methylation status of CpG islands. Proceedings of the National Academy of Sciences of the United States of America. 1996;93(18):9821-6.
92. Heyn H, Vidal E, Sayols S, Sanchez-Mut JV, Moran S, Medina I, et al. Whole-genome bisulfite DNA sequencing of a DNMT3B mutant patient. Epigenetics. 2012;7(6):542-50.
93. Zhang X, Yazaki J, Sundaresan A, Cokus S, Chan SW, Chen H, et al. Genome-wide high-resolution mapping and functional analysis of DNA methylation in arabidopsis. Cell. 2006;126(6):1189-201.
94. American Association for Cancer Research Human Epigenome Task F, European Union NoESAB. Moving AHEAD with an international human epigenome project. Nature. 2008;454(7205):711-5.
95. Berger SL, Kouzarides T, Shiekhattar R, Shilatifard A. An operational definition of epigenetics. Genes & development. 2009;23(7):781-3.
96. Đerke F. Psychoneuroimmunology and Epigenetics2015. 54-7 p.
97. Esteller M. Epigenetics in cancer. The New England journal of medicine. 2008;358(11):1148-59.
98. Klose RJ, Bird AP. Genomic DNA methylation: the mark and its mediators. Trends in biochemical sciences. 2006;31(2):89-97.
99. Aapola U, Liiv I, Peterson P. Imprinting regulator DNMT3L is a transcriptional repressor associated with histone deacetylase activity. Nucleic acids research. 2002;30(16):3602-8.
100. Deplus R, Brenner C, Burgers WA, Putmans P, Kouzarides T, de Launoit Y, et al. Dnmt3L is a transcriptional repressor that recruits histone deacetylase. Nucleic acids research. 2002;30(17):3831-8.
101. Reik W, Dean W, Walter J. Epigenetic reprogramming in mammalian development. Science. 2001;293(5532):1089-93.
102. Bell O, Tiwari VK, Thoma NH, Schubeler D. Determinants and dynamics of genome accessibility. Nature reviews Genetics. 2011;12(8):554-64.
103. Elango N, Yi SV. DNA methylation and structural and functional bimodality of vertebrate promoters. Molecular biology and evolution. 2008;25(8):1602-8.
104. Fernandez AF, Assenov Y, Martin-Subero JI, Balint B, Siebert R, Taniguchi H, et al. A DNA methylation fingerprint of 1628 human samples. Genome research. 2012;22(2):407-19.
105. Ball MP, Li JB, Gao Y, Lee JH, LeProust EM, Park IH, et al. Targeted and genome-scale strategies reveal gene-body methylation signatures in human cells. Nature biotechnology. 2009;27(4):361-8.
106. Feng S, Cokus SJ, Zhang X, Chen PY, Bostick M, Goll MG, et al. Conservation and divergence of methylation patterning in plants and animals. Proceedings of the National Academy of Sciences of the United States of America. 2010;107(19):8689-94.
107. Rauch TA, Zhong X, Wu X, Wang M, Kernstine KH, Wang Z, et al. High-resolution mapping of DNA hypermethylation and hypomethylation in lung cancer. Proceedings of the National Academy of Sciences of the United States of America. 2008;105(1):252-7.
108. Jjingo D, Conley AB, Yi SV, Lunyak VV, Jordan IK. On the presence and role of human gene-body DNA methylation. Oncotarget. 2012;3(4):462-74.
109. Jones PA. The DNA methylation paradox. Trends in genetics : TIG. 1999;15(1):34-7.
110. Straussman R, Nejman D, Roberts D, Steinfeld I, Blum B, Benvenisty N, et al. Developmental programming of CpG island methylation profiles in the human genome. Nature structural & molecular biology. 2009;16(5):564-71.

111. Kulis M, Heath S, Bibikova M, Queiros AC, Navarro A, Clot G, et al. Epigenomic analysis detects widespread gene-body DNA hypomethylation in chronic lymphocytic leukemia. Nature genetics. 2012;44(11):1236-42.
112. Yao HW, Li J. Epigenetic modifications in fibrotic diseases: implications for pathogenesis and pharmacological targets. J Pharmacol Exp Ther. 2015;352(1):2-13.
113. He L, Hannon GJ. MicroRNAs: small RNAs with a big role in gene regulation. Nature reviews Genetics. 2004;5(7):522-31.
114. Esquela-Kerscher A, Slack FJ. Oncomirs - microRNAs with a role in cancer. Nature reviews Cancer. 2006;6(4):259-69.
115. Bartel DP. MicroRNAs: target recognition and regulatory functions. Cell. 2009;136(2):215-33.
116. Esteller M. Non-coding RNAs in human disease. Nature reviews Genetics. 2011;12(12):861-74.
117. Lujambio A, Lowe SW. The microcosmos of cancer. Nature. 2012;482(7385):347-55.
118. Li L, Liu Y. Diverse small non-coding RNAs in RNA interference pathways. Methods in molecular biology. 2011;764:169-82.
119. Feinberg AP, Vogelstein B. Hypomethylation distinguishes genes of some human cancers from their normal counterparts. Nature. 1983;301(5895):89-92.
120. Davalos V, Martinez-Cardus A, Esteller M. The Epigenomic Revolution in Breast Cancer: From Single-Gene to Genome-Wide Next-Generation Approaches. The American journal of pathology. 2017;187(10):2163-74.
121. Widschwendter M, Siegmund KD, Muller HM, Fiegl H, Marth C, Muller-Holzner E, et al. Association of breast cancer DNA methylation profiles with hormone receptor status and response to tamoxifen. Cancer research. 2004;64(11):3807-13.
122. Martinez-Galan J, Torres-Torres B, Nunez MI, Lopez-Penalver J, Del Moral R, Ruiz De Almodovar JM, et al. ESR1 gene promoter region methylation in free circulating DNA and its correlation with estrogen receptor protein expression in tumor tissue in breast cancer patients. BMC cancer. 2014;14:59.
123. Pourteimoor V, Mohammadi-Yeganeh S, Paryan M. Breast cancer classification and prognostication through diverse systems along with recent emerging findings in this respect; the dawn of new perspectives in the clinical applications. Tumour biology : the journal of the International Society for Oncodevelopmental Biology and Medicine. 2016;37(11):14479-99.
124. Esteller M, Silva JM, Dominguez G, Bonilla F, Matias-Guiu X, Lerma E, et al. Promoter hypermethylation and BRCA1 inactivation in sporadic breast and ovarian tumors. Journal of the National Cancer Institute. 2000;92(7):564-9.
125. Veeck J, Ropero S, Setien F, Gonzalez-Suarez E, Osorio A, Benitez J, et al. BRCA1 CpG island hypermethylation predicts sensitivity to poly(adenosine diphosphate)-ribose polymerase inhibitors. Journal of clinical oncology : official journal of the American Society of Clinical Oncology. 2010;28(29):e563-4; author reply e5-6.
126. Gaudet MM, Campan M, Figueroa JD, Yang XR, Lissowska J, Peplonska B, et al. DNA hypermethylation of ESR1 and PGR in breast cancer: pathologic and epidemiologic associations. Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology. 2009;18(11):3036-43.
127. Jezkova E, Kajo K, Zubor P, Grendar M, Malicherova B, Mendelova A, et al. Methylation in promoter regions of PITX2 and RASSF1A genes in association with clinicopathological features in breast cancer patients. Tumour biology : the journal of the International Society for Oncodevelopmental Biology and Medicine. 2016.
128. Sunami E, Shinozaki M, Sim MS, Nguyen SL, Vu AT, Giuliano AE, et al. Estrogen receptor and HER2/neu status affect epigenetic differences of tumor-related genes in primary breast tumors. Breast cancer research : BCR. 2008;10(3):R46.

129. Fackler MJ, McVeigh M, Mehrotra J, Blum MA, Lange J, Lapides A, et al. Quantitative multiplex methylation-specific PCR assay for the detection of promoter hypermethylation in multiple genes in breast cancer. Cancer research. 2004;64(13):4442-52.
130. Huang KT, Mikeska T, Li J, Takano EA, Millar EK, Graham PH, et al. Assessment of DNA methylation profiling and copy number variation as indications of clonal relationship in ipsilateral and contralateral breast cancers to distinguish recurrent breast cancer from a second primary tumour. BMC cancer. 2015;15:669.
131. Kajabova V, Smolkova B, Zmetakova I, Sebova K, Krivulcik T, Bella V, et al. RASSF1A Promoter Methylation Levels Positively Correlate with Estrogen Receptor Expression in Breast Cancer Patients. Translational oncology. 2013;6(3):297-304.
132. Radpour R, Kohler C, Haghighi MM, Fan AX, Holzgreve W, Zhong XY. Methylation profiles of 22 candidate genes in breast cancer using high-throughput MALDI-TOF mass array. Oncogene. 2009;28(33):2969-78.
133. Stefansson OA, Villanueva A, Vidal A, Marti L, Esteller M. BRCA1 epigenetic inactivation predicts sensitivity to platinum-based chemotherapy in breast and ovarian cancer. Epigenetics. 2012;7(11):1225-9.
134. Kloten V, Becker B, Winner K, Schrauder MG, Fasching PA, Anzeneder T, et al. Promoter hypermethylation of the tumor-suppressor genes ITIH5, DKK3, and RASSF1A as novel biomarkers for blood-based breast cancer screening. Breast cancer research : BCR. 2013;15(1):R4.
135. Jeronimo C, Monteiro P, Henrique R, Dinis-Ribeiro M, Costa I, Costa VL, et al. Quantitative hypermethylation of a small panel of genes augments the diagnostic accuracy in fine-needle aspirate washings of breast lesions. Breast cancer research and treatment. 2008;109(1):27-34.
136. Toyooka KO, Toyooka S, Virmani AK, Sathyanarayana UG, Euhus DM, Gilcrease M, et al. Loss of expression and aberrant methylation of the CDH13 (H-cadherin) gene in breast and lung carcinomas. Cancer research. 2001;61(11):4556-60.
137. Feng W, Shen L, Wen S, Rosen DG, Jelinek J, Hu X, et al. Correlation between CpG methylation profiles and hormone receptor status in breast cancers. Breast cancer research : BCR. 2007;9(4):R57.
138. Roll JD, Rivenbark AG, Sandhu R, Parker JS, Jones WD, Carey LA, et al. Dysregulation of the epigenome in triple-negative breast cancers: basal-like and claudin-low breast cancers express aberrant DNA hypermethylation. Experimental and molecular pathology. 2013;95(3):276-87.
139. Chimonidou M, Tzitzira A, Strati A, Sotiropoulou G, Sfikas C, Malamos N, et al. CST6 promoter methylation in circulating cell-free DNA of breast cancer patients. Clinical biochemistry. 2013;46(3):235-40.
140. Spitzwieser M, Entfellner E, Werner B, Pulverer W, Pfeiler G, Hacker S, et al. Hypermethylation of CDKN2A exon 2 in tumor, tumor-adjacent and tumor-distant tissues from breast cancer patients. BMC cancer. 2017;17(1):260.
141. Shan M, Yin H, Li J, Li X, Wang D, Su Y, et al. Detection of aberrant methylation of a six-gene panel in serum DNA for diagnosis of breast cancer. Oncotarget. 2016;7(14):18485-94.
142. Harbeck N, Nimmrich I, Hartmann A, Ross JS, Cufer T, Grutzmann R, et al. Multicenter study using paraffin-embedded tumor tissue testing PITX2 DNA methylation as a marker for outcome prediction in tamoxifen-treated, node-negative breast cancer patients. Journal of clinical oncology : official journal of the American Society of Clinical Oncology. 2008;26(31):5036-42.
143. Nimmrich I, Sieuwerts AM, Meijer-van Gelder ME, Schwope I, Bolt-de Vries J, Harbeck N, et al. DNA hypermethylation of PITX2 is a marker of poor prognosis in untreated lymph node-negative hormone receptor-positive breast cancer patients. Breast cancer research and treatment. 2008;111(3):429-37.

144. Gobel G, Auer D, Gaugg I, Schneitter A, Lesche R, Muller-Holzner E, et al. Prognostic significance of methylated RASSF1A and PITX2 genes in blood- and bone marrow plasma of breast cancer patients. Breast cancer research and treatment. 2011;130(1):109-17.
145. Huang R, Ding P, Yang F. Clinicopathological significance and potential drug target of CDH1 in breast cancer: a meta-analysis and literature review. Drug design, development and therapy. 2015;9:5277-85.
146. Liu J, Sun X, Qin S, Wang H, Du N, Li Y, et al. CDH1 promoter methylation correlates with decreased gene expression and poor prognosis in patients with breast cancer. Oncology letters. 2016;11(4):2635-43.
147. Naghitorabi M, Mohammadi-Asl J, Sadeghi HM, Rabbani M, Jafarian-Dehkordi A, Javanmard SH. Quantitation of CDH1 promoter methylation in formalin-fixed paraffin-embedded tissues of breast cancer patients using differential high resolution melting analysis. Advanced biomedical research. 2016;5:91.
148. Bibikova M, Lin Z, Zhou L, Chudin E, Garcia EW, Wu B, et al. High-throughput DNA methylation profiling using universal bead arrays. Genome research. 2006;16(3):383-93.
149. Bibikova M, Le J, Barnes B, Saedinia-Melnyk S, Zhou L, Shen R, et al. Genome-wide DNA methylation profiling using Infinium(R) assay. Epigenomics. 2009;1(1):177-200.
150. Bibikova M, Barnes B, Tsan C, Ho V, Klotzle B, Le JM, et al. High density DNA methylation array with single CpG site resolution. Genomics. 2011;98(4):288-95.
151. Sandoval J, Heyn H, Moran S, Serra-Musach J, Pujana MA, Bibikova M, et al. Validation of a DNA methylation microarray for 450,000 CpG sites in the human genome. Epigenetics. 2011;6(6):692-702.
152. Moran S, Arribas C, Esteller M. Validation of a DNA methylation microarray for 850,000 CpG sites of the human genome enriched in enhancer sequences. Epigenomics. 2016;8(3):389-99.
153. Grossman RL, Heath AP, Ferretti V, Varmus HE, Lowy DR, Kibbe WA, et al. Toward a Shared Vision for Cancer Genomic Data. The New England journal of medicine. 2016;375(12):1109-12.
154. Moran S, Vizoso M, Martinez-Cardus A, Gomez A, Matias-Guiu X, Chiavenna SM, et al. Validation of DNA methylation profiling in formalin-fixed paraffin-embedded samples using the Infinium HumanMethylation450 Microarray. Epigenetics. 2014;9(6):829-33.
155. Flower KJ, Shenker NS, El-Bahrawy M, Goldgar DE, Parsons MT, Investigators KC, et al. DNA methylation profiling to assess pathogenicity of BRCA1 unclassified variants in breast cancer. Epigenetics. 2015;10(12):1121-32.
156. Bradford MM. A rapid and sensitive method for the quantitation of microgram quantities of protein utilizing the principle of protein-dye binding. Analytical biochemistry. 1976;72:248-54.
157. Hutzinger R, Mrazek J, Vorwerk S, Huttenhofer A. NcRNA-microchip analysis: a novel approach to identify differential expression of noncoding RNAs. RNA Biol. 2010;7(5):586-95.
158. Tormo E, Pineda B, Serna E, Guijarro A, Ribas G, Fores J, et al. MicroRNA Profile in Response to Doxorubicin Treatment in Breast Cancer. Journal of cellular biochemistry. 2015;116(9):2061-73.
159. Irizarry RA, Bolstad BM, Collin F, Cope LM, Hobbs B, Speed TP. Summaries of Affymetrix GeneChip probe level data. Nucleic acids research. 2003;31(4):e15.
160. Pena-Chilet M, Martinez MT, Perez-Fidalgo JA, Peiro-Chova L, Oltra SS, Tormo E, et al. MicroRNA profile in very young women with breast cancer. BMC cancer. 2014;14:529.
161. Weisenberger DJ. Characterizing DNA methylation alterations from The Cancer Genome Atlas. The Journal of clinical investigation. 2014;124(1):17-23.
162. Sandoval J, Mendez-Gonzalez J, Nadal E, Chen G, Carmona FJ, Sayols S, et al. A prognostic DNA methylation signature for stage I non-small-cell lung cancer. Journal of clinical oncology : official journal of the American Society of Clinical Oncology. 2013;31(32):4140-7.

163. Villanueva A, Portela A, Sayols S, Battiston C, Hoshida Y, Mendez-Gonzalez J, et al. DNA methylation-based prognosis and epidrivers in hepatocellular carcinoma. Hepatology. 2015;61(6):1945-56.
164. Heyn H, Li N, Ferreira HJ, Moran S, Pisano DG, Gomez A, et al. Distinct DNA methylomes of newborns and centenarians. Proceedings of the National Academy of Sciences of the United States of America. 2012;109(26):10522-7.
165. Heyn H, Moran S, Hernando-Herraez I, Sayols S, Gomez A, Sandoval J, et al. DNA methylation contributes to natural human variation. Genome research. 2013;23(9):1363-72.
166. Stewart SK, Morris TJ, Guilhamon P, Bulstrode H, Bachman M, Balasubramanian S, et al. oxBS-450K: a method for analysing hydroxymethylation using 450K BeadChips. Methods. 2015;72:9-15.
167. Field SF, Beraldi D, Bachman M, Stewart SK, Beck S, Balasubramanian S. Accurate measurement of 5-methylcytosine and 5-hydroxymethylcytosine in human cerebellum DNA by oxidative bisulfite on an array (OxBS-array). PloS one. 2015;10(2):e0118202.
168. Shlyueva D, Stampfel G, Stark A. Transcriptional enhancers: from properties to genome-wide predictions. Nat Rev Genet. 2014;15(4):272-86.
169. Whyte WA, Orlando DA, Hnisz D, Abraham BJ, Lin CY, Kagey MH, et al. Master transcription factors and mediator establish super-enhancers at key cell identity genes. Cell. 2013;153(2):307-19.
170. Hnisz D, Abraham BJ, Lee TI, Lau A, Saint-Andre V, Sigova AA, et al. Super-enhancers in the control of cell identity and disease. Cell. 2013;155(4):934-47.
171. Perini G, Diolaiti D, Porro A, Della Valle G. In vivo transcriptional regulation of N-Myc target genes is controlled by E-box methylation. Proceedings of the National Academy of Sciences of the United States of America. 2005;102(34):12117-22.
172. Consortium EP. An integrated encyclopedia of DNA elements in the human genome. Nature. 2012;489(7414):57-74.
173. Siggens L, Ekwall K. Epigenetics, chromatin and genome organization: recent advances from the ENCODE project. Journal of internal medicine. 2014;276(3):201-14.
174. Lizio M, Harshbarger J, Shimoji H, Severin J, Kasukawa T, Sahin S, et al. Gateways to the FANTOM5 promoter level mammalian expression atlas. Genome biology. 2015;16:22.
175. Aryee MJ, Jaffe AE, Corrada-Bravo H, Ladd-Acosta C, Feinberg AP, Hansen KD, et al. Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. Bioinformatics. 2014;30(10):1363-9.
176. McCartney DL, Walker RM, Morris SW, McIntosh AM, Porteous DJ, Evans KL. Identification of polymorphic and off-target probe binding sites on the Illumina Infinium MethylationEPIC BeadChip. Genomics data. 2016;9:22-4.
177. Fortin JP, Labbe A, Lemire M, Zanke BW, Hudson TJ, Fertig EJ, et al. Functional normalization of 450k methylation array data improves replication in large cancer studies. Genome biology. 2014;15(12):503.
178. Horvath S, Zhang Y, Langfelder P, Kahn RS, Boks MP, van Eijk K, et al. Aging effects on DNA methylation modules in human brain and blood tissue. Genome biology. 2012;13(10):R97.
179. Bell JT, Tsai PC, Yang TP, Pidsley R, Nisbet J, Glass D, et al. Epigenome-wide scans identify differentially methylated regions for age and age-related phenotypes in a healthy ageing population. PLoS genetics. 2012;8(4):e1002629.
180. Rodriguez-Rodero S, Fernandez-Morera JL, Fernandez AF, Menendez-Torre E, Fraga MF. Epigenetic regulation of aging. Discov Med. 2010;10(52):225-33.
181. Murgatroyd C, Wu Y, Bockmuhl Y, Spengler D. The Janus face of DNA methylation in aging. Aging. 2010;2(2):107-10.
182. Teschendorff AE, Menon U, Gentry-Maharaj A, Ramus SJ, Weisenberger DJ, Shen H, et al. Age-dependent DNA methylation of genes that are suppressed in stem cells is a hallmark of cancer. Genome research. 2010;20(4):440-6.

183. Bollati V, Schwartz J, Wright R, Litonjua A, Tarantini L, Suh H, et al. Decline in genomic DNA methylation through aging in a cohort of elderly subjects. Mechanisms of ageing and development. 2009;130(4):234-9.
184. Christensen BC, Houseman EA, Marsit CJ, Zheng S, Wrensch MR, Wiemels JL, et al. Aging and environmental exposures alter tissue-specific DNA methylation dependent upon CpG island context. PLoS genetics. 2009;5(8):e1000602.
185. Fraga MF, Agrelo R, Esteller M. Cross-talk between aging and cancer: the epigenetic language. Ann N Y Acad Sci. 2007;1100:60-74.
186. Fraga MF, Esteller M. Epigenetics and aging: the targets and the marks. Trends Genet. 2007;23(8):413-8.
187. Li Y, Zhu J, Tian G, Li N, Li Q, Ye M, et al. The DNA methylome of human peripheral blood mononuclear cells. PLoS Biol. 2010;8(11):e1000533.
188. Illingworth R, Kerr A, Desousa D, Jorgensen H, Ellis P, Stalker J, et al. A novel CpG island set identifies tissue-specific methylation at developmental gene loci. PLoS Biol. 2008;6(1):e22.
189. Bernstein BE, Stamatoyannopoulos JA, Costello JF, Ren B, Milosavljevic A, Meissner A, et al. The NIH Roadmap Epigenomics Mapping Consortium. Nat Biotechnol. 2010;28(10):1045-8.
190. Thompson RF, Atzmon G, Gheorghe C, Liang HQ, Lowes C, Greally JM, et al. Tissue-specific dysregulation of DNA methylation in aging. Aging cell. 2010;9(4):506-18.
191. Numata S, Ye T, Hyde TM, Guitart-Navarro X, Tao R, Wininger M, et al. DNA methylation signatures in development and aging of the human prefrontal cortex. American journal of human genetics. 2012;90(2):260-72.
192. Koch CM, Wagner W. Epigenetic-aging-signature to determine age in different tissues. Aging. 2011;3(10):1018-27.
193. Hernandez DG, Nalls MA, Gibbs JR, Arepalli S, van der Brug M, Chong S, et al. Distinct DNA methylation changes highly correlated with chronological age in the human brain. Human molecular genetics. 2011;20(6):1164-72.
194. Horvath S. DNA methylation age of human tissues and cell types. Genome biology. 2013;14(10):R115.
195. Horvath S. Erratum to: DNA methylation age of human tissues and cell types. Genome biology. 2015;16:96.
196. Hovestadt V ZM. conumee: Enhanced copy-number variation analysis using Illumina 450k methylation arrays., R package version 0.99.4. http://www.bioconductor.org/packages/release/bioc/html/conumee.html. 2015.
197. Sturm D, Witt H, Hovestadt V, Khuong-Quang DA, Jones DT, Konermann C, et al. Hotspot mutations in H3F3A and IDH1 define distinct epigenetic and biological subgroups of glioblastoma. Cancer Cell. 2012;22(4):425-37.
198. A SVaO. DNAcopy: DNA copy number data analysis. R package version 1.48.0. 2016.
199. Butcher LM, Beck S. Probe Lasso: a novel method to rope in differentially methylated regions with 450K DNA methylation data. Methods. 2015;72:21-8.
200. Mullis K, Faloona F, Scharf S, Saiki R, Horn G, Erlich H. Specific enzymatic amplification of DNA in vitro: the polymerase chain reaction. Cold Spring Harb Symp Quant Biol. 1986;51 Pt 1:263-73.
201. Higuchi R, Fockler C, Dollinger G, Watson R. Kinetic PCR analysis: real-time monitoring of DNA amplification reactions. Biotechnology (N Y). 1993;11(9):1026-30.
202. Livak KJ, Schmittgen TD. Analysis of relative gene expression data using real-time quantitative PCR and the 2(-Delta Delta C(T)) Method. Methods. 2001;25(4):402-8.
203. Vlachos IS, Kostoulas N, Vergoulis T, Georgakilas G, Reczko M, Maragkakis M, et al. DIANA miRPath v.2.0: investigating the combinatorial effect of microRNAs in pathways. Nucleic acids research. 2012;40(Web Server issue):W498-504.
204. Joshi-Tope G, Gillespie M, Vastrik I, D'Eustachio P, Schmidt E, de Bono B, et al. Reactome: a knowledgebase of biological pathways. Nucleic acids research. 2005;33(Database issue):D428-32.

205. Chen EY, Tan CM, Kou Y, Duan Q, Wang Z, Meirelles GV, et al. Enrichr: interactive and collaborative HTML5 gene list enrichment analysis tool. BMC Bioinformatics. 2013;14:128.
206. Hosack DA, Dennis G, Jr., Sherman BT, Lane HC, Lempicki RA. Identifying biological themes within lists of genes with EASE. Genome biology. 2003;4(10):R70.
207. Li A, Liu Z, Li M, Zhou S, Xu Y, Xiao Y, et al. HDAC5, a potential therapeutic target and prognostic biomarker, promotes proliferation, invasion and migration in human breast cancer. Oncotarget. 2016;7(25):37966-78.
208. Liu Q, Zheng JM, Chen JK, Yan XL, Chen HM, Nong WX, et al. Histone deacetylase 5 promotes the proliferation of glioma cells by upregulation of Notch 1. Molecular medicine reports. 2014;10(4):2045-50.
209. Costa BM, Smith JS, Chen Y, Chen J, Phillips HS, Aldape KD, et al. Reversing HOXA9 oncogene activation by PI3K inhibition: epigenetic mechanism and prognostic significance in human glioblastoma. Cancer research. 2010;70(2):453-62.
210. Qiu C, Bu X, Jiang Z. Protocadherin-10 acts as a tumor suppressor gene, and is frequently downregulated by promoter methylation in pancreatic cancer cells. Oncology reports. 2016;36(1):383-9.
211. Albino D, Longoni N, Curti L, Mello-Grand M, Pinton S, Civenni G, et al. ESE3/EHF controls epithelial cell differentiation and its loss leads to prostate tumors with mesenchymal and stem-like features. Cancer research. 2012;72(11):2889-900.
212. Shi J, Qu Y, Li X, Sui F, Yao D, Yang Q, et al. Increased expression of EHF via gene amplification contributes to the activation of HER family signaling and associates with poor survival in gastric cancer. Cell death & disease. 2016;7(10):e2442.
213. Li L, Yan J, Xu J, Liu CQ, Zhen ZJ, Chen HW, et al. CXCL17 expression predicts poor prognosis and correlates with adverse immune infiltration in hepatocellular carcinoma. PloS one. 2014;9(10):e110064.
214. Setta-Kaffetzi N, Simpson MA, Navarini AA, Patel VM, Lu HC, Allen MH, et al. AP1S3 mutations are associated with pustular psoriasis and impaired Toll-like receptor 3 trafficking. American journal of human genetics. 2014;94(5):790-7.
215. Lehrmann H, Pritchard LL, Harel-Bellan A. Histone acetyltransferases and deacetylases in the control of cell proliferation and differentiation. Advances in cancer research. 2002;86:41-65.
216. Agis-Balboa RC, Pavelka Z, Kerimoglu C, Fischer A. Loss of HDAC5 impairs memory function: implications for Alzheimer's disease. Journal of Alzheimer's disease : JAD. 2013;33(1):35-44.
217. Miller TA, Witter DJ, Belvedere S. Histone deacetylase inhibitors. Journal of medicinal chemistry. 2003;46(24):5097-116.
218. Khan O, La Thangue NB. HDAC inhibitors in cancer biology: emerging mechanisms and clinical applications. Immunology and cell biology. 2012;90(1):85-94.
219. Ohguchi H, Hideshima T, Anderson KC. The biological significance of histone modifiers in multiple myeloma: clinical applications. Blood cancer journal. 2018;8(9):83.
220. Hu C, Lv L, Peng J, Liu D, Wang X, Zhou Y, et al. MicroRNA-375 suppresses esophageal cancer cell growth and invasion by repressing metadherin expression. Oncology letters. 2017;13(6):4769-75.
221. Zhang Y, Xu D, Pan J, Yang Z, Chen M, Han J, et al. Dynamic monitoring of circulating microRNAs as a predictive biomarker for the diagnosis and recurrence of papillary thyroid carcinoma. Oncology letters. 2017;13(6):4252-66.
222. Gazdar AF, Kurvari V, Virmani A, Gollahon L, Sakaguchi M, Westerfield M, et al. Characterization of paired tumor and non-tumor cell lines established from patients with breast cancer. International journal of cancer. 1998;78(6):766-74.
223. Fan LZ, Cherian MG. Potential role of p53 on metallothionein induction in human epithelial breast cancer cells. British journal of cancer. 2002;87(9):1019-26.

224. Rakha EA, Ellis IO. Triple-negative/basal-like breast cancer: review. Pathology. 2009;41(1):40-7.
225. Rakha EA, Elsheikh SE, Aleskandarany MA, Habashi HO, Green AR, Powe DG, et al. Triple-negative breast cancer: distinguishing between basal and nonbasal subtypes. Clinical cancer research : an official journal of the American Association for Cancer Research. 2009;15(7):2302-10.
226. Dolle JM, Daling JR, White E, Brinton LA, Doody DR, Porter PL, et al. Risk factors for triple-negative breast cancer in women under the age of 45 years. Cancer epidemiology, biomarkers & prevention : a publication of the American Association for Cancer Research, cosponsored by the American Society of Preventive Oncology. 2009;18(4):1157-66.
227. Sasa M, Bando Y, Takahashi M, Hirose T, Nagao T. Screening for basal marker expression is necessary for decision of therapeutic strategy for triple-negative breast cancer. Journal of surgical oncology. 2008;97(1):30-4.
228. van de Rijn M, Perou CM, Tibshirani R, Haas P, Kallioniemi O, Kononen J, et al. Expression of cytokeratins 17 and 5 identifies a group of breast carcinomas with poor clinical outcome. The American journal of pathology. 2002;161(6):1991-6.
229. Nofech-Mozes S, Trudeau M, Kahn HK, Dent R, Rawlinson E, Sun P, et al. Patterns of recurrence in the basal and non-basal subtypes of triple-negative breast cancers. Breast cancer research and treatment. 2009;118(1):131-7.
230. Nielsen TO, Hsu FD, Jensen K, Cheang M, Karaca G, Hu Z, et al. Immunohistochemical and clinical characterization of the basal-like subtype of invasive breast carcinoma. Clinical cancer research : an official journal of the American Association for Cancer Research. 2004;10(16):5367-74.
231. Livasy CA, Karaca G, Nanda R, Tretiakova MS, Olopade OI, Moore DT, et al. Phenotypic evaluation of the basal-like subtype of invasive breast carcinoma. Mod Pathol. 2006;19(2):264-71.
232. Charafe-Jauffret E, Ginestier C, Monville F, Finetti P, Adelaide J, Cervera N, et al. Gene expression profiling of breast cell lines identifies potential new basal markers. Oncogene. 2006;25(15):2273-84.
233. Anders C, Carey LA. Understanding and treating triple-negative breast cancer. Oncology (Williston Park). 2008;22(11):1233-9; discussion 9-40, 43.
234. Armes JE, Trute L, White D, Southey MC, Hammet F, Tesoriero A, et al. Distinct molecular pathogeneses of early-onset breast cancers in BRCA1 and BRCA2 mutation carriers: a population-based study. Cancer research. 1999;59(8):2011-7.
235. Grushko TA, Blackwood MA, Schumm PL, Hagos FG, Adeyanju MO, Feldman MD, et al. Molecular-cytogenetic analysis of HER-2/neu gene in BRCA1-associated breast cancers. Cancer research. 2002;62(5):1481-8.
236. Umemura S, Takekoshi S, Suzuki Y, Saitoh Y, Tokuda Y, Osamura RY. Estrogen receptor-negative and human epidermal growth factor receptor 2-negative breast cancer tissue have the highest Ki-67 labeling index and EGFR expression: gene amplification does not contribute to EGFR expression. Oncology reports. 2005;14(2):337-43.
237. Jacquemier J, Ginestier C, Rougemont J, Bardou VJ, Charafe-Jauffret E, Geneix J, et al. Protein expression profiling identifies subclasses of breast cancer and predicts prognosis. Cancer research. 2005;65(3):767-79.
238. Sabatier R, Jacquemier J, Bertucci F, Esterni B, Finetti P, Azario F, et al. Peritumoural vascular invasion: a major determinant of triple-negative breast cancer outcome. European journal of cancer. 2011;47(10):1537-45.
239. Liu HT, Ma R, Yang QF, Du G, Zhang CJ. Lymphangiogenic characteristics of triple negativity in node-negative breast cancer. Int J Surg Pathol. 2009;17(6):426-31.
240. Phillips TM, McBride WH, Pajonk F. The response of CD24(-/low)/CD44+ breast cancer-initiating cells to radiation. Journal of the National Cancer Institute. 2006;98(24):1777-85.

241. Pajonk F, Vlashi E, McBride WH. Radiation resistance of cancer stem cells: the 4 R's of radiobiology revisited. Stem cells. 2010;28(4):639-48.
242. Li X, Lewis MT, Huang J, Gutierrez C, Osborne CK, Wu MF, et al. Intrinsic resistance of tumorigenic breast cancer cells to chemotherapy. Journal of the National Cancer Institute. 2008;100(9):672-9.
243. Bao S, Wu Q, McLendon RE, Hao Y, Shi Q, Hjelmeland AB, et al. Glioma stem cells promote radioresistance by preferential activation of the DNA damage response. Nature. 2006;444(7120):756-60.
244. Venkatesh H, Monje M. Neuronal Activity in Ontogeny and Oncology. Trends in Cancer.3(2):89-112.
245. Eissa S, Matboli M, Shehata HH. Breast tissue-based microRNA panel highlights microRNA-23a and selected target genes as putative biomarkers for breast cancer. Translational research : the journal of laboratory and clinical medicine. 2015;165(3):417-27.
246. Koller K, Das S, Leuschner I, Korbelius M, Hoefler G, Guertl B. Identification of the transcription factor HOXB4 as a novel target of miR-23a. Genes, chromosomes & cancer. 2013;52(8):709-15.
247. Chen B, Zhu A, Tian L, Xin Y, Liu X, Peng Y, et al. miR23a suppresses pancreatic cancer cell progression by inhibiting PLK1 expression. Molecular medicine reports. 2018.
248. Kuchiba A, Iwasaki M, Ono H, Kasuga Y, Yokoyama S, Onuma H, et al. Global methylation levels in peripheral blood leukocyte DNA by LUMA and breast cancer: a case-control study in Japanese women. British journal of cancer. 2014;110(11):2765-71.
249. van Veldhoven K, Polidoro S, Baglietto L, Severi G, Sacerdote C, Panico S, et al. Epigenome-wide association study reveals decreased average methylation levels years before breast cancer diagnosis. Clinical epigenetics. 2015;7:67.
250. Bardowell SA, Parker J, Fan C, Crandell J, Perou CM, Swift-Scanlan T. Differential methylation relative to breast cancer subtype and matched normal tissue reveals distinct patterns. Breast cancer research and treatment. 2013;142(2):365-80.
251. Stefansson OA, Moran S, Gomez A, Sayols S, Arribas-Jorba C, Sandoval J, et al. A DNA methylation-based definition of biologically distinct breast cancer subtypes. Molecular oncology. 2015;9(3):555-68.
252. Cancello G, Maisonneuve P, Mazza M, Montagna E, Rotmensz N, Viale G, et al. Pathological features and survival outcomes of very young patients with early breast cancer: how much is "very young"? Breast. 2013;22(6):1046-51.
253. Fredholm H, Magnusson K, Lindstrom LS, Tobin NP, Lindman H, Bergh J, et al. Breast cancer in young women and prognosis: How important are proliferation markers? European journal of cancer. 2017;84:278-89.
254. Bormann F, Rodriguez-Paredes M, Hagemann S, Manchanda H, Kristof B, Gutekunst J, et al. Reduced DNA methylation patterning and transcriptional connectivity define human skin aging. Aging cell. 2016;15(3):563-71.
255. Weidner CI, Wagner W. The epigenetic tracks of aging. Biological chemistry. 2014;395(11):1307-14.
256. Jung M, Pfeifer GP. Aging and DNA methylation. BMC biology. 2015;13:7.
257. Perna L, Zhang Y, Mons U, Holleczek B, Saum KU, Brenner H. Epigenetic age acceleration predicts cancer, cardiovascular, and all-cause mortality in a German case cohort. Clinical epigenetics. 2016;8:64.
258. Zheng Y, Joyce BT, Colicino E, Liu L, Zhang W, Dai Q, et al. Blood Epigenetic Age may Predict Cancer Incidence and Mortality. EBioMedicine. 2016;5:68-73.
259. Sehl ME, Henry JE, Storniolo AM, Ganz PA, Horvath S. DNA methylation age is elevated in breast tissue of healthy women. Breast cancer research and treatment. 2017;164(1):209-19.
260. Shen J, Wang S, Zhang YJ, Wu HC, Kibriya MG, Jasmine F, et al. Exploring genome-wide DNA methylation profiles altered in hepatocellular carcinoma using Infinium HumanMethylation 450 BeadChips. Epigenetics. 2013;8(1):34-43.

