AUTONOMOUS UNIVERSITY OF MADRID

BIOCHEMISTRY DEPARTMENT

Deregulated microRNAs in breast cancer and their potential role as diagnostic and prognostic biomarkers

Nerea Matamala Zamarro

MADRID, 2015

Cover design by **Matamala Diseño Gráfico**

Press financed by **Human Cancer Genetics Programme (CNIO)**

UNIVERSIDAD AUTONOMA DE MADRID

BIOCHEMISTRY DEPARTMENT

FACULTY OF MEDICINE

AUTONOMOUS UNIVERSITY OF MADRID

Deregulated microRNAs in breast cancer and their potential role as diagnostic and prognostic biomarkers

Doctoral thesis of

Nerea Matamala Zamarro,

M.Sc. in Biomedical Research

Thesis directors

Dr. Javier Benítez Ortiz

Dr. Beatriz Martínez Delgado

cnio

HUMAN GENETICS GROUP

HUMAN CANCER GENETICS PROGRAMME

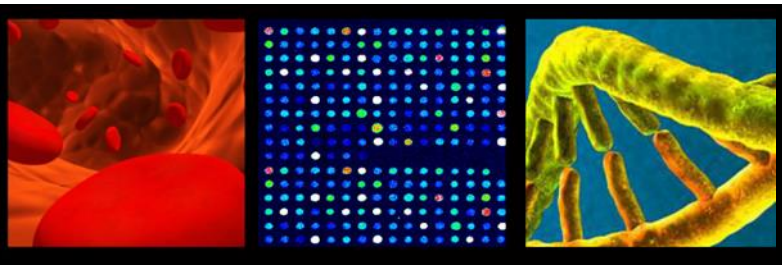
SPANISH NATIONAL CANCER RESEARCH CENTRE

This thesis, submitted for the degree of Doctor of Philosophy
at the Autonomous University of Madrid, has been elaborated in the
Human Cancer Genetics laboratory at the Spanish National Cancer Research Center (CNIO),
under the supervision of **Dr. Javier Benítez Ortiz** and **Dr. Beatriz Martínez Delgado.**

This work was supported by following grants and fellowships:

- "LIFE: desafío integral al cáncer de mama", funded by Centro para el Desarrollo Tecnológico Industrial (CDTI), Ministerio de Economía y Competitividad (MINECO), and Fondo Europeo de Desarrollo Regional (FEDER), 2012-2014; Dr. Javier Benítez Ortiz
- EACR Travel Fellowship Award, 2014; Nerea Matamala Zamarro

A mis padres, a mi hermana y a Mario



Acknowledgements

Esta tesis doctoral no hubiese sido posible sin la ayuda de un sinfín de personas. Espero saber transmitir mediante estas líneas mi inmensa gratitud a todas ellas.

En primer lugar me gustaría dar las gracias a Javier Benítez por su labor en la dirección de esta tesis y por darme la oportunidad de iniciarme como científica. Gracias por la confianza depositada, la orientación recibida, y la libertad para tomar decisiones en el proyecto. Por enseñarme a resolver los problemas, ayudarme a superar las situaciones difíciles y compartir la alegría de los buenos momentos. Por tu disponibilidad y por ser en gran parte el responsable del buen ambiente que se respira en el programa de Genética del Cáncer Humano, ha sido un auténtico placer trabajar en un lugar así.

Muchas gracias también a Beatriz Martínez por ser una excelente supervisora. Por resolver todas mis dudas, por tus sabios consejos y por estar siempre dispuesta a ayudarme. Por ser tan cercana y comprensiva, y transmitirme tu buena energía y optimismo.

Especial agradecimiento a las pacientes con cáncer de mama por su continua y excelente colaboración, sin ellas este libro estaría en blanco. De igual manera quisiera agradecer a todos los profesionales que han colaborado recogiendo muestras y datos clínicos de las pacientes: a José Ignacio Arias del Hospital Monte Naranco y a Pablo Isidro Marrón del Biobanco del Principado de Asturias (Oviedo), a Rebeca Miñambres de Sistemas Genómicos (Valencia), a Mª Ángeles Castilla del Hospital Virgen del Rocío (Sevilla) y, en especial, a Maite Vargas por acogerme en el laboratorio de Anatomía Patológica del Hospital Virgen de la Macarena, por tu cercanía y buena disposición, y por ayudarme a conseguir gran parte del material utilizado en esta tesis.

Igualmente quiero dar las gracias a los compañeros del CNIO que han colaborado directamente en este proyecto: a Edu y a Gonzalo de la Unidad de Bioinformática por resolverme muchas dudas, por vuestra paciencia y por enseñarme tanto; a Guada y a Puri de la Unidad de Genómica por vuestra ayuda con los arrays; a las chicas de la Unidad de Inmunohistoquímica por vuestra asistencia con las parafinas; a Kira y a Toya por echarme una mano con las interminables extracciones de RNA; y a Lucía Inglada por tus consejos en el análisis estadístico.

Muy especialmente quiero agradecer a los compañeros del grupo de Genética Humana toda su ayuda, cariño y buenos momentos vividos. A Ali por ser la primera en recibirme cuando llegué al CNIO con esa calurosa bienvenida: por fin una española!!! Por tener siempre una solución para todos los problemas, por tu cercanía y complicidad. A Maika por ser tan cariñosa siempre y haberme hecho reír tanto. A Samu por sorprenderme con tus preguntas y ocurrencias, y por animar a todos con tu energía. A Kira por ser siempre tan positiva y contagiarme tu optimismo. A Carlos por inundar el laboratorio con tu música y tu sentido del humor, y por tener siempre unas palabras de ánimo. Dale duro que eres el siguiente! A Javi por interesarte siempre por cómo estaba, por tus consejos, tus bromas y tu cercanía, se agradece mucho tener a alguien así cerca. A Ale por tu dulzura y simpatía, y por calmarme los nervios en aquel Progress en el que tú me presentabas. Aunque se te vaya un poco la cabra, has sido una compi de escritorio genial! A Bea por tu sinceridad y espontaneidad, y por aquel bombón que será difícil de olvidar... A Ana Osorio y María García por ser excelentes científicas y estupendas personas, por vuestra simpatía y por darme siempre buenos consejos. A Oriol y Fati por vuestro sentido del humor y vuestro ingenio. A Toya por echarme una mano cuando la he necesitado. Y a Miguel Urioste por acercarme a las pacientes y recordarme la razón por la que investigamos.

Y como no, dar las gracias a mis rusas por ser tan buena gente y ponerse tan bien los sombreros al estilo Muñoveros. A Miljana por introducirme en el maravilloso mundo de los miRNAs, a Marta por resolverme todas mis dudas informáticas, por tu dulzura y tus ánimos con la tesis (hay luz al final del túnel!), a Vero por tu alegría y por estar siempre dispuesta a ayudar, y sobre todo a Tere por arrancarme siempre una sonrisa, por tu amistad y por toda tu ayuda en el laboratorio, has sido un ejemplo a seguir!

Quiero agradecer a los estudiantes y residentes que han traído aire fresco al laboratorio, en especial a Natalia, Ana la serbia, Sergio, Sofía y Laura Pena, por vuestra simpatía y cariño. Asimismo agradezco al resto de grupos del programa de Genética del Cáncer Humano que hayan sido encantadores conmigo siempre que les he pedido ayuda. Al CEGEN: Sara, Charo, Nuria, Belén, Tais, Guille y Ana, por aguantar todas las cuantis que he puesto y por estar siempre dispuestos a dejarnos una caja de puntas. A los Endocrinos: Iñaki, María A, Aguirre, María C, Lara, Álvaro, Rocío, Lucía, Alberto, Cristina y Meme, por vuestros consejos y ayuda siempre que la he necesitado. A los Citogenéticos: Ana, Alba, Miguel, Carra, Sandra,

Mamen, Luis, Miriam, Rocío, Sara y Juan, por vuestra cercanía y aportaciones en los lab meetings. A los Epidemiólogos por vuestros consejos en los seminarios. A Irene por encargarte de trámites, llamadas y papeleos varios, y a Celia por hacerlo cuando Irene no está.

De igual modo me gustaría agradecer a todos mis compañeros del laboratorio de Genética del Hospital Virgen del Rocío, en especial a los becarios con los que tanto he compartido y con los que me he reído una jartá: Juancho, Rocío, María G, Belén y Mari Valle. Por los breaks, los delfines y las vacas, las escapadas a comer fuera, los vasos rotos en Pepe el muerto, la sangre de Cristo del Garlochi, las reuniones de castores, el notiweb, los rebujitos y las sevillanas. En especial nada hubiese sido lo mismo sin mi Mari Valle, porque fuiste un gran apoyo desde el primer día que llegué a Sevilla, por tu alegría y tu energía y porque tienes un corazón enorme.

Gracias a los que me dieron la primera oportunidad en el mundo de la investigación. Al laboratorio de Biología Molecular del IVO: María, Antonio, Zaida y José Antonio, por vuestra cercanía y generosidad (todavía guardo aquel balón…). A Daniela y Anne del laboratorio de Biología Molecular de la Universidad de York por vuestra gran simpatía y cariño, que tanto se agradecen cuando vives en el extranjero.

Quiero agradecer a mis chicas del Erasmus: Maca, Isa y Claudia, que durante un tiempo fueran mi familia, por todo lo que vivimos juntas y por seguir ahí a pesar de la distancia. En especial a mi Clau porque no sé qué hubiese hecho sin ti, por tu comprensión, sinceridad y sensibilidad.

También quiero dar las gracias a mi gente de Valencia. A los biolokos, por haber compartido muchas horas de clase, de exámenes y de fiestas. En especial a José y Cris, por tener siempre una sonrisa en la cara y un rato para un café. A las pelus: Emma, Amparo, Laura y Carmen, por las cenitas, los bailes, las charlas y los viajes. Mi agradecimiento va sobre todo para Amparo porque desde el primer día conectamos y por haber compartido tantas cosas: carreras en la playa, marchas en la montaña, paseos en bici, clases de tango, conciertos y salidas nocturnas. A Kike, por seguir siendo el mismo chico con el que iba en

autobús a clase, por defenderme de los malhechores de la noche, por tus CDs de música y las tardes de piscina.

A la gente de ese gran pueblo que es Muñoveros por ser sencillamente geniales, por vuestra cercanía, alegría y generosidad, y por demostrar que juntos podemos hacer grandes cosas.

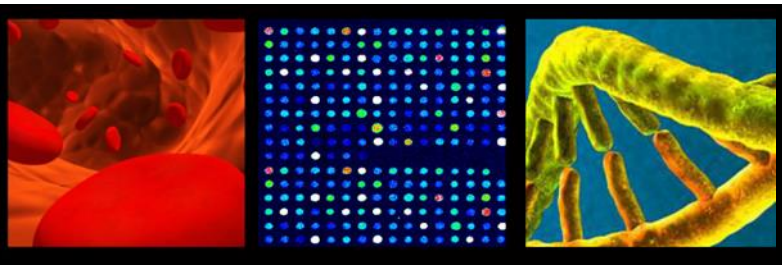
Muchas gracias a mis tíos y primos porque aunque somos muchos y estemos desperdigados siempre hacemos el esfuerzo por reunirnos, por vuestro apoyo y vuestro cariño. En especial gracias a Patri, Laura, Pablo, Cris y María, por compartir veranos, navidades y demás fiestas de guardar. Mil gracias a Pablo por ser el artista de la familia y haberme hecho una portada tan chula.

Me gustaría agradecer a mis abuelos, estén donde estén, por haberse preocupado tanto por mí y haberme cuidado. En especial a Fernanda por ser la mejor amama que se puede tener.

A mis padres por haberme apoyado siempre, por vuestro amor incondicional y esfuerzo para que no nos falte nada. Gracias por cuidarme, por confiar en mí y por estar orgullosos de todo lo que hago. A Elena por ser mi hermanita del alma, porque contigo he crecido en muchos aspectos y porque aunque estemos lejos eres un gran apoyo.

Finalmente a Mario por hacerme feliz cada día. Gracias por tu comprensión y ternura, por cuidarme con esas comidas tan ricas, por preocuparte por mí y hacerme reír, por todo lo que hemos compartido y lo que nos queda por compartir. Por hacerme sentir muy especial. Porque sencillamente te quiero.

A todos ellos, y a todos los que se me olvidan, ***Gracias!***



Summary / Resumen

SUMMARY

During the past few years, a large number of studies have explored the value of gene expression profiling in breast cancer (BC), revealing the existence of at least four major molecular subtypes with different clinical features (luminal A, luminal B, Her2 and triple negative (TN)). However, few reports have investigated the usefulness of microRNA (miRNA) expression profiling in BC management. Given the important role that miRNAs play in tumorigenesis and their great potential as novel clinical biomarkers, we aimed to investigate their deregulation in BC and their diagnostic and prognostic utility. We performed miRNA expression profiling in a large series of primary breast tumors and normal breast tissues, and identified a number of miRNAs commonly and specifically deregulated in the four subtypes, many of them not previously reported. In addition, five miRNA signatures that discriminate breast tumors and BC molecular subtypes with high sensitivity and specificity were defined. We hypothesize that these signatures might be informative for BC diagnosis. The most relevant tumoral miRNAs were analyzed in two independent series of plasma. MiR-505-5p, miR-125b-5p, miR-21-5p and miR-96-5p were confirmed to be overexpressed in the plasma of BC patients when compared with healthy women (AUC=0.61-0.72), and we found that the levels of miR-505-5p and miR-21-5p decreased in a group of treated patients. Our results demonstrate the potential utility of these miRNAs as non-invasive biomarkers for early BC detection. We have also defined a set of 17 miRNAs that are downregulated in breast tumors of node-positive TN patients with poor outcome. Moreover, we found that miR-30c-5p and miR-195-5p are associated with recurrence, and that miR-195-5p might be an independent prognostic marker in triple negative breast cancer (TNBC). Thus, analysis of miR-195-5p expression could serve to define a group of TN patients who may benefit from a more aggressive therapy. Finally, we found that two miRNAs that are specifically overexpressed in TN tumors, miR-498 and miR-187-5p, target the 3'UTR of the *BRCA1* gene. Furthermore, we demonstrated that miR-498 regulates *BRCA1* expression in BC cell lines and its inhibition leads to reduced proliferation in TNBC cells. These results shed light on the mechanisms behind the decreased expression of *BRCA1* in sporadic TNBC. In summary, our findings bring new insights in the deregulation of miRNAs in BC molecular subtypes and their potential use as BC biomarkers.

RESUMEN

En los últimos años numerosos estudios han investigado el valor de los perfiles de expresión génica en el cáncer de mama (CM), revelando la existencia de al menos cuatro subtipos moleculares con distintas características clínicas (luminal A, luminal B, Her2 y triple negativo (TN)). Sin embargo, pocas investigaciones han explorado la utilidad de los perfiles de expresión de microRNAs (miRNA) en el manejo del CM. Teniendo en cuenta el importante papel que los miRNAs juegan en la carcinogénesis y su gran potencial como nuevos biomarcadores, el objetivo de esta tesis ha sido investigar su desregulación en CM y su utilidad diagnóstica y pronóstica. Hemos obtenido los perfiles de expresión de miRNAs de una gran serie de tumores primarios de mama y tejidos mamarios normales, y hemos identificado miRNAs comúnmente y específicamente desregulados en los cuatro subtipos, muchos de ellos no reportados hasta el momento. Además hemos definido cinco firmas de miRNAs que discriminan los tumores de mama y los subtipos moleculares de CM con gran sensibilidad y especificidad. Los miRNAs tumorales más relevantes fueron analizados en dos series independientes de plasmas, confirmándose la sobreexpresión de los miR-505-5p, miR-125b-5p, miR-21-5p y miR-96-5p en el plasma de pacientes con CM en comparación con mujeres sanas (AUC=0.61-0.72). Además, los niveles de expresión de los miR-505-5p y miR-21-5p disminuyeron en un grupo de pacientes tratadas. Nuestros resultados demuestran la posible utilidad de estos miRNAs como biomarcadores no invasivos para la detección temprana del CM. También hemos definido un conjunto de 17 miRNAs que están infraexpresados en tumores de mama de pacientes TN con ganglios positivos y mal pronóstico. Además, hemos encontrado que los miR-30c-5p y miR-195-5p están asociados con recurrencia y que el miR-195-5p podría ser un marcador de pronóstico independiente en cáncer de mama triple negativo (CMTN). De esta forma, el análisis de la expresión del miR-195-5p en tumores podría servir para definir un grupo de pacientes TN que podrían beneficiarse de una terapia más agresiva. Finalmente, hemos encontrado que dos miRNAs que están específicamente sobreexpresados en tumores TN, miR-498 y miR-187-5p, tienen como diana la región 3'UTR de *BRCA1*. Además, hemos demostrado que el miR-498 regula la expresión de *BRCA1* en líneas celulares de CM, y que su inhibición da lugar a una reducción en la proliferación de las células TN. Estos resultados arrojan luz sobre los mecanismos responsables de la disminución en la expresión de *BRCA1* en CMTN esporádico. En resumen,

nuestros descubrimientos aportan nuevos conocimientos sobre la desregulación de los miRNAs en los subtipos moleculares de CM y sobre su posible uso como biomarcadores.

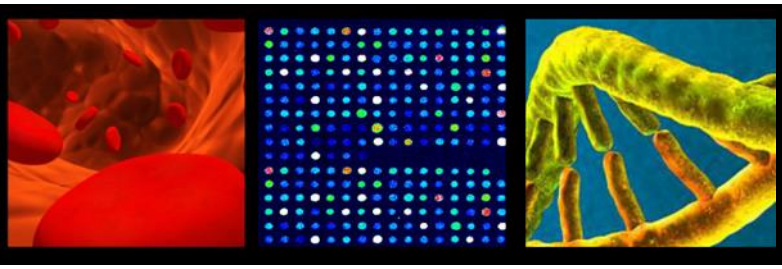
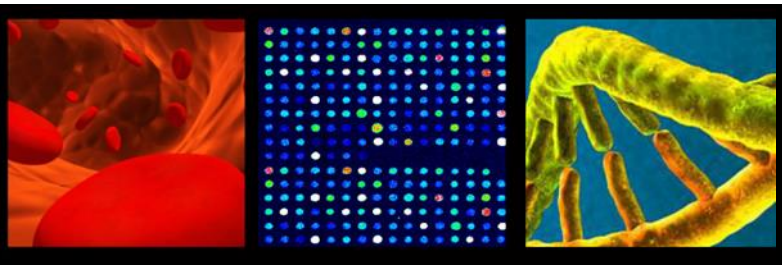


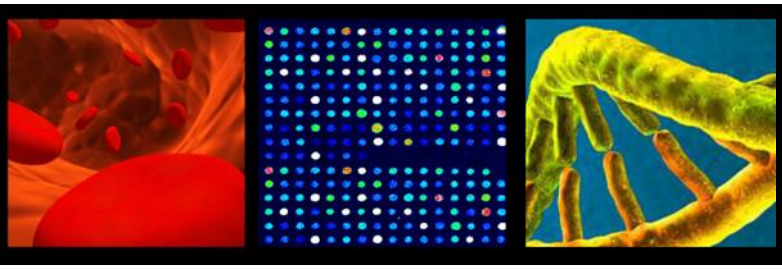
Table of contents



Abbreviations

ACTB	Actin, beta
AGO2	Argonaute RISC catalytic component 2
ASCO	American society of clinical oncology
ATM	Ataxia telangiectasia mutated
AUC	Area under the curve
BCL2	B-cell lymphoma 2
CI	Confidence interval
BRCA1	Breast cancer susceptibility gene 1
BRCA2	Breast cancer susceptibility gene 2
BRCAX	Breast cancer susceptibility gene X
BRIP1	BRCA-interacting protein C-terminal helicase 1
BSA	Bovine serum albumin
CA 15-3	Cancer antigen 15-3
CDH1	Cadherin 1, type 1, E-Cadherin (epithelial)
cDNA	Complementary DNA
CEA	Carcinoembryonic antigen
CHEK2	Checkpoint kinase 2
CIP	Calf intestinal phosphatase
CK	Cytokeratin
CLL	Chronic lymphocytic leukemia
Ct	Cycle threshold
DMEM	Dulbecco's modified Eagle's medium
DMSO	Dimethyl sulfoxide
DNA	Deoxyribonucleic acid
dNTP	Deoxynucleotide triphosphate
ECL	Enhanced chemiluminescence
ECM	Extracellular matrix
EDTA	Ethylenediaminetetraacetic acid
EGFR	Epidermal growth factor receptor
ER	Estrogen receptor
FC	Fold change
FDR	False discovery rate
FFPE	Formalin fixed paraffin embedded
GATA3	GATA binding protein 3
GTP	Guanosine-5'-triphosphate
HER2	Human epidermal growth factor receptor 2
HIF-1	Hypoxia-inducible factor-1
HMGA1	High mobility group AT-hook 1
HR	Hazard ratio
HRP	Horseradish peroxidase
HSP70	Heat shock 70kD protein
H2	Her2
ID4	Inhibitor of DNA binding 4
IHC	Immunohistochemistry
KEGG	Kyoto encyclopedia of genes and genomes
KNN	k-nearest neighbor
KRAS	Kirsten rat sarcoma viral oncogene homolog
LNA	Locked nucleic acid
lnRNA	long non-coding RNA
LA	Luminal A
LB	Luminal B

LOH	Loss of heterozygosity
MAPK	Mitogen-activated protein (MAP) kinase
miRNA	microRNA, micro ribonucleic acid
mRNA	Messenger ribonucleic acid
MRPL19	Mitochondrial ribosomal protein L19
mTOR	Mechanistic target of rapamycin (serine/threonine kinase)
NBS1	Nijmegen breakage syndrome 1
NPM1	Nucleophosmin (Nucleolar phosphoprotein B23, Numatrin)
PALB2	Partner and localizer of BRCA2
PARP	Poly (ADP-ribose) polymerase
PBS	Phosphate buffer saline
PCR	Polymerase chain reaction
PIK3CA	Phosphatidylinositol-4,5-bisphosphate 3-kinase, catalytic subunit alpha
PI3K	Phosphatidylinositol 3-kinase
PR	Progesterone receptor
PTEN	Phosphatase and tensin homolog
qRT-PCR	Quantitative real-time PCR
RAD50	RAD50 homolog (S. cerevisiae)
RAD51	Recombination protein A 51
RF	Random forest
RFS	Relapse-free survival
RISC	RNA-induced silencing complex
RNA	Ribonucleic acid
RNase	Ribonuclease
ROC	Receiver operating characteristic
ROR	Risk of relapse
RPM	Rotations per minute
RS	Recurrence score
RT	Reverse transcriptase
RT	Room temperature
SDS	Sodium dodecyl sulfate
SDS-PAGE	Sodium dodecyl sulfate polyacrylamide gel electrophoresis
SNP	Single nucleotide polymorphism
SRC	v-src avian sarcoma (Schmidt-Ruppin A-2) viral oncogene homolog
STK11	Serine/threonine kinase 11
SVM	Support vector machines
TBS	Tris buffered saline
TERT	Telomerase reverse transcriptase
TGF-β	Transforming growth factor beta
TN	Triple negative
TNM	Tumor, node, metastasis
TOP2A	Topoisomerase (DNA) II alpha 170kDa
TP53	Tumor protein p53
UTR	Untranslated region
VEGF	Vascular endothelial growth factor
WHO	World health organization
WST-1	Water-soluble tetrazolium salt
XRCC2	X-ray repair complementing defective repair in Chinese hamster cells 2
YWHAZ	Tyrosine 3-monooxygenase/tryptophan 5-monooxygenase activation protein, zeta



Introduction

1. BREAST CANCER

1.1. Mammary gland and breast cancer

The mammary gland is a highly specialized organ that is responsible for lactating. Both males and females have glandular tissue within the breasts, however, after puberty the glandular tissue begins to develop in response to estrogen release in females. The biological role of the mammary glands is to produce milk to nourish a newborn infant and to pass antibodies needed for infant's protection against infections while the immature immune system is initiating its function.

The mammary gland is formed by fifteen to twenty lobes that are arranged radially and delimited by septa of conjunctive tissue and adipose tissue in the subcutaneous layer (**Figure 1A**) (Ali and Coombes, 2002). Each lobe is formed by smaller functional units, the lobules, from which ducts converge towards the main duct of the lobe: the lactiferous duct. The lactiferous ducts are responsible for delivering the milk to the surface of the skin and out of the mother through tiny pores in the nipple (Hassiotou and Geddes, 2013).

Ductal mammary epithelium is comprised of two layers of cells: a luminal/inner layer of secretory epithelial cells that enclose the ductal lumen, and a basal/outer layer of contractile myoepithelial cells that surround the luminal layer (**Figure 1B**) (Visvader, 2009). The basal layer lies on the basement membrane and is thought to contain multipotent mammary stem cells. The epithelial ductal tree is embedded within a complex stroma, the mammary fat pad, which contains fibroblasts, adipocytes, blood vessels, nerves and various immune cells, all of which are important for normal mammary development and function (Hassiotou and Geddes, 2013).

Breast cancer is a complex disease resulting from abnormal and disorganized proliferation of cells that compose breast tissue. About 95% of malignant breast tumors are carcinomas, which originate from the epithelium of the mammary gland. Carcinomas developing from the ducts are known as ductal carcinomas (the most common ones), while those developing from lobules are known as lobular carcinomas. Neoplastic transformation typically proceeds from a benign, well-differentiated localized tumor, carcinoma *in situ*, to

invasive cancer that penetrates basal membrane infiltrating the fatty tissue of the breast, and ultimately to metastatic tumor that disseminates to other parts of the body through lymphatic and blood vessels.

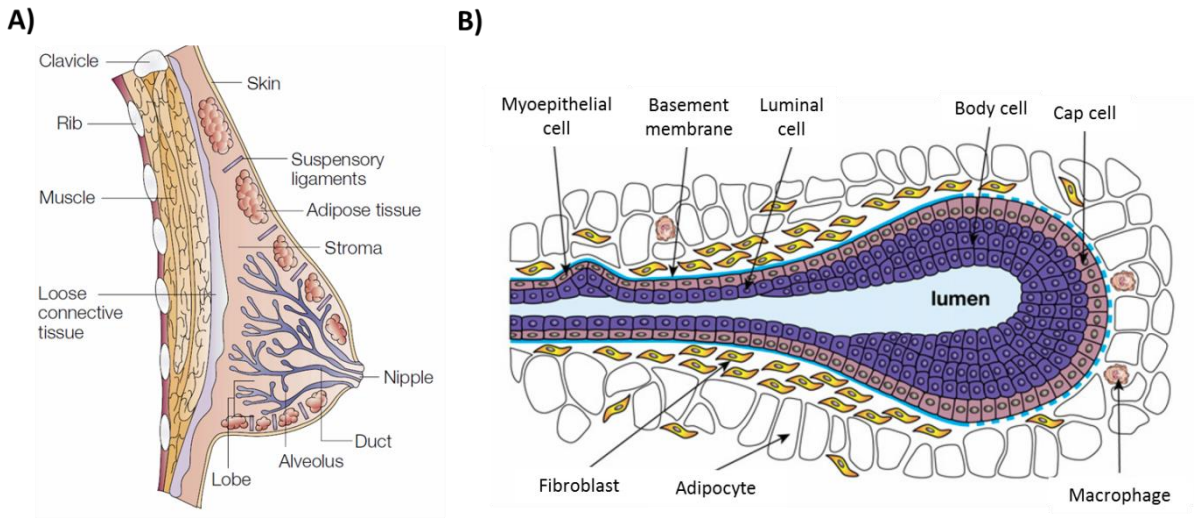


Figure 1. A) Anatomy of the human mammary gland. Each mammary gland contains 15–20 lobes, each lobe containing a series of branched ducts that drain into the nipple (adapted from Ali and Coombes, 2002). **B)** Schematic representation of a terminal duct (adapted from Visvader, 2009).

1.2. Epidemiology

Worldwide, breast cancer is the second most frequent cancer and, by far, the most common cancer among women. In 2012 it was estimated that 1.67 million of women were diagnosed with breast cancer, accounting for one quarter of the total new cancer cases in women (Ferlay et al., 2014). Breast cancer is the most common cancer in women in both developing and developed regions, although incidence rates vary considerably across the world, with the highest rates in Western Europe (96 per 100,000) and the lowest in Middle Africa (27 per 100,000) (**Figure 2**) (Servick, 2014). This variation is likely due to differences in reproductive and hormonal factors and the availability of early detection services.

Regarding mortality, breast cancer ranks as the fifth cause of death from cancer overall with 522,000 deaths estimated in 2012. It is the most deadly cancer in women in developing regions and the second cause in developed regions. The decrease in breast cancer death rates in developed countries over the last 25 years is a result of early detection through mammography, improvements in treatments and the implement of genetic testing (Ferlay et al., 2014).

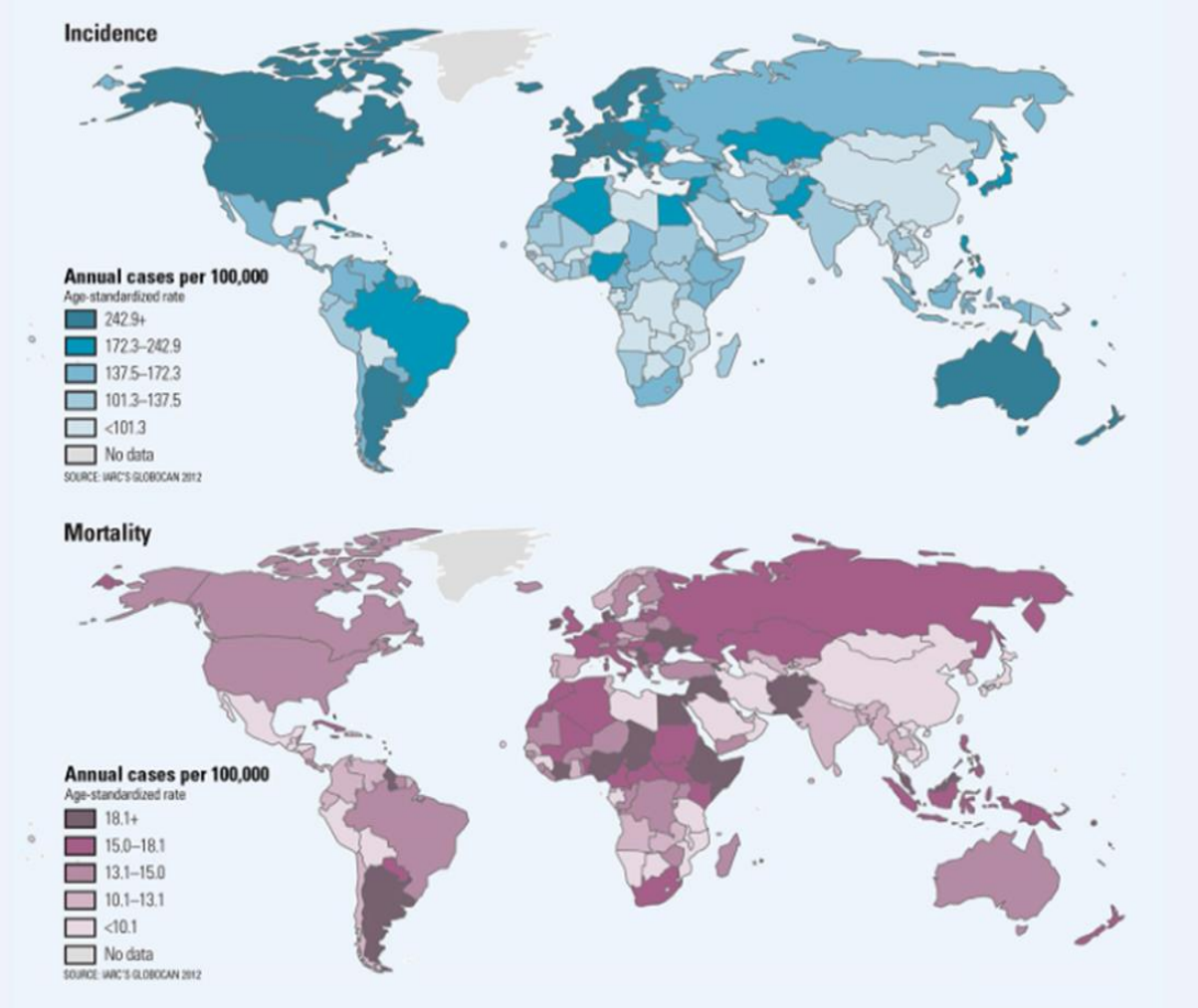


Figure 2. Estimated age-standarized incidence and mortality rates of breast cancer per 100,000 women in 2012 (adapted from Servick, 2014).