261. Marzese DM, Scolyer RA, Huynh JL, Huang SK, Hirose H, Chong KK, et al. Epigenome-wide DNA methylation landscape of melanoma progression to brain metastasis reveals aberrations on homeobox D cluster associated with prognosis. Human molecular genetics. 2014;23(1):226-38.
262. Lv X, Li L, Lv L, Qu X, Jin S, Li K, et al. HOXD9 promotes epithelial-mesenchymal transition and cancer metastasis by ZEB1 regulation in hepatocellular carcinoma. Journal of experimental & clinical cancer research : CR. 2015;34:133.
263. Tang X, Yin X, Xiang T, Li H, Li F, Chen L, et al. Protocadherin 10 is frequently downregulated by promoter methylation and functions as a tumor suppressor gene in non-small cell lung cancer. Cancer biomarkers : section A of Disease markers. 2012;12(1):11-9.
264. Shi D, Murty VV, Gu W. PCDH10, a novel p53 transcriptional target in regulating cell migration. Cell cycle. 2015;14(6):857-66.
265. Haberland M, Montgomery RL, Olson EN. The many roles of histone deacetylases in development and physiology: implications for disease and therapy. Nat Rev Genet. 2009;10(1):32-42.
266. Dawson MA, Kouzarides T. Cancer epigenetics: from mechanism to therapy. Cell. 2012;150(1):12-27.
267. Li A, Liu Z, Li M, Zhou S, Xu Y, Xiao Y, et al. Correction: HDAC5, a potential therapeutic target and prognostic biomarker, promotes proliferation, invasion and migration in human breast cancer. Oncotarget. 2017;8(18):30619-20.
268. Klieser E, Urbas R, Stattner S, Primavesi F, Jager T, Dinnewitzer A, et al. Comprehensive immunohistochemical analysis of histone deacetylases in pancreatic neuroendocrine tumors: HDAC5 as a predictor of poor clinical outcome. Human pathology. 2017.
269. de Cremoux P, Dalvai M, N'Doye O, Moutahir F, Rolland G, Chouchane-Mlik O, et al. HDAC inhibition does not induce estrogen receptor in human triple-negative breast cancer cell lines and patient-derived xenografts. Breast cancer research and treatment. 2015;149(1):81-9.
270. Jang ER, Lim SJ, Lee ES, Jeong G, Kim TY, Bang YJ, et al. The histone deacetylase inhibitor trichostatin A sensitizes estrogen receptor alpha-negative breast cancer cells to tamoxifen. Oncogene. 2004;23(9):1724-36.
271. Raha P, Thomas S, Thurn KT, Park J, Munster PN. Combined histone deacetylase inhibition and tamoxifen induces apoptosis in tamoxifen-resistant breast cancer models, by reversing Bcl-2 overexpression. Breast cancer research : BCR. 2015;17:26.
272. Thomas S, Thurn KT, Raha P, Chen S, Munster PN. Efficacy of histone deacetylase and estrogen receptor inhibition in breast cancer cells due to concerted down regulation of Akt. PloS one. 2013;8(7):e68973.
273. Calin GA, Dumitru CD, Shimizu M, Bichi R, Zupo S, Noch E, et al. Frequent deletions and down-regulation of micro- RNA genes miR15 and miR16 at 13q14 in chronic lymphocytic leukemia. Proceedings of the National Academy of Sciences of the United States of America. 2002;99(24):15524-9.
274. Baffa R, Fassan M, Volinia S, O'Hara B, Liu CG, Palazzo JP, et al. MicroRNA expression profiling of human metastatic cancers identifies cancer gene targets. The Journal of pathology. 2009;219(2):214-21.
275. Di Leva G, Briskin D, Croce CM. MicroRNA in cancer: new hopes for antineoplastic chemotherapy. Upsala journal of medical sciences. 2012;117(2):202-16.
276. Suzuki H, Maruyama R, Yamamoto E, Kai M. DNA methylation and microRNA dysregulation in cancer. Molecular oncology. 2012;6(6):567-78.
277. Piletic K, Kunej T. MicroRNA epigenetic signatures in human disease. Archives of toxicology. 2016;90(10):2405-19.
278. Cockerill PN. Structure and function of active chromatin and DNase I hypersensitive sites. The FEBS journal. 2011;278(13):2182-210.

279. Kaur S, Lotsari-Salomaa JE, Seppanen-Kaijansinkko R, Peltomaki P. MicroRNA Methylation in Colorectal Cancer. Advances in experimental medicine and biology. 2016;937:109-22.
280. Lujambio A, Ropero S, Ballestar E, Fraga MF, Cerrato C, Setien F, et al. Genetic unmasking of an epigenetically silenced microRNA in human cancer cells. Cancer research. 2007;67(4):1424-9.
281. Agirre X, Vilas-Zornoza A, Jimenez-Velasco A, Martin-Subero JI, Cordeu L, Garate L, et al. Epigenetic silencing of the tumor suppressor microRNA Hsa-miR-124a regulates CDK6 expression and confers a poor prognosis in acute lymphoblastic leukemia. Cancer research. 2009;69(10):4443-53.
282. Oltra SS, Pena-Chilet M, Vidal-Tomas V, Flower K, Martinez MT, Alonso E, et al. Methylation deregulation of miRNA promoters identifies miR124-2 as a survival biomarker in Breast Cancer in very young women. Scientific reports. 2018;8(1):14373.

ANNEXES

ANNEXES

ANNEX I. Significant miRNAs differentially expressed between breast cancer cell lines from young women vs. older counterparts.

	Mean expression differences (CLO-CLVY)	p-value	FDR adjusted p-value
hsa-miR-423-3p_st	-3,14	4,83E-13	5,57E-10
hsa-miR-24-2-star_st	3,67	2,49E-12	1,18E-09
hsa-miR-423-5p_st	-3,94	3,08E-12	1,18E-09
hsa-miR-126_st	4,09	2,70E-10	7,79E-08
hsa-miR-671-5p_st	-2,42	2,07E-09	4,78E-07
hsa-miR-27a_st	4,74	3,22E-09	6,18E-07
hsa-miR-19b_st	6,32	7,30E-09	1,20E-06
hsa-miR-324-3p_st	-2,04	9,83E-09	1,42E-06
hsa-miR-574-3p_st	-2,02	1,25E-08	1,60E-06
hsa-miR-874_st	-3,14	2,60E-08	3,01E-06
hsa-miR-744_st	-2,86	6,42E-08	6,74E-06
hsa-miR-197_st	-2,99	9,64E-08	9,27E-06
hsa-miR-193a-5p_st	-6,33	1,44E-07	1,28E-05
hsa-miR-320a_st	-1,84	3,77E-07	2,90E-05
hsa-miR-615-3p_st	-5,46	3,57E-07	2,90E-05
hsa-miR-320b_st	-1,88	4,16E-07	3,00E-05
hsa-miR-320c_st	-1,87	5,56E-07	3,77E-05
v11_hsa-miR-768-3p_st	-3,26	9,58E-07	6,14E-05
hsa-miR-23a_st	2,60	1,09E-06	6,63E-05
hsa-miR-99a_st	4,80	1,51E-06	8,72E-05
hsa-miR-378-star_st	-1,93	1,76E-06	9,66E-05
hsa-miR-484_st	-3,87	1,85E-06	9,69E-05
hsa-miR-328_st	-2,93	2,25E-06	1,13E-04
hsa-miR-1260b_st	2,72	2,37E-06	1,14E-04
hsa-miR-125b_st	6,69	2,61E-06	1,16E-04
hsa-miR-99b-star_st	-2,21	2,58E-06	1,16E-04
hsa-miR-16_st	1,56	7,99E-06	3,41E-04
hsa-miR-1307_st	-2,63	9,99E-06	3,84E-04
hsa-miR-320e_st	-2,18	9,80E-06	3,84E-04
hsa-miR-601_st	-1,70	9,98E-06	3,84E-04
hsa-miR-27b_st	2,77	1,14E-05	4,19E-04
hsa-miR-3198_st	-0,77	1,16E-05	4,19E-04
hsa-miR-188-5p_st	-3,00	1,24E-05	4,35E-04
hsa-miR-20b_st	3,21	1,62E-05	5,51E-04
hsa-miR-1975_st	-1,80	1,85E-05	6,10E-04

hsa-miR-30b_st	3,48	1,94E-05	6,22E-04
hsa-miR-27a-star_st	3,46	2,03E-05	6,34E-04
hsa-miR-15a_st	3,67	2,39E-05	7,06E-04
v11_hsa-miR-768-5p_st	-2,39	2,36E-05	7,06E-04
hsa-miR-422a_st	-1,86	3,41E-05	9,84E-04
hsa-miR-1180_st	-1,66	3,91E-05	1,10E-03
hsa-miR-1280_st	2,24	5,33E-05	1,46E-03
hsa-miR-29a_st	4,38	5,51E-05	1,47E-03
hsa-miR-4317_st	2,23	5,61E-05	1,47E-03
hsa-miR-1296_st	-2,69	6,00E-05	1,54E-03
hsa-miR-3178_st	1,73	6,26E-05	1,57E-03
hsa-miR-4298_st	-2,00	6,38E-05	1,57E-03
hsa-miR-619_st	-0,43	6,60E-05	1,59E-03
hsa-miR-2110_st	-2,42	6,79E-05	1,60E-03
hsa-miR-371-5p_st	-0,56	7,31E-05	1,69E-03
hsa-miR-15b_st	1,61	8,32E-05	1,88E-03
hsa-miR-125a-3p_st	-2,48	8,72E-05	1,94E-03
hsa-miR-24_st	1,20	9,66E-05	2,10E-03
hsa-miR-720_st	-1,81	1,04E-04	2,22E-03
hsa-miR-21_st	5,91	1,17E-04	2,40E-03
hsa-miR-224-star_st	-1,44	1,18E-04	2,40E-03
hsa-miR-324-5p_st	-1,70	1,18E-04	2,40E-03
hsv2-miR-H13_st	0,51	1,28E-04	2,54E-03
hsa-miR-595_st	-0,65	1,60E-04	3,12E-03
hsa-let-7a_st	1,12	1,90E-04	3,64E-03
hsa-miR-18b_st	2,64	1,97E-04	3,73E-03
hsa-let-7b_st	-1,68	2,02E-04	3,76E-03
hsa-miR-1306_st	-1,22	2,08E-04	3,81E-03
hsa-miR-20a_st	1,81	2,18E-04	3,81E-03
hsa-miR-30c_st	2,66	2,18E-04	3,81E-03
hsa-miR-769-3p_st	-2,59	2,13E-04	3,81E-03
hsa-miR-1298_st	-0,65	2,27E-04	3,92E-03
hsa-miR-92b_st	-2,36	2,35E-04	3,99E-03
hsa-miR-1275_st	-1,52	2,49E-04	4,16E-03
hsa-miR-4270_st	-2,94	2,60E-04	4,29E-03
hsa-miR-320d_st	-1,67	3,11E-04	4,98E-03
hsa-miR-760_st	-1,78	3,09E-04	4,98E-03
hsa-miR-378b_st	-0,72	3,58E-04	5,66E-03
hsa-miR-561_st	-0,42	3,73E-04	5,82E-03
hsa-miR-660_st	1,28	4,07E-04	6,26E-03
hsa-let-7f_st	3,65	4,24E-04	6,43E-03
hsa-miR-3127_st	-1,20	5,14E-04	7,70E-03
hsa-miR-1274a_st	2,15	5,21E-04	7,71E-03
hsa-let-7c_st	-1,36	5,53E-04	8,08E-03
hsa-miR-103_st	-0,45	5,84E-04	8,43E-03

hsa-miR-1287_st	0,91	6,02E-04	8,58E-03
hsa-miR-639_st	-0,69	6,42E-04	9,03E-03
hsa-miR-877_st	-1,97	6,91E-04	9,61E-03
hsa-miR-106b_st	1,80	7,25E-04	9,85E-03
hsa-miR-513c_st	-0,62	7,26E-04	9,85E-03
hsa-miR-424_st	0,82	7,63E-04	1,02E-02
hsa-miR-1273d_st	-0,85	9,20E-04	1,22E-02
hsa-miR-532-3p_st	-2,71	9,33E-04	1,22E-02
hsa-miR-589-star_st	0,78	9,60E-04	1,24E-02
hsa-miR-125a-5p_st	-1,44	1,12E-03	1,39E-02
hsa-miR-27b-star_st	1,69	1,10E-03	1,39E-02
hsa-miR-30e_st	1,59	1,10E-03	1,39E-02
hsa-miR-92a_st	-1,35	1,12E-03	1,39E-02
hsa-miR-99b_st	-1,37	1,16E-03	1,42E-02
hsa-miR-886-5p_st	5,50	1,22E-03	1,49E-02
hsa-miR-365-star_st	-1,91	1,32E-03	1,57E-02
hsa-miR-378_st	-1,87	1,31E-03	1,57E-02
hsa-miR-1207-5p_st	-1,99	1,51E-03	1,77E-02
hsa-miR-3130-3p_st	-0,49	1,52E-03	1,77E-02
hsa-miR-3141_st	-1,78	1,54E-03	1,77E-02
hsa-miR-100_st	6,70	1,60E-03	1,82E-02
hsa-miR-2276_st	-0,50	1,63E-03	1,84E-02
hsa-miR-210_st	-2,23	1,65E-03	1,85E-02
hsa-miR-545-star_st	0,53	1,69E-03	1,87E-02
hsa-miR-26a-2-star_st	-0,28	1,94E-03	2,14E-02
hsa-miR-4284_st	1,60	1,99E-03	2,16E-02
hsa-miR-886-3p_st	2,48	2,01E-03	2,16E-02
hsa-miR-378c_st	-1,61	2,21E-03	2,36E-02
hsa-miR-501-5p_st	-1,34	2,31E-03	2,44E-02
hsa-miR-221-star_st	0,63	2,51E-03	2,64E-02
hsa-miR-107_st	-0,46	2,57E-03	2,67E-02
hsa-miR-220b_st	-0,31	2,61E-03	2,69E-02
hsa-miR-193b-star_st	-3,24	2,70E-03	2,76E-02
hsa-miR-148b_st	1,61	2,79E-03	2,83E-02
hsa-miR-191-star_st	0,84	3,18E-03	3,19E-02
hsa-miR-206_st	-0,38	3,23E-03	3,21E-02
hsa-miR-935_st	1,30	3,32E-03	3,28E-02
hsa-miR-192_st	2,02	3,40E-03	3,33E-02
hsa-miR-3136_st	-1,05	3,58E-03	3,47E-02
hsa-miR-106a_st	1,34	3,73E-03	3,54E-02
hsa-miR-1274b_st	0,84	3,71E-03	3,54E-02
hsa-miR-18a_st	2,10	3,75E-03	3,54E-02
hsa-miR-301a_st	2,75	3,80E-03	3,54E-02
hsa-miR-586_st	0,40	3,83E-03	3,54E-02
hsa-miR-938_st	-0,37	3,80E-03	3,54E-02

hsa-miR-337-3p_st	-0,56	3,99E-03	3,65E-02
hsa-miR-10a_st	2,36	4,07E-03	3,70E-02
hsa-miR-493_st	0,32	4,11E-03	3,71E-02
hsa-miR-1229_st	-1,27	4,21E-03	3,74E-02
hsa-miR-766_st	-1,84	4,18E-03	3,74E-02
hsa-miR-1539_st	-0,56	4,26E-03	3,75E-02
hsa-miR-4288_st	0,92	4,35E-03	3,81E-02
hsa-miR-4296_st	-0,62	4,51E-03	3,91E-02
hsa-miR-1292_st	-1,37	4,61E-03	3,94E-02
hsa-miR-139-5p_st	1,69	4,58E-03	3,94E-02
hsa-miR-17_st	1,28	4,72E-03	4,00E-02
hsa-miR-340-star_st	-0,29	4,78E-03	4,03E-02
hsa-let-7e_st	-1,09	4,93E-03	4,09E-02
hsa-miR-146a_st	5,02	4,90E-03	4,09E-02
hsa-miR-19a_st	1,69	5,23E-03	4,29E-02
hsa-miR-2052_st	-0,35	5,24E-03	4,29E-02
hsa-let-7g_st	2,06	5,31E-03	4,31E-02
hsa-miR-1973_st	-1,52	5,38E-03	4,34E-02
hsa-miR-1226_st	-1,44	5,64E-03	4,43E-02
hsa-miR-1277_st	0,43	5,58E-03	4,43E-02
hsa-miR-3149_st	-0,33	5,53E-03	4,43E-02
hsa-miR-345_st	-2,52	5,64E-03	4,43E-02
hsa-miR-181b_st	-1,32	5,86E-03	4,57E-02
hsa-miR-519b-5p_st	-0,38	5,98E-03	4,63E-02
hsa-miR-3173_st	-0,89	6,11E-03	4,70E-02
hsa-miR-3162_st	-2,47	6,19E-03	4,73E-02
hsa-let-7i_st	2,40	6,55E-03	4,90E-02
hsa-miR-3126-3p_st	0,35	6,59E-03	4,90E-02
hsa-miR-421_st	2,37	6,48E-03	4,90E-02
hsv2-miR-H6_st	-1,71	6,58E-03	4,90E-02

CLVY: breast cancer cell lines from young women; CLO: breast cancer cell lines from old women.

ANNEX II. List of miRNAs from selected sub-nodes in breast cancer cell lines study**.**

Node 1	Node 2	Node 3	Node 4
hsa-miR-193b-5p	hsa-miR-146a-5p	hsa-miR-92a-3p	hsa-miR-1292-5p
hsa-miR-3162-5p	hsa-miR-100-5p	hsa-let-7c-5p	hsa-miR-188-5p
hsa-miR-125a-3p	hsa-miR-21-5p	hsa-let-7e-5p	hsa-miR-378a-5p
hsa-miR-639	hsa-miR-19b-3p	hsa-miR-320c	hsa-miR-484
hsa-miR-378b	hsa-miR-125b-5p	hsa-miR-320b	hsa-miR-532-3p
hsa-miR-2110	hsa-miR-10a-5p	hsa-miR-320a	hsa-miR-766-3p
hsa-miR-422a	hsa-miR-24-2-5p	hsa-miR-181b-5p	hsa-miR-365b-5p
hsa-miR-760	hsa-miR-126-3p	hsa-miR-320d	hsa-miR-1973
hsa-miR-345-5p	hsa-miR-4317	hsa-miR-99b-5p	hsa-miR-1296-5p
hsa-miR-501-5p	hsa-miR-18b-5p		hsa-miR-769-3p
hsa-miR-378c	hsa-miR-27a-5p		hsa-miR-874-3p
hsa-miR-197-3p	hsa-miR-660-5p		hsa-miR-328-3p
hsa-miR-99b-3p	hsa-miR-1260b		
hsa-miR-324-3p	hsa-miR-99a-5p		
hsa-miR-671-5p	hsa-miR-15a-5p		
hsa-miR-3141	hsa-miR-30b-5p		
hsa-miR-4298	hsa-miR-18a-5p		
hsa-miR-574-3p	hsa-miR-20b-5p		
hsa-miR-744-5p	hsa-miR-29a-3p		
hsa-miR-423-3p	hsa-miR-421		
hsa-miR-423-5p	hsa-miR-4284		
hsa-miR-210-3p	hsa-let-7f-5p		
hsa-miR-320e	hsa-miR-27b-3p		
hsa-miR-92b-3p	hsa-let-7g-5p		
hsa-miR-1207-5p	hsa-miR-30c-5p		
hsa-miR-4270			
hsa-miR-1275			
hsa-miR-877-5p			
hsa-miR-1180-3p			
hsa-miR-1307-3p			
hsa-miR-324-5p			

ANNEX III. Pathway enrichment results for significant miRNA by selected sub-nodes from the breast cancer cell lines study.

Node 1		
KEGG pathway	**p-value**	**miRNAs count**
Glycosaminoglycan biosynthesis - heparan sulfate heparin	5,96E-04	12
TGF-beta signaling pathway	5,96E-04	19
Glioma	5,96E-04	20
Signaling pathways regulating pluripotency of stem cells	5,96E-04	22
Regulation of actin cytoskeleton	5,96E-04	22
Endocytosis	5,96E-04	23
Proteoglycans in cancer	5,96E-04	23
Thyroid hormone signaling pathway	6,59E-04	20
Adherens junction	1,32E-03	20
FoxO signaling pathway	1,51E-03	18
Viral carcinogenesis	1,72E-03	20
Glycosphingolipid biosynthesis - ganglio series	2,82E-03	6
Thyroid cancer	4,04E-03	13
Melanoma	4,04E-03	19
Focal adhesion	4,04E-03	23
Pathways in cancer	4,04E-03	24
Chronic myeloid leukemia	4,52E-03	18
Prostate cancer	4,52E-03	20
Phosphatidylinositol signaling system	4,52E-03	22
MAPK signaling pathway	4,52E-03	24
Choline metabolism in cancer	4,59E-03	18
Bacterial invasion of epithelial cells	4,59E-03	21
Oxytocin signaling pathway	5,43E-03	21
Adrenergic signaling in cardiomyocytes	5,65E-03	17
ErbB signaling pathway	9,13E-03	19
Central carbon metabolism in cancer	1,43E-02	16
Calcium signaling pathway	1,43E-02	18
Synaptic vesicle cycle	1,73E-02	14
Melanogenesis	2,06E-02	16
HIF-1 signaling pathway	2,06E-02	18
Nicotine addiction	2,10E-02	15
Pancreatic cancer	2,22E-02	16
Long-term potentiation	2,65E-02	15
Ras signaling pathway	3,48E-02	23
Insulin signaling pathway	3,49E-02	17
Node 2		
KEGG pathway	**p-value**	**miRNAs count**
ECM-receptor interaction	6,10E-27	19
Fatty acid biosynthesis	5,64E-11	8

Proteoglycans in cancer	1,43E-10	23
Mucin type O-Glycan biosynthesis	2,36E-08	14
Lysine degradation	3,88E-08	17
Signaling pathways regulating pluripotency of stem cells	2,33E-07	23
Glioma	5,18E-06	23
Focal adhesion	8,66E-06	24
mTOR signaling pathway	9,18E-06	22
PI3K-Akt signaling pathway	9,18E-06	24
Glutamatergic synapse	1,95E-05	20
MAPK signaling pathway	8,29E-05	22
Estrogen signaling pathway	8,77E-05	20
Neurotrophin signaling pathway	2,02E-04	25
Morphine addiction	2,20E-04	19
Wnt signaling pathway	2,64E-04	24
Amoebiasis	3,28E-04	18
Renal cell carcinoma	3,41E-04	22
Melanoma	5,13E-04	19
Rap1 signaling pathway	5,64E-04	23
Ras signaling pathway	6,95E-04	24
Axon guidance	8,88E-04	19
Platelet activation	9,85E-04	22
Pathways in cancer	1,35E-03	24
FoxO signaling pathway	1,85E-03	21
Biotin metabolism	1,86E-03	2
TGF-beta signaling pathway	2,63E-03	18
Insulin signaling pathway	2,63E-03	24
Retrograde endocannabinoid signaling	3,07E-03	21
Glycosaminoglycan biosynthesis - heparan sulfate	3,90E-03	11
GABAergic synapse	3,90E-03	19
T cell receptor signaling pathway	3,90E-03	20
Hippo signaling pathway	3,90E-03	23
AMPK signaling pathway	4,34E-03	24
ErbB signaling pathway	4,83E-03	22
Prostate cancer	5,42E-03	22
Non-small cell lung cancer	8,34E-03	20
Alanine aspartate and glutamate metabolism	9,26E-03	15
Oxytocin signaling pathway	9,26E-03	21
Adrenergic signaling in cardiomyocytes	9,67E-03	19
Choline metabolism in cancer	1,00E-02	22
Gap junction	1,03E-02	17
Progesterone-mediated oocyte maturation	1,03E-02	19
N-Glycan biosynthesis	1,05E-02	14
Cholinergic synapse	1,05E-02	19
Regulation of actin cytoskeleton	1,05E-02	21
cGMP-PKG signaling pathway	1,13E-02	21

Prolactin signaling pathway	1,34E-02	21
B cell receptor signaling pathway	1,48E-02	19
Small cell lung cancer	1,48E-02	19
p53 signaling pathway	2,16E-02	18
Long-term depression	2,20E-02	16
Long-term potentiation	2,20E-02	22
Thyroid hormone signaling pathway	2,20E-02	22
Circadian entrainment	2,58E-02	19
Protein processing in endoplasmic reticulum	2,82E-02	22
Fc epsilon RI signaling pathway	2,89E-02	19
GnRH signaling pathway	3,19E-02	20
Chronic myeloid leukemia	3,37E-02	20
Melanogenesis	3,52E-02	18
Glycosaminoglycan biosynthesis - keratan sulfate	3,73E-02	7
Fatty acid metabolism	3,73E-02	13
Sphingolipid signaling pathway	3,73E-02	16
Dilated cardiomyopathy	3,73E-02	16
Protein digestion and absorption	4,50E-02	17
Endometrial cancer	4,78E-02	19
Pancreatic cancer	4,84E-02	19
Node 3		
KEGG pathway	**p-value**	**miRNAs count**
Thyroid cancer	9,21E-03	7
ErbB signaling pathway	9,21E-03	11
Hippo signaling pathway	1,12E-02	8
N-Glycan biosynthesis	1,48E-02	6
Cell adhesion molecules (CAMs)	2,34E-02	10
Amphetamine addiction	3,04E-02	8
Phosphatidylinositol signaling system	4,68E-02	8
Cluster 4		
KEGG pathway	**p-value**	**miRNAs count**
ECM-receptor interaction	3,59E-07	8
Glioma	6,97E-06	8
Proteoglycans in cancer	1,25E-05	8
Signaling pathways regulating pluripotency of stem cells	5,05E-04	8
Chronic myeloid leukemia	6,15E-04	8
FoxO signaling pathway	1,07E-03	8
PI3K-Akt signaling pathway	1,07E-03	8
Transcriptional misregulation in cancer	1,07E-03	8
Focal adhesion	1,83E-03	8
Pathways in cancer	2,04E-03	9
Thyroid cancer	2,05E-03	8
Endometrial cancer	2,63E-03	8

Prostate cancer	3,13E-03	8
Melanoma	5,62E-03	8
mTOR signaling pathway	9,93E-03	8
Wnt signaling pathway	1,08E-02	8
Acute myeloid leukemia	1,10E-02	8
Prolactin signaling pathway	1,21E-02	8
Colorectal cancer	1,21E-02	8
Circadian rhythm	1,53E-02	8
Estrogen signaling pathway	1,54E-02	8
Viral carcinogenesis	1,64E-02	8
Glycosaminoglycan biosynthesis - heparan sulfate heparin	2,12E-02	8
Melanogenesis	2,12E-02	8
Hypertrophic cardiomyopathy (HCM)	2,15E-02	7
AMPK signaling pathway	2,62E-02	8
Regulation of actin cytoskeleton	2,62E-02	8
Non-small cell lung cancer	2,80E-02	8
Ras signaling pathway	2,83E-02	9
ErbB signaling pathway	3,41E-02	8
Long-term potentiation	3,48E-02	8
Dorso-ventral axis formation	3,59E-02	8
Renal cell carcinoma	3,59E-02	8
Axon guidance	3,61E-02	8
Neurotrophin signaling pathway	3,61E-02	8
Pancreatic cancer	3,80E-02	8
Adherens junction	4,60E-02	8

ANNEX IV. Significant miRNAs differentially expressed between breast cancer age groups obtained in the combined study with breast cancer samples from cell lines and FFPE patient tumours.

	Mean expression differences (Old -Young)	p-value	FDR adjusted p-value
hsa-miR-23a_st	1,27	2,15E-08	1,24E-05
hsa-miR-24-2-star_st	2,17	1,89E-08	1,24E-05
hsa-miR-1207-5p_st	-1,62	3,93E-07	1,03E-04
hsa-miR-27b-star_st	1,25	4,62E-07	1,03E-04
hsa-miR-4270_st	-1,99	6,25E-07	1,03E-04
hsa-miR-4317_st	1,22	3,04E-07	1,03E-04
v11_hsa-miR-768-3p_st	-1,72	5,89E-07	1,03E-04
v11_hsa-miR-923_st	-1,57	1,18E-06	1,70E-04
hsa-miR-4281_st	-1,91	1,93E-06	2,47E-04
hsa-miR-132_st	1,57	3,05E-06	3,52E-04
hsa-miR-149-star_st	-1,78	5,59E-06	5,87E-04
hsa-miR-27a-star_st	1,95	6,90E-06	6,64E-04
hsa-miR-23b_st	0,62	8,98E-06	7,41E-04
hsa-miR-29b-1-star_st	2,10	8,89E-06	7,41E-04
hsa-miR-24_st	0,77	1,54E-05	1,04E-03
hsa-miR-27a_st	2,03	1,53E-05	1,04E-03
hsa-miR-27b_st	1,58	1,54E-05	1,04E-03
hsa-miR-423-5p_st	-1,39	1,94E-05	1,24E-03
hsa-miR-762_st	-1,59	2,29E-05	1,39E-03
hsa-miR-193a-5p_st	-2,35	3,31E-05	1,91E-03
hsa-miR-220b_st	-0,39	3,76E-05	1,96E-03
hsa-miR-3141_st	-1,35	3,90E-05	1,96E-03
hsa-miR-92b-star_st	-1,18	3,83E-05	1,96E-03
hsa-let-7a-2-star_st	0,62	6,60E-05	3,08E-03
hsa-miR-660_st	1,06	6,68E-05	3,08E-03
hsv1-miR-H6-5p_st	-1,77	8,22E-05	3,65E-03
hsa-miR-30c_st	1,71	8,61E-05	3,68E-03
hsa-miR-3162_st	-1,62	9,17E-05	3,76E-03
hsa-miR-548x_st	-0,38	9,45E-05	3,76E-03
hsa-miR-939_st	-1,27	1,16E-04	4,45E-03
hsa-miR-30e-star_st	1,07	1,23E-04	4,58E-03
hsa-miR-1275_st	-1,22	1,45E-04	4,82E-03
hsa-miR-183_st	2,15	1,50E-04	4,82E-03
hsa-miR-1973_st	-1,42	1,47E-04	4,82E-03
hsa-miR-518b_st	-0,24	1,43E-04	4,82E-03
hsa-miR-615-5p_st	-0,57	1,43E-04	4,82E-03
hsa-miR-15b_st	1,31	1,94E-04	5,45E-03
hsa-miR-20b_st	1,62	1,94E-04	5,45E-03
hsa-miR-545-star_st	0,25	1,88E-04	5,45E-03

hsv2-miR-H22_st	-1,48	1,90E-04	5,45E-03
hsv2-miR-H6_st	-1,26	1,83E-04	5,45E-03
hsa-miR-19a_st	1,03	2,11E-04	5,80E-03
hsa-miR-92a-1-star_st	1,17	2,41E-04	6,47E-03
hsa-miR-421_st	1,65	2,57E-04	6,75E-03
hsa-miR-720_st	-2,40	2,68E-04	6,87E-03
hsa-miR-3153_st	-0,47	2,89E-04	7,26E-03
hsa-miR-4299_st	-1,27	3,15E-04	7,74E-03
hsa-miR-1287_st	0,56	3,25E-04	7,82E-03
hsa-miR-1202_st	-1,34	3,48E-04	8,20E-03
hsa-miR-1228_st	-0,81	3,82E-04	8,65E-03
hsa-miR-3196_st	-1,25	3,75E-04	8,65E-03
hsa-miR-1268_st	-1,27	3,92E-04	8,70E-03
hsa-miR-19b_st	2,21	4,12E-04	8,97E-03
hsa-miR-153_st	0,28	4,43E-04	9,47E-03
hsa-miR-940_st	-0,89	5,14E-04	1,06E-02
hsv2-miR-H10_st	-1,23	5,11E-04	1,06E-02
hsv1-miR-H17_st	-1,16	5,25E-04	1,06E-02
hsa-miR-135a-star_st	-0,48	5,34E-04	1,06E-02
hsa-miR-30a-star_st	2,16	5,66E-04	1,10E-02
hsv2-miR-H5_st	-1,65	5,69E-04	1,10E-02
hsa-miR-148b_st	1,03	6,07E-04	1,13E-02
hsa-miR-15a_st	1,77	6,25E-04	1,13E-02
hsa-miR-1975_st	-1,28	6,32E-04	1,13E-02
hsa-miR-29a_st	1,84	6,37E-04	1,13E-02
hsa-miR-3175_st	-1,42	6,10E-04	1,13E-02
hsa-miR-1228-star_st	-1,43	6,49E-04	1,14E-02
hsa-miR-589-star_st	0,55	7,14E-04	1,23E-02
hsa-miR-18b_st	1,46	7,42E-04	1,26E-02
hsa-miR-301a_st	1,45	7,70E-04	1,29E-02
hsa-miR-3147_st	-0,78	7,99E-04	1,32E-02
hsa-miR-1207-3p_st	-0,25	8,36E-04	1,34E-02
hsa-miR-208b_st	0,24	8,32E-04	1,34E-02
hsa-miR-936_st	-0,68	8,63E-04	1,36E-02
hsa-miR-128_st	1,30	9,00E-04	1,38E-02
hsa-miR-1288_st	-0,21	8,92E-04	1,38E-02
hsa-miR-216b_st	0,45	9,36E-04	1,42E-02
hsa-miR-4292_st	-0,38	1,05E-03	1,58E-02
hsa-miR-192_st	1,23	1,11E-03	1,64E-02
hsa-miR-513a-5p_st	-0,81	1,27E-03	1,85E-02
hsa-miR-139-5p_st	1,30	1,35E-03	1,94E-02
hsa-miR-424_st	0,37	1,36E-03	1,94E-02
hsa-miR-3198_st	-0,31	1,44E-03	2,03E-02
hsa-miR-548a-5p_st	0,37	1,49E-03	2,08E-02
hsa-miR-1249_st	-0,48	1,54E-03	2,12E-02

hsa-miR-125b-1-star_st	1,17	1,65E-03	2,23E-02
hsa-miR-181d_st	1,22	1,88E-03	2,47E-02
hsa-miR-1908_st	-1,12	1,90E-03	2,47E-02
hsa-miR-499-3p_st	0,25	1,86E-03	2,47E-02
hsa-miR-665_st	-1,00	1,90E-03	2,47E-02
hsa-miR-3156_st	-0,34	1,95E-03	2,50E-02
hsa-miR-28-5p_st	0,88	1,99E-03	2,53E-02
hsa-miR-3154_st	-0,80	2,06E-03	2,55E-02
hsa-miR-3197_st	-1,43	2,04E-03	2,55E-02
hsa-miR-16-1-star_st	0,38	2,12E-03	2,61E-02
hsa-miR-2861_st	-1,15	2,23E-03	2,65E-02
hsa-miR-371-5p_st	-0,67	2,21E-03	2,65E-02
hsv2-miR-H3_st	-0,42	2,23E-03	2,65E-02
hsa-miR-299-5p_st	-0,27	2,38E-03	2,80E-02
hsa-miR-575_st	-0,42	2,48E-03	2,84E-02
hsa-miR-638_st	-1,08	2,47E-03	2,84E-02
hsv2-miR-H25_st	-1,12	2,45E-03	2,84E-02
hsa-miR-29c_st	0,87	2,62E-03	2,97E-02
hsa-miR-93-star_st	1,31	2,72E-03	3,05E-02
hsa-miR-494_st	-1,26	2,88E-03	3,20E-02
hsa-miR-100_st	2,44	3,04E-03	3,30E-02
hsa-miR-4276_st	-0,24	3,06E-03	3,30E-02
hsa-miR-4297_st	-0,27	3,05E-03	3,30E-02
hsa-miR-29b_st	0,87	3,12E-03	3,33E-02
hsa-miR-3177_st	-0,50	3,29E-03	3,48E-02
hsv2-miR-H23-star_st	-0,24	3,36E-03	3,53E-02
hsa-miR-1322_st	-0,18	3,55E-03	3,69E-02
hsa-miR-146a-star_st	0,20	3,63E-03	3,70E-02
hsa-miR-4288_st	0,44	3,59E-03	3,70E-02
hsa-miR-192-star_st	0,26	3,75E-03	3,80E-02
hsa-miR-1225-5p_st	-1,16	3,80E-03	3,81E-02
hsa-miR-23a-star_st	1,33	3,88E-03	3,86E-02
hsa-miR-1182_st	-0,53	4,16E-03	4,10E-02
hsa-miR-30b_st	1,42	4,26E-03	4,14E-02
hsa-miR-4323_st	-0,37	4,27E-03	4,14E-02
hsa-miR-1915_st	-1,01	4,37E-03	4,17E-02
hsa-miR-223-star_st	0,26	4,37E-03	4,17E-02
hsa-miR-29c-star_st	0,61	4,45E-03	4,21E-02
hsa-miR-4310_st	-0,41	4,52E-03	4,24E-02
hsa-miR-18a_st	1,73	4,81E-03	4,48E-02
hsa-miR-188-3p_st	-0,32	5,03E-03	4,57E-02
hsa-miR-571_st	0,17	5,02E-03	4,57E-02
hsv2-miR-H20_st	0,33	4,96E-03	4,57E-02
hsa-miR-374b_st	0,51	5,23E-03	4,64E-02
hsa-miR-4296_st	-0,28	5,23E-03	4,64E-02

hsa-miR-935_st	0,67	5,21E-03	4,64E-02
hsa-miR-505_st	0,54	5,32E-03	4,69E-02
hsa-miR-150-star_st	-0,92	5,69E-03	4,98E-02

ANNEX V. List of miRNAs from selected sub-nodes in the combined study with breast cancer cell lines and patients.