1.3. Risk factors

As breast cancer is a multifactorial disease, both genetic and non-genetic factors are involved in its development (Cuzick, 2008; McPherson et al., 2000). Being a woman is the main risk factor for breast cancer. Men can develop breast cancer, but this disease is about 100 times more common among women than men. As with many other cancers, the incidence of breast cancer increases with age. About 2 out of 3 invasive breast cancers occur in women over the age of 55. Mammography density is an important factor in terms of breast cancer risk. Women with dense breasts (higher percentage of non-fatty tissue) are up to 5 times more likely to develop breast cancer compared with women with less dense breasts. In addition, dense breast tissue can make mammograms less accurate.

Family history and genetics are linked with increased breast cancer risk. Up to 7% of breast cancer cases are thought to be hereditary, with mutations in *BRCA1* and *BRCA2* genes accounting for a substantial proportion of high risk families. Having one first-degree relative (mother, sister or daughter) with breast cancer doubles a woman's risk. The risk is further increased with a larger number of affected first-degree relatives or relatives who developed the disease before the age of 50. In addition, a woman with cancer in one breast has 3- to 4-fold increased risk of developing a new cancer in the other breast or in another part of the same breast, and some types of benign breast disease are linked with increased breast cancer risk. Race and ethnicity also play a role in breast cancer incidence: studies show that white women are more likely to develop breast cancer than African American women, although African American women develop more aggressive breast tumors.

Reproductive factors are well established risk factors for breast cancer. The modification of sex hormones levels (mainly estrogen exposure) may explain the link between these factors and breast cancer risk. Nulliparous women or women who have their first child after the age of 30 have a higher risk of breast cancer compared to women who gave birth before age 30. The relative risk increases by about 3% for each year older a woman is when she first gives birth. Subsequent births reduce relative risk by about 7% per birth. Breastfeeding is protective, especially if a woman breastfeeds for longer than 1 year. Breast cancer risk increases for each year younger at menarche and for each year older at menopause. Women using oral contraceptives and women using hormone replacement therapy for menopausal symptoms have slightly greater risk of breast cancer. Finally, lifestyle can influence the chances to develop breast cancer. Obesity is associated with a twofold increase in the risk of breast cancer in postmenopausal women. Saturated fat intake, lack of physical exercise, alcohol use and tobacco smoking are probable causes of breast cancer.

1.4. Genetic susceptibility

As mentioned before, positive family history is one of the most important risk factors for developing breast cancer. A large majority of breast cancer cases are sporadic, usually detected in older patients (>55 years), while approximately 5-7% of breast cancer cases arise in patients with strong familial aggregation of breast tumors with various affected members throughout several generations (**Figure 3A**) (Melchor and Benitez, 2013). These

families show an apparently dominant inheritance pattern and are characterized by an early age of onset, overrepresentation of ovarian cancers, bilateral breast cancers and/or male breast cancers.

Current genetic landscape of breast cancer susceptibility consists of two rare high-penetrance (>10-fold risk) susceptibility genes, several rare moderate-penetrance (2-4 fold risk) genes, and a large number of common low-penetrance (<1.5-fold risk) alleles (Ghoussaini et al., 2013). Family-based linkage analysis and positional cloning led to the identification of high-penetrance genes *BRCA1* (Miki et al., 1994) and *BRCA2* (Wooster et al., 1994), two tumor-suppressor genes involved in DNA repair that may explain around 25% of familial breast cancer risk. Lifetime risks of breast and ovarian cancer by the age 70 years are 65 and 39%, respectively, among women carrying *BRCA1* mutations, and 39 and 11% for *BRCA2*-mutation carriers. As tumor suppressor genes, mutation of both alleles is required for neoplastic transformation to occur. The mutation inherited through the germ line is often small and causes premature protein truncation, while the wild-type allele is usually lost somatically in the tumor cell (loss of heterozygosity). Cancer predisposition syndromes due to mutations in *PTEN* (Cowden syndrome), *STK11* (Peutz-Jeghers syndrome), *TP53* (Li-Fraumeni syndrome) and *CDH1* are also associated with high risk of breast cancer and account for 5% of the familial risk. Moderate-penetrance genes have been identified through their involvement in biological pathways that include BRCA1 and BRCA2, and have also been reported in about 5% of familial breast cancers. Such is the case of *CHEK2, ATM, BRIP1, PALB2, RAD50, RAD51B, RAD51C, RAD51D, XRCC2 and NBS1*. Finally, genome-wide tag SNP association studies have led to the identification of nearly one hundred common low-susceptibility loci (Michailidou et al., 2015). Unlike high-susceptibility genes, most of the variants identified in these loci are found in non-coding regions of the genome and are likely to involve regulation of genes in multiple pathways. These genes explain altogether around 14% of familial cancer risk.

Still, approximately half of familial breast cancer cases show no mutations in any of these genes, and are classified as BRCAX families (**Figure 3B**) (Melchor and Benitez, 2013). These families may either carry a mutation in a gene still not associated with breast cancer or be explained by additive low-penetrance loci (polygenic model). Modern sequencing technologies, analysis of non-coding RNA expression (such as microRNA and lncRNA) and

epigenetic studies will add more details to the description of the complex genetic architecture underlying hereditary breast cancer.

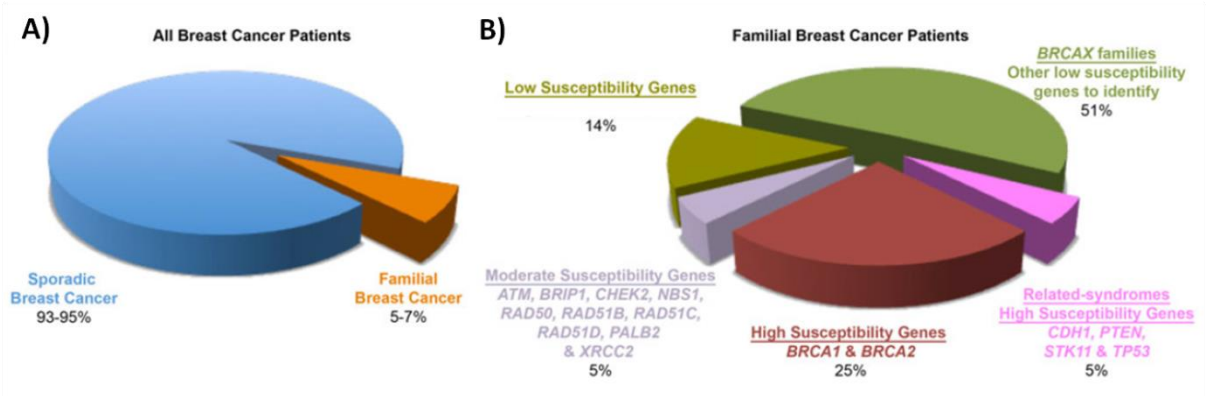


Figure 3. Genetic landscape of breast cancer susceptibility. **A)** A minor fraction of breast cancer cases are hereditary. **B)** Proportion of familial breast cancer patients explained by the identified susceptibility genes (adapted from Melchor and Benitez, 2013).

2. BREAST CANCER HETEROGENEITY

Breast cancer is a highly heterogeneous disease including a number of different entities with specific histopathological features, biological behaviors, clinical outcomes and responses to therapies. The identification of these entities is essential for cancer management since it allows for the categorization of patients into clinically relevant subgroups to aid prognostication and determine the appropriate therapy. Historically, breast cancer classification has been addressed with different perspectives, from the more traditional histopathological subgroups to the newer molecular subtypes.

2.1. Histopathological classification

Breast cancer is classified by pathologists on its histological appearance together with clinical and pathological factors. The histological classification of breast carcinoma is based on the wide range of morphological phenotypes that tumors exhibit. Many different histological types are described in the latest edition of the WHO classification of breast tumors (Lakhani et al., 2012), including invasive breast carcinomas, precursor lesions, lesions of low malignant potential, benign epithelial proliferations, fibroepithelial, myoepithelial and mesenchymal neoplasms, among others. Invasive breast carcinoma comprises 70-80%

of all cases and is a group of malignant epithelial tumors characterized by invasion of adjacent tissues and a marked tendency to metastasize to distant sites. Invasive carcinoma of no special type (previously known as invasive ductal carcinoma not otherwise specified) comprises the largest group of invasive breast cancers. It is a heterogeneous group of tumors that fail to exhibit sufficient characteristics to achieve classification as a specific histological type such as lobular, tubular, cribiform or mucinous carcinomas. A major disadvantage of this classification is that it is unable to reflect the much wider heterogeneity of breast cancer, because it groups within the same class tumors that have a very different biological and clinical profile. As a result, a variety of clinical and pathological factors are routinely used to categorize patients with breast cancer.

Tumor grade is one of the most important tumor intrinsic characteristics that can be determined by histopathological analysis of breast cancer. The grade of a breast cancer is representative of the aggressive potential of the tumor, with low grade cancers tending to be less aggressive than high grade cancers. The grading system most widely used is the Nottingham Histologic Score system (the Elston-Ellis modification of Scarff-Bloom-Richardson grading system) (Bloom and Richardson, 1957; Elston and Ellis, 1991). In this scoring system pathologists assess the degree of differentiation (tubule formation and nuclear pleomorphism) and the proliferative activity (mitotic index) of a tumor. High-grade breast cancers tend to recur and metastasize early while patients with low-grade tumors generally have a very good clinical outcome. However, many breast cancers fell into the intermediate grade category that is very heterogeneous with varying prognosis.

Incorporation of tumor stage allows a more accurate prediction of patient prognosis and has led to the development of the Nottingham Prognostic Index (Galea et al., 1992). The staging system for breast cancer is based on the TNM (Tumor, Node, Metastasis)-classification and reflects the extent of spread of the cancer when it is first diagnosed. The TNM staging takes into account the size of the tumor (T), the lymph node involvement (N) and the presence or absence of distant metastases (spread to distant organs) (M). Once the T, N and M are determined, a stage of 0, I, II, III, or IV is assigned, with stage 0 being in situ, stage I being early stage invasive cancer, and stage IV being the most advanced. In general, TNM stage is inversely correlated with the prognosis.

Immunohistochemical (IHC) biomarkers such as the expression of estrogen receptor (ER), progesterone receptor (PR) and human epidermal growth factor receptor (HER2) provide additional therapeutic predictive value. In general, ER expression is associated with a favorable prognosis and response to endocrine therapy while HER2 amplification suggests aggressive behavior and response to anti-Her2 therapy (trastuzumab) (National Institutes of Health Consensus Development, 2001). Therefore, the current approach incorporates IHC biomarkers with tumor grade, tumor stage, presence of extensive vascular invasion and patient's age. Most of these variables are combined in Adjuvant! Online (www.adjuvantonline.com), a free web-based tool that predicts breast cancer outcomes and the efficacy of adjuvant therapy in patients with breast cancer.

However, although the existing classification has been fundamental for prognostic and predictive evaluation in groups of patients, its role in evaluating risk in an individual patient with breast cancer is more limited, as tumors with apparent similarities in clinical and pathological characteristics may have different responses to therapy and clinical outcomes. This inaccuracy leads to overtreatment of some patients with unnecessarily toxic therapies and to undertreatment of others who receive false assurance of a favorable prognosis (Bergh and Holmquist, 2001). In addition, this classification provides limited insight into the complex underlying biology and the molecular pathways driving the disease in different subtypes. As a result, a molecular classification of breast cancers based on their gene expression profile has been proposed in recent years.

2.2. Molecular classification: intrinsic subtypes

Gene expression profiling using microarray technology allows simultaneous measurement of the expression of thousands of genes in a single tissue sample and represents a valuable tool to assess molecular and potential biological differences in breast cancers. Using this technology, Perou, Sorlie and colleagues (Perou et al., 2000; Sorlie et al., 2001) demonstrated the stratification of breast tumors into several major subtypes beyond the traditional hormone receptor-positive and hormone receptor-negative subgroups. The most reproducibly identified molecular subtypes among the hormone receptor-positive tumors are the luminal A and luminal B groups. The HER2 and basal-like groups are the major molecular subtypes identified among hormone receptor-negative tumors. Other

molecular subtypes such as luminal C and normal breast-like groups have also been identified in some studies, but are less well characterized than the luminal A, luminal B, HER2, and basal-like types. These breast cancer molecular subtypes not only have distinct gene expression profiles, but also have unique clinical features, prognosis and response to therapy, as summarized in **Table 1** (Schnitt, 2010).

Table 1. Clinical and pathological features of major molecular subtypes of breast cancer determined by gene expression profiling (adapted from Schnitt, 2010).

	Luminal	*HER2*	*Basal*
Gene expression pattern	High expression of hormone receptors and associated genes (luminal A > luminal B)	High expression of HER2 and other genes in amplicon Low expression of ER and associated genes	High expression of basal epithelial genes, basal cytokeratins Low expression of ER and associated genes Low expression of HER2
Clinical features	~70% of invasive breast cancers ER/PR positive Luminal B tend to be higher histological grade than luminal A Some overexpress HER2 (luminal B)	~15% of invasive breast cancers ER/PR negative More likely to be high grade and node positive	~15% of invasive breast cancers Most ER/PR/HER2 negative ('triple negative') *BRCA1* dysfunction (germline, sporadic) Particularly common in African-American women
Treatment response and outcome	Respond to endocrine therapy (but response to tamoxifen and aromatase inhibitors may be different for luminal A and luminal B) Response to chemotherapy variable (greater in luminal B than in luminal A) Prognosis better for luminal A than luminal B	Respond to trastuzumab (Herceptin) Respond to anthracycline-based chemotherapy Generally poor prognosis	No response to endocrine therapy or trastuzumab (Herceptin) Appear to be sensitive to platinum-based chemotherapy and PARP inhibitors Generally poor prognosis (but not uniformly poor)

Luminal A subclass are mostly ER-positive low-grade tumors characterized by a high expression level of luminal cytokeratins (CKs 8/18/19), ER, PR, BCL2 and P27KIP1, and a low expression of TP53 and HER2. Luminal B subtype share many of these characteristics but tend to have higher grade and lower expression of hormone receptors and may overexpress HER2. Basal-like tumors are high grade ER- and PR-negative tumors with low levels of luminal CKs, BCL2, P27 and HER2, and a high expression of basal CKs 5/6/14/17, TP53 and epidermal growth factor receptor (EGFR). HER2 tumors are more likely to be high grade and are characterized by a low, if any, expression level of ER, PR and P53 and a high expression of HER2.

The different gene expression patterns observed in breast tumors reflect their biological diversity and are associated with distinct prognosis (**Figure 4**) (Sorlie et al., 2001; Sorlie et al., 2003). An important finding is the separation of ER-positive tumors into at least two distinctive groups with a different disease course: luminal A with the best prognosis and

luminal B with a worse outcome, in particular with respect to relapse. Basal-like and HER2 tumors have the worst prognosis, presenting the shortest overall survival and relapse-free survival.

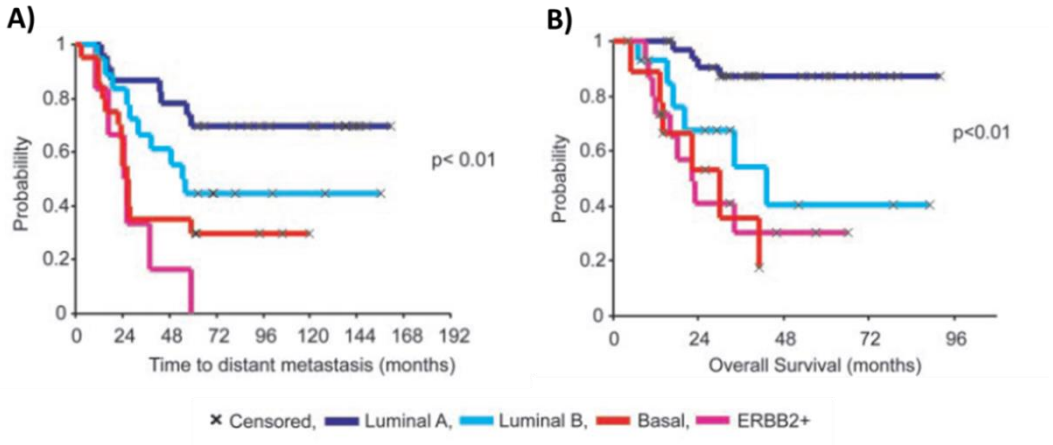


Figure 4. Kaplan–Meier analysis of disease outcome in breast cancer patients stratified according to the intrinsic subtypes. **A)** Time to development of distant metastasis in 97 sporadic cases. **B)** Overall survival for 72 patients with locally advanced breast cancer (adapted from Sorlie et al., 2003).

Molecular subtypes have also proved to have significant predictive value for therapeutic response of breast cancer. **Table 2** summarizes the treatment recommendations adopted by the 13th St Gallen International Breast Cancer Conference (2013) Expert Panel (Goldhirsch et al., 2013). Luminal cancers are generally hormone receptor-positive and appropriate for endocrine therapy. Luminal A subtype is less responsive to chemotherapy than luminal B tumors and therefore chemotherapy is not usually recommended for this subgroup of patients. HER2 positive tumors are suitable for targeted therapy such as trastuzumab with great clinical success. Chemotherapy is also appropriate for this group except for patients at very low risk. Basal-like subtype is resistant to current targeted therapies for breast cancer but benefits from chemotherapy much more than ER-positive tumors (Colleoni et al., 2000).

Table 2. 2013 St Gallen consensus definition of intrinsic subtypes of breast cancer and recommendations of systemic treatment (adapted from Goldhirsch et al., 2013).

Intrinsic subtype	IHC definition	Type of adjuvant therapy
Luminal A	ER+,PR+,HER2-,Ki67low	Endocrine therapy alone*
Luminal B	ER+,HER2-,Ki67high/PRlow	Endocrine + cytotoxic therapy
Luminal B	ER+,HER2+,anyKi67,anyPR	Endocrine + cytotoxics + anti-HER2 therapy
HER2	ER-,PR-,HER2+	Cytotoxics + anti-HER2 therapy
Basal-like	ER-,PR-,HER2-	Cytotoxic therapy

*Cytotoxics may be added in patients at high risk (high 21-gene RS, high 70-gene RS, grade 3, high nodal status)

2.3. Surrogate classifications

The main problem of the original molecular classification is that it has been derived from investigations on fresh frozen tissue, and it is not applicable to formalin-fixed and paraffin-embedded (FFPE) material, limiting its use in the clinical practice. More recently, a gene expression assay using 50 genes (PAM50) has been developed for use on FFPE tissue (Parker et al., 2009). The assay is based on quantitative real-time reverse transcription-polymerase chain reaction (qRT-PCR) of genes involved in proliferation, ER and ER-regulated genes, HER2, and basal and myoepithelial characteristics. The predictor accurately identifies the major molecular subtypes of breast cancer and generates risk-of-relapse (ROR) scores. The ability of the ROR score to predict prognosis has been confirmed in several retrospective investigations using tumor samples of patients with long-term follow-up data and of patients enrolled in randomized, clinical trials (Ellis et al., 2011; Nielsen et al., 2010). The PAM50 test is currently being developed for clinical use on the NanoString nCounter Analysis System (Prosigna™ Breast Cancer Gene Signature Assay, NanoString Technologies, Seattle) (Geiss et al., 2008; Reis et al., 2011).

Another attempt to bring the molecular classification of breast cancer into the clinical practice has been the use of the more familiar immunohistochemical markers. Accordingly, the combined evaluation of ER, PR, HER2 and Ki67 immunoreactivity would approximate the molecular classification of luminal A, luminal B, HER2-enriched and basal-like breast cancers. In fact, the panelists of the last St. Gallen Conference have endorsed the use of this markers to identify breast cancer subtypes and to inform the choice of the systemic treatments (**Table 2**) (Goldhirsch et al., 2013). However, recent studies have indicated that other markers in addition to ER, PR, HER2 and Ki67 are required to more accurately approximate the molecular subtypes. As an example, some basal-like breast cancers will not show the expected triple negative (ER, PR and HER2 negative) immunophenotype, and vice versa not all the immunohistochemically triple negative breast cancer will be classified as basal-like by gene expression profiling (Carey et al., 2010). The basal-like group can be defined more precisely using, in addition to ER, PR, and HER2, antibodies to CK5/6 and EGFR as basal-like cancers are most often triple negative and also express CK5/6 and/or EGFR. IHC-based definition of luminal A and luminal B breast tumors is also imperfect when compared with

gene expression profiling. As a result, a cut-point of >20% for PR has been proposed to define luminal A subtype (Prat et al., 2013).

2.4. Triple negative breast cancer: a clinical challenge

Triple negative breast cancers are defined as tumors that lack ER and PR expression and HER2 overexpression. These tumors represent an important clinical challenge because they do not respond to current targeted therapies for breast cancer, being chemotherapy the only option for triple negative patients. Moreover, these tumors are associated with the most aggressive clinical behavior and poorest prognosis in breast cancer.

Around 15% of invasive breast cancers are triple negative tumors. Even though there is considerable overlap between the profiles of triple negative and basal-like tumors, not all basal-like cancers determined by gene expression profiling lack ER, PR and HER2, and conversely, not all triple negative breast cancers express basal markers at the protein level. It is estimated that nearby 80% of basal-like cancers are triple negative and around 70% of triple negative tumors show a basal-like phenotype. The remainder 30% of triple negative tumors consists of a variety of molecular subtypes that are biologically distinct (Badve et al., 2011). In addition to this molecular heterogeneity, there is a histologic diversity. Although most of triple negative cancers are invasive carcinomas of no special type, other rare types are also included, ranging from those with an excellent prognosis to aggressive metaplastic carcinomas. Therefore, there is a clinical need to identify prognostic and predictive markers to substratify patients with triple negative cancers into groups that can be managed more efficaciously with specific therapies.

2.4.1. Similarities with basal-like breast cancers

Due to their large overlap, triple negative tumors show many characteristics that are associated with basal-like cancers. Typically they are grade 3 carcinomas with elevated mitotic count, high apoptotic rate, geographic or central tumor necrosis or fibrosis, a pushing border of invasion and a stromal lymphocytic response. In addition to the lack of expression of ER, PR and HER2, they frequently express basal cytokeratins (particularly CKs 5, 14 and 17) and the EGFR (Her1). Compared with other subtypes, they are more likely to express myoepithelial markers, such as caveolins, c-kit and P-cadherin, and less likely to

express epitelial markers, such as e-cadherin. They also have high expression of genes associated with proliferation (Ki67 and TOP2A) and up to 70% of them show TP53 gene mutations and/or nuclear accumulation. Expression of p53 homolog p63 is also upregulated (Carey et al., 2010).

Triple negative and basal-like cancers occur more frequently in younger patients (<50 years old) and generally behave aggressively. Women of African ancestry have been shown to have higher rates of triple negative disease, probably due to mutations that predispose to this subtype. The pattern of spread of tumors with a basal-like phenotype is different from that of other subtypes: while luminal tumors typically cause late bone metastases, triple negative breast cancer is more likely to cause early visceral metastases, fundamentally in brain and lungs. In addition, triple negative breast cancers are associated with a higher recurrence rate after diagnosis, a shorter disease-free interval, a shorter period from the time of recurrence until death and a shorter overall survival. The peak risk of recurrence is between the first and third years and the majority of deaths occur in the first 5 years following therapy (Dent et al., 2007).

2.4.2. Triple negative breast cancer and the BRCA1 pathway

There is increasing evidence to suggest a link between BRCA1 pathway and triple negative breast cancers. The majority of tumors in *BRCA1* mutation carriers are triple negative and show morphological and immunohistochemical similarities to basal-like cancers. Both triple negative and *BRCA1*-mutated tumors are characterized by high histological grade, atypically medullary features, high proliferation indices, pushing borders and lymphocytic infiltrate. Both lack ER, PR and HER2 expression and show p53 immunoexpression and TP53 somatic mutations, EGFR expression, peculiar patterns of cell-cycle protein expression and characteristic copy number aberrations. All this characteristics has led to the definition of the BRCAness phenotype (Turner and Reis-Filho, 2006). Moreover, it has been shown that *BRCA1*-mutated tumors consistently segregate with sporadic basal-like breast cancers in hierarchical clustering analysis using microarray expression profiling data. By contrast, tumors from *BRCA2* mutation carriers are predominantly hormone receptor positive and show similar gene expression profiles to sporadic luminal cancers (Sorlie et al., 2003).

Even though *BRCA1* somatic mutations have not been identified in sporadic triple negative tumors, a reduced expression of the *BRCA1* gene has been observed in most of the cases, which indicates a central role of BRCA1 in the development of basal-like carcinomas. This low expression has been associated with *BRCA1* promoter hypermethylation, loss of heterozygosity (LOH) at the *BRCA1* locus and overexpression of two proteins: HMGA1 and ID4 (Mueller and Roskelley, 2003; Turner et al., 2007). Nevertheless, it seems that other mechanisms might also be involved in the inactivation of *BRCA1* in sporadic triple negative tumors as those already described cannot account for the entire reduction of BRCA1 in these tumors.

2.4.3. Treatment strategies

Given the lack of hormone receptors and HER2 overexpression, chemotherapy is the only possibility for patients with triple negative disease. Regimens based on anthracyclines or taxanes are effective with high in-breast response rates. However, relapse rates are high in patients who do not achieve a pathologic complete response, resulting in a short disease-free survival and overall survival. Hence, other agents such as EGFR inhibitors, Src inhibitors, anti-angiogenic agents (Bevacizumab), androgen receptor targeted agents, poly(ADP-ribose) polymerase (PARP) inhibitors and platinum salts are currently being evaluated. In particular, platinum compounds and PARP inhibitors are effective in tumors with a dysfunctional BRCA1 pathway (Hudis and Gianni, 2011). The ongoing trials will show if these agents are more effective than conventional therapy and whether they are able to improve outcomes in this poor-prognosis group of patients.

2.5. Gene expression predictors of breast cancer outcomes

During the last decade, several gene signatures have been described for predicting outcome in patients with breast cancer. Two of them, Oncotype DX (Paik et al., 2004) and Mammaprint (Glas et al., 2006), have been validated with consistent results across multiple studies and have been shown to provide independent prognostic information beyond standard clinicopathological variables. These predictors have been endorsed by the 2013 St Gallen International Breast Cancer Expert Panel (Goldhirsch et al., 2013) and are the most widely used clinical gene-expression assays.

Oncotype is a 21-gene signature generated to predict prognosis of node-negative patients treated with tamoxifen. This assay uses qRT-PCR to measure the expression of 5 reference genes and 16 genes of interest related to proliferation, HER2 and ER signalling. The resulting recurrence score (RS) is 0 to 100, which translates into three risk-group categories: low (RS <18), intermediate (RS from 18 to <31) and high (RS ≥31). An interesting feature of this test is the fact that frozen tissues are not needed since the score is generated using RNA from FFPE tumors. Oncotype has been approved to identify a subgroup of patients within the ER-positive, node-negative breast cancer group who would benefit from addition of chemotherapy and more importantly to identify patients who could possibly be spared cytotoxic therapy. Currently, the RS is undergoing prospective validation as part of the Trial Assigning Individualized Options for Treatment (TAILORx) trial in order to establish if adjuvant chemotherapy improves survival in the group of patients with the intermediate score (Sparano and Paik, 2008).

Mammaprint is a 70-gene signature established to predict outcome in node-negative patients irrespective of ER status. The assay is based on a microarray platform that measures the expression of 70 genes associated with proliferation, invasion and angiogenesis. Initially the test could only be performed on fresh tumor (rarely available), but improvements in RNA processing have enabled its use in FFPE tissue. MammaPrint divides node-negative patients in two groups of low and high risk of recurrence. Currently this assay is undergoing prospective validation as part of the Microarray for Node-Negative Disease Avoids Chemotherapy (MINDACT) trial in order to establish if lymph node–negative breast cancer patients with low risk of recurrence according to Mammaprint but at high risk of recurrence based on clinicopathological factors can be safely spared adjuvant chemotherapy without affecting survival outcomes (Cardoso et al., 2008).

Although promising, these gene expression assays assign almost all patients with hormone receptor negative disease as high risk. They have been shown to made robust prognostic predictions within the group of ER-positive patients but not within the ER-negative disease (Fan et al., 2011). Hence, other signatures able to predict survival within the HER2-positive and triple negative cancers are urgently needed.

2.6. New technologies and novel subgroups

A new breast cancer subtype, known as claudin-low, has been identified in human tumors, in mouse tumors (Herschkowitz et al., 2007) and in a panel of breast cancer cell lines (Prat et al., 2010). These tumors are poor prognosis ER-, PR- and HER2- invasive ductal carcionomas characterized by low expression of genes involved in tight junctions and cell-cell adhesions including claudins 3, 4, and 7, occludin and E cadherin, showing high expression of epitelial to mesenchymal transition genes and stem cell features. As a consequence of the identification of this novel subgroup, triple negative breast cancers would be further subdivided into basal-like and claudin-low tumors.

More recently, the emergence of next generation sequencing has allowed the characterization of the mutational landscape of breast cancer. These analysis have identified likely genomic drivers of the four classical subtypes by focusing on the detection of genes more frequently mutated than expected by chance (Banerji et al., 2012; Stephens et al., 2012; The Cancer Genome Atlas, 2012). Somatic mutations in only three genes (*TP53*, *PIK3CA* and *GATA3*) have been shown to occur at >10% incidence across all breast cancers, although their frequency is different among the intrinsic subtypes.

The same technology has been used in the Molecular Taxonomy of Breast Cancer International Consortium (METABRICK) study in which the integrated analysis of both genomic and transcriptomic data across 2000 breast tumors has revealed ten different subtypes of breast cancer (Curtis et al., 2012a). In this analysis, germline variants and somatic aberrations were found to be associated with alterations in gene expression, although somatic copy number alterations accounted for the greatest variability in gene expression. Unsupervised analysis of joint copy number and gene expression data revealed ten novel subgroups with distinct clinical outcomes and patterns of chemosensitivity (IntClust 1-10). This genome-driven integrated classification has just been simplified into a gene-expression based method that has been validated in 7500 breast tumors (Ali et al., 2014). The clinical relevance of the IntClust classification and its implications for the development of new targeted therapies will be disclosed in the next years.