Node 1	Node 2
hsa-miR-18a-5p	hsa-miR-301a-3p
hsa-miR-183-5p	hsa-miR-148b-3p
hsa-miR-15a-5p	hsa-miR-192-5p
hsa-miR-421	hsa-miR-29b-3p
hsa-miR-181d-5p	hsa-miR-29c-5p
hsa-miR-4317	hsa-miR-29c-3p
hsa-miR-20b-5p	hsa-miR-30e-3p
hsa-miR-19b-3p	hsa-miR-935
hsa-miR-30b-5p	hsa-miR-505-3p
hsa-miR-660-5p	hsa-miR-1287-5p
hsa-miR-132-3p	hsa-miR-374b-5p
hsa-miR-28-5p	hsa-miR-589-3p
hsa-miR-100-5p	hsa-let-7a-2-3p
hsa-miR-15b-5p	hsa-miR-4288
hsa-miR-27a-3p	hsa-miR-216b-5p
hsa-miR-29a-3p	hsa-miR-16-1-3p
hsa-miR-27b-3p	
hsa-miR-30c-5p	

ANNEX VI. Pathway enrichment results for significant miRNA by selected sub-nodes in the breast cancer cell lines study.

Node 1		
KEGG pathway	**p-value**	**miRNAs count**
ECM-receptor interaction	4,94E-35	15
Proteoglycans in cancer	7,82E-12	18
Fatty acid biosynthesis	1,64E-11	8
Focal adhesion	4,09E-07	18
Morphine addiction	4,25E-07	17
Signaling pathways regulating pluripotency of stem cells	1,53E-06	18
Wnt signaling pathway	3,59E-06	18
Lysine degradation	6,35E-06	14
Glutamatergic synapse	7,69E-06	17
GABAergic synapse	9,24E-06	17
Gap junction	1,08E-05	17
PI3K-Akt signaling pathway	1,29E-05	18
Estrogen signaling pathway	2,20E-05	17
Glioma	2,29E-05	17
mTOR signaling pathway	2,88E-05	17
Pathways in cancer	2,88E-05	18
Renal cell carcinoma	3,19E-05	18
Amoebiasis	8,36E-05	14
Axon guidance	8,83E-05	15
Retrograde endocannabinoid signaling	1,30E-04	17
TGF-beta signaling pathway	1,47E-04	13
FoxO signaling pathway	2,35E-04	17
Hippo signaling pathway	2,35E-04	18
Melanoma	3,41E-04	15
Circadian entrainment	7,09E-04	17
Rap1 signaling pathway	1,23E-03	18
Mucin type O-Glycan biosynthesis	1,46E-03	11
Cholinergic synapse	1,50E-03	16
AMPK signaling pathway	1,50E-03	18
Platelet activation	1,50E-03	18
Prostate cancer	1,88E-03	17
Melanogenesis	2,82E-03	15
Regulation of actin cytoskeleton	2,86E-03	17
Alanine aspartate and glutamate metabolism	2,96E-03	19
cGMP-PKG signaling pathway	3,56E-03	17
Non-small cell lung cancer	3,81E-03	17
Progesterone-mediated oocyte maturation	3,99E-03	16
Long-term depression	4,78E-03	16
Prolactin signaling pathway	4,78E-03	17
ErbB signaling pathway	4,78E-03	18

Choline metabolism in cancer	5,10E-03	18
Long-term potentiation	6,29E-03	18
Protein digestion and absorption	7,10E-03	14
Vascular smooth muscle contraction	7,10E-03	17
Ras signaling pathway	7,10E-03	18
Adrenergic signaling in cardiomyocytes	7,56E-03	16
GnRH signaling pathway	7,56E-03	17
Endometrial cancer	8,44E-03	17
MAPK signaling pathway	8,94E-03	18
T cell receptor signaling pathway	9,24E-03	17
Neurotrophin signaling pathway	9,24E-03	18
Small cell lung cancer	1,06E-02	15
Oocyte meiosis	1,14E-02	17
Oxytocin signaling pathway	1,27E-02	17
Ubiquitin mediated proteolysis	1,81E-02	17
Thyroid hormone signaling pathway	1,84E-02	18
Glycosaminoglycan biosynthesis - heparan sulfate heparin	1,95E-02	9
Cocaine addiction	2,05E-02	17
B cell receptor signaling pathway	2,33E-02	16
Fatty acid metabolism	2,35E-02	10
Gastric acid secretion	2,50E-02	16
Dopaminergic synapse	2,52E-02	17
Colorectal cancer	2,54E-02	15
Protein processing in endoplasmic reticulum	2,64E-02	17
Acute myeloid leukemia	2,65E-02	17
cAMP signaling pathway	2,91E-02	18
p53 signaling pathway	3,09E-02	16
Insulin signaling pathway	3,12E-02	18
Sphingolipid signaling pathway	3,37E-02	14
Pancreatic cancer	3,87E-02	15
Node 2		
KEGG pathway	**p-value**	**miRNAs count**
ECM-receptor interaction	3,14E-34	13
Prion diseases	2,42E-10	9
Proteoglycans in cancer	1,35E-09	13
TGF-beta signaling pathway	5,65E-08	12
Pathways in cancer	5,93E-08	14
Lysine degradation	7,17E-08	13
Focal adhesion	2,91E-07	14
Signaling pathways regulating pluripotency of stem cells	3,54E-07	15
Amoebiasis	5,06E-06	12
GABAergic synapse	5,06E-06	14
Wnt signaling pathway	1,38E-05	15
Prostate cancer	1,73E-05	14

Arrhythmogenic right ventricular cardiomyopathy (ARVC)	8,44E-05	14
Morphine addiction	9,46E-05	13
PI3K-Akt signaling pathway	1,80E-04	14
Glioma	3,00E-04	11
Estrogen signaling pathway	3,00E-04	13
Colorectal cancer	3,25E-04	13
Circadian rhythm	4,48E-04	12
FoxO signaling pathway	4,96E-04	13
Ras signaling pathway	3,03E-03	13
Small cell lung cancer	3,66E-03	14
Choline metabolism in cancer	4,09E-03	13
Rap1 signaling pathway	5,94E-03	14
Sulfur metabolism	7,22E-03	6
Hedgehog signaling pathway	1,14E-02	10
Melanoma	1,14E-02	12
Pancreatic cancer	1,14E-02	13
Renal cell carcinoma	1,29E-02	13
Transcriptional misregulation in cancer	1,29E-02	14
Chronic myeloid leukemia	1,31E-02	13
Glycosaminoglycan biosynthesis - heparan sulfate	1,34E-02	8
Regulation of actin cytoskeleton	1,49E-02	14
MAPK signaling pathway	1,49E-02	15
cAMP signaling pathway	1,84E-02	14
Endocytosis	1,86E-02	14
Hippo signaling pathway	1,89E-02	13
Basal cell carcinoma	2,70E-02	12
Long-term potentiation	3,20E-02	12
p53 signaling pathway	3,70E-02	13
Endometrial cancer	4,32E-02	13

ANNEX VII. Statistical studies for distinctive hypermethylated CpGs in BCVY from metEPICVal study.

	Mean methylation BCVY	Mean methylation BCO	Wilcoxon rank sum test BCVY vs BCO p-value	Generalized linear model ~ age BCVY vs BCO p-value	Gene region
cg00049892	0,54	0,42	4,61E-02	6,14E-02	*ENPP1*
cg03340334	0,42	0,54	2,18E-02	3,33E-02	*MIR1268A*
cg19844232	0,62	0,74	1,04E-02	4,31E-02	*ASAP1*
cg23093939	0,36	0,48	2,65E-02	4,34E-02	
cg16731083	0,43	0,57	2,41E-02	3,65E-02	
cg12493813	0,67	0,54	1,44E-02	2,23E-02	*CCDC170*
cg08048895	0,60	0,49	2,18E-02	6,10E-02	*AUTS2*
cg15564444	0,45	0,57	2,65E-02	5,65E-02	*IGSF21*
cg20956407	0,71	0,58	3,52E-02	2,70E-02	*NDST1*
cg04903269	0,59	0,72	4,22E-02	5,57E-02	*CELF6*
cg06573181	0,63	0,52	2,65E-02	2,96E-02	*KCNQ5*
cg07215395	0,69	0,56	1,04E-02	2,95E-02	*CDO1*
cg10006933	0,21	0,37	3,86E-02	3,83E-02	*BRINP3*
cg13284574	0,46	0,58	2,41E-02	3,60E-02	*ESPN*
cg25563198	0,51	0,40	3,21E-02	7,62E-02	*FKBP5; LOC285847*
cg14155265	0,64	0,51	1,97E-02	2,65E-02	*AK4*
cg15988937	0,73	0,62	4,61E-02	4,47E-02	
cg10165543	0,38	0,50	2,65E-02	4,39E-02	*KALRN*
cg22514173	0,64	0,76	1,78E-02	3,93E-02	
cg16716542	0,64	0,51	1,60E-02	2,54E-02	*CDC42BPB*
cg22006449	0,77	0,65	2,65E-02	5,00E-02	*PPP2R1B*
cg12426652	0,47	0,60	4,22E-02	5,57E-02	*CLDN14*
cg08206482	0,36	0,52	4,02E-03	1,05E-02	*ITGAD*
cg25093492	0,47	0,34	2,92E-02	4,56E-02	*PLCH1*
cg04478905	0,54	0,64	4,22E-02	7,01E-02	*CLDN14*
cg17367688	0,72	0,61	3,52E-02	5,37E-02	
cg07255909	0,63	0,49	1,16E-02	1,90E-02	
cg07070848	0,83	0,72	1,44E-02	3,83E-02	*PPP3CA*
cg22652022	0,55	0,42	3,52E-02	5,31E-02	*EXOC4*
cg22542373	0,40	0,51	9,30E-03	2,24E-02	
cg24724990	0,71	0,60	1,97E-02	5,62E-02	*TMEM123*
cg12033458	0,75	0,60	2,39E-03	2,33E-02	*UNC5C*
cg15863598	0,72	0,59	9,30E-03	1,64E-02	*MYST4*
cg10537261	0,46	0,36	9,30E-03	1,82E-02	
cg06620896	0,20	0,32	1,44E-02	2,51E-02	
cg17371203	0,69	0,59	2,18E-02	3,55E-02	*SEMA6B*
cg19070079	0,67	0,56	1,16E-02	1,79E-02	*LOC101926975*
cg04315336	0,64	0,52	2,92E-02	3,11E-02	*THSD4*
cg10777462	0,57	0,43	3,11E-03	1,37E-02	

cg27211179	0,58	0,47	3,86E-02	1,90E-02	*EXD2*
cg13314012	0,72	0,60	3,86E-02	5,78E-02	*LINC01170*
cg10953326	0,66	0,54	7,39E-03	1,59E-02	
cg19910976	0,63	0,52	1,60E-02	3,01E-02	
cg24868292	0,63	0,50	1,30E-02	2,85E-02	*DNASE2B*
cg24765446	0,55	0,44	3,21E-02	4,15E-02	*WFDC6*
cg13244293	0,40	0,30	1,02E-03	7,91E-03	*NGLY1; OXSM*
cg19664838	0,81	0,67	3,52E-02	3,49E-02	
cg15522953	0,74	0,58	5,52E-04	3,25E-03	*ZMIZ1*
cg10807084	0,44	0,56	4,61E-02	3,97E-02	*AK5*
cg20045300	0,60	0,46	1,97E-02	2,30E-02	*FAM114A1*
cg05545915	0,56	0,72	1,30E-02	3,09E-02	
cg10350446	0,65	0,54	9,30E-03	1,79E-02	
cg07035145	0,07	0,24	1,82E-03	3,18E-03	*PYGL*
cg16247085	0,47	0,37	3,52E-02	7,78E-02	*LSAMP*
cg02855302	0,63	0,52	4,22E-02	4,79E-02	*CDC42BPB*
cg14022913	0,44	0,31	2,65E-02	6,26E-02	*NEIL3*
cg09039163	0,59	0,70	4,22E-02	6,63E-02	*ESPNL*
cg16333846	0,70	0,59	2,18E-02	2,43E-02	*TIAM2*
cg16172329	0,76	0,62	6,57E-03	1,64E-02	*DPP4*
cg23633671	0,64	0,52	2,39E-03	1,25E-02	
cg21108805	0,39	0,27	1,58E-03	9,57E-03	*WIF1*
cg01334432	0,63	0,52	4,61E-02	9,04E-02	*LOC642587; MIR205*
cg14678654	0,47	0,58	1,60E-02	3,44E-02	
cg03522150	0,64	0,53	1,60E-02	4,42E-02	*C11orf49*
cg02772619	0,72	0,61	9,30E-03	2,54E-02	
cg18818644	0,68	0,56	1,97E-02	3,35E-02	
cg02078468	0,81	0,69	9,30E-03	1,26E-02	*CERS2*
cg12728465	0,59	0,47	1,30E-02	1,98E-02	*ZKSCAN1*
cg05867326	0,52	0,41	3,86E-02	3,75E-02	
cg23925996	0,20	0,30	2,18E-02	1,96E-02	*ARL5B*
cg19839026	0,26	0,41	2,18E-02	2,20E-02	*PYGL*
cg13905475	0,73	0,61	3,52E-02	2,41E-02	
cg22663660	0,69	0,55	6,57E-03	1,42E-02	*LIPC*
cg09829164	0,50	0,38	3,52E-02	3,29E-02	*RADIL*
cg10443977	0,86	0,74	1,30E-02	2,58E-02	*ATP13A4*
cg12178501	0,70	0,55	3,54E-03	9,82E-03	*PLEKHS1*
cg08514385	0,55	0,66	7,39E-03	4,61E-02	
cg03185290	0,68	0,55	5,16E-03	1,54E-02	*CNOT4*
cg07220152	0,11	0,22	5,16E-03	6,08E-02	
cg10640852	0,32	0,20	2,73E-03	1,44E-02	
cg10310395	0,86	0,69	2,18E-02	1,44E-02	*TCF12*
cg01766699	0,80	0,59	2,73E-03	7,58E-03	*DNAH7*
cg13469034	0,78	0,68	2,92E-02	2,72E-02	*KIAA1026*
cg16928392	0,88	0,74	5,16E-03	2,62E-02	*DGKD*

cg13742924	0,68	0,57	1,04E-02	2,34E-02	*MARCH1*
cg27488252	0,64	0,49	7,55E-04	1,18E-02	*CLASP1*
cg02820058	0,64	0,52	1,44E-02	3,18E-02	*PRKG1*
cg09684557	0,77	0,64	1,16E-02	1,50E-02	*OR6C68*
cg15851515	0,69	0,58	1,16E-02	2,61E-02	*AIM1*
cg04955791	0,53	0,43	3,21E-02	7,26E-02	*SYNE1*
cg12945572	0,61	0,48	1,58E-03	9,25E-03	*PTGER3*
cg05073607	0,56	0,44	1,16E-02	1,55E-02	*POU6F2-AS1*
cg20494691	0,54	0,40	1,18E-03	5,33E-03	*MYCT1*
cg18720720	0,44	0,31	4,56E-03	2,06E-02	
cg10028440	0,82	0,68	3,11E-03	1,66E-02	
cg03969251	0,59	0,49	2,39E-03	7,54E-03	*DCLK2*
cg10159032	0,75	0,63	9,30E-03	1,92E-02	*LRRC37B*
cg07954193	0,79	0,64	3,52E-02	4,79E-02	*C1D*
cg17457832	0,86	0,75	3,52E-02	3,12E-02	*RPA3; UMAD1*
cg01322839	0,71	0,55	1,58E-03	5,15E-03	*CORO2A*
cg11192416	0,23	0,41	1,97E-02	2,64E-02	
cg19080846	0,85	0,75	1,18E-03	1,59E-02	*NRAP*
cg24156694	0,55	0,44	7,39E-03	1,91E-02	*UMODL1*
cg25598696	0,63	0,49	4,56E-03	9,33E-03	*MARCH1*
cg19459178	0,67	0,51	2,65E-02	2,40E-02	*PCSK6*
cg18184895	0,84	0,73	1,78E-02	1,72E-02	*MALRD1*
cg13525835	0,85	0,73	8,30E-03	1,50E-02	*CABLES1*
cg16284880	0,52	0,37	1,60E-02	2,96E-02	
cg21456581	0,67	0,54	4,61E-02	3,12E-02	
cg05952081	0,16	0,32	4,61E-02	1,62E-02	
cg10191005	0,56	0,44	4,22E-02	1,74E-02	*DIO3OS*
cg24168276	0,77	0,66	2,65E-02	4,84E-02	*BCL2*
cg01160882	0,45	0,59	1,18E-03	7,93E-03	*DKK1*
cg14836363	0,77	0,65	3,21E-02	1,90E-02	
cg24435747	0,56	0,43	2,92E-02	3,28E-02	*IL28RA*
cg20823324	0,75	0,63	1,04E-02	2,84E-02	*TMED3*
cg14707974	0,39	0,50	2,18E-02	2,98E-02	
cg07867420	0,62	0,48	1,78E-02	1,89E-02	*COL23A1*
cg00869372	0,67	0,57	2,92E-02	2,84E-02	
cg01942816	0,58	0,47	1,78E-02	3,36E-02	*MIR589*
cg07498415	0,35	0,46	1,78E-02	1,90E-02	
cg01713201	0,62	0,47	5,83E-03	5,71E-03	*SYNE1*
cg04402745	0,74	0,62	6,57E-03	1,62E-02	
cg06072309	0,51	0,41	4,61E-02	4,46E-02	*VIPR1*
cg18162875	0,68	0,57	7,39E-03	1,01E-02	*LOC100128531; KIAA1671*
cg06113001	0,55	0,44	2,92E-02	2,28E-02	*GLYCAM1*
cg10664737	0,57	0,69	3,86E-02	4,60E-02	
cg23167136	0,73	0,59	3,21E-02	2,46E-02	*CSRP1*

cg21203643	0,67	0,56	1,30E-02	1,31E-02	*DEGS2*
cg18401380	0,63	0,52	2,18E-02	4,18E-02	*MYO5B*
cg17934874	0,78	0,64	6,57E-03	1,28E-02	*AGAP1*
cg19644270	0,65	0,55	2,65E-02	2,15E-02	
cg04931238	0,47	0,33	2,73E-03	1,20E-02	
cg05607305	0,75	0,62	1,37E-03	6,93E-03	
cg04440724	0,83	0,70	2,92E-02	2,34E-02	*UBN1*
cg15944290	0,51	0,39	2,92E-02	6,37E-02	*MACROD2*
cg26503877	0,17	0,31	1,16E-02	1,47E-02	*ETS1*
cg03755420	0,51	0,40	2,18E-02	2,30E-02	
cg14599421	0,69	0,54	1,82E-03	8,56E-03	*DTNB*
cg11832805	0,54	0,69	7,39E-03	1,71E-02	*LOC284933; FAM19A5*
cg15736126	0,74	0,56	8,79E-04	5,55E-03	*ZNF296*
cg22625504	0,49	0,38	2,18E-02	4,49E-02	*TTI1*
cg05026097	0,83	0,73	6,57E-03	1,70E-02	*OTUD3*
cg20954746	0,82	0,66	7,55E-04	5,28E-03	*DGKD*
cg08410479	0,70	0,59	9,30E-03	1,41E-02	*LOC100128531; KIAA1671*
cg25472135	0,81	0,71	1,78E-02	1,67E-02	*SYNE3*
cg01868551	0,49	0,36	1,97E-02	1,93E-02	
cg05895353	0,74	0,84	2,41E-02	7,96E-02	
cg07035704	0,62	0,51	2,18E-02	5,07E-02	*MIR2052*
cg24446969	0,54	0,40	4,56E-03	1,31E-02	*NRXN3*
cg20943095	0,62	0,50	1,97E-02	2,43E-02	*CPLX3*
cg08712054	0,34	0,23	2,92E-02	3,16E-02	*HOXC4*
cg25460263	0,70	0,58	3,11E-03	7,68E-03	*XIRP2*
cg23905763	0,23	0,38	3,86E-02	2,64E-02	*SOGA3*
cg12393334	0,69	0,55	2,65E-02	2,26E-02	
cg09765987	0,70	0,54	5,83E-03	1,54E-02	
cg06204665	0,65	0,54	4,22E-02	4,69E-02	*NIPAL3*
cg00670438	0,27	0,44	2,65E-02	1,87E-02	*PYGL*
cg08494221	0,42	0,52	3,86E-02	2,68E-02	*PURG; WRN*
cg08144021	0,74	0,64	1,60E-02	3,54E-02	*CRTAC1*
cg14483244	0,16	0,26	3,52E-02	3,94E-02	*HDAC5*
cg13466956	0,46	0,60	1,78E-02	2,70E-02	*SLC19A3*
cg03458344	0,55	0,43	2,92E-02	4,94E-02	*C1orf129*
cg14251889	0,57	0,47	1,97E-02	4,82E-02	*KIF26B*
cg05499811	0,61	0,49	1,30E-02	1,56E-02	*DOCK8*
cg17130176	0,78	0,67	1,16E-02	4,35E-02	*BEST3*
cg18902109	0,66	0,56	4,22E-02	6,57E-02	
cg26400135	0,80	0,69	3,52E-02	8,44E-02	
cg15808668	0,42	0,55	6,57E-03	1,77E-02	*IP6K3*
cg05585556	0,80	0,66	1,30E-02	1,65E-02	*SNHG4*
cg18593045	0,60	0,49	2,18E-02	3,00E-02	
cg23161350	0,59	0,49	1,78E-02	2,67E-02	

cg10141151	0,43	0,31	2,65E-02	4,82E-02	
cg09129997	0,47	0,35	1,37E-03	1,02E-02	
cg22076912	0,61	0,49	1,30E-02	1,63E-02	
cg15285104	0,78	0,63	9,30E-03	1,58E-02	*LINC00160*
cg19505827	0,31	0,45	2,18E-02	3,51E-02	*TSPAN32*
cg23313755	0,64	0,53	9,30E-03	1,53E-02	
cg20276367	0,69	0,56	1,58E-03	1,21E-02	
cg18086978	0,47	0,57	4,22E-02	3,92E-02	*CLMN*
cg09872392	0,60	0,48	1,78E-02	1,74E-02	*PHKG1*
cg14273219	0,68	0,55	4,56E-03	1,46E-02	*PPP1R3F*
cg17436674	0,52	0,38	1,19E-02	3,13E-02	*ARHGEF38*
cg12802876	0,50	0,39	4,61E-02	6,10E-02	
cg10528395	0,61	0,48	2,41E-02	1,69E-02	
cg05006873	0,50	0,39	2,18E-02	4,19E-02	
cg19914513	0,82	0,69	5,83E-03	7,38E-03	*TTC12*
cg17237907	0,61	0,71	3,86E-02	9,04E-02	*MICAL1*
cg17975941	0,57	0,47	4,22E-02	4,62E-02	
cg05865493	0,67	0,54	7,39E-03	1,47E-02	*HDAC9*
cg13686214	0,63	0,50	7,55E-04	5,49E-03	
cg08817204	0,66	0,54	2,18E-02	4,97E-02	*NRG3*
cg11345968	0,83	0,72	3,86E-02	6,59E-02	*PRNCR1*
cg03079840	0,63	0,47	2,02E-04	5,73E-03	*SLC35F3*
cg02136949	0,72	0,57	5,83E-03	2,41E-02	*TRAPPC9*
cg13762927	0,72	0,61	8,30E-03	1,36E-02	*KIAA1456*
cg12026858	0,69	0,57	3,86E-02	4,45E-02	*ZBTB16*
cg14503204	0,75	0,64	2,65E-02	3,68E-02	*SGCZ*
cg11924659	0,56	0,67	7,39E-03	4,90E-02	
cg06787433	0,79	0,66	2,18E-02	2,96E-02	*SH3BP4*
cg14707834	0,47	0,62	1,60E-02	1,63E-02	*SIGIRR*
cg13563869	0,55	0,45	1,30E-02	3,02E-02	
cg23863184	0,84	0,72	2,41E-02	3,57E-02	*ZNF407*
cg02512586	0,53	0,41	1,97E-02	3,63E-02	
cg00339415	0,46	0,36	2,18E-02	2,43E-02	*OBP2B*
cg14850831	0,74	0,62	2,41E-02	2,94E-02	*GUCD1*
cg22387864	0,64	0,49	1,30E-02	1,70E-02	*ISPD*
cg09683533	0,76	0,63	2,65E-02	2,30E-02	*AFF1*
cg00851827	0,74	0,59	8,30E-03	2,17E-02	*ATP6V0D1*
cg01635228	0,70	0,58	2,41E-02	2,20E-02	
cg22859289	0,33	0,43	3,52E-02	5,03E-02	
cg06526020	0,87	0,76	4,22E-02	3,05E-02	*NUDT3*
cg10652722	0,30	0,44	2,41E-02	2,78E-02	
cg11988321	0,55	0,66	4,22E-02	7,92E-02	*MZB1*
cg07248586	0,50	0,38	2,41E-02	2,63E-02	
cg19309009	0,62	0,49	3,54E-03	1,24E-02	*MARK4*
cg11737879	0,78	0,67	5,83E-03	1,43E-02	*PEX3*

cg23263302	0,64	0,52	3,86E-02	3,43E-02	*LDB2*
cg01708617	0,62	0,72	5,16E-03	1,24E-02	*RAI14*
cg01039497	0,57	0,46	2,41E-02	3,36E-02	*TRAK1*
cg01526396	0,47	0,58	2,92E-02	3,30E-02	
cg13919631	0,69	0,56	2,41E-02	3,60E-02	
cg14185504	0,60	0,46	1,16E-02	1,70E-02	*FIG4*
cg19698472	0,70	0,57	2,41E-02	1,88E-02	
cg18170393	0,69	0,58	3,21E-02	3,32E-02	
cg06261500	0,74	0,60	5,16E-03	1,21E-02	*SND1*
cg18962562	0,56	0,44	1,44E-02	2,25E-02	*PPFIBP2*
cg05135791	0,56	0,44	1,97E-02	4,00E-02	*TRHDE*
cg26656452	0,81	0,71	9,30E-03	1,36E-02	*HABP2*
cg00864666	0,56	0,43	2,39E-03	6,39E-03	*RAPGEF4*
cg00364758	0,62	0,47	1,78E-02	1,03E-02	*EPDR1*
cg23204757	0,26	0,15	7,39E-03	2,22E-02	
cg03885271	0,14	0,31	4,02E-03	3,04E-02	*BHLHE23*
cg15387301	0,77	0,65	1,82E-03	1,52E-02	*RASGEF1B*
cg25641388	0,69	0,56	3,21E-02	2,45E-02	
cg23470770	0,64	0,52	3,21E-02	4,32E-02	*MAP3K5*
cg01913492	0,69	0,57	3,86E-02	2,47E-02	
cg06036471	0,33	0,46	9,30E-03	1,62E-02	*CPEB3*
cg12966404	0,62	0,52	2,65E-02	4,96E-02	*STRIP1*
cg24464067	0,83	0,72	1,04E-02	3,39E-02	*NET1*
cg08473326	0,45	0,57	9,30E-03	1,66E-02	
cg14221048	0,79	0,61	1,41E-04	7,02E-03	
cg10308445	0,79	0,63	6,57E-03	8,51E-03	*PGAP1*
cg25668563	0,63	0,47	9,30E-03	1,98E-02	*CCDC129*
cg10085153	0,81	0,67	2,39E-03	6,03E-03	*PAPSS2*
cg06571387	0,50	0,61	6,57E-03	2,07E-02	*HOXD12*
cg21635588	0,67	0,57	3,21E-02	3,59E-02	*IPCEF1*
cg09974610	0,76	0,56	1,18E-03	4,43E-03	*EHF*
cg20137977	0,65	0,54	4,56E-03	1,30E-02	*MCC*
cg27484582	0,66	0,55	3,52E-02	3,57E-02	
cg18702947	0,81	0,66	8,30E-03	1,16E-02	*PPFIA2*
cg04761450	0,63	0,52	2,41E-02	2,40E-02	*FOXO1*
cg11028738	0,57	0,40	8,30E-03	1,58E-02	
cg15783703	0,86	0,74	1,78E-02	1,47E-02	*AUTS2*
cg02017534	0,31	0,43	4,22E-02	3,13E-02	*SIAH3*
cg19801921	0,42	0,57	2,92E-02	3,81E-02	
cg02361126	0,62	0,50	3,21E-02	3,57E-02	
cg06588136	0,82	0,71	4,71E-04	1,79E-02	*CFH*
cg01827197	0,64	0,53	1,97E-02	5,47E-02	*A1CF*
cg18202122	0,54	0,39	1,82E-03	5,74E-03	
cg09324866	0,55	0,44	1,41E-04	4,89E-03	
cg10308680	0,54	0,41	1,60E-02	2,24E-02	*CCDC33*

cg26581860	0,64	0,48	3,86E-02	3,78E-02	*EHF*
cg22596683	0,69	0,51	4,56E-03	1,04E-02	*EHF*
cg25865597	0,48	0,36	1,78E-02	2,44E-02	*PRDM16*
cg15167311	0,58	0,46	1,82E-03	9,13E-03	*PRKD1*
cg06231121	0,45	0,35	2,73E-03	1,70E-02	
cg13138848	0,66	0,51	3,54E-03	5,63E-03	
cg02723558	0,75	0,57	1,04E-02	1,02E-02	*BDNFOS*
cg10142997	0,52	0,41	1,60E-02	3,62E-02	*PCDH9*
cg20433989	0,59	0,49	1,97E-02	2,16E-02	*FNDC3A*
cg05719445	0,52	0,41	7,55E-04	4,98E-03	
cg06559333	0,76	0,56	2,08E-03	5,74E-03	*AUTS2*
cg04335144	0,63	0,52	1,04E-02	1,68E-02	*PLEKHA7*
cg12741293	0,65	0,44	2,39E-03	4,99E-03	
cg12678421	0,73	0,59	4,71E-04	8,98E-03	
cg16129213	0,82	0,69	8,30E-03	1,56E-02	*SOD1*
cg21187597	0,67	0,55	1,78E-02	2,86E-02	
cg25029197	0,62	0,47	2,39E-03	1,30E-02	*KRT7*
cg01204296	0,51	0,39	1,16E-02	2,94E-02	
cg04788216	0,74	0,56	1,44E-02	1,28E-02	*EHF*
cg17937897	0,53	0,36	4,41E-05	4,39E-03	
cg27130993	0,51	0,38	1,16E-02	1,41E-02	*ABLIM3*
cg04061534	0,58	0,47	7,39E-03	3,32E-02	*RIN2*
cg11581475	0,64	0,53	4,61E-02	3,90E-02	*ZNF428*
cg22509679	0,70	0,59	7,39E-03	1,65E-02	
cg25801428	0,60	0,49	1,97E-02	2,21E-02	*LOC100132735*
cg19015537	0,58	0,44	6,47E-04	6,18E-03	
cg01687997	0,68	0,53	3,54E-03	1,41E-02	*C7orf50*
cg20562823	0,63	0,51	1,30E-02	2,47E-02	
cg19612574	0,72	0,61	8,30E-03	2,60E-02	*MAPK8*
cg24716193	0,77	0,59	5,52E-04	9,37E-03	*CASC21*
cg27391744	0,72	0,56	7,55E-04	3,67E-03	*AQP5; LOC101927318*
cg21356379	0,81	0,69	1,16E-02	1,66E-02	
cg06571982	0,72	0,59	9,30E-03	1,33E-02	*MRPL1*
cg17585138	0,64	0,53	4,02E-03	1,28E-02	
cg03604322	0,52	0,40	2,92E-02	8,29E-02	*SLC5A1*
cg23900752	0,79	0,67	1,04E-02	1,49E-02	
cg25647648	0,70	0,51	1,37E-03	6,60E-03	*MTHFD2P1*
cg15069417	0,69	0,50	2,73E-03	5,40E-03	*LOC100128531; KIAA1671*
cg11020360	0,66	0,55	2,08E-03	2,91E-02	*LCP2; C5orf58*
cg10316857	0,76	0,63	9,30E-03	1,23E-02	
cg09366829	0,61	0,47	1,04E-02	1,10E-02	*VAV3*
cg25754826	0,46	0,35	2,73E-03	9,24E-03	*AUTS2*
cg01212975	0,78	0,67	2,41E-02	2,02E-02	
cg08771884	0,66	0,54	9,30E-03	1,89E-02	*UBN1*

cg10549234	0,74	0,59	4,02E-03	8,76E-03	*ADGRL2*
cg01892746	0,49	0,36	2,65E-02	3,79E-02	*SYNE1*
cg03692576	0,52	0,41	1,82E-03	9,78E-03	
cg18041719	0,40	0,27	3,39E-04	6,15E-03	*SEMA3D*
cg21687351	0,61	0,50	1,78E-02	1,96E-02	*KERA*
cg26479374	0,49	0,37	3,21E-02	5,01E-02	*CTBP1*
cg13721302	0,77	0,59	1,58E-03	9,27E-03	*SRPK2*
cg17608436	0,60	0,48	7,39E-03	1,49E-02	
cg14930320	0,53	0,40	2,41E-02	2,86E-02	
cg18284319	0,71	0,56	3,52E-02	2,28E-02	*RNF4*
cg13523510	0,47	0,37	3,86E-02	3,66E-02	
cg15471172	0,70	0,55	1,44E-02	1,27E-02	
cg08614769	0,77	0,63	3,21E-02	2,16E-02	*GMDS*
cg09337641	0,70	0,59	3,52E-02	4,46E-02	
cg20636399	0,58	0,47	7,39E-03	1,86E-02	*SLC24A4*
cg13984883	0,65	0,54	3,11E-03	1,62E-02	*AP1B1*
cg18181545	0,50	0,39	1,30E-02	1,75E-02	*EDIL3*
cg00461650	0,45	0,34	2,41E-02	3,30E-02	
cg25134328	0,34	0,45	3,21E-02	3,83E-02	
cg05966408	0,56	0,45	5,83E-03	1,62E-02	*NRG2*
cg14613491	0,56	0,46	1,44E-02	2,80E-02	*CNTN5*
cg17043875	0,68	0,53	1,04E-02	1,33E-02	
cg07737938	0,60	0,47	5,83E-03	1,37E-02	
cg20350898	0,56	0,43	6,57E-03	1,13E-02	*SIKE1*
cg15270502	0,67	0,55	4,02E-03	1,42E-02	
cg00550493	0,68	0,56	2,65E-02	1,98E-02	*EYS*
cg13832372	0,39	0,53	1,04E-02	1,56E-02	*LHX6*
cg12121765	0,77	0,66	3,21E-02	3,24E-02	*DLC1*
cg09334240	0,58	0,46	2,65E-02	3,28E-02	*ATF7*
cg02909298	0,69	0,53	5,16E-03	8,07E-03	
cg08765573	0,49	0,37	1,30E-02	1,99E-02	*SLC35F3*
cg12452334	0,73	0,60	1,60E-02	2,10E-02	
cg05179455	0,49	0,38	4,22E-02	6,77E-02	*COL6A3*
cg05438832	0,63	0,53	1,30E-02	2,62E-02	
cg11266682	0,31	0,21	1,78E-02	2,54E-02	*SLC2A9*
cg11666559	0,27	0,40	1,44E-02	2,11E-02	
cg05801088	0,35	0,50	3,86E-02	2,17E-02	
cg26440512	0,63	0,52	7,39E-03	2,02E-02	*ABLIM2*
cg01541932	0,71	0,57	2,92E-02	3,36E-02	*CTNNA2*
cg19931596	0,66	0,51	3,54E-03	1,13E-02	*SLC25A12*
cg26640276	0,60	0,46	1,30E-02	1,83E-02	
cg03928546	0,19	0,35	1,97E-02	2,19E-02	*CCDC105*
cg12602197	0,47	0,36	3,11E-03	3,62E-02	
cg05293827	0,57	0,43	8,30E-03	2,86E-02	*PDE1C*
cg14219442	0,68	0,55	1,04E-02	1,11E-02	