3. MICRORNAs AS MASTER GENE REGULATORS

3.1. MicroRNA biogenesis and function

MicroRNAs (miRNA) are small (~22 nucleotides) single-stranded non-coding RNAs that have an important function in gene expression regulation. After the discovery of the first miRNA lin-4 in *Caenorhabditis elegans* in 1993 (Lee et al., 1993), these small RNAs have been found to be an abundant class of RNAs in plants, animals and DNA viruses. MiRNAs act as negative regulators at post-transcriptional level by binding at the 3' untranslated regions (3'UTRs) of their target messenger RNAs (mRNAs). Depending on the level of complementarity between miRNA "seed" sequence and its target, they trigger either translational repression or mRNA degradation (Bartel, 2009). The exact mechanisms of miRNA function are still far from being fully understood and are a matter of active research. With more than 2500 reported human miRNAs (miRBase 21.0 release, June 2014), and each one potentially regulating hundreds of mRNAs, miRNAs represent one of the largest classes of gene regulators. Since many of these miRNA targets are involved in various signaling pathways, their impact on gene expression can be significantly amplified. In addition, many miRNAs are evolutionarily conserved from worms to humans, which implies that they are essential both during development and in the adult body. MiRNAs can function as master gene regulators, play an important role in many cellular processes such as differentiation, proliferation, apoptosis and stress response, and their alteration contributes to a range of human diseases, including cancer.

The biogenesis of miRNAs involves a complex protein system (**Figure 5**). MiRNAs, which generally seem to be transcribed by RNA polymerase II, are initially made as large RNA precursors that are called pri-miRNAs. The pri-miRNAs are processed in the nucleus by Drosha, a member of the RNase III enzyme family, in conjunction with the double-stranded RNA-binding protein Pasha, into ~70-nucleotide pre-miRNAs, which fold into imperfect stem-loop structures. After exported from the nucleus in a GTP-dependent fashion by exportin 5, the pre-miRNAs are subsequently processed by a second RNase III endonuclease called Dicer, releasing mature double-stranded miRNAs (~22 nucleotides in length), which in turn are incorporated into the RNA-induced silencing complex (RISC). One of the strands is

preferentially incorporated while the other is degraded giving rise to a functional RISC complex that can target specific protein-coding mRNAs (Esquela-Kerscher and Slack, 2006).

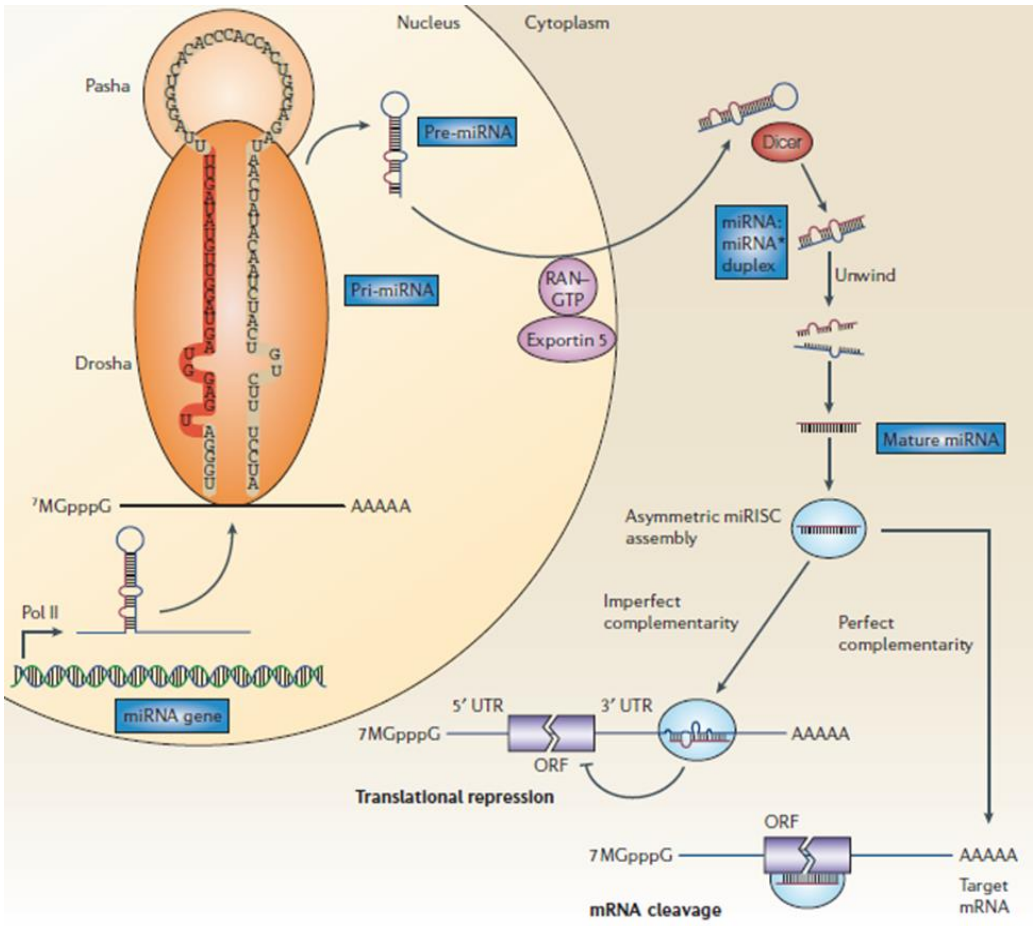


Figure 5. MiRNA biogenesis and functions (adapted from Esquela-Kerscher and Slack, 2006).

3.2. Alteration of miRNA expression in cancer

MiRNAs have been proposed to have a central role in controlling cellular transformation and tumor progression since they can function as tumor suppresors and oncogenes (oncomiRs). The downregulation or deletion of a miRNA that targets an oncogene leads to tumor formation, and vice versa, the amplification or overexpression of a miRNA that targets a tumor suppressor results in tumorigenesis (Esquela-Kerscher and Slack, 2006). MiRNA genes are usually located in small chromosomal alterations in tumors (in amplifications, deletions or linked to regions of loss of heterozygosity) or in common chromosomal-breakpoints that are associated with the development of cancer (Calin et al., 2004). In addition to structural genetic alterations, miRNAs can also be silenced by promoter

DNA methylation, loss of histone acetylation and abnormalities in miRNA-processing genes and proteins. Somatic mutations in miRNA seed sequence could lead to lack of repression of oncogenic mRNAs and/or aberrant downregulation of tumor supressive genes, but these seem to be infrequent (Diederichs and Haber, 2006).

Cancer cells show alterations in their miRNA expression profiles, and emerging data indicate that these patterns could be useful in improving the classification of cancers and predicting their behaviour. The first evidence of involvement of miRNAs in human cancer came from molecular studies characterizing the 13q14 deletion in human chronic lymphocytic leukemia (CLL). Two miRNAs, miR-15a and miR-16-1, located within this región were identified to be either deleted or downregulated in >50% CLL (Calin et al., 2002). Further studies have shown that miR-15a and miR-16-1 negatively regulate BCL2, an anti-apoptotic gene that is often overexpressed in many cancers, including leukaemias and lymphomas (Cimmino et al., 2005). Following this initial discovery, abnormal expression of miRNAs has been found in both solid and hematopoietic tumors by various genome-wide miRNA expression analysis techniques (Lu et al., 2005; Volinia et al., 2006). Cancer cells show distinct miRNA profiles compared with normal cells, and different miRNA expression profiles have been reported in tumors of different origin. Indeed, miRNA expression profiling seems to be a more accurate way of classifying tumors than gene expression profiling (Lu et al., 2005). Furthermore, miRNAs have one great practical advantage over mRNA: they are relatively well preserved in FFPE tissues presumably due to their small size and possibly a sheltered micro-environment (Hasemeier et al., 2008).

3.3. MiRNAs in breast cancer

Accumulating evidence demonstrates that aberrant expression of miRNAs is associated with breast cancer. The seminal study of miRNA expression in 76 breast tumors and 10 normal breast tissues led to the identification of 29 miRNAs whose expression is significantly dysregulated in breast cancer with the most consistently dysregulated miRNAs being miR-125b, miR-145, miR-10b, miR-21 and miR-155. Mir-10b, miR-125b and miR-145 were found to be downregulated, whereas miR-21 and miR-155 were upregulated, suggesting that they may potentially act as tumor suppressor genes or oncogenes, respectively (Iorio et al., 2005). Follow-up studies based on different technologies have validated some of these

miRNAs and identified new ones, suggesting reproducibility of miRNA deregulation in breast cancer (Farazi et al., 2011; Persson et al., 2011; Sempere et al., 2007; Volinia et al., 2006; Volinia et al., 2012). Since the number of annotated miRNAs increases every year, new miRNAs are still expected to be asociated with breast cancer.

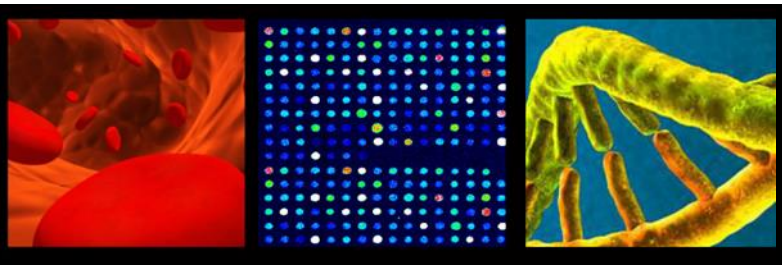
In addition, some studies have correlated miRNA expression with specific breast cancer histopathological features such as tumor stage, proliferation index, vascular invasion, ER, PR and HER2 status (Iorio et al., 2005; Lowery et al., 2009; Mattie et al., 2006; Volinia et al., 2012). Furthermore, preliminary studies suggest that miRNA signatures could define, similarly to what has been found by expression profiling of coding genes, the different intrinsic molecular subtypes (luminal A, luminal B, basal-like, HER2+). By analyzing the expression of 309 human miRNAs, Blenkiron and colleagues detected a number of miRNAs differentially expressed between these molecular subtypes (Blenkiron et al., 2007). Identification of miRNAs specific of breast cancer molecular subtypes would be of great relevance due to their potential influence on the different behavior of these tumors. Subtype-specific miRNAs could be used for classification purposes, as well as to provide better understanding of the biology of these groups of tumors, especially in the case of triple negative cancers, which are associated with the most aggressive clinical behavior and poorest prognosis in breast cancer and do not respond to current targeted therapies.

3.4. Circulating miRNAs as novel non-invasive biomarkers

Perhaps the most attractive application of miRNAs as cancer biomarkers comes from the finding of circulating miRNAs in different body fluids such as plasma, serum, urine, saliva, milk, etc. Tumor-specific miRNAs were first discovered in the serum of patients with diffuse large B-cell lymphoma (Lawrie et al., 2008). Since then, circulating miRNAs are attracting a great deal of attention as novel cancer biomarkers due to their ease of access and remarkable stability. It has been consistently shown that circulating miRNAs remain stable after being subjected to severe conditions that would normally degrade most RNAs, such as boiling, very low or high pH levels, extended storage, and 10 freeze–thaw cycles (Chen et al., 2008). This stability can be partially explained by two mechanisms: (i) protection of secreted miRNAs by the membrane of vesicles of endocytic origin called exosomes or microvesicles (30–100 nm) (Valadi et al., 2007), and (ii) stabilization of secreted miRNAs by their

association with RNA-binding proteins, such as AGO2 and NPM1 (Arroyo et al., 2011). Importantly, exosomes represent a newly discovered mechanism by which donor cells can communicate and influence the gene expression of recipient cells (Valadi et al., 2007), and studies have shown that tumor-derived exosomes can promote tumor progression (Skog et al., 2008). Although it is unclear how circulating miRNAs are liberated into body fluids, the packaging of specific miRNA populations into microvesicles appears to be a selective process. Studies in malignant mammary epitelial cells have demonstrated that the cellular and the extracellular miRNA profiles are different, suggesting that specific miRNAs are selected to be intracellularly retained or released by exosomes (Pigati et al., 2010). In addition, it seems that certain circulating miRNAs could be differentially expressed in the serum and plasma of breast cancer patients when compared with healthy individuals (Asaga et al., 2011; Cuk et al., 2013; Chan et al., 2013; Roth et al., 2010), indicating that these molecules may reflect the presence of a tumor. Nevertheless, few studies have been conducted in this area and further research is required.

In summary, the findings discussed above highlight the potential clinical utility of circulating miRNAs in breast cancer diagnosis. Studies in large cohort of patients with well-defined clinical data are needed in order to identify circulating miRNAs that could discriminate breast cancer patients from healthy individuals with robustness and reproducibility. In addition, their analysis in treated patients will shed light on their correlation with tumor dynamics.



Objectives

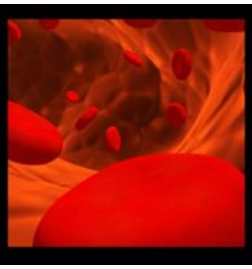
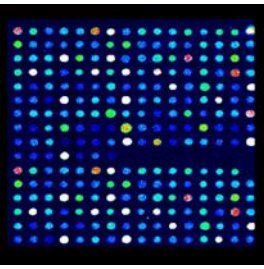
A large number of studies have explored the value of gene expression profiling in breast cancer, thus leading to some very interesting findings that have been successfully translated to the clinic. However, few reports have investigated the usefulness of miRNA expression profiling in breast cancer diagnosis, prognosis and treatment. Given the important role that miRNAs play in tumorigenesis and their great potential as novel clinical biomarkers, the specific goals of this thesis were:

1. To identify miRNAs differentially expressed in breast tumors and the main molecular subtypes of breast cancer (luminal A, luminal B, Her2 and triple negative) and to establish miRNA signatures for their discrimination.

2. To study in plasma the status of the most deregulated miRNAs identified in breast tumors and to analyze their utility as non-invasive biomarkers for early breast cancer detection.

3. To identify prognostic and/or predictive miRNAs in triple negative breast cancer patients.

4. To investigate the involvement of miRNAs in *BRCA1* regulation in sporadic triple negative breast cancer.


Materials and methods

1. PATIENTS AND SAMPLES

1.1. Types of samples and ethics statement

Samples used in this thesis were collected from breast cancer patients and healthy women of the same ethnicity (white Spaniards) in five Spanish institutions: Hospital Virgen de la Macarena and Hospital Virgen del Rocío (Sevilla), Hospital Monte Naranco and Biobanco del Principado de Asturias (Oviedo) and Sistemas Genómicos (Valencia). Two different types of samples were obtained:

Formalin-fixed paraffin embedded (FFPE) breast tumors were obtained from patients undergoing surgery for breast cancer. In addition, normal breast tissues were acquired after breast reduction surgery from healthy women with no family history of cancer and were used as control samples. FFPE tissues were stained by hematoxylin and eosin and examined by two pathologists (Ricardo González-Cámpora and Primitiva Menéndez). The tumoral area was identified and macrodissected from 3 sections of 30 µm thicknesses for subsequent RNA extraction.

Plasma samples were collected from breast cancer patients and healthy women. Plasma from breast cancer patients were divided into two groups: those obtained at the time-point of diagnosis before any treatment such as surgery, radiation or systemic therapy (pretreated) and those obtained after treatment (postreated). EDTA blood samples were processed for plasma within 1 hour of collection. Blood was centrifuged at 3000 g for 20 minutes at 10ºC followed by further centrifugation of the supernatant at 15500 g for 10 minutes at 10ºC to remove cell debris. The plasma was stored at -80ºC until use.

The study was performed in accordance with the principles of the Declaration of Helsinki. Written informed consent was obtained from all patients prior to sample collection, and the study was approved by the ethics committee of Instituto de Salud Carlos III (Madrid), Hospital Virgen del Rocío and Virgen de la Macarena (Sevilla) and Hospital Monte Naranco (Oviedo).

1.2. Patients' clinical data

Clinicopathologic features (age at diagnosis, histological type, tumor size, histologic grade, lymph node status, stage and expression of ER, PR, c-erb B2 and Ki-67) were retrieved from all patients. Tumors were classified as triple negative, Her2, luminal B or luminal A based on the expression of inmunohistochemical markers (ER, PR, c-erb B2 and Ki-67), following the criteria adopted in the 12th St Gallen International Breast Cancer Conference, 2011 (Goldhirsch et al., 2011) and the definition of luminal A tumors proposed by Prat et al. (Prat et al., 2013). A summary of these criteria is shown in **table 3**. In the case of a weak positive reaction of c-erb B2, fluorescent in situ hybridization was performed to confirm the overexpression of this receptor.

Table 3. Criteria used in this thesis to classify breast tumors into the main molecular subtypes.

Molecular subtype	IHC definition
Luminal A	ER+, PR high (≥20%), HER2-, Ki67 low (≤14%)
Luminal B	ER+, HER2-, Ki67 high (>14%) or PR low (<20%)
	ER+, HER2+, any Ki67, any PR
HER2	ER-, PR-, HER2+
Triple negative	ER-, PR-, HER2-

In addition, follow-up information was obtained from triple negative patients with the objective of performing survival analysis. Relapse-free survival (RFS) was defined as the time between initial diagnosis and relapse or death by the disease, with observations censored at last follow-up if no event had occurred. Median follow-up time of patients alive was 62 months (range: 57–99 months). Patients had not been treated with any systemic neoadjuvant therapy and had received adjuvant chemotherapy consisting in most cases in taxanes and/or anthracyclines.

1.3. Samples cohorts

During the elaboration of this thesis four different studies were performed including variable number of samples.

First study: one hundred and twenty-two FFPE breast tumors, as well as 11 normal breast tissues, were used for microarray profiling. Breast tumors were divided into a training (n=61) and a test set (n=61). Both series comprised a similar number of samples from each molecular subtype.

Second study: plasma samples were collected from 83 breast cancer patients and 26 healthy women for study of selected miRNAs in blood. Plasma from breast cancer patients were divided into two groups: those obtained at the time-point of diagnosis before any treatment (n=36) and those obtained after treatment (n=47). In addition, a validation set of plasma samples from 114 pretreated breast cancer patients and 116 healthy women was obtained.

Third study: twenty-one FFPE breast tumors from triple negative patients included in the first study were used here for survival analysis. A second series of 22 TN FFPE tumors was obtained for validation of selected miRNAs.

Fourth study: a panel of 6 human cell lines was used as a model to study functional effects of miRNA expression and inhibition. Five of them (MDA-MB-231, Hs578T, SKBR3, BT474 and MCF7) corresponded to sporadic breast tumors and were representative of the main molecular subtypes of breast cancer, and one (HEK-293T) was derived from an embryonic kidney.

2. RNA EXPRESSION ANALYSIS

2.1. RNA extraction and quantification

Total RNA was extracted from FFPE tissues using miRNeasy FFPE Kit (Qiagen, Valencia, CA, USA) according to the manufacturer's instructions. RNA from plasma was extracted from 250 ml of plasma using miRNeasy Mini Kit (Qiagen, Valencia, CA, USA) and the modified Exiqon protocol that includes the addition of MS2 RNA (Roche, Basel, Switzerland), a carrier RNA that ensures the highest and most consistent yield of RNA in the samples. The final elution volume was 50 µl. Finally, RNA from cell lines was extracted using miRNeasy Mini Kit

(Qiagen) according to the manufacturer's instructions. RNA quality and quantity were assessed by NanoDrop Spectophotometer (NanoDrop technologies, Wilmington, DE, USA).

2.2. Microarray hybridization

MicroRNA expression profiling was performed using miRCURY LNA™ microRNA Array 7th generation – hsa, mmu & rno (Exiqon A/S, Vedbaek, Denmark), in a single-color experimental design. The miRCURY LNA™ microRNA Array 7th generation – hsa, mmu & rno contains capture probes for all microRNAs in human, mouse, rat and their related viruses as annotated in miRBase Release v.18.0. This includes probes for 1919 human miRNAs in quadruplicate: 1894 miRNAs from miRBase Release 18.0 and 25 hsa-miRPlus not included in miRBase (Exiqon proprietary). In addition, 82 control probes are included: 52 spike-in control probes to ensure optimal labeling and hybridization, 7 negative control probes and 23 probes complementary to small nuclear RNAs.

Labeling and hybridization procedure was performed as recommended by manufacturer, using miRCURY LNA™ microRNA Hi-Power Labeling Kit (Exiqon). First, 300ng of total RNA was treated with Calf Intestinal Alkaline Phosphatase (CIP) to remove the 5'-phosphates from the microRNA termini. The 5 ul –reaction contained 0.5 ul of CIP buffer, 0.5 ul of CIP enzyme, 1 ul of synthetic RNA spike-in and 3 ul of RNA. The reaction was incubated at 37ºC for 30 min and 95ºC for 5 min. Second, a Hy3 fluorescent label was attached enzymatically to the 3'-end of the microRNAs in the total RNA sample. The 12.5 ul -reaction contained 3 ul of labeling buffer, 1 ul of labeling enzyme, 1.5 ul of Hy3 fluorescent dye, 2 ul of DMSO and 5 ul of CIP treated RNA. The reaction was incubated at 16ºC for 1h and heat inactivated by incubation at 65ºC for 15 minutes. Third, labeled samples were combined with 200 ul of hybridization buffer, denatured at 95ºC for 2 min and loaded onto a miRCURY LNA™ microRNA array slide. Hybridization took place over 16h at 56ºC using Agilent Hybridization chambers SureHyb and a rotating oven.

Arrays were then washed, dried and scanned with Agilent G2565AA Microarray Scanner System (Agilent Technologies, Santa Clara, CA, USA), with the laser set to 635nm, at Power 80 and PMT 70 setting, and a scan resolution of 10µm. To avoid ozone bleaching, microarrays were scanned in an ozone-free environment (less than 2 ppb ozone).

Fluorescence intensities on scanned images were measured with Agilent Feature Extraction software, version 10.7.3 (Agilent Technologies), using the modified Exiqon protocol. Reproducibility and reliability of each single microarray was assessed using Quality Control report data. Microarray dataset is publically available at NCBI's Gene Expression Omnibus database http://www.ncbi.nlm.nih.gov/geo/ under GEO accession number GSE58606.

2.3. Microarray data analysis

2.3.1. Normalization and pre-processing

Microarray background subtraction was carried out using normexp method. Processed intensity data were then log2 transformed and normalized using quantiles between arrays normalization. Replicate probes were merged by their mean profile and the data set was filtered to eliminate miRNAs with low expression variation across samples (VAR<0.03), reducing the number of miRNAs to 698.

2.3.2. Clustering

In order to obtain clustering of the data, unsupervised hierarchical clustering was performed using Gene Cluster software with average linkage clustering, Pearson correlation and uncentered metrics (http://rana.stanford.edu/software). Java Tree View was used for image visualization (http://jtreeview.sourceforge.net). The level of expression of each miRNA in each sample, relative to the median level of expression of that gene across all the samples was represented using a red-black-green color scale. Green corresponds to expression value below median, black equal to median, and red above the median.

2.3.3. Differential expression analysis

Differentially expressed miRNAs were obtained by applying linear models with R limma package (Smyth G) (Bioconductor project, http://www.bioconductor.org), implemented in the POMELOII tool (http://asterias.bioinfo.cnio.es/). To account for multiple hypotheses testing, the estimated significance level (p value) was adjusted using Benjamini & Hochberg False Discovery Rate (FDR) correction (Benjamini et al., 2001). FDR<0.05 was set as threshold to select significantly differentially expressed miRNAs.

2.3.4. Building miRNA microarray classifiers

In order to identify the smallest set of miRNAs that better discriminate breast tumors from normal breast tissues, and each molecular subtype from the rest of subtypes, five miRNA microarray classifiers were generated using samples from the training set. The predictors were built with the 698 miRNAs used in the differential expression analysis and the most relevant miRNAs were chosen using correlation feature selection, a method that evaluates a set of features on the basis of the following hypothesis: "A good feature subset is one that contains features highly correlated with the class, yet uncorrelated with each other".

We evaluated the performance of different methods that have been shown to function well with microarray data (Romualdi et al., 2003; Wessels et al., 2005): support vector machines (SVM), k-nearest neighbor (KNN) and random forest (RF), and that are included in the Prophet tool (http://babelomics.bioinfo.cipf.es/). These algorithms were applied to: i) 61 breast tumors and 7 normal breast tissues, ii) 15 triple negative and 46 non-triple negative tumors, iii) 13 Her2 and 48 non-Her2 tumors, iv) 17 luminal B and 44 non-luminal B tumors, and v) 16 luminal A and 45 non-luminal A tumors. The classification performance was evaluated by 5-fold cross validation repeated 10 times: samples were randomly divided into 5 sets with each set containing a fair representation of the classes to be learned. A predictor was built based on the data of 4 of these sets and tested in the remaining set of samples to determine its efficiency. This process was repeated 10 times with different combinations of samples and average classification efficiency was determined. Classifiers producing the minimal root median square error (RMSE) and maximal accuracy, Mathews correlation coefficient (MCC) and area under the curve (AUC) were selected.

To validate the performance of the selected classifiers, we used the samples from the test set: i) 61 breast tumors and 4 normal breast tissues, ii) 16 triple negative and 45 non-triple negative tumors, iii) 13 Her2 and 48 non-Her2 tumors, iv) 16 luminal B and 45 non-luminal B tumors, and v) 15 luminal A and 46 non-luminal A tumors. Sensitivity and specificity values were estimated based on the confusion matrix.

2.3.5. MiRNA target prediction

A number of computational prediction programs have been developed to identify putative miRNA targets based on sequence complementarity between the miRNA and its potential mRNA target 3' untranslated region (3'UTR) (Bartel, 2009). The most important factor for miRNA target prediction seems to be perfect complementarity to the 5' region of the miRNA centered on nucleotides 2-8, which is called "miRNA seed" (**Figure 6A**). Furthermore pairing to the 3' region of the miRNA can also compensate for a mismatch in the seed region. These so called "3'-compensatory sites" are centered on miRNA nucleotides 13-17 (**Figure 6B**). In an attempt to increase target prediction specificity some prediction algorithms are relying on target site evolutionary conservation and thermodynamic stability of the RNA-RNA duplex. To determine potential mRNA targets for specific miRNAs we have used several publically available target prediction algorithms, namely, TargetScan (http://www.targetscan.org/), miRanda (http://www.microRNA.org/), Pita (http://genie.weizmann.ac.il/pubs/mir07/mir07_dyn_data.html), MicroTar (http://tiger.dbs.nus.edu.sg/microtar/), RNAHybrid (http://bibiserv.techfak.uni-bielefeld.de/rnahybrid/) and DIANA microT (http://diana.imis.athena-innovation.gr/DianaTools/index.php?r=microT_CDS/index).

In addition, we have used a number of databases that compile experimentally validated miRNA-gene interactions, such as Diana TarBase (http://diana.imis.athena-innovation.gr/DianaTools/index.php?r=tarbase/index), OncomirDB (http://bioinfo.au.tsinghua.edu.cn/member/jgu/oncomirdb/index.php), and miRecords (http://mirecords.biolead.org/).

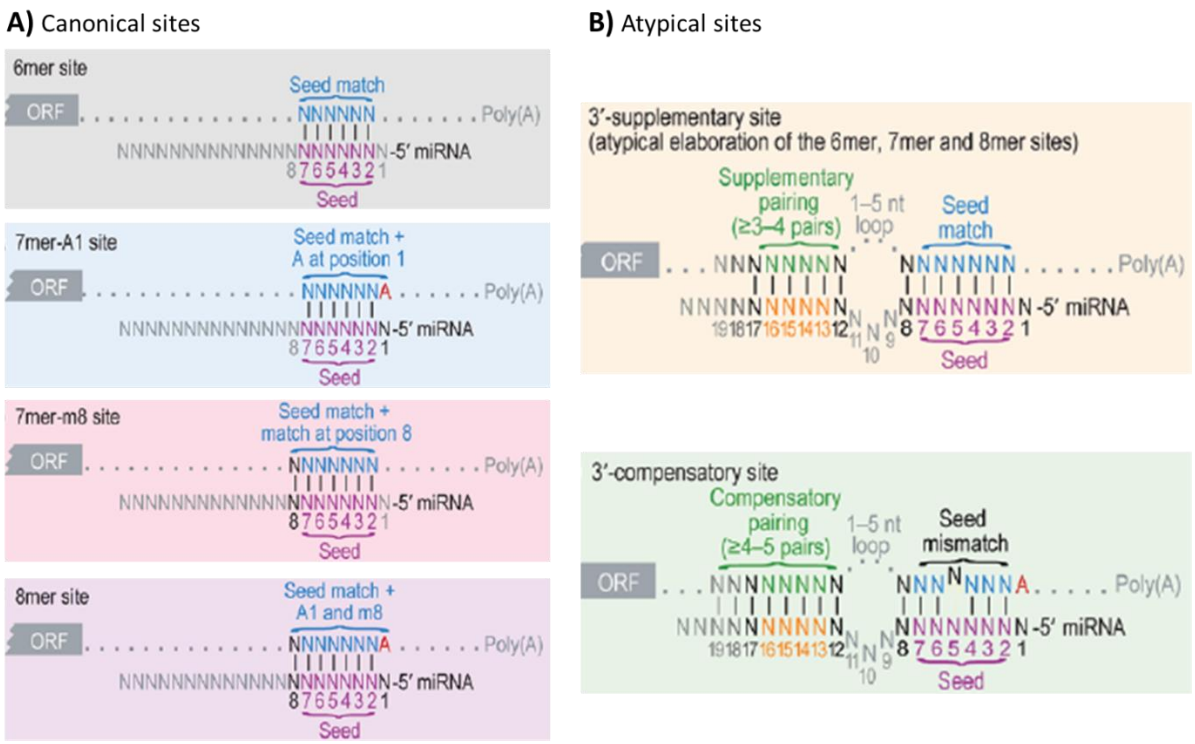


Figure 6. Types of miRNA target sites. **A)** Canonical, 6-8 nt seed-matched sites. **B)** Atypical 3' pairing. Vertical dashes indicate contiguous Watson-Crick pairing (adapted from Bartel, 2009).

2.3.6. MiRNA pathway analysis

DIANA miRPath pathway enrichment analysis (http://diana.imis.athena-innovation.gr/DianaTools/index.php?r=mirpath/index) was used to gain insight into global molecular networks and canonical pathways related to deregulated miRNAs. DIANA miRPath is a web-based computational tool developed to identify molecular pathways potentially altered by the expression of single or multiple microRNAs. The software performs an enrichment analysis of multiple microRNA target genes (predicted or experimentally validated) comparing each set of microRNA targets to all known KEGG pathways. Those pathways showing p-value <0.05 were considered significantly enriched between classes under comparison.

2.4. Real time quantitative PCR (qRT-PCR)

2.4.1. Detection of miRNA by qRT-PCR

Quantification of the expression of the most relevant miRNAs was performed by qRT-PCR using miRCURY LNA™ Universal RT microRNA PCR system (Exiqon) according to the manufacturer's protocol. In the case of FFPE tissues and cell lines, 12 ng of total RNA was reverse-transcribed with universal poly-T primers in 30 µl reactions. In the case of plasma samples and due to the low RNA concentrations, RNA amounts were used based on starting volume rather than RNA quantity, and 6 ul of total RNA was reverse-transcribed with universal poly-T primers in 20 ul reactions. In both cases, a 10 ul –reverse reaction contained 2 ul of Reaction Buffer, 1 ul of Enzyme Mix, 0.5 ul of synthetic RNA spike-in (Uni Sp6), nuclease-free water and 3 ul of RNA. The RNA spike-in is a synthetic template that allows the control of the quality of the cDNA synthesis. The reaction was carried out at 42ºC for 60 min and 95ºC for 5 min on a GeneAmp PCR System 9700 Thermal Cycler (Applied Biosystems). The cDNA samples were stored at -20ºC until further usage.