cg08995655	0,67	0,55	1,30E-02	2,64E-02	*NUDT19*
cg18290075	0,15	0,27	4,61E-02	7,42E-02	*SGEF*
cg07229402	0,58	0,44	5,83E-03	1,60E-02	
cg21948636	0,74	0,59	2,73E-03	1,13E-02	
cg01986648	0,56	0,42	1,16E-02	2,20E-02	
cg13075537	0,58	0,47	1,60E-02	2,65E-02	*TMEM57*
cg25266629	0,30	0,47	2,65E-02	3,32E-02	*TLX1*
cg07145899	0,71	0,61	4,02E-03	1,25E-02	*OTUD3*
cg15496665	0,69	0,57	1,97E-02	3,88E-02	*PAPPA*
cg10070202	0,64	0,53	9,30E-03	2,82E-02	*OSBP2*
cg04230537	0,82	0,71	6,57E-03	1,27E-02	*TRAP1*
cg19828122	0,64	0,51	1,16E-02	1,13E-02	*VAV3*
cg13169408	0,50	0,39	9,30E-03	2,00E-02	
cg09161311	0,69	0,55	9,30E-03	1,24E-02	
cg12204957	0,77	0,62	4,00E-04	9,32E-03	*IRAK2*
cg11305739	0,65	0,50	1,44E-02	1,64E-02	
cg02510157	0,65	0,51	7,39E-03	1,23E-02	*PDE4B*
cg08549781	0,61	0,48	2,41E-02	2,59E-02	*AP1S3*
cg11406977	0,57	0,45	2,65E-02	3,15E-02	
cg05562252	0,80	0,70	4,61E-02	2,65E-02	*CD164*
cg00574062	0,74	0,62	2,92E-02	4,81E-02	*ASIC2*
cg19177771	0,48	0,37	4,61E-02	3,18E-02	
cg26127201	0,70	0,57	3,21E-02	2,96E-02	
cg07461712	0,72	0,52	2,39E-03	6,33E-03	*LRRC37B*
cg09838568	0,20	0,34	3,21E-02	2,01E-02	
cg17053016	0,66	0,56	3,21E-02	5,13E-02	
cg17230414	0,66	0,53	2,65E-02	1,98E-02	
cg15937958	0,73	0,53	1,02E-03	8,32E-03	*CXCL17*
cg07721027	0,71	0,59	2,39E-03	1,70E-02	*IP6K3*
cg19958586	0,49	0,59	3,21E-02	3,76E-02	*CCRL2*
cg20791945	0,77	0,66	2,08E-03	8,23E-03	
cg27452462	0,78	0,68	2,65E-02	2,81E-02	*LINC01169*
cg15707869	0,66	0,55	9,30E-03	1,83E-02	*DCAF4L2*
cg23558626	0,75	0,63	4,61E-02	2,35E-02	
cg17726317	0,56	0,41	3,52E-02	2,51E-02	
cg13138556	0,74	0,62	1,44E-02	2,21E-02	*ENPP6*
cg14936543	0,54	0,43	2,65E-02	2,45E-02	*AGBL1*
cg26357982	0,66	0,54	2,41E-02	2,27E-02	*SLCO1B7*
cg25182785	0,67	0,51	1,82E-03	8,64E-03	*ARHGEF28*
cg20522398	0,52	0,39	1,97E-02	4,18E-02	
cg23010084	0,63	0,49	1,37E-03	7,08E-03	*ADGRB3*
cg14015502	0,55	0,42	3,35E-02	2,77E-02	*C10orf26*
cg06026636	0,70	0,55	4,56E-03	9,05E-03	*SND1*
cg25322618	0,49	0,37	2,65E-02	2,93E-02	*RAPGEF4*
cg14474138	0,71	0,53	9,30E-03	1,19E-02	

cg13884444	0,55	0,43	1,18E-03	8,74E-03	*SUN3*
cg03277898	0,77	0,65	2,92E-02	2,55E-02	
cg15800767	0,33	0,46	7,39E-03	2,35E-02	*LOC100128770; CCDC64B*
cg14927641	0,55	0,45	1,30E-02	5,40E-02	
cg16541681	0,65	0,50	5,40E-05	2,74E-03	*IPCEF1*
cg03140412	0,45	0,63	7,39E-03	2,26E-02	*SIGIRR*
cg16744531	0,39	0,21	3,54E-03	1,86E-02	*B3GNT3*
cg24011522	0,77	0,67	5,83E-03	2,73E-02	
cg03691573	0,24	0,39	4,61E-02	1,55E-02	*LOC100130451; SPAG16*
cg11898646	0,40	0,52	9,30E-03	3,51E-02	*EBF1*
cg20543968	0,64	0,48	2,39E-03	8,57E-03	
cg24335252	0,40	0,56	2,65E-02	2,34E-02	
cg14337655	0,24	0,38	1,18E-03	8,72E-03	*C19orf35*
cg09641663	0,82	0,72	1,18E-03	9,89E-03	*OSBPL10*
cg06058203	0,68	0,54	2,08E-03	8,28E-03	*MECP2*
cg16770641	0,61	0,49	2,18E-02	2,34E-02	
cg19261904	0,77	0,63	2,18E-02	1,72E-02	
cg15008149	0,57	0,42	1,18E-03	6,04E-03	*LOC100128531; KIAA1671*
cg03251454	0,51	0,41	4,22E-02	6,16E-02	
cg26198638	0,73	0,57	3,54E-03	1,12E-02	
cg19795338	0,58	0,43	1,18E-03	9,12E-03	*ATP13A4*
cg20343746	0,69	0,55	2,18E-02	2,71E-02	*GGACT*
cg08486363	0,59	0,47	1,30E-02	2,69E-02	*LINC01587*
cg04919036	0,53	0,43	5,16E-03	4,89E-02	
cg02271687	0,65	0,53	1,60E-02	2,45E-02	*PIK3R1*
cg25094495	0,50	0,39	1,60E-02	4,26E-02	
cg06638913	0,72	0,60	1,30E-02	1,73E-02	*SH3BP4*
cg07842960	0,57	0,42	1,04E-02	1,52E-02	*SH3BP4*
cg06477323	0,69	0,54	4,71E-04	6,63E-03	*PRKCH*
cg13660662	0,51	0,40	2,39E-03	1,86E-02	
cg07902983	0,30	0,42	2,41E-02	4,17E-02	*SYCN*
cg12766569	0,74	0,63	2,92E-02	2,67E-02	*KLK11*
cg02065027	0,80	0,69	4,61E-02	2,41E-02	*ARHGEF12*
cg26680428	0,50	0,37	4,61E-02	4,66E-02	
cg09802076	0,42	0,29	1,97E-02	3,20E-02	*SLC6A15*
cg25838080	0,51	0,61	2,65E-02	2,53E-02	*NR2E1*
cg13623999	0,66	0,55	3,21E-02	2,03E-02	
cg01178013	0,80	0,69	1,78E-02	1,84E-02	*TANC1*
cg07009554	0,45	0,56	1,30E-02	3,74E-02	
cg05181918	0,56	0,43	3,54E-03	8,62E-03	*LOC100128531; KIAA1671*
cg22800453	0,61	0,49	1,97E-02	1,05E-02	*MUCL1*
cg02320543	0,54	0,66	1,97E-02	3,99E-02	

cg20816107	0,54	0,40	4,02E-03	1,32E-02	*PLCL1*
cg17672195	0,50	0,37	1,97E-02	2,41E-02	*LOC100507205*
cg24546683	0,43	0,31	4,02E-03	2,44E-02	
cg02315271	0,41	0,54	4,61E-02	5,24E-02	
cg01915989	0,60	0,39	1,02E-03	9,41E-03	
cg04188392	0,61	0,48	2,39E-03	1,19E-02	*TUSC8*
cg18816129	0,71	0,59	2,73E-03	1,09E-02	*MEIS2*
cg11723489	0,54	0,43	1,30E-02	3,44E-02	
cg08349093	0,64	0,53	1,97E-02	2,84E-02	
cg00889540	0,53	0,43	1,44E-02	3,02E-02	
cg23321751	0,56	0,42	5,16E-03	1,75E-02	*KIAA1328*
cg27451581	0,67	0,51	1,44E-02	1,53E-02	*KLHDC7A*
cg06227673	0,65	0,53	1,37E-03	7,08E-03	
cg15948428	0,46	0,57	4,61E-02	6,06E-02	
cg08638124	0,71	0,59	3,21E-02	3,82E-02	*ASB2*
cg09766556	0,61	0,48	1,60E-02	2,31E-02	*SGCZ*
cg06957219	0,64	0,51	1,97E-02	2,99E-02	*DISC1*
cg25433259	0,43	0,32	2,39E-03	9,36E-03	*OSTCP1*
cg11137178	0,90	0,80	7,39E-03	4,84E-02	
cg01689653	0,67	0,54	1,58E-03	1,07E-02	*ABR*
cg13092992	0,74	0,60	4,56E-03	1,13E-02	
cg18034674	0,37	0,53	3,86E-02	2,19E-02	
cg07825247	0,71	0,58	1,30E-02	1,79E-02	*STAG1*
cg24843615	0,56	0,42	1,16E-02	1,68E-02	*MSGN1*
cg16698721	0,53	0,42	3,86E-02	5,88E-02	*CERS4*
cg15202552	0,44	0,33	3,54E-03	2,04E-02	*FCRL3*
cg13888854	0,48	0,59	1,97E-02	2,75E-02	*IL1R2*
cg13469457	0,61	0,49	2,92E-02	3,80E-02	
cg03389928	0,64	0,52	3,21E-02	2,99E-02	*KCNQ5*
cg16677112	0,62	0,46	2,08E-03	7,28E-03	*EFNA1*
cg25793657	0,77	0,66	3,86E-02	2,46E-02	*SNHG4; SNORA74A; MATR3*
cg01518970	0,69	0,57	6,57E-03	1,12E-02	
cg04675946	0,64	0,53	3,21E-02	4,29E-02	
cg23046990	0,72	0,57	2,92E-02	2,57E-02	*LINC00578*
cg03402807	0,76	0,62	2,41E-02	3,01E-02	*PRICKLE2*
cg25190663	0,63	0,51	7,39E-03	2,03E-02	
cg20988291	0,39	0,27	1,60E-02	3,09E-02	
cg09423378	0,77	0,60	1,58E-03	8,34E-03	*ARHGEF28*
cg10151281	0,64	0,47	8,30E-03	1,93E-02	*TPM4*
cg00223647	0,52	0,41	1,60E-02	2,26E-02	
cg05548017	0,62	0,52	3,86E-02	5,18E-02	*FGD5*
cg16688617	0,51	0,40	2,41E-02	3,68E-02	
cg01250864	0,59	0,44	1,37E-03	7,08E-03	*IFNGR1*
cg25378392	0,53	0,41	1,78E-02	2,84E-02	

cg11537026	0,74	0,57	2,41E-02	1,14E-02	
cg11351983	0,42	0,30	2,41E-02	1,99E-02	*ESRRG*
cg17401082	0,44	0,33	1,97E-02	3,51E-02	
cg17572723	0,67	0,57	1,97E-02	1,86E-02	*ZNF767P*
cg01214012	0,65	0,51	4,22E-02	3,53E-02	
cg12800585	0,58	0,48	8,30E-03	1,42E-02	*ARHGAP35*
cg15679007	0,62	0,50	2,92E-02	3,69E-02	
cg20832020	0,68	0,57	4,22E-02	4,78E-02	*TIGIT*
cg25765881	0,49	0,36	1,44E-02	2,31E-02	
cg18566313	0,50	0,33	4,02E-03	1,09E-02	*SORBS2*
cg09015280	0,10	0,21	3,21E-02	4,64E-02	
cg05881682	0,66	0,52	2,18E-02	3,43E-02	
cg19628619	0,66	0,51	4,02E-03	8,96E-03	*EFNA1*
cg25254263	0,81	0,65	1,82E-03	8,06E-03	*C12orf79*
cg26983411	0,64	0,52	3,21E-02	3,99E-02	
cg01613870	0,60	0,46	4,02E-03	1,06E-02	
cg15412918	0,64	0,53	2,65E-02	2,52E-02	
cg01439876	0,37	0,47	1,04E-02	1,93E-02	
cg12545864	0,48	0,35	1,44E-02	1,99E-02	
cg25640196	0,72	0,61	4,02E-03	8,64E-03	*RAD51B*

BCVY: breast cancer in very young women; BCO: breast cancer in old women.

ANNEX VIII. Significant CpG probes localized in promotor regions from hypomethylation global profile in BCVY from metEPICVal study

Probes	UCSC RefGene Name	Methylation differences (BCVY - BCO)	Gene expression differences (BCVY - BCO)	p-value	Adjusted p-value
cg20475650	*EHD2*	-0,14	0,47	1,10E-50	2,45E-49
cg01309726	*LOC100270710*	-0,13	0,46	1,40E-49	2,98E-48
cg10910579	*TMEM59L*	-0,16	0,79	2,25E-49	4,62E-48
cg11399100	*TMEM59L*	-0,15	0,79	2,25E-49	4,62E-48
cg17450859	*ASB5*	0,18	0,49	1,87E-48	3,74E-47
cg18342872	*ASB5*	0,18	0,49	1,87E-48	3,74E-47
cg08646125	*LRRC3B*	-0,12	1,04	9,24E-48	1,79E-46
cg26035366	*LRRC3B*	-0,17	1,04	9,24E-48	1,79E-46
cg06869039	*C8orf79*	0,13	1,05	2,70E-45	5,02E-44
cg01622379	*LHX8*	-0,17	-0,55	1,34E-43	2,25E-42
cg01185626	*LHX8*	-0,18	-0,55	1,34E-43	2,25E-42
cg25368122	*LHX8*	-0,24	-0,55	1,34E-43	2,25E-42
cg23952663	*LHX8*	-0,27	-0,55	1,34E-43	2,25E-42
cg19764599	*LHX8*	-0,21	-0,55	1,34E-43	2,25E-42
cg12764034	*LHX8*	-0,14	-0,55	1,34E-43	2,25E-42
cg20494691	*MYCT1*	0,14	0,58	6,87E-42	1,13E-40
cg23183497	*GRM3*	-0,15	0,46	3,20E-41	4,82E-40
cg17814180	*GRM3*	-0,18	0,46	3,20E-41	4,82E-40
cg00762869	*GRM3*	-0,25	0,46	3,20E-41	4,82E-40
cg15331781	*GRM3*	-0,14	0,46	3,20E-41	4,82E-40
cg11210312	*GRM3*	0,17	0,46	3,20E-41	4,82E-40
cg01005728	*ABCB1*	0,14	0,91	4,29E-40	5,69E-39
cg15759056	*ABCB1*	-0,29	0,91	4,29E-40	5,69E-39
cg25438493	*ABCB1*	-0,26	0,91	4,29E-40	5,69E-39
cg12270822	*ABCB1*	-0,20	0,91	4,29E-40	5,69E-39
cg00141548	*ABCB1*	-0,15	0,91	4,29E-40	5,69E-39
cg09105881	*ABCB1*	-0,17	0,91	4,29E-40	5,69E-39
cg27507700	*ABCB1*	-0,25	0,91	4,29E-40	5,69E-39
cg06562372	*ABCB1*	-0,19	0,91	4,29E-40	5,69E-39
cg00862116	*ABCB1*	-0,20	0,91	4,29E-40	5,69E-39
cg10126372	*PXDNL*	-0,10	-0,98	2,98E-38	3,69E-37
cg20383036	*RHOJ*	0,10	0,57	7,21E-38	8,78E-37
cg05350199	*RHOJ*	0,13	0,57	7,21E-38	8,78E-37
cg08479860	*ADCY8*	0,15	0,34	8,87E-38	1,06E-36
cg16696270	*ADCY8*	-0,13	0,34	8,87E-38	1,06E-36
cg19159488	*CLEC14A*	0,14	0,43	3,18E-34	3,22E-33
cg16482474	*WDR86*	-0,16	0,88	3,37E-34	3,37E-33
cg17509603	*WDR86*	-0,21	0,88	3,37E-34	3,37E-33
cg27304406	*BNC1*	-0,18	0,73	6,57E-30	6,02E-29
cg18560204	*BNC1*	-0,17	0,73	6,57E-30	6,02E-29

cg18952647	*BNC1*	-0,22	0,73	6,57E-30	6,02E-29
cg14385245	*BNC1*	-0,25	0,73	6,57E-30	6,02E-29
cg07224636	*TMTC1*	0,14	0,76	1,56E-28	1,38E-27
cg08729318	*TMTC1*	-0,12	0,76	1,56E-28	1,38E-27
cg01672943	*PAX9*	-0,27	-0,98	7,58E-25	5,97E-24
cg01972418	*PAX9*	-0,31	-0,98	7,58E-25	5,97E-24
cg01627823	*PAX9*	-0,35	-0,98	7,58E-25	5,97E-24
cg00970313	*PAX9*	-0,22	-0,98	7,58E-25	5,97E-24
cg23655970	*PAX9*	-0,19	-0,98	7,58E-25	5,97E-24
cg00388556	*PAX9*	-0,23	-0,98	7,58E-25	5,97E-24
cg21875330	*PAX9*	-0,24	-0,98	7,58E-25	5,97E-24
cg03938432	*BAI3*	-0,21	0,47	1,69E-24	1,30E-23
cg10244047	*BAI3*	-0,13	0,47	1,69E-24	1,30E-23
cg02653557	*ZIK1*	-0,17	0,62	4,05E-24	3,08E-23
cg01046104	*ZIK1*	-0,26	0,62	4,05E-24	3,08E-23
cg15343461	*TMEM90B*	-0,18	-0,98	1,62E-23	1,22E-22
cg12403137	*NAALAD2*	-0,23	0,48	2,11E-23	1,55E-22
cg05500015	*NAALAD2*	-0,24	0,48	2,11E-23	1,55E-22
cg09584827	*NAALAD2*	-0,25	0,48	2,11E-23	1,55E-22
cg14304817	*NAALAD2*	-0,19	0,48	2,11E-23	1,55E-22
cg10599900	*REM1*	-0,19	0,41	3,66E-23	2,67E-22
cg00741624	*KIAA1409*	-0,23	0,38	9,95E-23	7,19E-22
cg06622151	*SIX3*	-0,17	-0,94	2,93E-22	2,10E-21
cg21604358	*HPCA*	-0,19	0,44	4,43E-22	3,14E-21
cg11712990	*LOC645323*	-0,20	-0,77	3,14E-21	2,08E-20
cg19411025	*LOC645323*	-0,19	-0,77	3,14E-21	2,08E-20
cg04261138	*LOC645323*	-0,21	-0,77	3,14E-21	2,08E-20
cg12325536	*LOC645323*	-0,23	-0,77	3,14E-21	2,08E-20
cg08811309	*LOC645323*	-0,13	-0,77	3,14E-21	2,08E-20
cg16638920	*LOC645323*	-0,19	-0,77	3,14E-21	2,08E-20
cg24058566	*LOC645323*	-0,17	-0,77	3,14E-21	2,08E-20
cg09884341	*LOC645323*	-0,19	-0,77	3,14E-21	2,08E-20
cg01080998	*LOC645323*	-0,21	-0,77	3,14E-21	2,08E-20
cg20459527	*OR13C2*	0,18	0,06	8,13E-21	5,34E-20
cg23095584	*GBX2*	-0,21	-0,46	2,72E-20	1,77E-19
cg06953874	*HOXD9*	-0,14	0,43	3,31E-19	2,00E-18
cg10957151	*HOXD9*	-0,21	0,43	3,31E-19	2,00E-18
cg14142007	*HOXD9*	-0,23	0,43	3,31E-19	2,00E-18
cg22805974	*HOXD9*	-0,17	0,43	3,31E-19	2,00E-18
cg14991487	*HOXD9*	-0,21	0,43	3,31E-19	2,00E-18
cg09578028	*HOXD9*	-0,24	0,43	3,31E-19	2,00E-18
cg05167251	*HOXD9*	-0,24	0,43	3,31E-19	2,00E-18
cg19464917	*ISL2*	-0,11	-0,63	3,64E-19	2,18E-18
cg18961944	*GALR1*	-0,15	0,22	8,05E-19	4,55E-18
cg06562865	*GALR1*	-0,23	0,22	8,05E-19	4,55E-18

cg15484988	*GALR1*	-0,19	0,22	8,05E-19	4,55E-18
cg18664869	*PCDHB15*	-0,23	0,36	1,54E-17	8,04E-17
cg10757144	*PCDHB15*	-0,16	0,36	1,54E-17	8,04E-17
cg13298963	*ACCN1*	-0,10	0,97	1,90E-17	9,67E-17
cg25033990	*ACCN1*	-0,22	0,97	1,90E-17	9,67E-17
cg19942495	*ACCN1*	-0,19	0,97	1,90E-17	9,67E-17
cg24613080	*ACCN1*	-0,22	0,97	1,90E-17	9,67E-17
cg14603098	*ACCN1*	-0,20	0,97	1,90E-17	9,67E-17
cg17491850	*DAB1*	-0,16	0,30	1,97E-17	9,78E-17
cg24351410	*DAB1*	-0,17	0,30	1,97E-17	9,78E-17
cg21177673	*DAB1*	-0,16	0,30	1,97E-17	9,78E-17
cg26345888	*DAB1*	-0,16	0,30	1,97E-17	9,78E-17
cg11256607	*DAB1*	-0,20	0,30	1,97E-17	9,78E-17
cg06384463	*BARHL2*	-0,18	-0,19	5,39E-17	2,59E-16
cg06959142	*BARHL2*	-0,19	-0,19	5,39E-17	2,59E-16
cg13146465	*BARHL2*	-0,19	-0,19	5,39E-17	2,59E-16
cg17241310	*BARHL2*	-0,20	-0,19	5,39E-17	2,59E-16
cg12520861	*OTP*	-0,15	-0,44	1,15E-16	5,42E-16
cg21960680	*OTP*	-0,21	-0,44	1,15E-16	5,42E-16
cg08572394	*OTP*	-0,24	-0,44	1,15E-16	5,42E-16
cg12054820	*OTP*	-0,14	-0,44	1,15E-16	5,42E-16
cg05467523	*SV2A*	-0,13	0,54	3,34E-16	1,56E-15
cg15119027	*FGF3*	-0,26	0,03	6,21E-15	2,81E-14
cg24359323	*FGF3*	-0,21	0,03	6,21E-15	2,81E-14
cg15337897	*FGF3*	-0,24	0,03	6,21E-15	2,81E-14
cg20308540	*PLD5*	-0,11	0,63	1,06E-14	4,74E-14
cg22679728	*PLD5*	-0,18	0,63	1,06E-14	4,74E-14
cg12613383	*PLD5*	-0,18	0,63	1,06E-14	4,74E-14
cg03637878	*JAM3*	-0,20	0,36	4,31E-14	1,84E-13
cg00939226	*JAM3*	-0,23	0,36	4,31E-14	1,84E-13
cg24899571	*JAM3*	-0,16	0,36	4,31E-14	1,84E-13
cg02174225	*JAM3*	-0,17	0,36	4,31E-14	1,84E-13
cg14337655	*C19orf35*	-0,14	0,34	1,47E-13	6,22E-13
cg01970716	*NYNRIN*	-0,11	0,39	5,98E-13	2,50E-12
cg25824543	*ZNF175*	-0,19	0,35	6,54E-13	2,72E-12
cg15881486	*CBR3*	0,12	-0,38	6,81E-13	2,82E-12
cg16361890	*LTC4S*	-0,17	0,47	1,03E-12	4,24E-12
cg21176048	*PEX5L*	-0,25	0,39	2,42E-12	9,95E-12
cg20914464	*NOL4*	-0,15	0,69	5,92E-12	2,41E-11
cg22593533	*OLIG2*	-0,16	-0,19	1,07E-11	4,30E-11
cg20455092	*GHR*	-0,13	0,51	2,89E-11	1,14E-10
cg18409528	*TLX2*	-0,23	-0,22	3,99E-11	1,53E-10
cg03998104	*TLX2*	-0,25	-0,22	3,99E-11	1,53E-10
cg15112032	*PCDH17*	-0,22	-0,28	4,47E-11	1,70E-10
cg14427009	*PCDH17*	-0,17	-0,28	4,47E-11	1,70E-10

cg12398338	*PVRL3*	-0,19	0,47	5,51E-11	2,08E-10
cg00705142	*PVRL3*	-0,15	0,47	5,51E-11	2,08E-10
cg03654735	*RSPO1*	-0,21	0,66	7,11E-11	2,66E-10
cg01398050	*KCNH8*	-0,20	0,66	3,22E-10	1,18E-09
cg24997944	*KCNH8*	-0,21	0,66	3,22E-10	1,18E-09
cg24668089	*KCNH8*	-0,13	0,66	3,22E-10	1,18E-09
cg18500192	*KCNH8*	-0,27	0,66	3,22E-10	1,18E-09
cg04891086	*KCNA1*	-0,22	0,56	5,19E-10	1,87E-09
cg09487611	*KCNA1*	-0,13	0,56	5,19E-10	1,87E-09
cg15682043	*KCNA1*	-0,18	0,56	5,19E-10	1,87E-09
cg06204922	*KCNA1*	-0,26	0,56	5,19E-10	1,87E-09
cg16907029	*KCNA1*	-0,11	0,56	5,19E-10	1,87E-09
cg26162295	*GSDMA*	-0,11	-0,59	5,44E-10	1,95E-09
cg20699141	*DCUN1D1*	-0,14	-0,22	5,59E-10	2,00E-09
cg03895540	*OR52N2*	0,16	0,22	9,70E-10	3,45E-09
cg04360049	*CD1D*	-0,19	0,32	2,69E-08	9,08E-08
cg25683325	*CD1D*	-0,23	0,32	2,69E-08	9,08E-08
cg24900205	*KCNT2*	0,15	0,45	1,41E-07	4,57E-07
cg14664759	*KCNT2*	-0,16	0,45	1,41E-07	4,57E-07
cg09177518	*KCNT2*	-0,19	0,45	1,41E-07	4,57E-07
cg02858606	*CDH8*	-0,23	0,60	2,32E-07	7,14E-07
cg03671700	*CDH8*	-0,15	0,60	2,32E-07	7,14E-07
cg11323198	*CDH8*	-0,19	0,60	2,32E-07	7,14E-07
cg27120816	*CDH8*	-0,21	0,60	2,32E-07	7,14E-07
cg01718742	*CDH8*	-0,25	0,60	2,32E-07	7,14E-07
cg01007828	*CDH8*	-0,22	0,60	2,32E-07	7,14E-07
cg06830659	*CDH8*	-0,22	0,60	2,32E-07	7,14E-07
cg22465581	*CDH8*	-0,22	0,60	2,32E-07	7,14E-07
cg17995563	*CDH8*	-0,17	0,60	2,32E-07	7,14E-07
cg06321345	*FAM5B*	-0,13	0,89	3,62E-07	1,11E-06
cg19519310	*FOXF2*	-0,12	0,24	1,20E-06	3,63E-06
cg09246637	*SNTG1*	-0,19	0,01	1,56E-06	4,71E-06
cg00230631	*SNTG1*	-0,23	0,01	1,56E-06	4,71E-06
cg09238180	*HOXD8*	-0,21	0,37	2,48E-06	7,39E-06
cg20732478	*HOXD8*	-0,11	0,37	2,48E-06	7,39E-06
cg21815667	*HOXD8*	-0,21	0,37	2,48E-06	7,39E-06
cg27351358	*BDNF*	-0,17	0,54	3,44E-06	1,01E-05
cg16257091	*BDNF*	-0,20	0,54	3,44E-06	1,01E-05
cg03167496	*BDNF*	-0,19	0,54	3,44E-06	1,01E-05
cg22288103	*BDNF*	-0,22	0,54	3,44E-06	1,01E-05
cg25457956	*BDNF*	-0,26	0,54	3,44E-06	1,01E-05
cg15462887	*BDNF*	-0,23	0,54	3,44E-06	1,01E-05
cg25156688	*BDNF*	-0,25	0,54	3,44E-06	1,01E-05
cg01583131	*BDNF*	-0,15	0,54	3,44E-06	1,01E-05
cg17206583	*VGF*	-0,12	-0,38	9,19E-06	2,63E-05

cg12135573	*VGF*	-0,17	-0,38	9,19E-06	2,63E-05
cg12865837	*SIM1*	-0,20	-0,66	1,28E-05	3,62E-05
cg21063722	*SIM1*	-0,25	-0,66	1,28E-05	3,62E-05
cg11891393	*SIM1*	-0,17	-0,66	1,28E-05	3,62E-05
cg22826141	*SFRP4*	-0,15	-0,58	2,25E-05	6,28E-05
cg07669460	*SFRP4*	-0,18	-0,58	2,25E-05	6,28E-05
cg26678765	*C1orf161*	0,23	-0,40	2,77E-05	7,73E-05
cg07157893	*MARCH1*	0,16	-0,35	3,05E-05	8,46E-05
cg00032805	*MARCH1*	-0,14	-0,35	3,05E-05	8,46E-05
cg03557733	*MPP2*	-0,17	0,42	9,49E-05	2,60E-04
cg06304401	*NLGN1*	-0,23	0,29	9,68E-05	2,64E-04
cg24600895	*NLGN1*	-0,11	0,29	9,68E-05	2,64E-04
cg05754905	*F11R*	0,18	-0,17	1,15E-04	3,12E-04
cg04597433	*DRD5*	-0,15	0,02	1,18E-04	3,21E-04
cg14309111	*ODZ4*	-0,13	-0,36	1,37E-04	3,71E-04
cg10729426	*ZNF549*	-0,18	0,26	2,31E-04	6,12E-04
cg27304204	*TSPAN8*	0,19	0,80	3,39E-04	8,87E-04
cg15684563	*TSPAN8*	0,15	0,80	3,39E-04	8,87E-04
cg04950931	*TSPAN8*	0,12	0,80	3,39E-04	8,87E-04
cg19871235	*TSPAN8*	0,15	0,80	3,39E-04	8,87E-04
cg22387174	*CNPY1*	-0,20	0,05	3,63E-04	9,48E-04
cg09265529	*ZNF470*	-0,21	-0,19	5,06E-04	1,32E-03
cg27109600	*FLJ32063*	-0,16	0,09	7,48E-04	1,93E-03
cg22862420	*RGS17*	0,15	0,16	7,55E-04	1,94E-03
cg13178597	*RGS17*	0,11	0,16	7,55E-04	1,94E-03
cg06072662	*RGS17*	0,18	0,16	7,55E-04	1,94E-03
cg07811644	*NRSN1*	-0,15	-0,01	7,71E-04	1,97E-03
cg03284310	*HMX3*	-0,14	-0,07	9,90E-04	2,51E-03
cg21657577	*ZNF578*	-0,22	0,24	1,10E-03	2,74E-03
cg12665460	*ZNF578*	-0,20	0,24	1,10E-03	2,74E-03
cg11909748	*ZNF578*	-0,15	0,24	1,10E-03	2,74E-03
cg14582763	*ZNF578*	-0,20	0,24	1,10E-03	2,74E-03
cg13461241	*ZNF578*	-0,21	0,24	1,10E-03	2,74E-03
cg21553182	*ZNF578*	-0,18	0,24	1,10E-03	2,74E-03
cg25763393	*ZNF578*	-0,22	0,24	1,10E-03	2,74E-03
cg15577595	*ZNF804A*	-0,26	-0,37	2,00E-03	4,94E-03
cg03036557	*GPC5*	-0,19	0,16	2,07E-03	5,09E-03
cg12678562	*GPC5*	-0,17	0,16	2,07E-03	5,09E-03
cg10548038	*GPC5*	-0,18	0,16	2,07E-03	5,09E-03
cg06186301	*GPC5*	-0,14	0,16	2,07E-03	5,09E-03
cg00302983	*SACS*	-0,17	0,20	2,09E-03	5,09E-03
cg18836366	*SACS*	-0,18	0,20	2,09E-03	5,09E-03
cg07175985	*SACS*	-0,20	0,20	2,09E-03	5,09E-03
cg03734874	*TMEM179*	-0,18	0,17	3,29E-03	7,98E-03
cg12459037	*CNTN4*	-0,13	0,44	4,42E-03	1,06E-02

cg02748354	*EIF3A*	0,14	-0,10	4,69E-03	1,12E-02
cg11237426	*NPHS2*	-0,23	-0,12	5,00E-03	1,18E-02
cg19284211	*INSM1*	-0,23	-0,56	5,03E-03	1,19E-02
cg03498096	*INSM1*	-0,17	-0,56	5,03E-03	1,19E-02
cg06038655	*KLF14*	-0,20	0,10	6,88E-03	1,61E-02
cg08523865	*CNTNAP5*	-0,19	-0,16	7,77E-03	1,80E-02
cg01308258	*CNTNAP5*	-0,15	-0,16	7,77E-03	1,80E-02
cg17321954	*STXBP5L*	-0,18	-0,45	8,48E-03	1,94E-02
cg07806886	*STXBP5L*	-0,19	-0,45	8,48E-03	1,94E-02
cg03602730	*STXBP5L*	-0,14	-0,45	8,48E-03	1,94E-02
cg01855313	*STXBP5L*	-0,19	-0,45	8,48E-03	1,94E-02
cg07833019	*EGFLAM*	-0,15	0,25	1,07E-02	2,42E-02
cg00984146	*EGFLAM*	-0,20	0,25	1,07E-02	2,42E-02
cg24653263	*EGFLAM*	-0,23	0,25	1,07E-02	2,42E-02
cg06000994	*ZNF536*	-0,31	0,07	1,50E-02	3,29E-02
cg23331421	*ZNF536*	-0,24	0,07	1,50E-02	3,29E-02
cg08977371	*CDH13*	-0,19	0,21	1,86E-02	4,04E-02
cg09825093	*CDH13*	-0,16	0,21	1,86E-02	4,04E-02
cg08747377	*CDH13*	-0,27	0,21	1,86E-02	4,04E-02
cg05374412	*CDH13*	-0,29	0,21	1,86E-02	4,04E-02
cg00806490	*CDH13*	-0,16	0,21	1,86E-02	4,04E-02
cg06680481	*C20orf56*	-0,19	0,00	2,18E-02	4,68E-02
cg08582485	*C20orf56*	-0,23	0,00	2,18E-02	4,68E-02
cg05325193	*C20orf56*	-0,20	0,00	2,18E-02	4,68E-02
cg06226567	*C20orf56*	-0,17	0,00	2,18E-02	4,68E-02
cg18299578	*FOXG1*	-0,26	-0,36	2,72E-02	5,78E-02
cg02681442	*FOXG1*	-0,24	-0,36	2,72E-02	5,78E-02
cg19553563	*TRDN*	0,15	-0,43	2,94E-02	6,22E-02
cg16726875	*TRDN*	0,18	-0,43	2,94E-02	6,22E-02
cg14462830	*TRDN*	0,15	-0,43	2,94E-02	6,22E-02
cg17889149	*TRDN*	0,18	-0,43	2,94E-02	6,22E-02
cg02637078	*SYT5*	-0,19	0,40	3,77E-02	7,82E-02
cg14047019	*SYT5*	-0,12	0,40	3,77E-02	7,82E-02
cg11960711	*ITGB1BP3*	-0,12	-0,07	4,29E-02	8,86E-02
cg27601582	*ITGB1BP3*	-0,20	-0,07	4,29E-02	8,86E-02
cg12208256	*CSMD1*	-0,25	0,73	4,48E-02	9,23E-02
cg06484232	*OR52E4*	0,18	-0,07	5,49E-02	1,12E-01
cg27363327	*TTBK1*	-0,29	-0,32	1,00E-01	2,01E-01
cg09940360	*TTBK1*	-0,26	-0,32	1,00E-01	2,01E-01
cg07016914	*CARTPT*	-0,17	0,24	1,07E-01	2,13E-01
cg08743812	*CARTPT*	-0,19	0,24	1,07E-01	2,13E-01
cg26146287	*CARTPT*	-0,16	0,24	1,07E-01	2,13E-01
cg02762440	*CARTPT*	-0,19	0,24	1,07E-01	2,13E-01
cg01187920	*CARTPT*	-0,25	0,24	1,07E-01	2,13E-01
cg04251750	*CARTPT*	-0,26	0,24	1,07E-01	2,13E-01

cg07935012	*DCC*	-0,11	-0,17	1,42E-01	2,80E-01
cg26073933	*C1orf61*	-0,11	-0,24	2,03E-01	3,91E-01
cg09938227	*C1orf61*	-0,28	-0,24	2,03E-01	3,91E-01
cg15270145	*C1orf61*	-0,16	-0,24	2,03E-01	3,91E-01
cg23989821	*C14orf39*	-0,20	-0,19	2,57E-01	4,89E-01
cg18240143	*C14orf39*	-0,13	-0,19	2,57E-01	4,89E-01
cg02736505	*FAM150B*	-0,29	0,50	3,21E-01	6,08E-01
cg10269365	*CCDC140*	-0,20	-0,01	3,27E-01	6,16E-01
cg27164663	*LINGO2*	-0,17	0,13	3,31E-01	6,22E-01
cg12658269	*LINGO2*	-0,19	0,13	3,31E-01	6,22E-01
cg05279757	*LINGO2*	-0,25	0,13	3,31E-01	6,22E-01
cg26128092	*WDR8*	-0,25	0,11	3,56E-01	6,67E-01
cg15923139	*SORBS2*	0,16	0,28	3,83E-01	7,11E-01
cg09260773	*TBX15*	-0,13	0,18	5,21E-01	9,59E-01
cg19795338	*ATP13A4*	0,14	0,30	9,53E-01	1,00E+00
cg01250864	*IFNGR1*	0,14	-0,11	7,10E-01	1,00E+00
cg18041719	*SEMA3D*	0,13	0,21	9,99E-01	1,00E+00
cg15522953	*ZMIZ1*	0,16	0,02	9,97E-01	1,00E+00
cg21108805	*WIF1*	0,12	0,15	1,00E+00	1,00E+00
cg12221621	*SPSB1*	-0,12	-0,02	1,00E+00	1,00E+00
cg05764238	*ARHGEF19*	-0,10	0,09	9,57E-01	1,00E+00
cg10595388	*SYT6*	-0,17	0,04	8,31E-01	1,00E+00
cg01567634	*SYT6*	-0,18	0,04	8,31E-01	1,00E+00
cg03514545	*SYT6*	-0,17	0,04	8,31E-01	1,00E+00
cg08883066	*SYT6*	-0,11	0,04	8,31E-01	1,00E+00
cg03606772	*CRCT1*	-0,24	-0,16	9,99E-01	1,00E+00
cg00302521	*CRCT1*	-0,14	-0,16	9,99E-01	1,00E+00
cg09873164	*CRCT1*	-0,32	-0,16	9,99E-01	1,00E+00
cg03352287	*CRCT1*	-0,26	-0,16	9,99E-01	1,00E+00
cg01864564	*CRCT1*	-0,21	-0,16	9,99E-01	1,00E+00
cg19272720	*PLEKHA6*	0,18	-0,03	1,00E+00	1,00E+00
cg20902122	*DISP1*	0,13	0,14	8,79E-01	1,00E+00
cg23719318	*GDF7*	-0,38	0,00	5,67E-01	1,00E+00
cg18552861	*GDF7*	-0,23	0,00	5,67E-01	1,00E+00
cg05899618	*GDF7*	-0,14	0,00	5,67E-01	1,00E+00
cg06167664	*DPYSL5*	-0,24	-0,40	9,96E-01	1,00E+00
cg14483208	*DPYSL5*	-0,17	-0,40	9,96E-01	1,00E+00
cg10633958	*SLC8A1*	-0,11	-0,03	1,00E+00	1,00E+00
cg15523958	*PLEKHH2*	-0,17	0,07	9,95E-01	1,00E+00
cg00778995	*POU3F3*	-0,27	-0,13	9,81E-01	1,00E+00
cg18524739	*POU3F3*	-0,18	-0,13	9,81E-01	1,00E+00
cg02583418	*POU3F3*	-0,25	-0,13	9,81E-01	1,00E+00
cg24925163	*SFT2D3*	0,11	0,00	1,00E+00	1,00E+00
cg26660002	*GALNT13*	-0,24	-0,02	1,00E+00	1,00E+00
cg14630357	*GALNT13*	-0,20	-0,02	1,00E+00	1,00E+00

cg22373593	*GALNT13*	0,20	-0,02	1,00E+00	1,00E+00
cg23566791	*GALNT13*	0,19	-0,02	1,00E+00	1,00E+00
cg08813062	*SP9*	-0,25	-0,02	1,00E+00	1,00E+00
cg14459130	*SP9*	-0,20	-0,02	1,00E+00	1,00E+00
cg17964510	*SP9*	-0,24	-0,02	1,00E+00	1,00E+00
cg11528849	*NEUROD1*	-0,23	-0,14	1,00E+00	1,00E+00
cg16640855	*NEUROD1*	-0,22	-0,14	1,00E+00	1,00E+00
cg01789478	*NEUROD1*	-0,15	-0,14	1,00E+00	1,00E+00
cg22359606	*NEUROD1*	-0,21	-0,14	1,00E+00	1,00E+00
cg06963053	*NEUROD1*	-0,17	-0,14	1,00E+00	1,00E+00
cg19711579	*NEUROD1*	-0,14	-0,14	1,00E+00	1,00E+00
cg09942248	*NEUROD1*	-0,18	-0,14	1,00E+00	1,00E+00
cg06008912	*TMEFF2*	-0,27	-0,02	9,99E-01	1,00E+00
cg02170986	*RFTN2*	-0,10	0,10	9,78E-01	1,00E+00
cg01812328	*ACADL*	0,14	-0,18	1,00E+00	1,00E+00
cg19047707	*ECEL1*	-0,20	-0,10	1,00E+00	1,00E+00
cg01714811	*EOMES*	-0,12	-0,18	1,00E+00	1,00E+00
cg02694810	*EOMES*	-0,18	-0,18	1,00E+00	1,00E+00
cg07287167	*ROBO1*	-0,18	-0,01	1,00E+00	1,00E+00
cg07526135	*STAG1*	0,18	-0,13	8,47E-01	1,00E+00
cg26807939	*SOX14*	-0,18	0,06	1,00E+00	1,00E+00
cg16428251	*SOX14*	-0,24	0,06	1,00E+00	1,00E+00
cg04228468	*SLITRK3*	-0,19	0,02	1,00E+00	1,00E+00
cg18447727	*SLITRK3*	-0,16	0,02	1,00E+00	1,00E+00
cg13619915	*SLITRK3*	-0,20	0,02	1,00E+00	1,00E+00
cg14717170	*SLITRK3*	-0,23	0,02	1,00E+00	1,00E+00
cg27363829	*SLITRK3*	-0,29	0,02	1,00E+00	1,00E+00
cg09269007	*SLITRK3*	-0,27	0,02	1,00E+00	1,00E+00
cg14502484	*SLC6A3*	-0,29	-0,23	9,61E-01	1,00E+00
cg17818996	*CDH12*	0,16	0,12	9,99E-01	1,00E+00
cg00958768	*CDH12*	0,15	0,12	9,99E-01	1,00E+00
cg14505704	*CDH6*	-0,15	0,05	9,99E-01	1,00E+00
cg12238343	*RXFP3*	-0,18	0,03	1,00E+00	1,00E+00
cg00516449	*RXFP3*	-0,16	0,03	1,00E+00	1,00E+00
cg20963030	*ATP10B*	0,15	0,09	9,98E-01	1,00E+00
cg05157516	*HRH2*	-0,27	-0,11	9,85E-01	1,00E+00
cg06847624	*PFN3*	-0,21	0,01	1,00E+00	1,00E+00
cg00642460	*PFN3*	-0,18	0,01	1,00E+00	1,00E+00
cg14859460	*GRM6*	-0,24	0,04	1,00E+00	1,00E+00
cg18417237	*LRFN2*	-0,15	-0,26	9,85E-01	1,00E+00
cg13290564	*HCRTR2*	0,14	0,09	9,97E-01	1,00E+00
cg04703174	*KHDRBS2*	-0,12	-0,02	8,20E-01	1,00E+00
cg11315153	*KHDRBS2*	-0,16	-0,02	8,20E-01	1,00E+00
cg14224762	*RIMS1*	-0,15	0,00	1,00E+00	1,00E+00
cg12103152	*HTR1B*	-0,19	0,02	1,00E+00	1,00E+00

cg23424273	*HTR1B*	-0,20	0,02	1,00E+00	1,00E+00
cg08614481	*HTR1B*	-0,24	0,02	1,00E+00	1,00E+00
cg12215457	*HTR1B*	-0,19	0,02	1,00E+00	1,00E+00
cg13077519	*HTR1B*	-0,24	0,02	1,00E+00	1,00E+00
cg00183186	*HTR1B*	-0,24	0,02	1,00E+00	1,00E+00
cg13315970	*NT5E*	-0,23	-0,01	1,00E+00	1,00E+00
cg16387913	*NT5E*	0,17	-0,01	1,00E+00	1,00E+00
cg01489686	*FUT9*	-0,16	-0,44	6,14E-01	1,00E+00
cg20790058	*FUT9*	-0,19	-0,44	6,14E-01	1,00E+00
cg01815538	*PRDM13*	-0,16	-0,30	1,00E+00	1,00E+00
cg02400433	*PRDM13*	-0,20	-0,30	1,00E+00	1,00E+00
cg12227907	*PRDM13*	-0,18	-0,30	1,00E+00	1,00E+00
cg04974062	*PRDM13*	-0,19	-0,30	1,00E+00	1,00E+00
cg19223299	*PRDM13*	-0,23	-0,30	1,00E+00	1,00E+00
cg27125972	*PRDM13*	-0,12	-0,30	1,00E+00	1,00E+00
cg21239311	*NR2E1*	-0,26	-0,11	1,00E+00	1,00E+00
cg05135644	*NR2E1*	-0,21	-0,11	1,00E+00	1,00E+00
cg04407470	*NR2E1*	-0,34	-0,11	1,00E+00	1,00E+00
cg13776285	*NR2E1*	-0,28	-0,11	1,00E+00	1,00E+00
cg06103642	*NR2E1*	-0,20	-0,11	1,00E+00	1,00E+00
cg08929002	*NR2E1*	-0,16	-0,11	1,00E+00	1,00E+00
cg02149708	*T*	-0,20	-0,12	9,99E-01	1,00E+00
cg17188046	*T*	-0,23	-0,12	9,99E-01	1,00E+00
cg19675288	*T*	-0,22	-0,12	9,99E-01	1,00E+00
cg06463958	*T*	-0,19	-0,12	9,99E-01	1,00E+00
cg20870512	*UNCX*	-0,19	-0,01	9,98E-01	1,00E+00
cg14658067	*UNCX*	-0,15	-0,01	9,98E-01	1,00E+00
cg01202666	*TWIST1*	-0,20	-0,05	9,97E-01	1,00E+00
cg22498251	*TWIST1*	-0,16	-0,05	9,97E-01	1,00E+00
cg11124364	*TMEM196*	-0,13	-0,04	9,81E-01	1,00E+00
cg13298841	*HECW1*	-0,19	-0,16	1,00E+00	1,00E+00
cg08652532	*HECW1*	-0,24	-0,16	1,00E+00	1,00E+00
cg13027493	*HECW1*	-0,13	-0,16	1,00E+00	1,00E+00
cg25717239	*HECW1*	-0,17	-0,16	1,00E+00	1,00E+00
cg00974941	*C7orf57*	-0,19	-0,27	9,85E-01	1,00E+00
cg19919208	*MAGI2*	0,16	-0,07	1,00E+00	1,00E+00
cg11655410	*MAGI2*	0,12	-0,07	1,00E+00	1,00E+00
cg16111259	*MAGI2*	0,18	-0,07	1,00E+00	1,00E+00
cg02523844	*MAGI2*	-0,21	-0,07	1,00E+00	1,00E+00
cg24021956	*MAGI2*	-0,22	-0,07	1,00E+00	1,00E+00
cg10168361	*GNAT3*	0,13	-0,21	1,00E+00	1,00E+00
cg24734879	*SEMA3D*	0,13	0,21	9,99E-01	1,00E+00
cg01182980	*SEMA3D*	0,17	0,21	9,99E-01	1,00E+00
cg24277791	*LHFPL3*	-0,18	0,00	1,00E+00	1,00E+00
cg00903099	*HTR5A*	-0,19	-0,01	1,00E+00	1,00E+00

cg00595237	*GATA4*	-0,22	-0,03	9,86E-01	1,00E+00
cg27100236	*GATA4*	-0,18	-0,03	9,86E-01	1,00E+00
cg01509237	*NKAIN3*	-0,13	-0,03	1,00E+00	1,00E+00
cg07719492	*PRDM14*	-0,16	0,14	1,00E+00	1,00E+00
cg01610488	*TRPA1*	-0,16	0,19	1,00E+00	1,00E+00
cg20890210	*KCNB2*	-0,22	-0,38	8,84E-01	1,00E+00
cg19995049	*RPL30*	0,16	0,04	1,00E+00	1,00E+00
cg24627900	*HPYR1*	0,17	0,03	9,87E-01	1,00E+00
cg08326794	*BNC2*	0,16	-0,11	1,00E+00	1,00E+00
cg08826741	*BNC2*	0,16	-0,11	1,00E+00	1,00E+00
cg05674150	*BNC2*	-0,11	-0,11	1,00E+00	1,00E+00
cg17398595	*SH3GL2*	-0,18	-0,11	1,00E+00	1,00E+00
cg25440961	*SH3GL2*	-0,19	-0,11	1,00E+00	1,00E+00
cg11258943	*FOXE1*	-0,26	-0,10	7,06E-01	1,00E+00
cg13724160	*GRIN3A*	-0,14	-0,14	9,24E-01	1,00E+00
cg05868019	*OR13C8*	0,17	0,01	9,45E-01	1,00E+00
cg14080475	*AKNA*	-0,14	-0,02	1,00E+00	1,00E+00
cg11970741	*TNC*	0,11	0,30	5,98E-01	1,00E+00
cg25042430	*KIAA1217*	-0,20	-0,08	1,00E+00	1,00E+00
cg18673954	*INA*	-0,24	-0,52	6,24E-01	1,00E+00
cg23642747	*INA*	-0,27	-0,52	6,24E-01	1,00E+00
cg21384402	*INA*	-0,14	-0,52	6,24E-01	1,00E+00
cg24680586	*INA*	-0,25	-0,52	6,24E-01	1,00E+00
cg04142605	*INA*	-0,17	-0,52	6,24E-01	1,00E+00
cg13654588	*PRLHR*	-0,19	-0,03	7,04E-01	1,00E+00
cg12614318	*PRLHR*	-0,14	-0,03	7,04E-01	1,00E+00
cg02531437	*PRLHR*	-0,16	-0,03	7,04E-01	1,00E+00
cg19403534	*PRLHR*	-0,19	-0,03	7,04E-01	1,00E+00
cg14526953	*GPR123*	-0,23	-0,05	9,88E-01	1,00E+00
cg01178971	*KCNQ1*	-0,16	0,01	9,90E-01	1,00E+00
cg23178580	*OR2AG1*	0,17	0,00	1,00E+00	1,00E+00
cg05393484	*OR6A2*	0,14	-0,02	9,97E-01	1,00E+00
cg10157954	*ZNF215*	0,15	-0,11	9,98E-01	1,00E+00
cg04270835	*SLC17A6*	-0,13	-0,12	7,60E-01	1,00E+00
cg22371972	*SLC17A6*	-0,14	-0,12	7,60E-01	1,00E+00
cg03168749	*OR8B12*	0,12	0,02	9,99E-01	1,00E+00
cg12859211	*ROBO3*	-0,19	0,26	1,00E+00	1,00E+00
cg09251429	*ROBO3*	-0,19	0,26	1,00E+00	1,00E+00
cg01839688	*ROBO3*	-0,11	0,26	1,00E+00	1,00E+00
cg08833670	*ROBO3*	-0,18	0,26	1,00E+00	1,00E+00
cg11701621	*WNT1*	-0,22	0,03	5,48E-01	1,00E+00
cg16112157	*KRT6A*	0,15	-0,40	9,99E-01	1,00E+00
cg21383810	*WIF1*	-0,19	0,15	1,00E+00	1,00E+00
cg24719984	*PPFIA2*	-0,15	-0,23	9,55E-01	1,00E+00
cg24214152	*ALX1*	-0,26	-0,36	5,75E-01	1,00E+00

cg24083469	*ALX1*	-0,22	-0,36	5,75E-01	1,00E+00
cg14116122	*ALX1*	-0,22	-0,36	5,75E-01	1,00E+00
cg07780095	*ALX1*	-0,25	-0,36	5,75E-01	1,00E+00
cg06552356	*ALX1*	-0,19	-0,36	5,75E-01	1,00E+00
cg05677041	*SRRM4*	-0,17	0,06	9,77E-01	1,00E+00
cg02345991	*SRRM4*	-0,26	0,06	9,77E-01	1,00E+00
cg14843800	*SRRM4*	-0,20	0,06	9,77E-01	1,00E+00
cg13939792	*SIX6*	-0,22	-0,07	1,00E+00	1,00E+00
cg19456540	*SIX6*	-0,28	-0,07	1,00E+00	1,00E+00
cg14511698	*SIX6*	-0,17	-0,07	1,00E+00	1,00E+00
cg07862488	*SIX6*	-0,18	-0,07	1,00E+00	1,00E+00
cg00859478	*SIX6*	-0,20	-0,07	1,00E+00	1,00E+00
cg26517663	*FOXB1*	-0,19	0,02	1,00E+00	1,00E+00
cg07928083	*FOXB1*	-0,19	0,02	1,00E+00	1,00E+00
cg07060006	*FOXB1*	-0,16	0,02	1,00E+00	1,00E+00
cg20987924	*FOXB1*	-0,30	0,02	1,00E+00	1,00E+00
cg03059131	*FOXB1*	-0,32	0,02	1,00E+00	1,00E+00
cg03461851	*C2CD4B*	0,15	-0,18	1,00E+00	1,00E+00
cg01744019	*HBM*	-0,16	0,05	9,97E-01	1,00E+00
cg09854626	*SOX8*	-0,21	0,31	8,81E-01	1,00E+00
cg06490869	*CRYM*	-0,15	-0,23	1,00E+00	1,00E+00
cg20823471	*STX1B*	-0,19	0,20	5,57E-01	1,00E+00
cg04180299	*RLTPR*	-0,25	0,36	5,95E-01	1,00E+00
cg21309147	*STAC2*	-0,14	0,11	1,00E+00	1,00E+00
cg08504407	*HRNBP3*	-0,14	0,09	7,24E-01	1,00E+00
cg22105145	*LOC642597*	-0,17	0,05	8,83E-01	1,00E+00
cg00247557	*LOC642597*	-0,16	0,05	8,83E-01	1,00E+00
cg23234640	*LOC642597*	-0,12	0,05	8,83E-01	1,00E+00
cg02664349	*FAM38B*	-0,22	-0,16	1,00E+00	1,00E+00
cg04696345	*ZBTB7C*	-0,13	0,04	1,00E+00	1,00E+00
cg05928342	*ZNF177*	-0,20	0,13	1,00E+00	1,00E+00
cg08065231	*ZNF177*	-0,26	0,13	1,00E+00	1,00E+00
cg17283453	*ZNF177*	-0,17	0,13	1,00E+00	1,00E+00
cg16055378	*VSTM2B*	-0,28	0,03	9,90E-01	1,00E+00
cg01464835	*VSTM2B*	-0,20	0,03	9,90E-01	1,00E+00
cg10743390	*VSTM2B*	-0,19	0,03	9,90E-01	1,00E+00
cg05056953	*DPF1*	-0,12	-0,03	1,00E+00	1,00E+00
cg09568464	*ZNF582*	-0,17	0,06	1,00E+00	1,00E+00
cg22647407	*ZNF582*	-0,18	0,06	1,00E+00	1,00E+00
cg09861917	*FOXA2*	-0,23	-0,19	1,00E+00	1,00E+00
cg16963144	*FOXA2*	-0,24	-0,19	1,00E+00	1,00E+00
cg06672560	*CBLN4*	-0,15	0,17	1,00E+00	1,00E+00

BCVY: breast cancer in very young women; BCO: breast cancer in old women.

ANNEX IX. Significant CpG probes localized in enhancer regions from hypomethylation global profile in BCVY from metEPICVal study.

Probes	UCSC RefGene Name	Methylation differences (BCVY - BCO)	Gene expression differences (BCVY - BCO)	p-value	Adjusted p-value
cg02950416	*BCAN*	-0,29	0,79	6,67E-41	9,81E-40
cg20270188	*BCAN*	-0,22	0,79	6,67E-41	9,81E-40
cg07224636	*TMTC1*	0,14	0,76	1,56E-28	1,38E-27
cg08729318	*TMTC1*	-0,12	0,76	1,56E-28	1,38E-27
cg26365957	*PTPRS*	-0,18	-0,38	4,84E-28	4,21E-27
cg20548934	*CADM3*	-0,26	1,10	6,46E-28	5,47E-27
cg15891112	*CADM3*	-0,29	1,10	6,46E-28	5,47E-27
cg03789420	*CADM3*	-0,17	1,10	6,46E-28	5,47E-27
cg00957516	*CADM3*	-0,18	1,10	6,46E-28	5,47E-27
cg20724589	*HIVEP3*	-0,16	0,30	6,77E-25	5,55E-24
cg14839013	*HIVEP3*	-0,11	0,30	6,77E-25	5,55E-24
cg03938432	*BAI3*	-0,21	0,47	1,69E-24	1,30E-23
cg10244047	*BAI3*	-0,13	0,47	1,69E-24	1,30E-23
cg06953874	*HOXD9*	-0,14	0,43	3,31E-19	2,00E-18
cg10957151	*HOXD9*	-0,21	0,43	3,31E-19	2,00E-18
cg14142007	*HOXD9*	-0,23	0,43	3,31E-19	2,00E-18
cg22805974	*HOXD9*	-0,17	0,43	3,31E-19	2,00E-18
cg14991487	*HOXD9*	-0,21	0,43	3,31E-19	2,00E-18
cg09578028	*HOXD9*	-0,24	0,43	3,31E-19	2,00E-18
cg05167251	*HOXD9*	-0,24	0,43	3,31E-19	2,00E-18
cg03637878	*JAM3*	-0,20	0,36	4,31E-14	1,84E-13
cg00939226	*JAM3*	-0,23	0,36	4,31E-14	1,84E-13
cg24899571	*JAM3*	-0,16	0,36	4,31E-14	1,84E-13
cg02174225	*JAM3*	-0,17	0,36	4,31E-14	1,84E-13
cg05312305	*ZNF655*	-0,10	0,36	2,62E-10	9,74E-10
cg11164088	*SHANK1*	-0,10	0,59	2,80E-10	1,03E-09
cg15805568	*SHANK1*	-0,11	0,59	2,80E-10	1,03E-09
cg17787134	*IGF2BP1*	-0,18	-0,51	6,70E-10	2,39E-09
cg27256435	*LRCH2*	-0,18	0,45	1,17E-08	4,00E-08
cg06321345	*FAM5B*	-0,13	0,89	3,62E-07	1,11E-06
cg21852439	*GRID2*	-0,10	0,47	2,05E-06	6,15E-06
cg18629427	*GRID2*	-0,16	0,47	2,05E-06	6,15E-06
cg20057676	*ZIC2*	-0,12	-0,83	9,06E-06	2,60E-05
cg24742746	*ZIC2*	-0,17	-0,83	9,06E-06	2,60E-05
cg17206583	*VGF*	-0,12	-0,38	9,19E-06	2,63E-05
cg12135573	*VGF*	-0,17	-0,38	9,19E-06	2,63E-05
cg21657577	*ZNF578*	-0,22	0,24	1,10E-03	2,74E-03
cg12665460	*ZNF578*	-0,20	0,24	1,10E-03	2,74E-03
cg11909748	*ZNF578*	-0,15	0,24	1,10E-03	2,74E-03
cg14582763	*ZNF578*	-0,20	0,24	1,10E-03	2,74E-03

cg13461241	*ZNF578*	-0,21	0,24	1,10E-03	2,74E-03
cg21553182	*ZNF578*	-0,18	0,24	1,10E-03	2,74E-03
cg25763393	*ZNF578*	-0,22	0,24	1,10E-03	2,74E-03
cg16210670	*SNUPN*	-0,10	0,08	6,60E-03	1,55E-02
cg07958937	*GRB10*	-0,10	-0,12	8,06E-03	1,85E-02
cg25848339	*GRB10*	0,15	-0,12	8,06E-03	1,85E-02
cg03370077	*GRB10*	0,18	-0,12	8,06E-03	1,85E-02
cg11321337	*LRRC7*	0,16	0,00	1,32E-02	2,90E-02
cg17163179	*LRRC7*	0,14	0,00	1,32E-02	2,90E-02
cg08977371	*CDH13*	-0,19	0,21	1,86E-02	4,04E-02
cg09825093	*CDH13*	-0,16	0,21	1,86E-02	4,04E-02
cg08747377	*CDH13*	-0,27	0,21	1,86E-02	4,04E-02
cg05374412	*CDH13*	-0,29	0,21	1,86E-02	4,04E-02
cg00806490	*CDH13*	-0,16	0,21	1,86E-02	4,04E-02
cg18299578	*FOXG1*	-0,26	-0,36	2,72E-02	5,78E-02
cg02681442	*FOXG1*	-0,24	-0,36	2,72E-02	5,78E-02
cg12204957	*IRAK2*	0,15	0,22	6,06E-02	1,23E-01
cg17410236	*FLRT2*	-0,14	0,40	1,73E-01	3,38E-01
cg01711160	*FLRT2*	-0,23	0,40	1,73E-01	3,38E-01
cg18267374	*NEFM*	-0,20	0,12	1,82E-01	3,56E-01
cg07552803	*NEFM*	-0,19	0,12	1,82E-01	3,56E-01
cg20701183	*LMO4*	-0,13	-0,26	3,57E-01	6,68E-01
cg08603642	*ROR1*	0,15	0,00	1,00E+00	1,00E+00
cg06167664	*DPYSL5*	-0,24	-0,40	9,96E-01	1,00E+00
cg14483208	*DPYSL5*	-0,17	-0,40	9,96E-01	1,00E+00
cg15706621	*SPTBN1*	-0,12	-0,03	1,00E+00	1,00E+00
cg05276829	*SPTBN1*	-0,11	-0,03	1,00E+00	1,00E+00
cg00778995	*POU3F3*	-0,27	-0,13	9,81E-01	1,00E+00
cg18524739	*POU3F3*	-0,18	-0,13	9,81E-01	1,00E+00
cg02583418	*POU3F3*	-0,25	-0,13	9,81E-01	1,00E+00
cg08813062	*SP9*	-0,25	-0,02	1,00E+00	1,00E+00
cg14459130	*SP9*	-0,20	-0,02	1,00E+00	1,00E+00
cg17964510	*SP9*	-0,24	-0,02	1,00E+00	1,00E+00
cg20415809	*ITGA4*	-0,20	-0,11	9,86E-01	1,00E+00
cg06952671	*ITGA4*	-0,26	-0,11	9,86E-01	1,00E+00
cg25024074	*ITGA4*	-0,22	-0,11	9,86E-01	1,00E+00
cg16057262	*ITGA4*	-0,14	-0,11	9,86E-01	1,00E+00
cg06008912	*TMEFF2*	-0,27	-0,02	9,99E-01	1,00E+00
cg05911659	*CPLX2*	-0,20	0,01	9,94E-01	1,00E+00
cg23495748	*CPLX2*	-0,18	0,01	9,94E-01	1,00E+00
cg19885761	*CPLX2*	-0,22	0,01	9,94E-01	1,00E+00
cg16590189	*CPLX2*	-0,25	0,01	9,94E-01	1,00E+00
cg26982364	*CPLX2*	-0,14	0,01	9,94E-01	1,00E+00
cg23409774	*CPLX2*	-0,21	0,01	9,94E-01	1,00E+00
cg14224762	*RIMS1*	-0,15	0,00	1,00E+00	1,00E+00

cg21717549	*VGLL2*	-0,20	0,02	1,00E+00	1,00E+00
cg08983097	*VGLL2*	-0,26	0,02	1,00E+00	1,00E+00
cg13371788	*VGLL2*	-0,20	0,02	1,00E+00	1,00E+00
cg20870512	*UNCX*	-0,19	-0,01	9,98E-01	1,00E+00
cg14658067	*UNCX*	-0,15	-0,01	9,98E-01	1,00E+00
cg16842053	*CARD11*	-0,19	0,01	1,00E+00	1,00E+00
cg00595237	*GATA4*	-0,22	-0,03	9,86E-01	1,00E+00
cg27100236	*GATA4*	-0,18	-0,03	9,86E-01	1,00E+00
cg01610488	*TRPA1*	-0,16	0,19	1,00E+00	1,00E+00
cg05191879	*C8orf73*	-0,17	0,23	9,97E-01	1,00E+00
cg12859211	*ROBO3*	-0,19	0,26	1,00E+00	1,00E+00
cg09251429	*ROBO3*	-0,19	0,26	1,00E+00	1,00E+00
cg01839688	*ROBO3*	-0,11	0,26	1,00E+00	1,00E+00
cg08833670	*ROBO3*	-0,18	0,26	1,00E+00	1,00E+00
cg09143195	*CNTN1*	-0,16	-0,11	1,00E+00	1,00E+00
cg15087347	*CNTN1*	-0,16	-0,11	1,00E+00	1,00E+00
cg19987210	*AVPR1A*	-0,29	-0,11	1,00E+00	1,00E+00
cg09040797	*AVPR1A*	-0,25	-0,11	1,00E+00	1,00E+00
cg10862431	*AVPR1A*	-0,25	-0,11	1,00E+00	1,00E+00
cg27032502	*AVPR1A*	-0,19	-0,11	1,00E+00	1,00E+00
cg13631391	*AVPR1A*	-0,21	-0,11	1,00E+00	1,00E+00
cg17771605	*LOC283392*	-0,21	0,40	9,85E-01	1,00E+00
cg01030121	*LOC283392*	-0,29	0,40	9,85E-01	1,00E+00
cg09972192	*LOC283392*	-0,14	0,40	9,85E-01	1,00E+00
cg00580978	*LOC283392*	-0,16	0,40	9,85E-01	1,00E+00
cg13099890	*LOC283392*	-0,25	0,40	9,85E-01	1,00E+00
cg19516457	*LOC283392*	-0,14	0,40	9,85E-01	1,00E+00
cg26517663	*FOXB1*	-0,19	0,02	1,00E+00	1,00E+00
cg07928083	*FOXB1*	-0,19	0,02	1,00E+00	1,00E+00
cg07060006	*FOXB1*	-0,16	0,02	1,00E+00	1,00E+00
cg20987924	*FOXB1*	-0,30	0,02	1,00E+00	1,00E+00
cg03059131	*FOXB1*	-0,32	0,02	1,00E+00	1,00E+00
cg04180299	*RLTPR*	-0,25	0,36	5,95E-01	1,00E+00
cg02701059	*HNF1B*	-0,22	-0,02	9,98E-01	1,00E+00
cg14054928	*CA10*	-0,24	-0,16	7,80E-01	1,00E+00
cg14073722	*CA10*	-0,23	-0,16	7,80E-01	1,00E+00
cg14231749	*CA10*	-0,13	-0,16	7,80E-01	1,00E+00
cg04946749	*MGAT5B*	-0,24	0,09	9,99E-01	1,00E+00
cg09568464	*ZNF582*	-0,17	0,06	1,00E+00	1,00E+00
cg22647407	*ZNF582*	-0,18	0,06	1,00E+00	1,00E+00
cg10439765	*SLC12A5*	-0,16	0,07	9,95E-01	1,00E+00

BCVY: breast cancer in very young women; BCO: breast cancer in old women.

ANNEX X. Significant CpG probes localized in enhancer regions from hypermethylation distinctive signature in BCVY from metEPICVal study.

Probes	UCSC RefGene Name	Methylation differences (BCVY - BCO)	Gene expression differences (BCVY - BCO)	p-value	Adjusted p-value
cg08494221	*PURG*	-0,10	0,54	1,01E-49	4,57E-48
cg14185504	*FIG4*	0,13	-0,24	1,53E-13	9,92E-13
cg00670438	*PYGL*	-0,18	-0,33	5,60E-06	2,06E-05
cg19839026	*PYGL*	-0,15	-0,33	5,60E-06	2,06E-05
cg07035145	*PYGL*	-0,17	-0,33	5,60E-06	2,06E-05
cg14483244	*HDAC5*	-0,10	0,23	1,05E-04	3,39E-04
cg10142997	*PCDH9*	0,11	0,36	8,54E-04	2,47E-03
cg12204957	*IRAK2*	0,15	0,23	2,45E-02	6,23E-02
cg17237907	*MICAL1*	-0,10	0,08	3,78E-01	7,56E-01
cg10807084	*AK5*	-0,13	0,39	1,00E+00	1,00E+00
cg26440512	*ABLIM2*	0,10	-0,04	9,71E-01	1,00E+00
cg01708617	*RAI14*	-0,10	-0,10	9,52E-01	1,00E+00
cg18181545	*EDIL3*	0,11	-0,49	6,56E-01	1,00E+00
cg10070202	*OSBP2*	0,10	-0,01	9,38E-01	1,00E+00

BCVY: breast cancer in very young women; BCO: breast cancer in old women.

ANNEX XI. Significant CpG probes localized in TFBS regions from hypermethylation distinctive signature in BCVY from metEPICVal study.

Probes	UCSC RefGene Name	Methylation differences (BCVY - BCO)	Gene expression differences (BCVY - BCO)	p-value	Adjusted p-value
cg26581860	*EHF*	0,16	-1,18	1,49E-26	1,76E-25
cg04788216	*EHF*	0,18	-1,18	1,49E-26	1,76E-25
cg09974610	*EHF*	0,20	-1,18	1,49E-26	1,76E-25
cg22596683	*EHF*	0,17	-1,18	1,49E-26	1,76E-25
cg20823324	*TMED3*	0,12	-0,29	2,50E-24	2,72E-23
cg12033458	*UNC5C*	0,15	0,73	6,45E-20	6,75E-19
cg06571387	*HOXD12*	-0,11	-0,07	1,12E-19	1,13E-18
cg23167136	*CSRP1*	0,14	0,35	8,41E-18	7,63E-17
cg19844232	*ASAP1*	-0,12	-0,27	6,67E-17	5,67E-16
cg06261500	*SND1*	0,14	-0,12	3,01E-14	2,10E-13
cg06026636	*SND1*	0,15	-0,12	3,01E-14	2,10E-13
cg14599421	*DTNB*	0,15	-0,35	1,81E-13	1,14E-12
cg14022913	*NEIL3*	0,13	-0,62	1,85E-13	1,14E-12
cg27211179	*EXD2*	0,10	0,14	1,40E-10	7,05E-10
cg05293827	*PDE1C*	0,14	0,65	3,21E-10	1,59E-09
cg03522150	*C11orf49*	0,11	0,26	8,75E-10	4,25E-09
cg25029197	*KRT7*	0,14	-0,50	2,50E-04	7,80E-04
cg05966408	*NRG2*	0,10	0,61	2,60E-04	8,03E-04
cg15851515	*AIM1*	0,10	-0,32	4,60E-04	1,36E-03
cg08995655	*NUDT19*	0,12	-0,22	7,82E-03	2,05E-02
cg12204957	*IRAK2*	0,15	0,23	2,45E-02	6,23E-02
cg20137977	*MCC*	0,12	0,25	2,58E-02	6,50E-02
cg19958586	*CCRL2*	-0,11	-0,16	3,62E-02	8,50E-02
cg22663660	*LIPC*	0,15	0,36	6,02E-02	1,35E-01
cg00550493	*EYS*	0,12	-0,27	2,52E-01	5,39E-01
cg03140412	*SIGIRR*	-0,18	0,21	3,19E-01	6,68E-01
cg14707834	*SIGIRR*	-0,15	0,21	3,19E-01	6,68E-01
cg13525835	*CABLES1*	0,12	-0,11	3,67E-01	7,40E-01
cg27451581	*KLHDC7A*	0,17	-0,21	9,99E-01	1,00E+00
cg24435747	*IL28RA*	0,13	-0,18	9,81E-01	1,00E+00
cg06957219	*DISC1*	0,13	-0,04	1,00E+00	1,00E+00
cg07954193	*C1D*	0,14	-0,05	1,00E+00	1,00E+00
cg05548017	*FGD5*	0,10	-0,20	9,76E-01	1,00E+00
cg13244293	*NGLY1*	0,10	-0,06	8,11E-01	1,00E+00
cg10165543	*KALRN*	-0,12	0,05	1,00E+00	1,00E+00
cg18290075	*SGEF*	-0,12	0,18	7,42E-01	1,00E+00
cg20045300	*FAM114A1*	0,14	-0,07	1,00E+00	1,00E+00
cg20956407	*NDST1*	0,13	-0,10	9,67E-01	1,00E+00
cg23470770	*MAP3K5*	0,12	0,15	1,00E+00	1,00E+00
cg01250864	*IFNGR1*	0,14	-0,11	6,99E-01	1,00E+00

cg01687997	*C7orf50*	0,15	-0,01	1,00E+00	1,00E+00
cg18041719	*SEMA3D*	0,13	0,29	9,78E-01	1,00E+00
cg19612574	*MAPK8*	0,11	-0,02	1,00E+00	1,00E+00
cg02820058	*PRKG1*	0,13	0,13	1,00E+00	1,00E+00
cg20433989	*FNDC3A*	0,11	-0,05	1,00E+00	1,00E+00
cg10151281	*TPM4*	0,17	-0,07	1,00E+00	1,00E+00
cg11581475	*ZNF428*	0,10	0,06	1,00E+00	1,00E+00
cg06058203	*MECP2*	0,14	0,04	1,00E+00	1,00E+00

BCVY: breast cancer in very young women; BCO: breast cancer in old women.

ANNEX XII. DNA methylation age results for metEPICVal samples

Sample	Chronological Age	DNAmAge	AgeAccelerationDiff	AgeAccelerationResidual
Met32	63	87,32	24,32	23,06
Met31	62	52,73	-9,27	-11,1
Met20	33	27,72	-5,28	-23,73
Met33	69	83,21	14,21	16,39
Met55	34	54,32	20,32	2,44
Met36	51	38,14	-12,86	-21
Met22	35	48,44	13,44	-3,87
Met21	34	53,86	19,86	1,98
Met38	59	60,99	1,99	-1,56
Met15	33	43,19	10,19	-8,27
Met37	66	85,17	19,17	19,64
Met23	28	43,47	15,47	-5,85
Met24	34	53,35	19,35	1,47
Met57	34	38,74	4,74	-13,14
Met39	67	72,46	5,46	6,5
Met42	72	54,15	-17,85	-13,95
Met26	34	40,58	6,58	-11,3
Met43	61	61,11	0,11	-2,28
Met28	31	61,39	30,39	10,79
Met46	81	54,41	-26,59	-17,52
Met34	69	67,99	-1,01	1,18
Met56	26	36,65	10,65	-11,82
Met29	33	36,68	3,68	-14,77
Met30	28	50,54	22,54	1,22
Met50	34	86,36	52,36	34,48
Met54	33	40,54	7,54	-10,92
Met53	34	59,98	25,98	8,1
Met51	34	44,98	10,98	-6,9
Met49	54	55,88	1,88	-4,54
Met61	28	64,57	36,57	15,25
Met60	35	70,69	35,69	18,39
Met63	33	75,23	42,23	23,77
Met40	77	71,2	-5,8	0,97
Met62	35	49,16	14,16	-3,15

ANNEX XIII. List of 193 CpG probes that were significantly differentially methylated in BCVY (p-value < 0.05 and methylation differences ± 0.1) obtained by Wilcoxon rank sum test. Table includes p-values, methylation differences between BCVY and BCO, miRNAs regulated by significant probe, probe position relative to gene and CpG island.