The rest of steps were common for FFPE tissues, cell lines and plasma. cDNA was 10x diluted and amplified by qPCR with miRNA-specific primers optimized with LNA. Briefly, in a 10 ul –reaction, 4 ul of diluted cDNA were mixed with 5 ul SYBR Green master mix and 1 ul PCR primer mix. The amplification conditions consisted of an initial step at 95ºC for 10 minutes, followed by 50 cycles of 10 seconds at 95ºC and 1 minute at 60ºC. MiRNA expression levels were detected using ABI Prism Sequence Detection System 7900HT (Applied Biosystems). All reactions were performed in triplicate and no-template controls were included in each run. In order to assess amplification specificity, dissociation curves and replicate assays were examined and those miRNAs whose dissociation curve showed unspecific amplification or inconsistent replicate Cq values were removed from further analysis. MiR-103a-3p was used to normalize miRNA expression as it appears in the literature as widely-used endogenous control for miRNA qRT-PCR and was stably expressed among our samples. Relative expression was calculated using the comparative cycle threshold (ΔΔCt) method implemented in qBasePLUS software (Biogazelle).

2.4.2. Detection of mRNA by qRT-PCR

Five hundred nanograms of total RNA was reverse transcribed using the High-Capacity cDNA Reverse Transcription Kit (Applied Biosystems) and random primers following manufacturer's instructions. Briefly, in a 20 ul –reaction, 10 ul of template RNA were mixed with 2 ul of RT Buffer (10x), 0.8 ul of dNTP mix (100 mM), 2 ul of random primers (10x), 1 ul of MultiScribe Reverse Transcriptase (50 U/ul), 1 ul of RNAse inhibitor (20 U/ul) and nuclease-free water. The reaction mixture was incubated at 25ºC for 10 min, 37ºC for 2h and 85ºC for 5 min. The cDNA was then 10x diluted and amplified by qPCR with the use of FAM/NFQ fluorescently labeled probes TaqMan (Roche Universal Probe library, Roche), specific primers (Sigma-Aldrich, St. Louis, MO, USA) and TaqMan Universal PCR Maser Mix (Applied Biosystems). The primers and probes used are listed in **Table 4**. In brief, the 12.5 ul –reaction contained 1.25 ul of forward and reverse primers (10 uM), 0.125 ul of TaqMan probe (10 uM), water and 6.25 ul of TaqMan Universal PCR Master Mix (2x). The amplifications conditions consisted of an initial step at 95ºC for 10 minutes, followed by 45 cycles of 15 seconds at 95ºC and 1 minute at 60ºC. mRNA expression levels were detected using ABI Prism Sequence Detection System 7900HT (Applied Biosystems). All reactions were performed in triplicate and no-template controls were included in each run. In addition, each set of primers was tested for efficacy using serial dilutions of a control cDNA sample. Normalization of mRNA expression was carried out using *ACTB* and *MRLP19* as reference genes. Relative expression was calculated using the comparative cycle threshold (ΔΔCt) method implemented in qBasePLUS software (Biogazelle).

Table 4. Oligonucleotide primers and probes used for mRNA qRT-PCR.

Gene name	Primers	Sequence (5'-3')
BRCA1	Forward primer	ttaaagaaagaaaaatgctga
	Reverse primer	ggtggtttcttccattgacc
	Universal Probe Library	#82
ACTB	Forward primer	ccaaccgcgagaagatga
	Reverse primer	ccagaggcgtacagggatag
	Universal Probe Library	#64
MRLP19	Forward primer	ggaatgttatcgaaggacaag
	Reverse primer	caggaagggcatctcgtaag
	Universal Probe Library	#42

2.5. Statistical analysis

Statistical analysis was performed using GraphPad PRISM 5 software (GraphPad Software, La Jolla, CA) and SPSS software package, version 17.0 (IBM). In all the analysis, a two-tailed p-value<0.05 was considered statistically significant.

2.5.1. Differential expression and ROC curve analysis

Kolmogorov–Smirnov test was used to analyze the normal distribution of the miRNA expression levels, and unpaired t test or Mann–Whitney test were applied when appropriate to evaluate differences in miRNA expression between two groups. MiRNA discrimination potential was analyzed by computing receiver operating characteristic (ROC) curves and calculating areas under the curves (AUC) with corresponding 95% confidence intervals (CI), as well as the optimal specificity and sensitivity values. For associations between miRNA expression levels and clinicopathologic characteristics of the patients at the time of breast cancer diagnosis, non-parametric Mann–Whitney test (two groups' comparison) and Kruskal Wallis test (multiple groups' comparison) were applied.

2.5.2. Survival analysis

Estimation of survival time distribution was performed using Kaplan-Meier method and differences between survival curves were assessed for statistical significance with log-rank test if the proportional hazard assumption was valid, or Gehan–Breslow–Wilcoxon test otherwise. A univariate Cox proportional hazards regression analysis was conducted to determine the impact of miRNA expression status on RFS. To adjust for other prognostic factors potentially acting as confounding variables (tumors size, age at diagnosis, Ki-67 expression levels and nodal status), we used multivariate Cox proportional hazards regression model. The prognostic value of the miRNAs analyzed was tested by comparing patients with expression levels ≥median versus those with expression levels <median.

3. CELL LINES AND FUNCTIONAL STUDIES

3.1. Breast cancer cell lines

A panel of 6 human cell lines was used as a model to study functional effects of miRNA expression and inhibition. Five of them (MDA-MB-231, Hs578T, SKBR3, BT474 and MCF7) corresponded to sporadic breast tumors and were representative of the main molecular subtypes of breast cancer (**Table 5**). These cell lines were obtained from the Cancer Epigenetics Group at the Bellvitge Institute for Biomedical Research (Barcelona, Spain). The sixth cell line (HEK-293T) was derived from an embryonic kidney and was obtained from the Cytogenetics Group at the Spanish National Cancer Research Centre (Madrid, Spain).

Table 5. Breast cancer cell lines used in this thesis.

Breast cancer cell line	Molecular subtype	IHC markers
MDA-MB-231	Triple negative	ER-, PR-, Her2-
Hs578T	Triple negative	ER-, PR-, Her2-
SKBR3	Her2	ER-, PR-, Her2+
BT474	Luminal B	ER-, PR+, Her2+
MCF7	Luminal A	ER+, PR+, Her2-

3.2. Manteinance and subculturing of cells

Cell lines were grown in RPMI-1640 (Sigma-Aldrich) or Dulbecco's Modified Eagle's Medium (DMEM, Sigma-Aldrich) containing 10% fetal bovine serum, 1% penicillin/streptomycin and 0.5% fungizone (Gibco, Life Technologies). In the case of BT474 cells, the medium was completed with 0.01 mg/ml of insulin (Sigma-Aldrich). Cells were maintained in an atmosphere of 5% CO2 in air at 37ºC and were passaged at approximately 80-90% confluence.

3.3. Luciferase reporter assay

To verify direct binding of miR-498 and miR-187-5p to *BRCA1*, we performed target in vitro assays using luciferase reporter system (**Figure 7**). Pre-miRNA oligonucleotides (pre-miR-498, pre-miR-187-5p, pre-miR-146a-5p and non-targeting control) were purchased from Ambion (Life Technologies). In brief, 100 ng of Firefly Luciferase-*BRCA1* 3'UTR

construct, together with 7.5 ng of Renilla Luciferase vector and 6 pmol (50 nM) of individual pre-miRNA oligonucleotides or mock transfection control, were transfected using Lipofectamine 2000 reagent (Invitrogen, San Diego, CA) into 293T cells in a 96-well plate format following manufacturer's instructions. Cells were grown for 48h, after which cells were harvested and luciferase activity was assayed with Dual-Glo Luciferase Assay System (Promega) according to manufacturer's instructions. Experiments were performed in triplicate and normalization was achieved using Renilla luciferase activity.

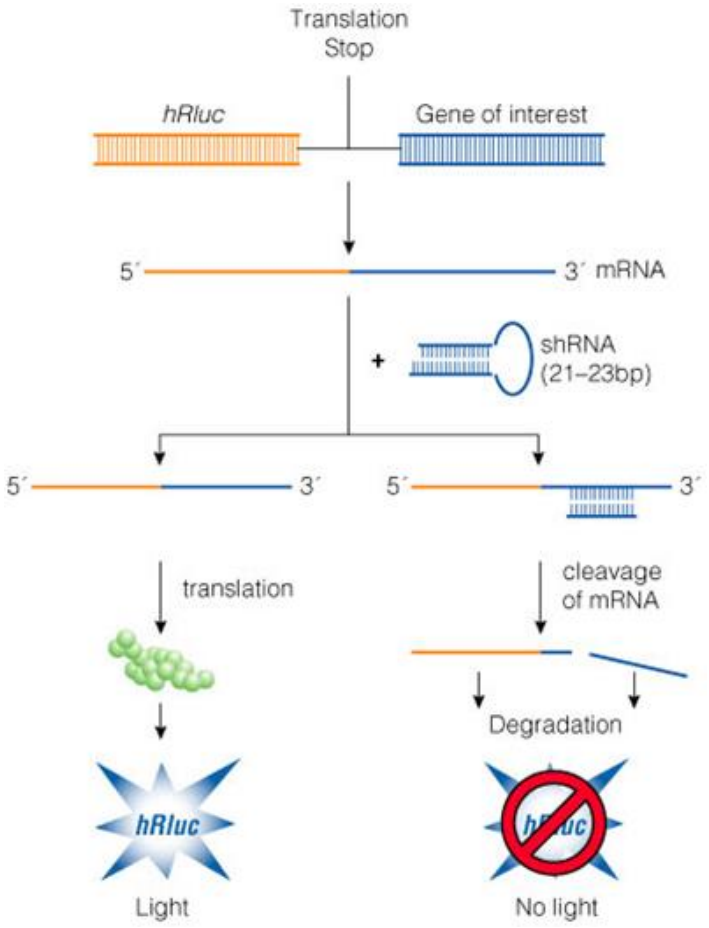


Figure 7. Outline of the luciferase reporter assay. Firefly luciferase gene, with 3'UTR region of interest cloned immediately downstream of the stop codon, codes for an oxidative enzyme that converts luciferin substrate into oxiluciferin in a reaction that emits light. Photon emission is detected by luminometer, and the signal intensity is directly proportional to the amount of the enzyme. In the presence of a miRNA that binds to the 3'UTR and induces either mRNA degradation or translational inhibition, production of luciferase enzyme is reduced/abolished resulting in lower signal emission.

3.4. MiRNA transfection

In order to express or inhibit miR-498, synthetic pre-miR-498 or anti-miR-498 oligonucleotides were transfected into MCF7 or HS678T cells, respectively, using Oligofectamine reagent (Invitrogen). Pre-miRNA oligonucleotides were purchased from Ambion (Life Technologies) and anti-miRNA oligonucleotides from Exiqon. One day before transfection, cells were seeded in 6-well plates with 2 ml of growth medium without antibiotics to a density of 60%. Stock transfection mixes were made according to manufacturer's instructions. Briefly, 5 ul of Oligofectamine reagent was diluted in 15 ul of Opti-MEM I Medium with L-glutamine (Invitrogen, CA, USA) and incubated 5 minutes at RT. In another tube, pre-miRNA or anti-miRNA oligonucleotides were diluted with Opti-MEM to a final concentration of 25nM. Both mixes were incubated together for 20 minutes at RT to allow complex formation between miRNA and lipids. To transfect cells, the growth media was removed and replaced with 800 ul of Opti-MEM media, and 200 ul of the appropriate transfection mix. Control cells were treated with a non-targeting miRNA or with no miRNA precursor or inhibitor (mock). Cells were incubated at 37ºC for 6 h, after which media was replaced with the same volume of fresh full growth media. At 48 hours after transfection cells were harvested for mRNA and/or protein analysis.

3.5. Cell proliferation assay

Cell proliferation assay was assessed by using water-soluble tetrazolium salt (WST-1) assay. HS578T cells were seeded in 96-well plates one day before transfection at 30-40% confluence in antibiotic-free media. Cells were transfected using Oligofectamine (Invitrogen) with 25 nM of anti-miR-498, non-targeting miRNA or mock transfection control. Cells were incubated in 10µl of WST-1 (Roche) diluted in 200 µl normal culture medium at 37ºC for 2h. The assay is based on the cleavage of the red WST-1 tetrazolium salt to yellow soluble formazan by metabolically active cells. The formazan dye is quantified by optical density at 450 nm by means of a multi-well spectrophotometer (Perkin Elmer, Massachusetts, USA). Cell viability was determined at 24, 48, 72, 96, 120 or 144 hours after transfection. Each value represents the average of six independent replicates.

4. PROTEIN-BASED ASSAYS

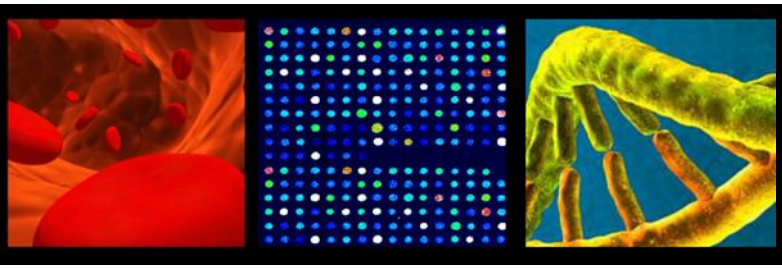
4.1. Protein extraction and quantification

Since BRCA1 protein is predominantly expressed in the nucleus, nuclear protein extraction was performed. Cells were trypsinized and washed twice with PBS 1X. Cell lysates were prepared by resuspending cell pellets in 200 ul of RSB buffer (Tris 10 mM pH 7.5, NaCl 10 mM and MgCl2 3 mM,) with protease inhibitor (Roche) per 10^6 of cells, and by incubating 20 min on ice. After centrifugation at 2000 rpm for 5 min at 4ºC, supernatant contains the cytoplasmic fraction and the remaining pellet the nuclear fraction. In order to remove any contaminating cytoplasm, pellets were again resuspended in 200 ul of RSB buffer containing protease inhibitor and centrifuged, and the remaining pellet was washed twice with 200 ul of RSB buffer without protease inhibitor. The pellet was then dissolved in 20 ul of NB buffer (Tris 10 mM pH 7.5, NaCl 0.4 mM and EDTA 1 mM) containing protease inhibitor, shaked in a rotor for 15 min at 4ºC and centrifuged at 11000 rpm for 5 min at 4ºC. Supernatant was collected and protein concentration was measured by Lowry assay method (Bio-Rad laboratories, Hercules, CA, USA) using bovine serum albumin (BSA) (Sigma-Aldrich) to create a standard curve with known concentrations of protein.

4.2. Western blot analysis

Equal amounts of protein (50µg) were separated by SDS-PAGE on 6% home-made gels at 75 V for 1.5 hour using a Mini-PROTEAN Tetra cell electrophoresis chamber (Bio-Rad). Briefly, 10 ml of resolving gel contained 2 ml of 30% acrylamide, 2.5 ml of 1.5 M Tris (Ph 8.8), 0.1 ml of 10% SDS, 0.1 ml of 10% ammonium persulfate, 0.008 ml of TEMED and water; and 5 ml of 5% stacking gel contained 0.83 ml of 30% acrylamide, 0.63 ml of 1 M Tris (Ph 6.8), 0.05 ml of 10% SDS, 0.05 ml of 10% ammonium persulfate, 0.005 ml of TEMED and water. After electrophoresis, proteins separated on the gel were electrotransferred during 1h at 60V to nitrocellulose membrane (Whatman) using 1x NuPAGE Transfer buffer (Invitrogen) with 10% methanol and a Mini Trans-Blot Cell (Bio-Rad). The membranes were blocked overnight at 4ºC with 5% non-fat dry milk in 1xTBS with 0.05% Tween (TBS-T) detergent, washed with TBS-T and incubated with primary antibody overnight at 4ºC. For BRCA1 detection, mouse antibody against BRCA1 (OP92, Calbiochem, Darmstadt, Germany)

at 1/100 dilution was used. For HSP70/HSC70 detection (loading control), mouse antibody against HSP70/HSC70 (ADI-SPA-820, Enzo Life Sciences) at 1/2000 dilution was used. Next, membranes were washed three times for 10 minutes with TBS-T and incubated with the corresponding horseradish peroxidase (HRP) conjugated (Dako, Glostrup, Denmark) secondary antibody at 1/10000 dilution for 2 hours at room temperature. After washing the membranes three times for 10 minutes with TBS-T, the antibody visualization was carried out with Amersham ECL Western Blotting Detection Reagent (GE Healthcare Life Sciences Buckinghamshire, UK) and Ortho CP-G Plus x-ray films (AGFA, Mortsel, Belgium). Films were scanned and signal was quantified using Image J program. BRCA1 protein content was determined relative to HSP70/HSC70 protein content.