Probe	p-value	Methylation differences (BCVY - BCO)	UCSC RefGene Name	UCSC RefGene Group	Relation to Island
cg07234865	4,41E-05	-1,71E-01	MIR9-1	Body;5'UTR	N_Shore
cg25229253	5,40E-05	1,99E-01	MIR548F1	Body	OpenSea
cg08737296	8,02E-05	-2,86E-01	MIR124-3	TSS1500	Island
cg26258950	9,72E-05	1,81E-01	MIR548F1	Body	OpenSea
cg07642566	3,39E-04	-1,37E-01	MIR7-3	TSS1500;Body	OpenSea
cg15270145	4,00E-04	-1,63E-01	MIR9-1	5'UTR;TSS200	Island
cg06138966	4,71E-04	2,12E-01	MIR2113	TSS1500	OpenSea
cg06372015	4,71E-04	-1,25E-01	MIR129-2	TSS1500	N_Shore
cg23570321	5,52E-04	-1,28E-01	MIR3692	TSS1500;Body	OpenSea
cg26371731	6,47E-04	-1,58E-01	MIR9-1	TSS200;5'UTR	N_Shore
cg21099163	6,47E-04	1,96E-01	MIR548F1	ExonBnd;Body	OpenSea
cg21927946	6,47E-04	-1,07E-01	MIR7-3	TSS1500;Body	OpenSea
cg14301438	7,55E-04	1,21E-01	MIR1273H	Body	OpenSea
cg17192679	7,55E-04	-1,59E-01	MIR1256	1stExon;5'UTR;Body	Island
cg20582655	1,02E-03	-1,89E-01	MIR9-1	TSS200;5'UTR	N_Shore
cg11638181	1,37E-03	-1,34E-01	MIR129-2	TSS1500	N_Shore
cg07796261	1,58E-03	1,47E-01	MIR548F1	Body	OpenSea
cg01336796	1,58E-03	1,35E-01	MIR4439	TSS1500;Body	OpenSea
cg05775586	1,58E-03	1,83E-01	MIR873	TSS1500;5'UTR	OpenSea
cg03475101	1,58E-03	1,42E-01	MIR204	TSS1500;Body	OpenSea
cg26005082	1,58E-03	-1,30E-01	MIR7-3	TSS1500;Body	OpenSea
cg22333214	1,82E-03	-2,26E-01	MIR137	TSS200	S_Shore
cg09910993	1,82E-03	1,37E-01	MIR548F1	Body	OpenSea
cg10688790	1,82E-03	1,71E-01	MIR548F3	TSS1500;Body	OpenSea
cg15791171	2,08E-03	1,68E-01	MIR548F3	Body	OpenSea
cg07494702	2,08E-03	1,49E-01	MIR548F3	Body	OpenSea
cg06855983	2,08E-03	-1,27E-01	MIR181C;MIR181D	TSS1500	N_Shore
cg22807241	2,39E-03	1,62E-01	MIR548F3	Body	OpenSea
cg16256490	2,39E-03	1,58E-01	MIR383	TSS1500;Body	OpenSea
cg17528257	2,73E-03	1,12E-01	MIR551B	TSS200;Body	OpenSea
cg17396522	3,11E-03	1,12E-01	MIR548N	ExonBnd;Body	OpenSea
cg24720939	3,11E-03	1,65E-01	MIR548H2	Body	OpenSea
cg03976364	3,11E-03	1,73E-01	MIR5684	TSS1500	OpenSea
cg19004110	3,11E-03	1,40E-01	MIR383	TSS200;Body	OpenSea
cg17874329	3,54E-03	1,62E-01	MIR1273H	Body	OpenSea
cg03220633	3,54E-03	-1,52E-01	MIR2277	Body;TSS1500	S_Shore
cg01397141	3,54E-03	-1,92E-01	MIR184	Body	OpenSea

cg27559408	3,54E-03	1,65E-01	MIR330	TSS1500;Body	N_Shore
cg05423529	4,02E-03	-1,13E-01	MIR137	TSS200	S_Shore
cg18471021	4,02E-03	1,11E-01	MIR7853	Body	OpenSea
cg13285968	4,02E-03	-1,02E-01	MIR1258	5'UTR;TSS200	N_Shore
cg02174748	4,02E-03	1,24E-01	MIR300	TSS200	OpenSea
cg13031987	4,02E-03	1,12E-01	MIR548AE2	TSS1500;Body	N_Shore
cg13811969	4,02E-03	-1,49E-01	MIR1268A	TSS200;Body	N_Shore
cg19261968	4,56E-03	1,19E-01	MIR548F1	Body	OpenSea
cg03949274	4,56E-03	1,46E-01	MIR5684	TSS1500	OpenSea
cg04947764	4,56E-03	-1,98E-01	MIR184	TSS200	OpenSea
cg12116566	4,56E-03	-1,90E-01	MIR1256	5'UTR;Body	Island
cg02903877	5,16E-03	-1,29E-01	MIR1268A	Body	OpenSea
cg16407471	5,16E-03	-2,12E-01	MIR129-2	TSS200	Island
cg09243104	5,16E-03	1,16E-01	MIR380;MIR1197;MIR323A;MIR758	TSS1500	OpenSea
cg04927004	5,16E-03	-1,67E-01	MIR124-3	TSS1500	Island
cg26222940	5,83E-03	1,16E-01	MIR548N	5'UTR;Body	S_Shore
cg11074814	5,83E-03	-1,41E-01	MIR129-2	TSS1500	N_Shore
cg20260028	5,83E-03	1,66E-01	MIR211	TSS1500;Body	OpenSea
cg00722320	5,83E-03	-2,10E-01	MIR184	TSS200	OpenSea
cg02065637	5,83E-03	-2,31E-01	MIR124-3	TSS1500	Island
cg15854570	6,57E-03	1,19E-01	MIR548F1	TSS1500;Body	OpenSea
cg04355791	6,57E-03	-1,31E-01	MIR129-2	TSS1500	N_Shore
cg17297071	6,57E-03	-1,52E-01	MIR155HG	TSS200	Island
cg14315558	6,57E-03	-2,14E-01	MIR155HG	Body	Island
cg02176951	7,39E-03	1,34E-01	MIR548F1	Body	OpenSea
cg02758410	7,39E-03	1,30E-01	MIR548F1	Body	OpenSea
cg15164782	7,39E-03	1,13E-01	MIR548G	5'UTR;Body	OpenSea
cg06112642	7,39E-03	-1,20E-01	MIR6130	TSS1500;Body	S_Shore
cg14048837	7,39E-03	1,27E-01	MIR548Q	Body;5'UTR	OpenSea
cg01257685	7,39E-03	-1,06E-01	MIR548AZ	Body	OpenSea
cg06625777	7,39E-03	1,52E-01	MIR762HG	TSS1500;Body	S_Shore
cg05945782	7,39E-03	-1,07E-01	MIR212	TSS1500	Island
cg04936377	7,39E-03	1,12E-01	MIR1268A	Body	OpenSea
cg02405538	8,30E-03	-1,90E-01	MIR7853	TSS200;Body	N_Shore
cg00105060	8,30E-03	1,20E-01	MIR548T	Body	OpenSea
cg14416371	8,30E-03	-1,74E-01	MIR129-2	TSS200	Island
cg08096702	8,30E-03	1,72E-01	MIR211	TSS1500;Body	OpenSea
cg02921294	8,30E-03	-1,32E-01	MIR1256	Body;TSS200	N_Shore
cg06928021	9,30E-03	1,33E-01	MIR921	TSS1500;Body	OpenSea
cg00978822	9,30E-03	1,01E-01	MIR449B;MIR449A	TSS1500Body	N_Shore
cg26188365	9,30E-03	1,14E-01	MIR490	TSS1500;5'UTR;Body	OpenSea
cg16341266	9,30E-03	1,35E-01	MIR548I4	TSS1500;Body	OpenSea
cg15556502	9,30E-03	-2,15E-01	MIR129-2	TSS200	Island
cg02473781	9,30E-03	-1,01E-01	MIR130B;MIR301B	TSS1500;TSS200	S_Shore
cg19277411	1,04E-02	1,29E-01	MIR5684	Body	OpenSea

cg23545774	1,04E-02	1,46E-01	MIR3163	Body	OpenSea
cg04845520	1,04E-02	1,38E-01	MIR7641-2	TSS1500	OpenSea
cg05965535	1,04E-02	-1,14E-01	MIR1268A	ExonBnd;Body	OpenSea
cg03673191	1,04E-02	1,43E-01	MIR5009	Body	OpenSea
cg17762933	1,16E-02	-1,43E-01	MIR7853	TSS200;Body	N_Shore
cg24990327	1,16E-02	1,19E-01	MIR548F3	TSS1500;Body	OpenSea
cg03469437	1,16E-02	1,24E-01	MIR383	TSS200;Body	OpenSea
cg03678049	1,16E-02	-1,63E-01	MIR1268A	Body	OpenSea
cg14944647	1,16E-02	-1,81E-01	MIR129-2	TSS200	Island
cg02650317	1,16E-02	-1,72E-01	MIR124-3	TSS1500	Island
cg15028514	1,16E-02	-1,09E-01	MIR124-3	TSS200	Island
cg06660530	1,16E-02	-1,44E-01	MIR124-3	Body	Island
cg04282607	1,16E-02	-1,20E-01	MIR301B;MIR130B	Body;TSS1500	S_Shore
cg13578899	1,30E-02	1,12E-01	MIR548F1	Body	OpenSea
cg02559101	1,30E-02	1,05E-01	MIR1273E	Body	OpenSea
cg23936106	1,30E-02	1,12E-01	MIR217HG	Body	OpenSea
cg09836205	1,30E-02	1,63E-01	MIR2113	TSS200	OpenSea
cg02373104	1,30E-02	1,35E-01	MIR548F3	Body	OpenSea
cg08260861	1,30E-02	1,39E-01	MIR548T	Body	OpenSea
cg18309927	1,30E-02	-1,43E-01	MIR124-2HG	TSS200;Body	N_Shore
cg07792478	1,30E-02	-1,19E-01	MIR124-2	TSS1500	Island
cg16189671	1,30E-02	-1,09E-01	MIR124-2	TSS200	S_Shore
cg02540736	1,30E-02	1,12E-01	MIR345	TSS1500;TSS200	S_Shore
cg11673244	1,30E-02	-1,55E-01	MIR130B;MIR301B	TSS1500;TSS200	S_Shore
cg14282114	1,44E-02	1,03E-01	MIR6130	Body	OpenSea
cg21021332	1,44E-02	1,39E-01	MIR6130	Body	OpenSea
cg22688428	1,44E-02	-1,47E-01	MIR1247	TSS1500;TSS200	Island
cg16436686	1,60E-02	1,06E-01	MIR548F1	Body	OpenSea
cg13391521	1,60E-02	1,19E-01	MIR548F3	Body	OpenSea
cg17918158	1,60E-02	1,05E-01	MIR548G	5'UTR;Body	OpenSea
cg23917868	1,60E-02	-1,19E-01	MIR145	TSS200;Body	OpenSea
cg01939477	1,60E-02	-2,05E-01	MIR129-2	TSS200	Island
cg00608334	1,60E-02	-1,06E-01	MIR4497	TSS1500;5'UTR	N_Shore
cg02523350	1,60E-02	1,44E-01	MIR548AZ	Body	OpenSea
cg23787364	1,78E-02	1,11E-01	MIR1273E	Body	OpenSea
cg01942816	1,78E-02	1,14E-01	MIR589	Body;Body	S_Shore
cg18627360	1,78E-02	-1,05E-01	MIR124-3	TSS200	Island
cg12749863	1,78E-02	-1,72E-01	MIR155HG	Body	Island
cg03221073	1,97E-02	1,09E-01	MIR548F1	Body	OpenSea
cg02023150	1,97E-02	-1,82E-01	MIR5096	TSS1500;TSS200;Body	S_Shore
cg20505332	1,97E-02	1,24E-01	MIR548AC	Body	OpenSea
cg03436967	1,97E-02	1,36E-01	MIR548I4	Body	OpenSea
cg02582781	1,97E-02	1,35E-01	MIR548H3	Body	OpenSea
cg05376374	1,97E-02	-1,61E-01	MIR129-2	TSS200	Island
cg18301891	1,97E-02	-1,01E-01	MIR1247	TSS1500	Island

cg25147193	1,97E-02	-1,54E-01	MIR181C	TSS1500	Island
cg16759218	2,18E-02	1,13E-01	MIR548G	5'UTR;Body	OpenSea
cg25199591	2,18E-02	1,12E-01	MIR2113	TSS1500	OpenSea
cg03340334	2,18E-02	-1,24E-01	MIR1268A	Body	OpenSea
cg07035704	2,18E-02	1,13E-01	MIR2052	TSS1500	OpenSea
cg24750854	2,18E-02	-1,07E-01	MIR548H4	5'UTR;1stExon;Body	Island
cg23176340	2,18E-02	-1,27E-01	MIR7-2	TSS1500	OpenSea
cg00576773	2,18E-02	-1,05E-01	MIR9-3	TSS1500	N_Shore
cg04735310	2,18E-02	-1,22E-01	MIR196A1	TSS1500	Island
cg20277905	2,18E-02	-1,60E-01	MIR124-3	TSS200	Island
cg09005612	2,41E-02	1,15E-01	MIR548F1	Body	OpenSea
cg23460578	2,41E-02	-1,02E-01	MIR10B	TSS1500	N_Shore
cg13519595	2,41E-02	1,19E-01	MIR548A2	Body	OpenSea
cg10456132	2,41E-02	1,16E-01	MIR6130	TSS1500;Body	OpenSea
cg07101541	2,41E-02	1,34E-01	MIR548W	Body	OpenSea
cg10248218	2,41E-02	1,46E-01	MIR4526	TSS1500;Body	OpenSea
cg24669741	2,65E-02	1,02E-01	MIR548F1	1stExon;5'UTR;Body	OpenSea
cg10384855	2,65E-02	1,03E-01	MIR548F3	Body	OpenSea
cg21895505	2,65E-02	-1,61E-01	MIR5096	TSS1500;TSS200;Body	S_Shore
cg13716848	2,65E-02	1,43E-01	MIR548AJ2	Body	OpenSea
cg13332552	2,65E-02	1,14E-01	MIR2054	TSS1500	OpenSea
cg07618155	2,65E-02	-1,33E-01	MIR124-2HG	TSS200;Body	N_Shore
cg02398371	2,65E-02	1,27E-01	MIR548H3	Body	OpenSea
cg13555855	2,65E-02	-1,39E-01	MIR3663HG;MIR3663	Body;TSS1500	S_Shore
cg00362690	2,65E-02	-1,16E-01	MIR548H4	Body	OpenSea
cg18824446	2,65E-02	-1,29E-01	MIR7-2	TSS1500	OpenSea
cg09066676	2,65E-02	-1,01E-01	MIR662	TSS1500	OpenSea
cg24169735	2,65E-02	1,07E-01	MIR762HG	TSS1500;Body	S_Shore
cg11392297	2,65E-02	-1,22E-01	MIR1914	Body	S_Shore
cg14451430	2,92E-02	1,12E-01	MIR548F1	Body	OpenSea
cg21341558	2,92E-02	1,06E-01	MIR548G	Body	OpenSea
cg16062053	2,92E-02	1,44E-01	MIR551B	TSS1500;Body	OpenSea
cg10698928	2,92E-02	-1,69E-01	MIR124-2	TSS1500	Island
cg16318949	2,92E-02	-1,14E-01	MIR184	TSS200;Body	OpenSea
cg02694017	2,92E-02	-1,11E-01	MIR935	Body;TSS200	Island
cg15699267	2,92E-02	-2,07E-01	MIR124-3	TSS1500	Island
cg08378275	3,21E-02	1,11E-01	MIR548AE2	Body	S_Shelf
cg19769982	3,21E-02	-1,24E-01	MIR155HG	TSS1500	N_Shore
cg21503989	3,52E-02	-1,02E-01	MIR7853	Body	OpenSea
cg19982471	3,52E-02	-1,01E-01	MIR1258	Body;5'UTR	N_Shore
cg00972731	3,52E-02	1,02E-01	MIR548H2	Body	OpenSea
cg04185799	3,52E-02	1,07E-01	MIR548AJ2	Body	OpenSea
cg23841711	3,52E-02	1,16E-01	MIR548H3	Body	OpenSea
cg23976185	3,52E-02	1,03E-01	MIR31HG	Body	OpenSea
cg11210410	3,52E-02	1,19E-01	MIR1268A	Body	OpenSea

cg21884062	3,52E-02	-1,19E-01	MIR548F5	Body;TSS200	N_Shore
cg15558727	3,52E-02	1,00E-01	MIR4500HG	Body	OpenSea
cg24229568	3,52E-02	1,12E-01	MIR365A	TSS1500	OpenSea
cg13933773	3,52E-02	-1,21E-01	MIR1256	Body;TSS200	N_Shore
cg24987741	3,86E-02	1,02E-01	MIR3714	TSS1500;Body	OpenSea
cg23480273	3,86E-02	1,03E-01	MIR548AC	Body	OpenSea
cg00351980	3,86E-02	1,23E-01	MIR548G	5'UTR;Body	OpenSea
cg15164276	3,86E-02	1,02E-01	MIR551B	TSS200;Body	OpenSea
cg10123619	3,86E-02	1,12E-01	MIR6082	TSS1500	OpenSea
cg07426000	3,86E-02	1,01E-01	MIR3163	Body	OpenSea
cg05474726	3,86E-02	-1,43E-01	MIR124-2	TSS200	S_Shore
cg25720803	3,86E-02	-1,01E-01	MIR619	Body;TSS1500	OpenSea
cg08162457	4,22E-02	1,23E-01	MIR1273H	Body	OpenSea
cg06036509	4,22E-02	1,06E-01	MIR3163	Body	OpenSea
cg07007506	4,22E-02	-1,50E-01	MIR155HG	TSS1500	N_Shore
cg01334432	4,61E-02	1,06E-01	MIR205	Body;TSS1500	OpenSea
cg26374481	4,61E-02	1,28E-01	MIR548G	Body	OpenSea
cg00210994	4,61E-02	-1,33E-01	MIR548G	5'UTR;Body	N_Shore
cg27083040	4,61E-02	-1,08E-01	MIR145	TSS200;Body	OpenSea
cg03387135	4,61E-02	-1,40E-01	MIR124-3	TSS200	Island
cg23628411	4,61E-02	-1,09E-01	MIR1256	5'UTR;Body	S_Shore

BCVY: breast cancer in very young women; BCO: breast cancer in old women.

ANNEX XIV. Related publications

1. **Sara S**. **Oltra**, Maria Peña-Chilet, Victoria Vidal-Tomas, Kirsty Flower, María Teresa Martinez, Elisa Alonso, Octavio Burgues, Ana Lluch, James M. Flanagan, and Gloria Ribas. Methylation deregulation of miRNAs promoters in Breast Cancer in very young women **(Scientific Reports, September 2018),** DOI:10.1038/s41598-018-32393-3.

2. **Sara S. Oltra**, Maria Peña-Chilet, Kirsty Flower, María Teresa Martinez, Elisa Alonso, Octavio Burgues, Ana Lluch, James M. Flanagan, and Gloria Ribas. Increased DNA methylation age acceleration in breast cancer tumours from very young women Submitted to (**Breast Cancer Research and Treatment, under revision since June 2018)**.

3. **Sara S. Oltra**, Maria Peña-Chilet, MT Martinez, Eduardo Tormo, Pilar Eroles, Joan Climent, Ana Lluch, Gloria Ribas. Suitability of HC1500 and HCC1937cell lines as models for breast cancer from young women when using global miRNA expression. (**Submitted to Cancer Epidemiology, Biomarkers & Prevention, September 2018**).

4. **Sara S. Oltra,** Marta Albanell**,** María Teresa Martinez, Ana Lluch, Gloria Ribas. *HDAC5* inhibitor as a potential treatment in breast cancer affecting very young women. (**In preparation**).

SCIENTIFIC REPORTS

OPEN

Methylation deregulation of miRNA promoters identifies miR124-2 as a survival biomarker in Breast Cancer in very young women

Received: 31 May 2018
Accepted: 2 August 2018
Published online: 26 September 2018

Sara S. Oltra[1], Maria Peña-Chilet[1], Victoria Vidal-Tomas[1], Kirsty Flower[2], María Teresa Martinez[1], Elisa Alonso[3], Octavio Burgues[3], Ana Lluch[1,4], James M. Flanagan[2] & Gloria Ribas[1,4]

MiRNAs are part of the epigenetic machinery, and are also epigenetically modified by DNA methylation. MiRNAs regulate expression of different genes, so any alteration in their methylation status may affect their expression. We aimed to identify methylation differences in miRNA encoding genes in breast cancer affecting women under 35 years old (BCVY), in order to identify potential biomarkers in these patients. In Illumina Infinium MethylationEPIC BeadChip samples (metEPICVal), we analysed the methylation of 9,961 CpG site regulators of miRNA-encoding genes present in the array. We identified 193 differentially methylated CpG sites in BCVY ($p\text{-value} < 0.05$ and methylation differences ± 0.1) that regulated 83 unique miRNA encoding genes. We validated 10 CpG sites using two independent datasets based on Infinium Human Methylation 450k array. We tested gene expression of miRNAs with differential methylation in BCVY in a meta-analysis using The Cancer Genome Atlas (TCGA), Clariom D and Affymetrix datasets. Five miRNAs (miR-9, miR-124-2, miR-184, miR-551b and miR-196a-1) were differently expressed (FDR $p\text{-value} < 0.01$). Finally, only miR-124-2 shows a significantly different gene expression by quantitative real-time PCR. MiR-124-hypomethylation presents significantly better survival rates for older patients as opposed to the worse prognosis observed in BCVY, identifying it as a potential specific survival biomarker in BCVY.

Breast cancer has the highest incidence rate of all cancers in women worldwide[1]. Although early breast cancer generally has an excellent prognosis, breast cancer in young women is associated with a high risk of systemic disease at long-term follow-up[2–7]. Young women tend to be diagnosed at a later stage with highly proliferative, high-grade tumours with the presence of lymphovascular invasion[3–5,8–12].

Hypomethylation may result in aberrant or inappropriate gene expression that contributes to neoplastic transformation, tumorigenesis or cancer progression (oncogenes)[13]. In addition, genome-wide loss of methylation contributes to chromosomal instability by destabilizing pericentromeric regions of certain chromosomes[14–16]. Gene-specific hypermethylation typically reflects hypermethylation of CpG-rich regions within gene promoter sequences that lead to gene silencing events[17].

MicroRNAs (miRNAs) are short non-coding RNAs that act as important regulators of gene expression as part of the epigenetic machinery. Epigenetic modifications were reported to play an important role in many disease onsets and progressions and can be used to explain several features of complex diseases, such as late onset and fluctuation of symptoms. In addition, miRNAs not only function as part of the epigenetic machinery but can also be epigenetically modified by DNA methylation and histone modifications themselves like any other gene. Methylation regulates the CpG islands on miRNA promoters altering their expression, the histone modifications affecting chromatin structure or changing the affinities for chromatin-associated proteins, thereby modulating gene expression, and therefore, miRNA gene expression. On the other hand, miRNAs can directly target

[1]INCLIVA Biomedical Research Institute, Hospital Clínico Universitario Valencia, University of Valencia, Valencia, Spain. [2]Department of Surgery and Cancer, Imperial College London, London, UK. [3]Pathology Department, Hospital Clínico Universitario Valencia, University of Valencia, Valencia, Spain. [4]Center for Biomedical Network Research on Cancer, Valencia, Spain. Correspondence and requests for materials should be addressed to G.R. (email: gribas@incliva.es)

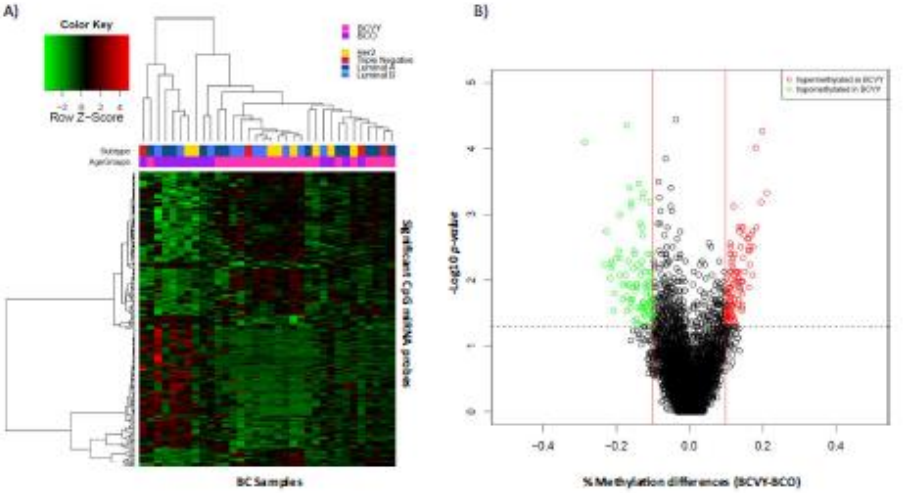


Figure 1. Differential miRNA methylation study in BCVY vs. BCO from metEPICVal samples. (**A**) Heatmap representing a supervised cluster centred on the median of the methylation levels at the 193 CpG sites that regulated miRNA genes distinctive in BCVY. Hypermethylated CpG probes in BCVY (red) and hypomethylated probes (green). Samples represented as BCVY (light pink) and BCO samples (purple). Molecular subtypes are indicated as: Her2 (yellow), triple negative (red), luminal A (dark blue) and luminal B (light blue). (**B**) Volcano-plot representation of methylation for significant CpG regulators of miRNA genes. Hypomethylated probes in BCVY are represented in green colour and hypermethylated probes in BCVY are represented in red. Red lines delimit ±0.1 methylation differences between BCVY vs. BCO and the dotted line represents a *p-value* threshold of 0.05.

epigenetic factors, such as DNA methyltransferases or histone deacetylases, thus regulating chromatin structure. Moreover, several studies have reported coordinated actions between miRNAs and other epigenetic mechanisms to reinforce the regulation of gene expression[18]. There is a strong connection between epigenome and miRNome, and any dysregulation of this complex system can result in various physiological and pathological conditions[19].

The aim of the current study is to analyse the methylation alterations of CpG associated with miRNA encoding genes in breast cancer tumours in very young women (≤35 years old) and older ones (>50 years old). DNA methylation was analysed using Illumina Infinium HumanMethylation EPIC array[20] (EPICarray); this array measures DNA methylation at approximately 850 000 CpG sites across the genome and replace the previous Infinium HumanMethylation450 BeadChip, which analyses 450,000 CpG sites. The EPICarray incorporates CpG sites located in enhancer regions identified by the ENCODE[21] and FANTOM5[22] projects. We hypothesized that methylation differences in miRNA-encoding genes could also represent gene expression differences. This could be a contributing factor in the poorer outcome of tumours in young women.

Results

Methylation differences in CpG probes regulating gene-encoding miRNAs. We analysed data from 6,567 CpG sites regulators of miRNA encoding genes present in the Illumina Infinium MethylationEPIC BeadChip in BC samples from Hospital Clínico Universitario of Valencia (metEPICVal). Wilcoxon rank sum test shown 193 CpG probes that were significantly differentially methylated in BCVY ($p\text{-value} < 0.05$ and methylation differences ±0.1) and that regulated 83 unique miRNA encoding genes. Among them, 90 were hypomethylated and 103 hypermethylated in BCVY (Fig. 1A,B). Significant differentially methylated CpG sites obtained are included in Supplementary Table 1. However, hierarchical clustering showed two principal sample groups: one consisting of BCVY samples and other including BCO, four BCO samples clustered with BCVY patients. Generalized linear model (GLM) showed that methylation differences distinguishing BCVY from BCO were not related to ER status ($p\text{-value} = 0.06$) or molecular subtypes ($p\text{-values} = 0.65$) and no molecular subtype clusters were identified in heatmap representation (Fig. 1A). All data has been analysed again with the addition of five samples from healthy female donors, showing highly similar results (data not shown) proving that differences found are those related to breast cancer affecting young women and not due just produced by age differences.

Genomic and functional context of significant CpG sites. In terms of CpG context, significant methylation differences for miRNA CpG probes were localized in islands ($p\text{-value} = 5.29 \times 10^{-7}$) and regions away from them called open sea ($p\text{-value} = 4.73 \times 10^{-5}$). Specifically, significant CpG probes localized in island regions were mainly hypomethylated in BCVY and those localized in open sea sites were hypermethylated in this age group (Fig. 2A). We analysed the functional localization of differentially methylated miRNA probes and the most

Figure 2. Genomic and functional context of significant CpG site regulators of genes encoding miRNAs which are differentially methylated in BCVY-metEPICVal samples. Percentage of methylation differences for statistically significant CpG sites from BCVY-BCO comparison according to location of the CpG relative to the island (**A**) and to the UCSC gene region feature category and regulatory elements (**B**). Red bars represent hypermethylation in BCVY and green bars represents hypomethylation. *Statistically significant *p-values* ($p < 0.001$). N/S: north/south; upstream or downstream to the CpG island.

Validated CpG probes	miRNA gene	TCGA		Combined study	
		% Methylation differences (BCVY - BCO)	*p-value*	% Methylation differences (BCVY - BCO)	*p-value*
cg04735310	MIR196A1	−11,13	4,13E-50	−14,77	5,57E-04
cg07234865	MIR9-1	−12,79	1,8E-45	−14,30	6,72E-05
cg08737296	MIR124-3	−13,29	7,3E-23	−15,73	3,84E-03
cg07792478	MIR124-2	−14,31	3,92E-21	−10,07	1,03E-02
cg05474726	MIR124-2	−11,78	2,66E-20	−14,27	4,21E-03
cg22333214	MIR137	−15,42	2,74E-20	−11,66	1,41E-02
cg04947764	MIR184	−10,56	5,67E-18	−17,85	4,73E-04
cg25147193	MIR181C	−10,32	5,61E-17	−13,96	8,03E-03
cg16407471	MIR129-2	−10,07	5,6E-12	−13,04	2,84E-02
cg00210994	MIR548G	−10,13	2,11E-07	−14,67	2,29E-02

Table 1. Table of CpG probes regulators of miRNA genes that were significantly differentially methylated in BCVY compared with BCO samples in the validation study using TCGA and the combined study data sets. This table includes a list of 10 significant CpG probes and the corresponding miRNA that were regulated by them.

important differences were identified in DNasa I hypersensitive sites (DHSs) ($p\text{-value} = 4 \times 10^{-3}$), which were generally hypermethylated in BCVY. Although not significant, transcription factor binding sites, promoters and gene body regions presented methylation differences of more than 4% between BCVY and BCO (Fig. 2B).

Methylation Validation in two populations. Both data validation sample sets were analysed using Infinium HumanMethylation450 BeadChip (HM450K), which includes around 50% of probes included in the EPICarray. Of 193 significant differentially methylated miRNA probes identified in the metEPICVal sample set, only 85 were shared with the HM450K array and could be analysed in the validation study, whereas the 108 remaining probes were exclusively for the EPICarray and could not be included in the validation study. We verified the methylation differences for 17 CpG sites between BCVY and BCO in TCGA data ($p\text{-value} < 0.05$ and methylation differences ± 0.1). Combined data showed 30 differentially methylated miRNA CpG probes among the 85 analysed in the validation study. Finally, we were able to validate a total of 10 methylation probes regulating miRNAs that were significant in both validation data sets, and all of them were hypomethylated in BCVY compared to older patients (Table 1).

Meta-analysis expression of miRNAs regulated by significant CpG probes. To elucidate whether differentially methylated CpG probes affected miRNA gene expression, we analysed expression of affected miRNA in a meta-analysis using three gene expression data sets (TCGA, Clariom D and Peña-Chilet *et al.*[23]). We analysed expression in miRNAs regulated by the 10 validated CpG sites. In addition, we included miRNAs regulated by those 108 CpG probes that were differently methylated in BCVY-BCO comparison and exclusive of the EPICarray, which were therefore excluded from the methylation validation study in the HM450K data sets. Finally, we had a total of 118 CpG probes that regulated 62 unique miRNAs. Gene expression for unique miRNAs was evaluated by t-test in the three data sets separately. Only 10 miRNAs were included in the three studies and could therefore be evaluated in the meta-analysis (Table 2), and 4 of them were significantly de-regulated with adjusted $p\text{-value} < 0.05$ (miR-9-1, miR-184, miR-551b and miR-196a-1).

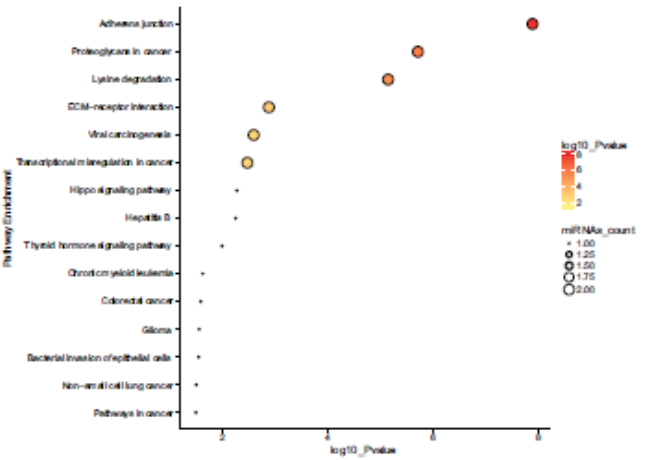


Figure 3. Pathway enrichment analysis results obtained by DIANA mirpath. Plot represents the main pathways in which miRNA regulated by significant CpG probes in BCVY were involved. Dot colour indicates *p-values* and miRNA count indicates the number of miRNAs involved in the represented pathways.

was regulated by open sea regions, we identified pathways related to cancer as previously observed for island miRNAs and, moreover, they were implicated in tumour-necrosis signalling pathways, extracellular matrix-receptor interaction and prolactin signalling pathways. Finally, for DHSs we obtained, again, most of the pathways previously observed in island regions. Pathway enrichment results by category region are reported in Supplementary Fig. 1.

qRT-PCR Validation of miRNA expression. Validation by quantitative real-time PCR (qRT-PCR) was performed on an independent set of samples with similar characteristics to those used in the metEPICVal study. We analysed the expression of the four significant miRNAs from the meta-analysis (miR-9-1, miR-184, miR-551b and miR-196a-1) and additionally, we analysed the expression of the precursor miR-124-2 that was hypomethylated in both methylation validation data sets but could not be evaluated in the expression study because it was not included in all data sets used in the meta-analysis expression. MiR-124-2 was significantly overexpressed in BCVY samples ($p\text{-value} < 0.05$) compared with older patients and this result correlated with the significant hypomethylation detected in BCVY (Fig. 4A). MiR-9-1 ($p\text{-value} = 0.07$), miR-196a-1 ($p\text{-value} = 0.36$) and miR-184 ($p\text{-value} = 0.7$) were overexpressed in BCVY, which are in agreement with the hypomethylation detected and miR-551b was repressed in BCVY ($p\text{-value} = 0.07$) that was consistent with the hypermethylation found. However, the differences did not reach statistical significance (Supplementary Fig. 2).

Clinical importance of miRNA methylation as an independent prognostic factor for relapse-free survival and overall survival in BCVY. To test the hypothesis of miRNA-methylation association to relapse-free survival (RFS) and overall survival (OS), we investigated the metEPICVal samples and methylation samples from TCGA that present follow-up data. We next performed a univariate cox regression study to determine whether hypo- or hypermethylation of significant miRNAs were correlated with underlying clinical conditions in BCVY and BCO. Whereas no significance was reached in the RFS analysis, miRNA hypomethylation was related with higher relapse risk in both BC groups (BCO and BCVY) for all miRNAs analysed with the exception of miR-184, in which hypomethylation reduces the risk of relapse in BCVY, contrary to the results for BCO (Supplementary Fig. 3). For OS studies, univariate cox analysis showed that miR-124-2 hypermethylation was significantly related with reducing survival in BCO ($p\text{-value} = 9 \times 10^{-4}$) (Fig. 4B). In order to evaluate whether other clinical factors were impacting in the survival in BCO patients apart from miR-124-2 methylation, we performed a multivariable Cox regression analysis including data from oestrogen receptor status and BC molecular subtypes. Multivariate analysis shown a significant relation with miR-124-2-hypermethylation in BCO and poor survival ($p\text{-value}\ 1.02 \times 10^{-3}$, hazard ratio HR = 3.23). However, oestrogen receptor status and molecular subtype were not contributing to survival in BCO patients. Although no significant results were obtained in univariate cox study among BCVY survival and miR-124-2 methylation, we observed a contrary effect on them, in which poorer survival was related with miR-124-2 hypomethylation (Fig. 4C). BCVY sample size is limited compared with older patients and further studies with larger dataset should be addressed to evaluate mir-124-2-hypomethylation and prognosis in this group. Additionally, OS analysis demonstrated a good prognostic factor for miR-184 hypomethylation in both BCVY and BCO groups, and a reduced survival rate for miR-196a-1-hypomethylation in BCVY: however, these results did not reach statistical significance (Supplementary Fig. 4).

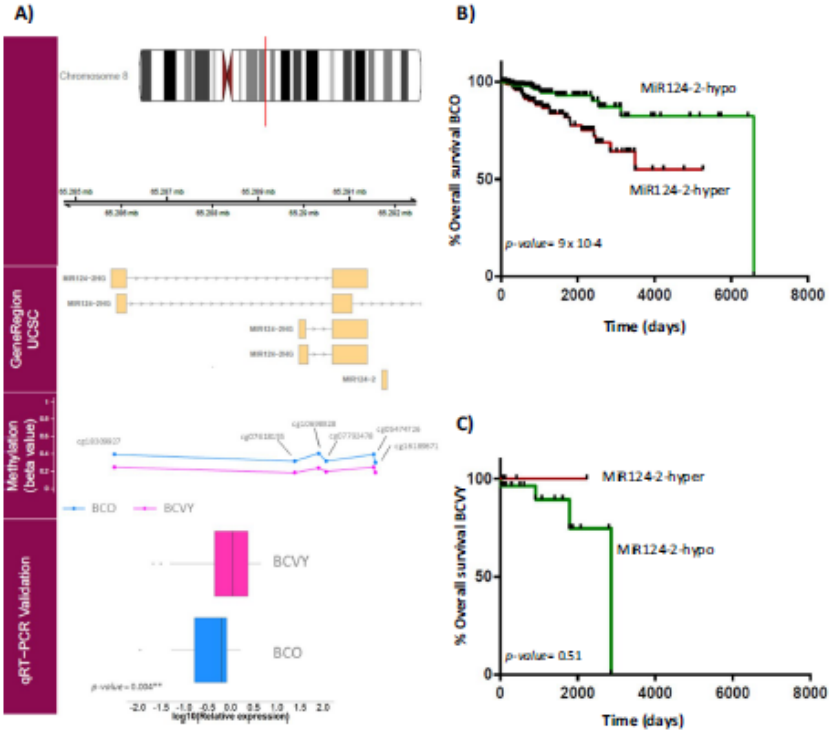


Figure 4. MiR124-2 expression validation by qRT-PCR and overall survival studies. First track represents the chromosome position of the different methylated region. Gene Region track represents genome position of the significant CpG regions obtained. Methylation track shows the β-values for the significant CpG probes regulating miR-124-2. Validation results for miR-124-2 expression using qRT-PCR are plotted in the qRT-PCR Validation track. Boxplots represent the relative mean expression for BCO and BCVY samples, *p-values* were obtained by Wilcoxon rank sum test (**A**); Representation of overall survival curves for miR-124-2 according to their methylation status in BCO (**B**) and BCVY (**C**). Green curves represent miRNA hypomethylation and red colour line hypermethylation. *P-values* were obtained by multivariate Cox analysis adjusting for oestrogen receptor status and molecular subtype.

Discussion

Several miRNAs have been involved in cancer pathogenesis[24]. Accordingly, altered miRNA expression profiles have been found in every type of human cancer including colon, brain, lung and breast cancer[25,26] suggesting miRNAs as possible biomarkers for early cancer detection. Furthermore, miRNAs not only function as part of the epigenetic machinery, but are also epigenetically modified by DNA methylation themselves. DNA methylation plays a key role in silencing numerous cancer-related genes affecting several processes, and there is considerable evidence supporting the idea that DNA methylation is actively involved in the dysregulation of miRNAs in cancer[27].

Breast cancer affecting young women has been previously associated with a more aggressive type of tumours than those diagnose in older women which could be responsible for the poor prognosis that characterizes these tumours. Previous group studies have shown differing methylation signatures for breast cancer in young women than in older patients[28]. Although the mechanism underlying miRNA dysregulation in cancer is not yet fully understood, recent studies have shown that epigenetic mechanisms play an important role in regulating miRNA expression[19].

In this study, we analysed the differences in DNA methylation of miRNA encoding genes between breast cancer samples in young and old women. Results from the methylation study in metEPICVal samples performed in the Illumina Infinium MethylationEPIC BeadChip showed 193 CpG sites that were significantly differentially methylated between groups. Most highlighted methylation differences were localized in open sea and island CpG regions. Specifically, open sea sites were hypermethylated and islands and regions near them were hypomethylated in BCVY. These results are in agreement with our previous global methylation study[28] that found hypomethylation localized in CpG island regions and hypermethylation in open sea and regions away from the

islands. Relative to that, it has been seen that several gene-encoding miRNAs are more frequently hypermethylated and consequently repressed in regions away from the islands than the CpG island itself[27], agreeing with our observations.

In terms of regulatory localization, most significantly hypomethylated probes in BCVY were located on DHSs. These regions of chromatin are sensitive to cleavage by DNase I enzyme, and chromatin has lost its condensed structure, making the DNA more accessible. This remodelled state is essential to increased transcriptional activity. DHSs therefore tended to be enriched on highly expressed genes throughout whole gene regions while not showing significant changes for low and silently expressed genes. Also, DHSs are enriched in regions away from CpG islands, suggesting preference to act within active chromatin domains that present low density CpG islands[29]. In contrast to the previously mentioned hypermethylation observed in regions away from islands, DHS regions present hypomethylation in BCVY, contributing to the more accessible DNA and consequently upregulating their expression.

In order to validate the results found in the metEPICVal sample, we used data from two previous methylation analysis (TCGA methylation and combined data) performed with HM450K array, with the limitation that we could only analyse 85 CpG probes that were shared between EPICarray and HM450K. The methylation validation study identified 10 CpG significant probes, all of them hypomethylated in young women with breast cancer. These results are in agreement with HM450K array content, given that the number of CpG islands probes is higher than open sea probes, the latter poorly presented in the HM450K array. Based on this we were able to validate some of the hypomethylation probes for BCVY localized in islands regions or near them, using the HM450K validation data sets. However, hypomethylation was concentrated mainly in open sea regions, which were underrepresented in HM450K data sets and could not be validated.

Advances in microarray and sequencing technologies have enabled comprehensive analysis of the epigenome and miRNA expression in cancer cells, which has led to the identification of miRNA which are frequent targets of aberrant DNA methylation in cancer[30]. In the present study we were able to validate significant methylation differences in a set of differentially methylated miRNAs in cancer such as: miR-181c, miR-129-2, miR-196a-1, miR-137, miR-9-1 and miR-124, of which methylation differences were observed in miR-124-2 and miR-124-3. Unfortunately, some differentially methylated miRNA could not be analysed in the expression meta-analysis because they were not included in the three data sets analysed.

We explored the expression of miRNAs regulated by differently methylated probes in BCVY taking into account their genomic/functional category. Some significant miRNAs were regulated by probes situated in more than one genomic/functional category. Differentially deregulated MiR-196a-1 and miR-212 in BCVY were in turn regulated by differently methylated regions located in islands and DHS sites. Another similar case is miR-211 and miR-184, which were regulated by probes situated in open sea and DHS sites and were deregulated in BCVY. These results suggest an important regulatory role for DNA methylation in the miRNA-encoding genes which are significantly deregulated in BCVY vs. BCO in CpG from different regions.

Next, meta-analysis of expression in miRNAs regulated by validated regulatory CpG probes and new significant probes included in the EPICarray, revealed 4 miRNAs (miR-9-1, miR-184, miR-551b and miR-196a-1) that were significantly differently expressed and methylated in BCVY. Although miR-124-2 expression could not be evaluated in the meta-analysis expression study, its hypomethylation was validated in both methylation data sets and qRT-PCR revealed significant miR-124-2 upregulation in BCVY. However, none of the four miRNAs identified (miR-9-1, miR-184, miR-551b and miR-196a-1) could be validated by qRT-PCR.

Enrichment analysis showed that the significant differentially methylated miRNAs have functions linked to cancer. Most of them were implicated in pathways related with adherent junctions, important in cancer initiation and progression. Furthermore, some miRNAs were involved in extracellular matrix organization through proteoglycans and extracellular matrix-receptor interaction.

MiRNA-methylation has been extensively investigated as a prognostic factor, and survival analysis performed with our data and TCGA revealed that hypermethylation of miR-124-2 was significantly associated with poorer OS in BCO, in contrast to BCVY, in which miR-124-2-hypomethylation presented the lowest survival rates. These results suggest that hypomethylation of miR-124-2 may be a potential prognostic risk factor specific to BCVY. Although no significant results were reached for the rest miRNAs, our study suggested that miR-184-hypomethylation could be a good prognostic factor in breast cancer patients, in which methylation loss increases survival and reduces relapse rates. Conversely, miR-9-1 and miR-196a-1 hypomethylation were associated with reduced survival and higher relapse rates.

Within the human genome, three independent loci (miR-124-1, miR-124-2 and miR-124-3) encode the identical mature miR-124, and all are associated with CpG islands, which have been described as targets of hypermethylation in colon, stomach, liver, leukaemia and cervix cancers[31]. MiR-124 exerts a tumour suppressor effect by targeting cyclin-dependent kinase 6, and epigenetic silencing of miR-124 leads to CDK6 activation and Rb phosphorylation[31,32]. Interestingly, our results show a hypomethylation of miR-124-2 gene in BCVY compared with older patients and a higher miRNA expression. Additionally, the miR-124-2 has been found to be the most abundant miRNA expressed in neuronal cells and their differentiation. Neuronal elements have been described that are related to cancer microenvironment promoting cell growth, although the mechanisms mediating neuronal influences on cancer growth and progression are likely incompletely understood[33]. Previous studies in BCVY have shown that most of the pathways deregulated in this group were related to neuronal-system processes (Oltra SS *et al.*, unpublished data). However, more work is needed to gain insight into the miRNA-124-2-hypomethylation role in breast cancer affecting young women and its potential role as a biomarker of poor survival in BCVY.

Several limitations of the present study need to be acknowledged. Methylation validation analysis has been possible for only 85 of the total 193 significant CpG probes due to the lack of methylation studies performed with the EPICarray. The TCGA dataset only provides data on methylation studies performed with HM450K or previous arrays. In the case of miRNA expression meta-analysis, the three datasets employed for the study do not

include all significantly differentially methylated miRNAs. Thus, the miRNAs that were significantly deregulated in some but not all of the three studies were not included in the meta-analysis and information about expression status could not be evaluated; despite obtaining the methylation status of significant CpG regions for BCVY, we could not translate them into differences in gene expression.

The present work is the first to analyse methylation of miRNA-encoding genes using the EPICarray in breast cancer samples from young women, and the association with miRNA expression. Our main finding is the relationship of miR-124-hypomethylation with significantly better survival rates for BCO as opposed to the worse prognosis observed in BCVY, identifying it as a potential specific survival biomarker in this group.

Material and Methods

Methylation and gene expression samples. *Illumina Infinium MethylationEPIC BeadChip samples (metEPICVal samples).* Samples included in the Illumina Infinium MethylationEPIC BeadChip experiment were archived in formalin fixed paraffin-embedded and all were stored at the Pathology Department at the Hospital Clínico Universitario, Valencia, Spain. We used 26 samples of BCVY and 15 samples from older women with breast cancer (BCO). Additionally, we have 5 samples from healthy female donors (3 from young and 2 from old women). The clinical characteristics of the patients whose samples were included in the study are shown in Supplementary Table 2. Informed consent was obtained from all subjects and the study was approved by the Hospital Clinico Ethical Committee and all research was performed in accordance with relevant guidelines.

DNA methylation quantification and normalization. Samples were extracted using a commercial kit (QIAamp DNA FFPE Tissue Kit, Qiagen, Hilden, Germany) following the manufacturer's instructions. DNA samples were quantified (Quant-iT PicoGreen dsDNA Assay, Life Technologies, CA, USA), and assessed for purity by NanoDrop (Thermo Scientific, MA, USA). Samples were checked for suitability for FFPE restoration, as indicated in the Infinium HD FFPE QC Assay (Illumina Inc.); 500 ng of FFPE DNA were bisulphite converted using the EZ-96 DNA Methylation-Gold™ kit (Zymo Research Corp., CA, USA). Bisulfite-converted DNA from FFPE samples was restored following instructions from Infinium assay. Next, DNA was hybridised to the Illumina Infinium MethylationEPIC BeadChip array following the Illumina protocol. BeadChips were washed and scanned using the Illumina HiScan SQ scanner, and the intensities were extracted from GenomeStudio (v.2011.1) and Methylation module (1.9.0) software which normalises within-sample data. Raw microarray data and processed normalized data are available from Gene Expression Omnibus (GEO) (GSE100850). Background subtraction and colour correction for the dye bias seen in Infinium II probes, as well as removal of bad quality samples were performed using *minfi* package implemented in R Bioconductor[34]. The β-values (indicating methylation level at each CpG) were calculated using *minfi* package. A total of 30,271 single nucleotide polymorphism loci were excluded from subsequent analysis. Probes that were not detected in more than 20% of the samples were excluded from the analyses. We selected samples with >98% of probes detected. Additionally, we removed cross reactive probes for the methylation EPICarray described by McCartney *et al.*[35]. After pre-processing, we analysed 793,483 CpG sites in 21 samples from BCVY and 13 from BCO.

Analysis of miRNA-associated probes. Our study focused on probes associated with gene encoding for miRNAs. According to the information available in the Illumina annotation file, EPICarray platform includes 9,961 probes that are linked to miRNAs. After the pre-processing described in previous section, we obtained a set of 6,567 CpG probes associated with 1,264 unique miRNAs which were used in the present study.

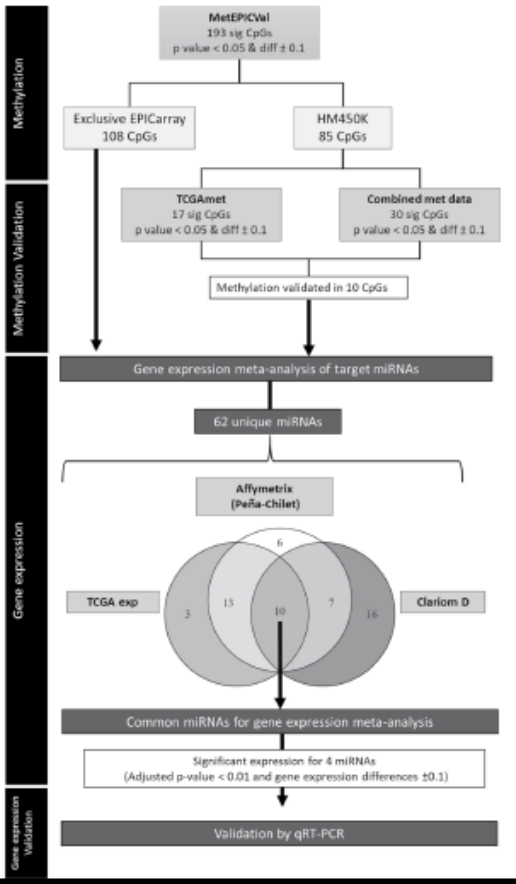
Analysis of probe distribution by gene-coding miRNAs revealed probes associated with one miRNA and probes associated with multiple miRNAs; the maximum number was for probes regulating 7 miRNAs. Information about correspondence between CpG probes and miRNAs regulated by them are included in Supplementary Table 3. Additionally, miRNAs may be regulated by unique or multiple probes (1–293). Probe count by miRNA is summarized in Supplementary Table 4.

Statistical analysis. We analysed methylation differences using the Wilcoxon Rank Sum test. CpG sites with both *p-values* < 0.05 and a minimum change of ±0.1 in β-values between BCVY and BCO were considered significant. Additionally, we performed a GLM analysis to assess whether specific methylation differences observed between BCVY and BCO were independent of molecular subtype and ER status. The study procedure diagram is included in Fig. 5.

MiRNA probes were classified into different categories according to their description in the Illumina manifest file, depending on function (promoter, gene body, TSS or UTRs), relation to the CpG context (N/S shore, N/S shelf and island), enhancer regions provided by ENCODE and FANTOM5 projects, and ENCODE regulatory elements (CpG sites in transcription factor binding sites [TFBS], open chromatin regions and DHSs). Methylation differences by categories in the significant probe set were assessed by Welch's t-test and *p-values* < 0.001 were considered to be statistically significant.

Univariate Cox analysis was used to analyse the effect of clinical variables and miRNA methylation on patients' relapse-free survival and overall survival. Multivariate Cox method was performed to adjust for clinical variables such as oestrogen receptor status and molecular subtype that could be influencing in relapse and/or survival.

Methylation Validation in two populations. We performed a combined study using methylation data from Flower *et al.*[36] analysed by Infinium HumanMethylation450 BeadChip and data from the metEPICVal study previously described. First, we analysed HM450K array samples according to the methods described in the data pre-processing section. We retained methylation levels for 405,068 probes present in both metEPICVal and Flower *et al*'s[36] HM450K study in a total of 64 samples (32 BCO samples, 32 from BCVY) that were used in the validation study.



	Methylation			miRNA expression			
	MetEPIC Val	TCGA met	Combined met data	TCGA exp	Affymetrix	Clariom D	qRT PCR samples
BCVY	N=21	N=39	N=32	N=39	N=21	N=27	N=27
BCO	N=13	N=50	N=32	N=50	N=12	N=13	N=13
Total Probes	6,567 miRNA CpGs	2,771 miRNA CpGs	2,931 miRNA CpGs	1,046 miRNA genes	768 miRNA precursors	1,217 miRNA genes	5 miRNAs

Figure 5. Workflow diagram of the procedure. The diagram shows the sample set analysed and the significant probes obtained in each step. The table includes the sample and probe size for each sample set. *diff: methylation differences between BCVY minus BCO; sig: significant.

Additionally, we used methylation data from independent breast tissue samples available from TCGA, that includes data for 485,577 probes in 720 BCO and 27 BCVY samples. Gene expression study for TCGA data was done with 50 permutations, and 50 samples were randomly selected and balanced by subtype.

Both studies were performed using the Illumina HM450K array, which includes only half of the probes present in the EPICarray. To validate our results from the BCVY-BCO study we selected differentially methylated miRNA probes that were included in HM450K and a Wilcoxon Rank Sum test was performed between BCVY and BCO tumour samples.

miRNA expression meta-analysis. We used our own gene expression data from 42 breast cancer patients (13 from BCO and 27 from BCVY) analysed by Clariom D array (Applied Biosystems by Life Technologies, Carlsbad, California, USA)[37]. Total RNA was isolated using RecoverAll Total Nucleic Acid Isolation Kit (Applied Biosystems by Life Technologies, Carlsbad, California, USA) following the manufacturer's protocol. RNA concentration was measured by Qubit® 3.0 (InvitrogenTM by Termo Fisher Scientific, Waltham, MA, USA) using the Qubit® RNA HS Assay Kit (Molecular Probes® by Life Technologies Carlsbad, California, USA). RNA quality was assessed by qRT-PCR; 25 ng of RNA were hybridized in the Clariom D array. Raw files (*.CEL) were normalized by Robust Multichip Average method using from the R Bioconductor *affy* package.

Expression of miRNAs regulated by differentially methylated probes were analysed using data from Clariom D samples. This study includes expression for 1,217 miRNA precursors. Additionally, in the meta-analysis we included previous miRNA expression published data from Peña-Chilet *et al.*[23], which were analysed by GeneChip® miRNA 2.0 Array (Affymetrix, Santa Clara,CA, USA) with GEO accession number GSE48088. This data set includes 21 BCVY and 12 BCO samples. Additionally, we downloaded gene expression from TCGA and we used data from 39 BCVY and 50 BCO samples, randomly selected and balanced by subtype.

We performed a meta-analysis study to combine *p-values* from three studies by Fisher method using the *metap* R package; *p-values* were adjusted by Benjamini-Hochberg FDR procedure[38]; genes with FDR < 0.01 and gene expression differences ± 0.1 were considered statistically significant.

Pathway enrichment analysis. DIANA miRPath pathway enrichment analysis was used to gain insight into global molecular networks and canonical pathways related to differentially expressed miRNAs (http://diana.imis.athena-innovation.gr/DianaTools/index.php?r=mirpath/index). The software performs an enrichment analysis of multiple miRNA target genes comparing each set of miRNA targets to all known KEGG pathways. Pathways showing a FDR *p-value* < 0.05 were considered significantly enriched between classes under comparison.

miRNA expression validation by quantitative real-time PCR. Validation of significant miRNAs was performed in a different sample set composed of 27 samples from BCVY and 13 from BCO using qRT-PCR (clinical characteristics in Supplementary Table 2) using Advance TaqMan Gene Expression Assays and TaqMan microRNA Assays (Applied Biosystems by Life Technologies, Carlsbad, California, USA) and normalising to hsa-mir-21 or RNU43 and RNU6B expression. The data were managed using the QuantiStudio Design & Analysis software (v1.4). Relative expression was calculated by using the comparative Ct method and obtaining the fold-change value ($\Delta\Delta$Ct). The Wilcoxon rank sum test was used for non-parametric samples, *p-value* < 0.05 were considered statistically significant.

References

1. Ferlay, J. *et al.* Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. *International journal of cancer* **136**, E359–386, https://doi.org/10.1002/ijc.29210 (2015).
2. Azim, H. A. *et al.* Elucidating prognosis and biology of breast cancer arising in young women using gene expression profiling. *Clinical cancer research: an official journal of the American Association for Cancer Research* **18**, 1341–1351, https://doi.org/10.1158/1078-0432.CCR-11-2599 (2012).
3. Fredholm, H. *et al.* Breast cancer in young women: poor survival despite intensive treatment. *PloS one* **4**, e7695, https://doi.org/10.1371/journal.pone.0007695 (2009).
4. Fredholm, H. *et al.* Long-term outcome in young women with breast cancer: a population-based study. *Breast cancer research and treatment* **160**, 131–143, https://doi.org/10.1007/s10549-016-3983-9 (2016).
5. Gnerlich, J. L. *et al.* Elevated breast cancer mortality in women younger than age 40 years compared with older women is attributed to poorer survival in early-stage disease. *Journal of the American College of Surgeons* **208**, 341–347, https://doi.org/10.1016/j.jamcollsurg.2008.12.001 (2009).
6. Kroman, N. *et al.* Factors influencing the effect of age on prognosis in breast cancer: population based study. *Bmj* **320**, 474–478 (2000).
7. Partridge, A. H. *et al.* Subtype-Dependent Relationship Between Young Age at Diagnosis and Breast Cancer Survival. *Journal of clinical oncology: official journal of the American Society of Clinical Oncology* **34**, 3308–3314, https://doi.org/10.1200/JCO.2015.65.8013 (2016).
8. Anders, C. K. *et al.* Young age at diagnosis correlates with worse prognosis and defines a subset of breast cancers with shared patterns of gene expression. *Journal of clinical oncology: official journal of the American Society of Clinical Oncology* **26**, 3324–3330, https://doi.org/10.1200/JCO.2007.14.2471 (2008).
9. Bharat, A., Aft, R. L., Gao, F. & Margenthaler, J. A. Patient and tumor characteristics associated with increased mortality in young women (< or = 40 years) with breast cancer. *Journal of surgical oncology* **100**, 248–251, https://doi.org/10.1002/jso.21268 (2009).
10. Colleoni, M. *et al.* Very young women (<35 years) with operable breast cancer: features of disease at presentation. Annals of oncology: official journal of the European Society for. *Medical Oncology* **13**, 273–279 (2002).
11. Sabiani, L. *et al.* Breast cancer in young women: Pathologic features and molecular phenotype. *Breast* **29**, 109–116, https://doi.org/10.1016/j.breast.2016.07.007 (2016).
12. Swain, S. M., Nunes, R., Yoshizawa, C., Rothney, M. & Sing, A. P. Quantitative Gene Expression by Recurrence Score in ER-Positive Breast Cancer, by Age. *Advances in therapy* **32**, 1222–1236, https://doi.org/10.1007/s12325-015-0268-3 (2015).
13. Feinberg, A. P. & Vogelstein, B. Hypomethylation of ras oncogenes in primary human cancers. *Biochemical and biophysical research communications* **111**, 47–54 (1983).
14. Eden, A., Gaudet, F., Waghmare, A. & Jaenisch, R. Chromosomal instability and tumors promoted by DNA hypomethylation. *Science* **300**, 455, https://doi.org/10.1126/science.1083557 (2003).
15. Gaudet, F. *et al.* Induction of tumors in mice by genomic hypomethylation. *Science* **300**, 489–492, https://doi.org/10.1126/science.1083558 (2003).
16. Narayan, A. *et al.* Hypomethylation of pericentromeric DNA in breast adenocarcinomas. *International journal of cancer* **77**, 833–838 (1998).
17. Herman, J. G. & Baylin, S. B. Gene silencing in cancer in association with promoter hypermethylation. *The New England journal of medicine* **349**, 2042–2054, https://doi.org/10.1056/NEJMra023075 (2003).
18. Bannister, A. J. & Kouzarides, T. Regulation of chromatin by histone modifications. *Cell research* **21**, 381–395, https://doi.org/10.1038/cr.2011.22 (2011).

19. Piletic, K. & Kunej, T. MicroRNA epigenetic signatures in human disease. *Archives of toxicology* **90**, 2405–2419, https://doi.org/10.1007/s00204-016-1815-7 (2016).
20. http://support.illumina.com. *Infinium MethylationEPIC BeadChip Kit Support*.
21. Consortium, E. P. An integrated encyclopedia of DNA elements in the human genome. *Nature* **489**, 57–74, https://doi.org/10.1038/nature11247 (2012).
22. Lizio, M. *et al.* Gateways to the FANTOM5 promoter level mammalian expression atlas. *Genome biology* **16**, 22, https://doi.org/10.1186/s13059-014-0560-6 (2015).
23. Pena-Chilet, M. *et al.* MicroRNA profile in very young women with breast cancer. *BMC cancer* **14**, 529, https://doi.org/10.1186/1471-2407-14-529 (2014).
24. Calin, G. A. *et al.* Frequent deletions and down-regulation of micro- RNA genes miR15 and miR16 at 13q14 in chronic lymphocytic leukemia. *Proceedings of the National Academy of Sciences of the United States of America* **99**, 15524–15529, https://doi.org/10.1073/pnas.242606799 (2002).
25. Baffa, R. *et al.* MicroRNA expression profiling of human metastatic cancers identifies cancer gene targets. *The Journal of pathology* **219**, 214–221, https://doi.org/10.1002/path.2586 (2009).
26. Di Leva, G., Briskin, D. & Croce, C. M. MicroRNA in cancer: new hopes for antineoplastic chemotherapy. *Upsala journal of medical sciences* **117**, 202–216, https://doi.org/10.3109/03009734.2012.660551 (2012).
27. Suzuki, H., Maruyama, R., Yamamoto, E. & Kai, M. DNA methylation and microRNA dysregulation in cancer. *Molecular oncology* **6**, 567–578, https://doi.org/10.1016/j.molonc.2012.07.007 (2012).
28. Oltra, S. S. *et al.* Increased DNA methylation age acceleration in breast cancer tumours from very young women. *Breast cancer research and treatment In revision* (2018).
29. Cockerill, P. N. Structure and function of active chromatin and DNase I hypersensitive sites. *The FEBS journal* **278**, 2182–2210, https://doi.org/10.1111/j.1742-4658.2011.08128.x (2011).
30. Kaur, S., Lotsari-Salomaa, J. E., Seppanen-Kaijansinkko, R. & Peltomaki, P. MicroRNA Methylation in Colorectal Cancer. *Advances in experimental medicine and biology* **937**, 109–122, https://doi.org/10.1007/978-3-319-42059-2_6 (2016).
31. Lujambio, A. *et al.* Genetic unmasking of an epigenetically silenced microRNA in human cancer cells. *Cancer research* **67**, 1424–1429, https://doi.org/10.1158/0008-5472.CAN-06-4218 (2007).
32. Agirre, X. *et al.* Epigenetic silencing of the tumor suppressor microRNA Hsa-miR-124a regulates CDK6 expression and confers a poor prognosis in acute lymphoblastic leukemia. *Cancer research* **69**, 4443–4453, https://doi.org/10.1158/0008-5472.CAN-08-4025 (2009).
33. Venkatesh, H. & Monje, M. Neuronal Activity in Ontogeny and Oncology. *Trends in Cancer* **3**, 89–112, https://doi.org/10.1016/j.trecan.2016.12.008.
34. Aryee, M. J. *et al.* Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. *Bioinformatics* **30**, 1363–1369, https://doi.org/10.1093/bioinformatics/btu049 (2014).
35. McCartney, D. L. *et al.* Identification of polymorphic and off-target probe binding sites on the Illumina Infinium MethylationEPIC BeadChip. *Genomics data* **9**, 22–24, https://doi.org/10.1016/j.gdata.2016.05.012 (2016).
36. Flower, K. J. *et al.* DNA methylation profiling to assess pathogenicity of BRCA1 unclassified variants in breast cancer. *Epigenetics* **10**, 1121–1132, https://doi.org/10.1080/15592294.2015.1111504 (2015).
37. Vidal-Tomas, V. *et al.* Global transcriptome deregulation of Breast Cancer in Very Young Women samples. *Annals of Oncology* **28**, 7, https://doi.org/10.1093/annonc/mdx361 (2017).
38. Benjamini, Y. & Y., H. Controlling the false discovery rate: a practical and powerful approach to multiple testing. *Journal of the Royal Statistical Society* **57**, 289–300 (1995).

Acknowledgements

We would like to thank all the patients and volunteers who gave their consent for inclusion in this study, the medical staff in the Medical Oncology Unit at the INCLIVA Biomedical Research Institute in Valencia (Spain) for collecting the samples and providing the material for all the analyses, and the expert personnel in the Genotyping and Epigenetics Laboratory at the Central Biomedical Research Unit (UCIM) in the University of Valencia. We would like to thank the thorough English revision by an expert colleague. This study was supported by grants from The Ministry of Economy and Competitiveness and the Carlos III Health Institute (PI13/00606) and FEDER. SSO is funded on a FPU pre-doctoral fellowship (FPU13/04976) from MECD, (Spanish Government); G.R. is funded on a Miguel Servet II contract (CPII14-00013) from the Carlos III Health Institute; M.P.C. is funded by the private patients Foundation LeCado. CIBERONC (CB16/12/00481-CB16/12/00473) is an initiative of the Carlos III Health Institute. J.M.F. and K.J.F. acknowledge funding support from Breast Cancer Now, the Cancer Research UK (CRUK) Centre, the Experimental Cancer Medicine Centre (ECMC), and the National Institute for Health Research (NIHR), Biomedical Research Centre (BRC) based at Imperial College Healthcare NHS Trust and Imperial College London.

Author Contributions

G.R. and A.L. conceived the study. G.R. and M.P. supervised the study. G.R. and J.F. coordinate the work. J.F. and K.F. contributed in the statistical analysis of Illumina methylation data. G.R. and S.O. designed all aspects of the research. S.O. collected the data and performed computational analysis. E.A. and O.B. contribute in the pathology aspects. A.L. and M.M. contributed in the clinical patient data collection. V.V.T. contributed in the gene expression study by Clariom D array. S.O. and G.R. wrote the manuscript. All authors revised, read, and approved the final manuscript.

Additional Information

Supplementary information accompanies this paper at https://doi.org/10.1038/s41598-018-32393-3.

Competing Interests: The authors declare no competing interests.

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Breast Cancer Research and Treatment

Increased DNA methylation age acceleration in breast cancer tumours from very young women

--Manuscript Draft--

Manuscript Number:	BREA-D-18-00789R3
Full Title:	Increased DNA methylation age acceleration in breast cancer tumours from very young women
Article Type:	Preclinical study
Keywords:	Breast cancer in young women; Epigenetics; Illumina Infinium MethylationEPIC BeadChip; DNA methylation age; age acceleration
Corresponding Author:	Gloria Ribas, Ph.D. INCLIVA Biomedical Research Institute Valencia, Valencia SPAIN
Corresponding Author Secondary Information:	
Corresponding Author's Institution:	INCLIVA Biomedical Research Institute
Corresponding Author's Secondary Institution:	
First Author:	Sara S Oltra, PhD student
First Author Secondary Information:	
Order of Authors:	Sara S Oltra, PhD student
	Maria Peña-Chilet
	Kirsty Flower
	Maria Teresa Martinez
	Elisa Alonso
	Octavio Burgues
	Ana Lluch
	James M Flanagan
	Gloria Ribas, Ph.D.
Order of Authors Secondary Information:	
Abstract:	Purpose: Breast cancer in very young women (≤ 35 years) presents more aggressive and complex biological features than their older counterparts. Our aim was to evaluate methylation differences between young and older patients and their DNA epigenetic age. Methods: EPIC and 450k Illumina methylation arrays were used in 67 breast cancer tumours, including 32 from young patients, for methylation study and additionally we analysed their epigenetic age. Results: We identified 2 219 CpG sites differently-methylated in young vs. old breast cancer patients (FDR< 0.05; β-value difference ± 0.1). The signature showed a general hypomethylation profile with a selective small hypermethylation profile located in open-sea regions in young against older patients and normal tissue. Strikingly, tumours from young patients presented a significant increased epigenetic age-acceleration compared with older women. The affected genes were enriched for pathways in neuronal-system pathways, cell communication, and matrix organisation. Validation in an independent sample highlighted consistent higher expression of HOXD9, and PCDH10 genes in young patients. Regions implicated in the hypermethylation profile were involved in Notch signalling pathways, the immune system or DNA repair. We further validated HDAC5 expression in tumours from young patients.

	Conclusion: We have identified a DNA methylation signature that is specific to the tumours from the young patients and have shown that epigenetic age-acceleration is increased in them.
Response to Reviewers:	Dear Editorial Office, We have removed from the text all possible ACROMYMS, making special efforts in those more commonly used, such as BCVY, BCO, BC, NVY and NO and others. Under our point of view is it extremely difficult to remove those with widely use meaning, and we have left the terms, DNA, RNA, CpG islands, PCR, TCGA (we have reduce its use though), statistical terms such as FDR and GLM, that are cited few times, as well as the names of genes and kits used in the analysis. We have also left the acronyms from the two studies which to our opinion help the reading. Those are metEPICval, our study, from IC-HM450K, the study of our collaborators in England. When possible we have made the changes in Figures and Tables, otherwise we have explained the terms used. We hope to have complied with the demands of the journal/editor team and the manuscript would now be suitable for revision. Do not hesitate to contact us again for any further suggestion. Sincerely Gloria Ribas
Funding Information:	Institute of Health Carlos III (PI13/00606) — Dr. Gloria Ribas Ministerio de Educacion Cultura y Deporte (FPU13/04976) — Miss Sara S Oltra Institute of Health Carlos III (CPII14-00013) — Dr. Gloria Ribas CIBERONC (CB16/12/00481-CB16/12/00473) — Dr Ana Lluch, Dr. Gloria Ribas

Increased DNA methylation age acceleration in breast cancer tumours from very young women

Sara S. Oltra[1], Maria Peña-Chilet[1], Kirsty Flower[2], María Teresa Martinez[1], Elisa Alonso[3], Octavio Burgues[3], Ana Lluch[1]¥, James M. Flanagan[2], and Gloria Ribas[1]¥*

1. Biomedical Research Institute INCLIVA, Hospital Clínico Universitario Valencia, University of Valencia, Spain.

2 Department of Surgery and Cancer, Imperial College London, London, UK.

3. Pathology Department, Hospital Clínico Universitario Valencia, University of Valencia, Spain.

¥ Center for Biomedical Network Research on Cancer (CIBERONC)

* Corresponding author address:

Medical Oncology and Haematology Unit,

Biomedical Research Institute – INCLIVA,

Av. Menendez Pelayo, accesorio 4

46010 Valencia, Spain.

Tel.: +34 963862894.

E-mail address: gribas@incliva.es

Abstract

Purpose: Breast cancer in very young women (≤ 35 years) presents more aggressive and complex biological features than their older counterparts. Our aim was to evaluate methylation differences between young and older patients and their DNA epigenetic age.

Methods: EPIC and 450k Illumina methylation arrays were used in 67 breast cancer tumours, including 32 from young patients, for methylation study and additionally we analysed their epigenetic age.

Results: We identified 2 219 CpG sites differently-methylated in young vs. old breast cancer patients (FDR< 0.05; β-value difference ± 0.1). The signature showed a general hypomethylation profile with a selective small hypermethylation profile located in open-sea regions in young against older patients and normal tissue. Strikingly, tumours from young patients presented a significant increased epigenetic age-acceleration compared with older women.

The affected genes were enriched for pathways in neuronal-system pathways, cell communication, and matrix organisation. Validation in an independent sample highlighted consistent higher expression of *HOXD9*, and *PCDH10* genes in young patients. Regions implicated in the hypermethylation profile were involved in Notch signalling pathways, the immune system or DNA repair. We further validated *HDAC5* expression in tumours from young patients.

Conclusion: We have identified a DNA methylation signature that is specific to the tumours from the young patients and have shown that epigenetic age-acceleration is increased in them.

Keywords: Breast cancer in young women, Epigenetics, Illumina Infinium MethylationEPIC BeadChip, DNA methylation age, age acceleration

INTRODUCTION

Breast cancer is the most common malignancy in women worldwide [1]. Approximately 6.6% is diagnosed in women aged 40 or younger, and of all cancers diagnosed in this age group, 40% is breast cancer; the average risk of developing breast cancer by age 40 is one in 173 [2]. Unfortunately, they are not included in mammography screening programmes. Breast cancer in very young women (<35 years old) are typically aggressive, in part owing to the over-representation of high-grade, triple-negative tumours, although young age is also an independent negative predictor of cancer-specific survival [3].. Different studies suggested that breast cancer in young women present a different molecular biology to older patients and so, should be treated differently [3,4]. Moreover, previous studies from our group [5] suggest that miRNA expression is different in breast cancer in young women and that this may be involved in the increased aggressiveness of tumours in this age group. However, these alterations do not fully explain carcinogenesis and subsequent progression in these young women.

Epigenetic modifications are reported to play an important role in the onset and progression of many diseases, it can be used to explain several features of complex diseases, such as the late onset and fluctuation of symptoms [6]. Most or all studies on DNA-methylation support the hypothesis that aging is associated with a relaxation of epigenetic control. This relaxation could contribute to some of the well-documented age-related phenotypes, such as functional decline and the development of age-related diseases. The functional relevance of age-related epigenetic changes remains largely unknown, possibly with the exception of cancer. The number one age-related disease is cancer and indeed one of the best predictors of cancer risk is age [7,8].

In this study we focused on the differences in genomic methylation between breast cancer tumours of young versus old patients. We hypothesised that alterations in the epigenome of tumours from young women might contribute to the high aggressiveness of breast cancer in this age group. Understanding the role of the epigenome in young tumours may lead to the development of novel epigenetic-based diagnostic strategies, taking the age of these patients into account.

MATERIALS AND METHODS

Sample datasets for DNA-methylation assays

Illumina Infinium MethylationEPIC BeadChip samples from Hospital Clínico Universitario of Valencia (metEPICVal)

All the samples were archived formalin fixed paraffin-embedded stored at the *Hospital Clínico Universitario Valencia*, Spain. The clinical characteristics of the patients are shown in Table 1. This study was approved by the Institutional Health Incliva-Hospital Clinico Ethics Committee.

Infinium HumanMethylation450 BeadChip samples

We included methylation data from a previous study [9] which used Infinium HumanMethylation450 BeadChip (IC-HM450K) (GSE72277), retaining 11 samples from young women and 22 from old women .

The Cancer Genome Atlas data

We accessed methylation and gene expression data from The Cancer Genome Atlas (TCGA). Methylation study was performed by HM450K array and includes data for 485 577 probes in 720 breast cancer samples of old women and 27 from young patients. Gene expression by RNAseq (IlluminaHiSeq) includes data for 20 530 genes in 1102 breast cancer samples (35 from young and 924 from old women).

DNA extraction, methylation measurements and data pre-processing

The samples were extracted using QIAamp DNA FFPE Tissue Kit (Qiagen, Hilden, Germany). DNA samples were quantified (Quant-iT PicoGreen dsDNA Assay, Life Technologies, CA, USA), and assessed for purity by NanoDrop (Thermo Scientific, MA, USA); 500 ng DNA were bisulphite converted using the EZ-96 DNA Methylation-Gold™ kit (Zymo Research Corp., CA, USA) and restored following manufacturer's instructions. Next, DNA was hybridised to the Illumina Infinium MethylationEPIC BeadChip array (EPIC850K) [10]. Raw and normalized microarray data will be available from Gene Expression Omnibus (GSE100850). Post-array sample quality control was implemented removing bad quality samples and β-values calculation were performed using *minfi* R package [11]. Single nucleotide polymorphism and cross reactive probes [12] were removed. After pre-processing, we analysed 793 483 CpG sites in 21 samples from young women with breast cancer,13 from older patients, and 5 normal tissues from healthy female donors, 2 of whom are from young and 3 from old women.

Statistical analysis

The methylation differences were assessed using Wilcoxon rank sum test and generalized linear model (GLM), *p-values* were adjusted by Benjamini-Hochberg procedure to generate a false discovery rate (FDR). We considered significant CpG

sites those with both adjusted FDR *p-value* ≤ 0.05 and a minimum change of ±0.1 in β-values. Flow-diagram of the analysis is shown in Supplementary Figure S1. Chi-squared test and GLM analysis were used to check differences in the proportion of molecular subtypes and ER status between breast cancer age groups.

Probes were classified depending on relation to the CpG context (N/S shore, N/S shelf and island) according to their description in the Illumina manifest file. Differences were assessed by Welch's t-test and *p-value* ≤ 0.001 were considered to be statistically significant.

Estimation of DNAmethylation-Age

DNAmethylation-Age was calculated for metEPICVal samples. To this end, we used a multivariable linear model based on the DNA-methylation levels of 353 CpGs described by Horvath [13,14]. We calculated both age-acceleration difference and age-acceleration residual [13]. Differences in age-acceleration between young and were analysed by Wilcoxon rank sum test. We used TCGA methylation data for validation study of DNAmethylation-Age.

Validation study of methylation with combined cohort and TCGA data

We performed a combined study using methylation data from IC-HM450K and metEPICVal studies. Retaining methylation levels for 405 068 probes present in both studies in a total of 72 samples (35 from old breast cancer patients, 32 from young and 5 normal tissues from healthy female donors, 2 of whom are from young and 3 from old women). TCGA methylation data includes β-values for 485 577 probes in 720 from old patients and 27 samples from young patients.

TCGA gene expression

Gene expression regulated by the CpGs sites differentially methylated in young patients was analysed using RNAseq expression data from TCGA. RNAseq expression study was done with 50 permutations, and samples were randomly selected and balanced by subtype. We used *metap* R package to combine *p-values* for each gene in each study; *p-values* were adjusted by Benjamini-Hochberg FDR procedure; genes with FDR < 0.01 and gene expression differences ±0.1 were considered statistically significant (Supplementary Figure S1).

Pathway enrichment analysis

Enrichr (http://amp.pharm.mssm.edu/Enrichr), was used to gain insight into the pathways and biological processes. This software performs an enrichment analysis of multiple genes using different gene-set libraries such as Reactome, Gene Ontology or KEGG. Pathways showing a *p-value* ≤ 0.05 were considered significantly enriched.

Gene expression validation by quantitative real time-PCR

Gene validation was performed in a different sample set composed of 27 samples from breast cancer young patients and 13 from old counterparts using Quantitative real time-PCR (qRT-PCR) (clinical characteristics in Table 1). RNA from selected samples was isolated using RecoverAll Total Nucleic Acid Isolation Kit (Life Technologies, Carlsbad, California, USA), measured by NanoDrop ND 2000 UV–vis Spectrophotometer (Thermo Fisher Scientific Inc., Wilmington DE, USA). We obtain cDNA using the High Capacity cDNA transcription Kit (Life Technologies, Carlsbad, California, USA). TaqMan Gene Expression Assays (Life Technologies, Carlsbad, California, USA) was used for qRT-PCR for selected genes. Normalisation was done with GAPDH expression. We use the Thermofisher Scientific software QuantiStudio Desing and

Analysis software (v1.4). Relative expression was calculated by using the comparative Ct method and obtaining the fold-change value (ΔΔCt). The Wilcoxon rank sum test was used for non-parametric samples, *p-value* ≤ 0.05 were considered statistically significant.

RESULTS

Global hypomethylation in breast cancer of young women

Statistical analysis of the metEPICVal samples revealed 44 032 CpG sites differently-methylated between old breast cancer tumours vs normal tissue samples and 36 628 CpG for young breast cancer tumours vs. normal tissues. After removing 18 150 different methylated CpGs that were common to both the control and experimental groups, we analysed the differences in methylation; we identified 2219 CpG sites significantly differentially methylated in breast cancer depending on patient age (Supplementary Figure S1). There was a general hypomethylation profile in young patients in which 69% of significant CpG sites were hypomethylated compared to older ones (Figure 1a). In terms of CpG context, the global methylation differences found in the 2219 CpG probes between young and old patients were significant for CpG islands and the adjacent regions (N/S Shore) and for non-CpG islands (open-sea). However, their specific profiles were different: while CpG islands and near regions (N/S shore) were hypomethylated, open-sea regions were hypermethylated in samples of young patients (Figure 2a).

Distinctive hypermethylation profile in tumours from breast cancer in young patients

We identified 502 CpG sites that were significantly differently methylated in young patients versus old ones (Supplementary Figure S1), 92% were also significantly associated with age by GLM analysis (Supplementary Table S1). Only 16% of distinctive CpG sites were hypomethylated in young patients. Hierarchical clustering does not show molecular subtype clusters (Figure 1b). Thus, although samples from young patients are globally hypomethylated, we identified a group of CpG sites with higher methylation compared to the rest. We had not an overrepresentation of any subtype among any cancer group (breast cancer of young women *p-value* = 0.098 and breast cancer of old women *p-value* = 0.472). GLM analysis showed that methylation differences that distinguish young from old patients were specific of age and not related with ER status (*p-value* = 0.21) or molecular subtypes (*p-value* = 0.8). The 502 CpG probes were localized mainly in open-sea and regions distant from islands (Figure 2b).

Accelerated epigenetic aging of breast cancer in young women

We estimated the DNAmethylation-Age for metEPICVal samples (Supplementary Table S2) and observed a significant correlation between chronological age and DNAmethylation-Age for tumours of old women (r = 0.52, *p-value* = 0.039). Although, no significant age correlation was found for tumours of young patients (r = 0.14, *p-value* = 0.52), results shown higher age-acceleration significantly different to tumours from older ones (*p-value* = 5.5×10^{-4}) (Figure 3a-b-c). We have reproduced the same analyses to TCGA methylation for young breast cancer samples (*p-value* = 6.5×10^{-3}) (Figure 3d-e-f). We also examined DNAmethylation-Age vs. chronological age in normal samples from TCGA data, identifying an extremely good correlation decreasing in women with advanced age (r = 0.88, *p-value* = 8.4×10^{-33}) (Figure 3g-h). Despite homogeneously low age-accelerated values for healthy tissues compared with a more dispersed data in cancer samples, non-significance was reached (Figure 3i).

Strikingly, we detected an increased age-acceleration in young tumors-ER+ for metEPICVal set (*p-value* = 0.001), with a similar trend in the young breast cancer samples in the TCGA cohort (*p-value* = 0.08). Nevertheless, we found an enrichment of ER+ tumours samples with increased DNAmethylation-Age acceleration and we corroborated this association with TCGA (*p-value* < 2.2 x 10^{-16}) (Figure 3j-k-l). Age-acceleration was not associated with relapse or clinical subtypes. Nonetheless, all tumours from young samples with relapse presented increased age-acceleration compared with older patients (results not shown). However, there is a limitation in the sample size of young patients in TCGA data and no conclusions can be drawn.

Validation of methylation differences in combined and TCGA data

We validated our DNA-methylation findings in two 450K data sets, combined data and TCGA, including 72 and 747 samples, respectively. Considering significant metEPICVal probes present in HM450K (1432 probes) for the global hypomethylation profile in young patients, 657 were also significant in the combined study (45.8%) (Supplementary Figure S2a). A minor fraction was also obtained in the TCGA population (116 out of 1254 CpG sites β-values available, 9.25%) (Supplementary figure S2b). In both validation cohorts, the vast majority of significant probes were hypomethylated in young tumours (99% in combined data and 95,7% in TCGA data).

Additionally, 35.06 % and 25.29 % of the total significant probes in the distinctive methylation profile of young samples were also included in HM450K array in combined and TCGA data, respectively. After performing the Wilcoxon rank sum test restricted to significant CpG sites from distinctive hypermethylation signature in young samples included in HM450K validation sets, we observed 15 significant CpG sites of 176 for the combined data and 10 significant CpG sites of 127 for TCGA.

Gene expression TCGA

From the 2219 significant CpG probes from the hypomethylation profile, 573 target genes were both significantly differentially expressed and regulated in samples from young and old women with breast cancer. Next, we filtered these significant genes according to their correlation between methylation and gene expression, retaining overexpressed genes in young women regulated by hypomethylated probes as well as repressed genes regulated by hypermethylated probes, to obtain a total of 186 genes with an expression–methylation correlation. Gene expression for genes regulating CpGs sites from young samples distinctive profile were analysed following the same procedure. We could identify a set of 44 genes differentially expressed in young samples and regulated by regions differentially methylated.

Pathway enrichment analysis

Significant genes from the global hypomethylation profile in young breast cancer samples were involved in neuronal system pathways related with transmission across synapsis, GABA receptor activation and different neurotransmitter cycles among others. Additionally, 22% of pathways were related with extracellular matrix organization including sulfate/heparin metabolism, proteoglycans metabolism or fibroblasts activation. Pathways related with cell adhesion represented 16% of the significant pathways (Figure 4a). Complete pathway enrichment is included in Supplementary Table S3. The enrichment pathway analysis for significant genes that characterize exclusively young women with breast cancer are included in Supplementary Table S4. One of the most noteworthy was related with Notch and Notch1 signalling (18%). Immune system pathways represented 11% of total in which

Major Histocompatibility Complex I and II were important. Pathways related with DNA repair and vesicular traffic and biogenesis were also highly represented (Figure 4b).

Validation step by qRT-PCR

We selected three genes from the global hypomethylation profile (*FOXI2, HOXD9* and *PCDH10*). Among them, HOXD9 (*p-value* = $6x10^{-3}$) and PCDH10 (*p-value* = $4.7x10^{-2}$) were significantly upregulated in tumours of young women (Figure 5a and 5b). FOXI2 gene probe failed in most of the samples tested and was excluded from this study.

In addition, for the distinctive hypermethylation profile in young breast cancer samples, we selected 11 genes that were involved in the most significant deregulated pathways: *APIS3*, *APIB1*, *TIGIT*, *CXCL17*, *HDAC5*, *HDAC9*, *NEIL3*, *EHF*, *PAPSS2*, *PEX*3 and *OTUD3*. Six of them were confirmed to have a differential expression across groups: *APIB1* (*p-value* = $3.7x10^{-2}$), *TIGIT* (*p-value* = $8.1x10^{-2}$), *HDAC5* (*p-value* = $1.0x10^{-2}$), *HDAC9* (*p-value* = $3.7x10^{-2}$), *NEIL3* (*p-value* < $1x10^{-4}$) and *PAPSS2* (*p-value* = $1x10^{-3}$). Only *HDAC5* expression was consistently upregulated in both the qRT-PCR validation and TCGA data (Figure 5c). The mean expression values for each gene obtained in the validation step are included in Supplementary Table S5. The putative functional implications of the genes are summarised in Table 2.

DISCUSSION

The overall pattern shows a global hypomethylation signature in young women with breast cancer. Several epidemiological case-control studies have reported global hypomethylation in peripheral blood of cancer patients, suggesting systematic effect of hypomethylation on disease predisposition [15,16]. In addition, we have identified a

distinctive signature of hypermethylation characteristic of young tumours, localized mainly in regions around CpG islands and open-sea. Previous works revealed distinctive methylation patterns among subtypes, with hypermethylation for luminal B and hypomethylation for Triple-negative tumours [17,18]. Triple-negative subtype is more frequent in young women [19,20], however these are not significantly overrepresented among our young patients and GLM analysis demonstrate that methylation differences observed are characteristic of young women with breast cancer and independent of molecular subtypes and ER status.

According to the literature, with few exceptions, mammalian aging is more commonly associated with hypomethylation, especially at repetitive DNA sequences [21]. Thus, while general and localised hypomethylation is detected during aging, hypermethylation would simultaneously occur at specific CpG sites with the latter presumably repressing the expression of particular genes [22,13,14]. Contrary, our results show stronger CpG island hypomethylation in young breast cancer samples.

The global hypomethylation identified in the younger group, leads us to infer a higher gene deregulation in addition to a greater genomic instability, that could justify the aggressiveness of these tumours [21]. Additionally, we have detected noticeable discrepancies between chronological and DNAmethylation-Age in young tumours, which results in a significant increased age-acceleration compared with older samples in both datasets analysed.

Previous studies suggest that epigenetic age-acceleration might also be an indicator for a more aggressive course of tumour disease [23,24]. This age-acceleration has been seen in a previous study in breast tissue from healthy females respect its paired blood sample [25]. However, normal breast tissue from TCGA shown a strong correlation between chronological age and DNAmethylation-Age. This correlation is better than the one obtained for cancer tissues, suggesting that acceleration observed in young women with breast cancer is disease specific and not tissue related.

Furthermore, we observed a higher age-acceleration for ER+ tumours in both datasets. These results support previous observations from Horvath and colleagues [13,14]. Although we should expect higher acceleration in older women, because their major proportion of ER+ tumours and negative acceleration values for young patients because aggregates major proportion of triple-negative subtypes [13,14], our results strongly indicate a higher age-acceleration in young patients with breast cancer. Additionally, we found that young samples from metEPICVal that were ER+ presented a significant higher epigenetic age, a trend seen also in those in TCGA dataset. Breast cancer risk is linked with endogenous hormone exposure which is reduced in the post-menopausal status. Thus, we hypothesized that the acceleration found in breast cancer tumours from young patients may be partially influenced by ER status, but another factors such as hormonal levels could be involved in the increased DNAmethylation-Age and contributing with their poor outcome.

We could reproduce global hypomethylation found in young patients, both in combined study and TCGA datasets with 99% and 95,7% significant probes respectively. We could not detect the distinctive hypermethylated profile for young samples in any of the two validation populations, due to the low proportion of open-sea probes in the HM450K array. Our study suggests that open-sea regions may play an important role in gene regulation in different cancer types and specifically in young women with breast cancer, but because these regions are underrepresented in HM450K studies, further research will be required to test this hypothesis.

Furthermore, we showed that global hypomethylation in young women with breast cancer was significantly overrepresented in genes involved in neuronal-system processes, as well as, in extracellular matrix organisation and cell communication. In agreement with these results, previous articles have highlighted the importance of neural elements in the cancer microenvironment in which promotes cell growth.

However, the mechanisms mediating neuronal influences on cancer growth and progression are likely incompletely understood [26].

Gene expression analysis using TCGA identified 186 genes whose expression correlated with the global hypomethylation signature found in young women with breast cancer. As expected, two of three genes included in our validation analysis in a different sample set, *HOXD9* and *PCDH10*, were also overexpressed in the young group. Previous works showed that silencing *HOXD9* in liver cancer inhibited epithelial–mesenchymal transition, migration, and invasion in vitro and decreased tumorigenic and metastatic capacities in vivo [27]. This phenomenon is essential for tumour cells dissemination and metastasis and fits with the major invasive ability of tumours from young women.

While Protocadherin 10 (*PCDH10*) overexpression has been related to inhibition of invasion and metastasis [28]. Our results showed hypomethylation in a *PCDH10* gene-regulation region and correlates with higher expression of this gene in young patients, as also validated by qRT-PCR.

Further, we found that hypermethylation in young breast cancer patients modulates characteristic pathways related with immune system, DNA repair, vesicular trafficking and biogenesis as well as with Notch/Notch1 signalling, and probably has a role in the aggressive pathophysiology of the disease. *HDAC5* CpG regulatory regions, were hypomethylated and was coincident with overexpression validated in young patients. *HDAC5* is a histone deacetylase that enhances Notch1 expression in glioma cell lines promoting cell proliferation [29]. Notch signalling can be either oncogenic or tumour suppressive in distinct human cancers, or may even adopt both roles in the same tumour type [30]. Interestingly, previous work found expression signatures related to Notch signalling pathways in breast cancer patients younger than 45 years old [31]. As for *HDAC5*, high expression has been correlated with poorer prognosis in patients with breast cancer or pancreatic neuroendocrine tumours and has been attributed with

oncogenic effects [32,33]. Moreover, *in vitro* studies suggest that the HDAC5 inhibitor, LMK-235, could be a novel therapeutic strategy for treating breast cancer [34].

CONCLUSION

Our observations suggest that the methylation pattern seen in breast cancer affecting young women differs to that observed in older patients. Hypomethylation signature found may be contributing to the increased DNA-methylation age-acceleration and this genome instability could explain the aggressiveness seen in those tumours. Moreover, hormonal factors affecting young women could be also responsible of the accelerated DNAmethylation-Age and may be associated with their poor prognosis. Therefore, the differences observed in young tumours in our work and others, strengthen the hypothesis that breast cancer arising in young women has a potentially unique, aggressive and complex biological features and more efforts should be done to reach a more personalized treatment.

ACKNOWLEDGMENTS

We would like to thank all the patients and volunteers who gave their consent for inclusion in this study, the medical staff in the Medical Oncology Unit at the INCLIVA Biomedical Research Institute in Valencia (Spain) for collecting the samples and providing the material for all the analyses, and the expert personnel in the Genotyping and Epigenetics Laboratory at the Central Biomedical Research Unit (UCIM) in the University of Valencia. We would like to thank the thorough English revision by an expert colleague.

FUNDING

This study was supported by grants from The Ministry of Economy and Competitiveness and the Carlos III Health Institute (PI13/00606) and Fondo Europeo DEsarrollo Regional (FEDER). SSO is funded on a Formacion Profesorado Universitario (FPU) pre-doctoral fellowship (FPU13/04976) from Ministerio de Educación Ciencia y Deporte (MECD), (Spanish Government); GR is funded on a Miguel Servet II contract (CPII14-00013) from the Carlos III Health Institute; MPC is funded by the private patients Foundation LeCado. Centro de Investigacion Biomedica en Red Cancer (CIBERONC) (CB16/12/00481-CB16/12/00473) is an initiative of the Carlos III Health Institute. JMF and KJF acknowledge funding support from Breast Cancer Now, the Cancer Research UK (CRUK) Centre, the Experimental Cancer Medicine Centre (ECMC), and the National Institute for Health Research (NIHR), Biomedical Research Centre (BRC) based at Imperial College Healthcare NHS Trust and Imperial College London.

COMPLIANCE WITH ETHICAL STANDARDS

Conflicts of interest

The authors declare they have no conflicts of interest to declare.

Ethical approval

All procedures performed in the study involving human participants were in accordance with the current laws of the country in which it was performed.

REFERENCES

1. Global Burden of Disease Cancer C, Fitzmaurice C, Allen C, Barber RM, Barregard L, Bhutta ZA, Brenner H, Dicker DJ, Chimed-Orchir O, Dandona R, Dandona L, Fleming T, Forouzanfar MH, Hancock J, Hay RJ, Hunter-Merrill R, Huynh C, Hosgood HD, Johnson CO, Jonas JB, Khubchandani J, Kumar GA, Kutz M, Lan Q, Larson HJ, Liang X, Lim SS, Lopez AD, MacIntyre MF, Marczak L, Marquez N, Mokdad AH, Pinho C, Pourmalek F, Salomon JA, Sanabria JR, Sandar L, Sartorius B, Schwartz SM, Shackelford KA, Shibuya K, Stanaway J, Steiner C, Sun J, Takahashi K, Vollset SE, Vos T, Wagner JA, Wang H, Westerman R, Zeeb H, Zoeckler L, Abd-Allah F, Ahmed MB, Alabed S, Alam NK, Aldhahri SF, Alem G, Alemayohu MA, Ali R, Al-Raddadi R, Amare A, Amoako Y, Artaman A, Asayesh H, Atnafu N, Awasthi A, Saleem HB, Barac A, Bedi N, Bensenor I, Berhane A, Bernabe E, Betsu B, Binagwaho A, Boneya D, Campos-Nonato I, Castaneda-Orjuela C, Catala-Lopez F, Chiang P, Chibueze C, Chitheer A, Choi JY, Cowie B, Damtew S, das Neves J, Dey S, Dharmaratne S, Dhillon P, Ding E, Driscoll T, Ekwueme D, Endries AY, Farvid M, Farzadfar F, Fernandes J, Fischer F, TT GH, Gebru A, Gopalani S, Hailu A, Horino M, Horita N, Husseini A, Huybrechts I, Inoue M, Islami F, Jakovljevic M, James S, Javanbakht M, Jee SH, Kasaeian A, Kedir MS, Khader YS, Khang YH, Kim D, Leigh J, Linn S, Lunevicius R, El Razek HM, Malekzadeh R, Malta DC, Marcenes W, Markos D, Melaku YA, Meles KG, Mendoza W, Mengiste DT, Meretoja TJ, Miller TR, Mohammad KA, Mohammadi A, Mohammed S, Moradi-Lakeh M, Nagel G, Nand D, Le Nguyen Q, Nolte S, Ogbo FA, Oladimeji KE, Oren E, Pa M, Park EK, Pereira DM, Plass D, Qorbani M, Radfar A, Rafay A, Rahman M, Rana SM, Soreide K, Satpathy M, Sawhney M, Sepanlou SG, Shaikh MA, She J, Shiue I, Shore HR, Shrime MG, So S, Soneji S, Stathopoulou V, Stroumpoulis K, Sufiyan MB, Sykes BL, Tabares-Seisdedos R, Tadese F, Tedla BA, Tessema GA, Thakur JS, Tran BX, Ukwaja KN, Uzochukwu BS, Vlassov VV, Weiderpass E, Wubshet Terefe M, Yebyo HG, Yimam HH, Yonemoto N, Younis MZ, Yu C, Zaidi Z, Zaki ME, Zenebe ZM, Murray CJ, Naghavi M (2016) Global, Regional, and National Cancer Incidence, Mortality, Years of Life Lost, Years Lived With Disability, and Disability-Adjusted Life-years for 32 Cancer Groups, 1990 to 2015: A Systematic Analysis for the Global Burden of Disease Study. JAMA oncology. doi:10.1001/jamaoncol.2016.5688

2. Ferlay J, Shin HR, Bray F, Forman D, Mathers C, Parkin DM (2010) Estimates of worldwide burden of cancer in 2008: GLOBOCAN 2008. International journal of cancer 127 (12):2893-2917. doi:10.1002/ijc.25516

3. Narod SA (2012) Breast cancer in young women. Nature reviews Clinical oncology 9 (8):460-470. doi:10.1038/nrclinonc.2012.102

4. Anders CK, Fan C, Parker JS, Carey LA, Blackwell KL, Klauber-DeMore N, Perou CM (2011) Breast carcinomas arising at a young age: unique biology or a surrogate for aggressive intrinsic subtypes? Journal of clinical oncology : official journal of the American Society of Clinical Oncology 29 (1):e18-20. doi:10.1200/JCO.2010.28.9199

5. Pena-Chilet M, Martinez MT, Perez-Fidalgo JA, Peiro-Chova L, Oltra SS, Tormo E, Alonso-Yuste E, Martinez-Delgado B, Eroles P, Climent J, Burgues O, Ferrer-Lozano J, Bosch A, Lluch A, Ribas G (2014) MicroRNA profile in very young women with breast cancer. BMC cancer 14:529. doi:10.1186/1471-2407-14-529

6. Piletic K, Kunej T (2016) MicroRNA epigenetic signatures in human disease. Archives of toxicology 90 (10):2405-2419. doi:10.1007/s00204-016-1815-7
7. Campisi J, Vijg J (2009) Does damage to DNA and other macromolecules play a role in aging? If so, how? The journals of gerontology Series A, Biological sciences and medical sciences 64 (2):175-178. doi:10.1093/gerona/gln065
8. Gravina S, Vijg J (2010) Epigenetic factors in aging and longevity. Pflugers Archiv : European journal of physiology 459 (2):247-258. doi:10.1007/s00424-009-0730-7
9. Flower KJ, Shenker NS, El-Bahrawy M, Goldgar DE, Parsons MT, Investigators KC, Affect Study G, Spurdle AB, Morris JR, Brown R, Flanagan JM (2015) DNA methylation profiling to assess pathogenicity of BRCA1 unclassified variants in breast cancer. Epigenetics 10 (12):1121-1132. doi:10.1080/15592294.2015.1111504
10. http://support.illumina.com Infinium MethylationEPIC BeadChip Kit Support.
11. Aryee MJ, Jaffe AE, Corrada-Bravo H, Ladd-Acosta C, Feinberg AP, Hansen KD, Irizarry RA (2014) Minfi: a flexible and comprehensive Bioconductor package for the analysis of Infinium DNA methylation microarrays. Bioinformatics 30 (10):1363-1369. doi:10.1093/bioinformatics/btu049
12. McCartney DL, Walker RM, Morris SW, McIntosh AM, Porteous DJ, Evans KL (2016) Identification of polymorphic and off-target probe binding sites on the Illumina Infinium MethylationEPIC BeadChip. Genomics data 9:22-24. doi:10.1016/j.gdata.2016.05.012
13. Horvath S (2013) DNA methylation age of human tissues and cell types. Genome biology 14 (10):R115. doi:10.1186/gb-2013-14-10-r115
14. Horvath S (2015) Erratum to: DNA methylation age of human tissues and cell types. Genome biology 16:96. doi:10.1186/s13059-015-0649-6
15. Kuchiba A, Iwasaki M, Ono H, Kasuga Y, Yokoyama S, Onuma H, Nishimura H, Kusama R, Tsugane S, Yoshida T (2014) Global methylation levels in peripheral blood leukocyte DNA by LUMA and breast cancer: a case-control study in Japanese women. British journal of cancer 110 (11):2765-2771. doi:10.1038/bjc.2014.223
16. van Veldhoven K, Polidoro S, Baglietto L, Severi G, Sacerdote C, Panico S, Mattiello A, Palli D, Masala G, Krogh V, Agnoli C, Tumino R, Frasca G, Flower K, Curry E, Orr N, Tomczyk K, Jones ME, Ashworth A, Swerdlow A, Chadeau-Hyam M, Lund E, Garcia-Closas M, Sandanger TM, Flanagan JM, Vineis P (2015) Epigenome-wide association study reveals decreased average methylation levels years before breast cancer diagnosis. Clinical epigenetics 7:67. doi:10.1186/s13148-015-0104-2
17. Bardowell SA, Parker J, Fan C, Crandell J, Perou CM, Swift-Scanlan T (2013) Differential methylation relative to breast cancer subtype and matched normal tissue reveals distinct patterns. Breast cancer research and treatment 142 (2):365-380. doi:10.1007/s10549-013-2738-0
18. Stefansson OA, Moran S, Gomez A, Sayols S, Arribas-Jorba C, Sandoval J, Hilmarsdottir H, Olafsdottir E, Tryggvadottir L, Jonasson JG, Eyfjord J, Esteller M (2015) A DNA methylation-based definition of biologically distinct breast cancer subtypes. Molecular oncology 9 (3):555-568. doi:10.1016/j.molonc.2014.10.012
19. Cancello G, Maisonneuve P, Mazza M, Montagna E, Rotmensz N, Viale G, Pruneri G, Veronesi P, Luini A, Gentilini O, Goldhirsch A, Colleoni M (2013) Pathological features and survival outcomes of very young patients with early breast cancer: how much is "very young"? Breast 22 (6):1046-1051. doi:10.1016/j.breast.2013.08.006
20. Fredholm H, Magnusson K, Lindstrom LS, Tobin NP, Lindman H, Bergh J, Holmberg L, Ponten F, Frisell J, Fredriksson I (2017) Breast cancer in young women and prognosis: How important are proliferation markers? European journal of cancer 84:278-289. doi:10.1016/j.ejca.2017.07.044
21. Bormann F, Rodriguez-Paredes M, Hagemann S, Manchanda H, Kristof B, Gutekunst J, Raddatz G, Haas R, Terstegen L, Wenck H, Kaderali L, Winnefeld M, Lyko F (2016) Reduced DNA

methylation patterning and transcriptional connectivity define human skin aging. Aging cell 15 (3):563-571. doi:10.1111/acel.12470
22. Jung M, Pfeifer GP (2015) Aging and DNA methylation. BMC biology 13:7. doi:10.1186/s12915-015-0118-4
23. Perna L, Zhang Y, Mons U, Holleczek B, Saum KU, Brenner H (2016) Epigenetic age acceleration predicts cancer, cardiovascular, and all-cause mortality in a German case cohort. Clinical epigenetics 8:64. doi:10.1186/s13148-016-0228-z
24. Zheng Y, Joyce BT, Colicino E, Liu L, Zhang W, Dai Q, Shrubsole MJ, Kibbe WA, Gao T, Zhang Z, Jafari N, Vokonas P, Schwartz J, Baccarelli AA, Hou L (2016) Blood Epigenetic Age may Predict Cancer Incidence and Mortality. EBioMedicine 5:68-73. doi:10.1016/j.ebiom.2016.02.008
25. Sehl ME, Henry JE, Storniolo AM, Ganz PA, Horvath S (2017) DNA methylation age is elevated in breast tissue of healthy women. Breast cancer research and treatment 164 (1):209-219. doi:10.1007/s10549-017-4218-4
26. Venkatesh H, Monje M Neuronal Activity in Ontogeny and Oncology. Trends in Cancer 3 (2):89-112. doi:10.1016/j.trecan.2016.12.008
27. Lv X, Li L, Lv L, Qu X, Jin S, Li K, Deng X, Cheng L, He H, Dong L (2015) HOXD9 promotes epithelial-mesenchymal transition and cancer metastasis by ZEB1 regulation in hepatocellular carcinoma. Journal of experimental & clinical cancer research : CR 34:133. doi:10.1186/s13046-015-0245-3
28. Shi D, Murty VV, Gu W (2015) PCDH10, a novel p53 transcriptional target in regulating cell migration. Cell cycle 14 (6):857-866. doi:10.1080/15384101.2015.1004935
29. Liu Q, Zheng JM, Chen JK, Yan XL, Chen HM, Nong WX, Huang HQ (2014) Histone deacetylase 5 promotes the proliferation of glioma cells by upregulation of Notch 1. Molecular medicine reports 10 (4):2045-2050. doi:10.3892/mmr.2014.2395
30. Espinoza I, Pochampally R, Xing F, Watabe K, Miele L (2013) Notch signaling: targeting cancer stem cells and epithelial-to-mesenchymal transition. OncoTargets and therapy 6:1249-1259. doi:10.2147/OTT.S36162
31. Azim HA, Jr., Nguyen B, Brohee S, Zoppoli G, Sotiriou C (2015) Genomic aberrations in young and elderly breast cancer patients. BMC medicine 13:266. doi:10.1186/s12916-015-0504-3
32. Li A, Liu Z, Li M, Zhou S, Xu Y, Xiao Y, Yang W (2017) Correction: HDAC5, a potential therapeutic target and prognostic biomarker, promotes proliferation, invasion and migration in human breast cancer. Oncotarget 8 (18):30619-30620. doi:10.18632/oncotarget.17542
33. Klieser E, Urbas R, Stattner S, Primavesi F, Jager T, Dinnewitzer A, Mayr C, Kiesslich T, Holzmann K, Di Fazio P, Neureiter D, Swierczynski S (2017) Comprehensive immunohistochemical analysis of histone deacetylases in pancreatic neuroendocrine tumors: HDAC5 as a predictor of poor clinical outcome. Human pathology. doi:10.1016/j.humpath.2017.02.009
34. Li A, Liu Z, Li M, Zhou S, Xu Y, Xiao Y, Yang W (2016) HDAC5, a potential therapeutic target and prognostic biomarker, promotes proliferation, invasion and migration in human breast cancer. Oncotarget 7 (25):37966-37978. doi:10.18632/oncotarget.9274
35. Costa BM, Smith JS, Chen Y, Chen J, Phillips HS, Aldape KD, Zardo G, Nigro J, James CD, Fridlyand J, Reis RM, Costello JF (2010) Reversing HOXA9 oncogene activation by PI3K inhibition: epigenetic mechanism and prognostic significance in human glioblastoma. Cancer research 70 (2):453-462. doi:10.1158/0008-5472.CAN-09-2189
36. Qiu C, Bu X, Jiang Z (2016) Protocadherin-10 acts as a tumor suppressor gene, and is frequently downregulated by promoter methylation in pancreatic cancer cells. Oncology reports 36 (1):383-389. doi:10.3892/or.2016.4793
37. Albino D, Longoni N, Curti L, Mello-Grand M, Pinton S, Civenni G, Thalmann G, D'Ambrosio G, Sarti M, Sessa F, Chiorino G, Catapano CV, Carbone GM (2012) ESE3/EHF controls epithelial cell differentiation and its loss leads to prostate tumors with mesenchymal and stem-like features. Cancer research 72 (11):2889-2900. doi:10.1158/0008-5472.CAN-12-0212

38. Shi J, Qu Y, Li X, Sui F, Yao D, Yang Q, Shi B, Ji M, Hou P (2016) Increased expression of EHF via gene amplification contributes to the activation of HER family signaling and associates with poor survival in gastric cancer. Cell death & disease 7 (10):e2442. doi:10.1038/cddis.2016.346
39. Li L, Yan J, Xu J, Liu CQ, Zhen ZJ, Chen HW, Ji Y, Wu ZP, Hu JY, Zheng L, Lau WY (2014) CXCL17 expression predicts poor prognosis and correlates with adverse immune infiltration in hepatocellular carcinoma. PloS one 9 (10):e110064. doi:10.1371/journal.pone.0110064
40. Setta-Kaffetzi N, Simpson MA, Navarini AA, Patel VM, Lu HC, Allen MH, Duckworth M, Bachelez H, Burden AD, Choon SE, Griffiths CE, Kirby B, Kolios A, Seyger MM, Prins C, Smahi A, Trembath RC, Fraternali F, Smith CH, Barker JN, Capon F (2014) AP1S3 mutations are associated with pustular psoriasis and impaired Toll-like receptor 3 trafficking. American journal of human genetics 94 (5):790-797. doi:10.1016/j.ajhg.2014.04.005