Results

1. DEREGULATED MICRORNAs IN BREAST CANCER MOLECULAR SUBTYPES

A large number of studies have explored the value of gene expression profiling in breast cancer, thus leading to the stratification of breast tumors into at least four major subtypes: luminal A, luminal B, Her2 and triple negative (Hu et al., 2006; Perou et al., 2000; Sorlie et al., 2001; Sorlie et al., 2003). However, few reports have investigated the usefulness of miRNA expression profiling in breast cancer diagnosis. Taking into account the important role that miRNAs play in tumorigenesis and their ability to classify human tumors accurately (Lu et al., 2005), the first objective of this thesis was to identify miRNAs associated with breast tumors and with the main molecular subtypes of breast cancer, and to explore their discriminatory potential.

1.1. miRNA expression profiles in breast tumors and normal breast tissues

We explored the expression of 1919 human miRNAs in 122 primary breast tumors (31 triple negative, 27 Her2, 33 luminal B and 31 luminal A) and 11 normal breast tissues by using LNA based microarrays. After filtering the data to remove miRNAs with low expression variation across samples (VAR<0.03), we obtained 698 miRNAs for further analysis. Unsupervised hierarchical clustering (**Figure 8**) showed that miRNA expression profiling clearly separated breast tumors from normal breast tissues, although a perfect stratification was not observed according to the molecular subtype of the samples. While most triple negative and luminal B tumors clustered as two homogenous groups, luminal A and Her2 tumors were dispersed across the cluster exhibiting heterogeneous miRNA expression profiles. The right branch of the cluster, in which normal breast tissues were included, consisted of a higher proportion of ER-positive and low-moderate grade tumors, while the left branch was enriched in ER-negative and high grade tumors.

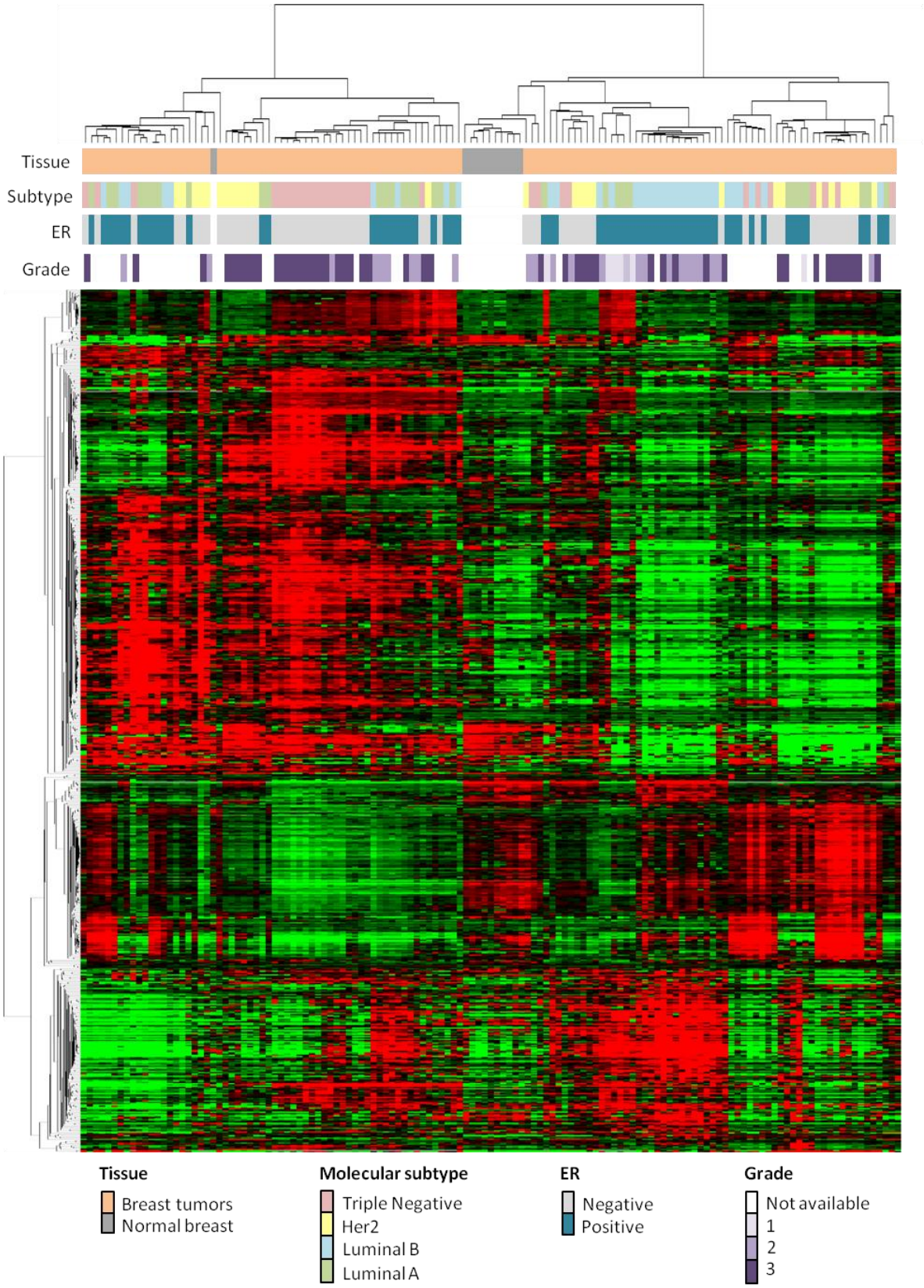


Figure 8. Unsupervised hierarchical clustering (Pearson correlation, average linkage clustering) over 698 miRNAs in 133 breast samples. Each column represents a breast sample and each row corresponds to a miRNA. Heatmap colors indicate relative miRNA expression: over-expression is represented in red and repression in green. Molecular subtype of the tumors, inmunohistochemical marker for ER and histologic grade are represented by color labels.

1.2. Identification of miRNAs specifically and commonly deregulated in breast cancer molecular subtypes

In order to detect significantly deregulated miRNAs in breast tumors and breast cancer molecular subtypes, supervised analysis was performed. A total of 194 miRNAs showed significant differential expression (FDR<0.05) between breast tumors and normal breast tissues: 117 were upregulated in breast tumors while 77 were downregulated. Eleven of these differentially expressed miRNAs had a fold change (FC)≥2 (**Table 6**).

Table 6. Top 11 miRNAs with the greatest difference in expression between breast tumors and normal breast tissues (FDR<0.05, FC≥2).

Gene Name	FDR	Fold Change	Median Normal Breast	Median Breast Tumors
miR-125b-5p	1E-07	4.1 ↓	10.9	8.6
miR-21-5p	9.07E-05	3.2 ↑	6.5	8.3
miR-3613-3p	7.16E-05	3.1 ↑	7.3	8.7
miR-4668-5p	0.000194	2.8 ↑	6.6	8.2
miR-4516	1.77E-05	2.5 ↓	11.0	9.7
miR-548as-3p	0.000167	2.4 ↑	6.4	7.7
miR-3656	2.00E-07	2.2 ↓	9.4	8.4
miR-4488	1.80E-06	2.2 ↓	8.8	7.7
miR-5704	5.04E-05	2.2 ↑	7.3	8.3
miR-141-3p	0.003198	2.0 ↑	7.3	8.0
miR-638	0.000536	2.0 ↓	8.7	7.9

FDR: False Discovery Rate adjusted p-value.

Comparison of the miRNA expression profile of each molecular subtype with the normal breast tissues led to the identification of 335, 98, 157 and 249 differentially expressed miRNAs in triple negative, Her2, luminal B and luminal A tumors, respectively (**Figure 9A**). In order to detect miRNAs specifically deregulated in each molecular subtype, we compared these results by using a Venn diagram. We identified 105, 1, 39 and 17 miRNAs specific for triple negative, Her2, luminal B and luminal A tumors, respectively, as well as 52 miRNAs commonly deregulated in the four molecular subtypes (**Figure 9B and Supplementary Table S1**).

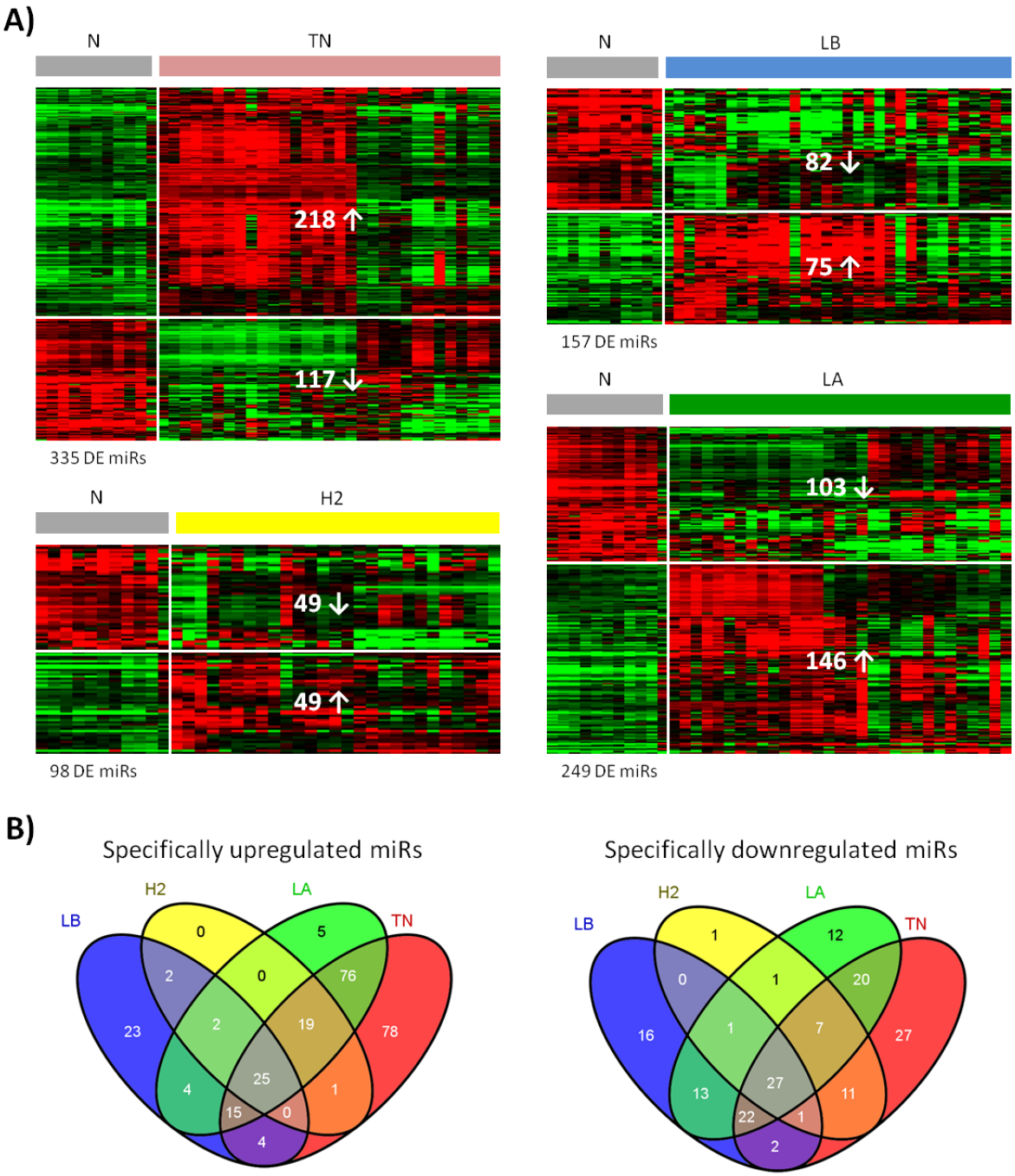


Figure 9. Deregulated miRNAs in breast cancer molecular subtypes. **A)** miRNAs differentially expressed (FDR<0.05) between normal breast tissues and each molecular subtype. **B)** Venn diagram showing miRNAs specifically up or downregulated in each molecular subtype, and those commonly deregulated in all the subtypes. TN: triple negative, H2: Her2, LB: luminal B, LA: luminal A, N: normal breast tissues.

1.3. miRNA signatures for breast tumor and breast cancer molecular subtype prediction

To identify the smallest set of miRNAs discriminating breast tumors from normal breast tissues, and each molecular subtype from the rest of subtypes, five miRNA microarray classifiers were generated. Breast tissues were divided into two groups: a training set (61 breast tumors and 7 normal breast tissues) used for the discovery phase and a test set (61 breast tumors and 4 normal breast tissues) used for the validation phase. Both series comprised a similar number of samples from each molecular subtype. Clinicopathologic characteristics of patients included in each series are shown in **Table 7**.

The predictors were built with samples from the training set and the 698 miRNAs used in the differential expression analysis. For the identification of the most representative miRNAs we used correlation feature selection and employed 5-fold cross-validation repeated 10 times to estimate how accurately the predictive model will perform in an independent data set. We used different algorithms that have been shown to function well with microarray data: support vector machines (SVM), k-nearest neighbor (KNN) and Random Forest (RF), and we selected SVM because it showed the best performance for the five different situations, producing the minimal root median square error (RMSE) and maximal accuracy, Mathews correlation coefficient (MCC) and AUC (**Supplementary Table S2**). Using this algorithm, we generated a 25-miRNA signature for breast tumor prediction with 100% sensitivity and 83% specificity, and a 8, 7, 16 and 3-miRNA signature for triple negative, Her2, luminal B and luminal A prediction, respectively, with high sensitivity (>93%) and specificity (99%). The list of miRNAs that compose each signature is shown in **Table 8**. Some of the miRNAs identified were detected as specifically deregulated (in the case of molecular subtypes) or commonly deregulated (in the case of breast tumors) in the differential expression analysis.

Table 7. Patients' clinicopathologic characteristics at the time of breast cancer diagnosis.

Parameter	Training set (n=61)	Test set (n=61)
Age, years	n=45	n=45
Mean	62	59
Range	33-89	28-82
Tumor size	n=42	n=39
T1 (≤ 2cm)	23 (55%)	21 (54%)
T2 (2.1-5cm)	14 (33%)	15 (38%)
T3/T4 (>5cm)	5 (12%)	3 (8%)
Grade	n=39	n=38
I	4 (10%)	1 (3%)
II	10 (26%)	15 (39%)
III	25 (64%)	22 (58%)
Nodes	n=41	n=39
negative	16 (39%)	16 (41%)
positive	25 (61%)	23 (59%)
Stage	n=40	n=36
I	12 (30%)	11 (30%)
II	17 (43%)	15 (42%)
III	8 (20%)	9 (25%)
IV	3 (7%)	1 (3%)
ER	n=61	n=61
negative	28 (46%)	30 (49%)
positive	33 (54%)	31 (51%)
PR	n=61	n=61
negative	30 (49%)	32 (52%)
positive	31 (51%)	29 (48%)
KI-67	n=60	n=59
low (<14%)	15 (25%)	17 (29%)
intermediate-high (≥14%)	45 (75%)	42 (71%)
Her2	n=60	n=61
negative	45 (75%)	45 (74%)
positive	15 (25%)	16 (26%)
Subtype	n=61	n=61
Luminal A	16 (26%)	15 (25%)
Luminal B	17 (28%)	16 (26%)
Her2	13 (21%)	14 (23%)
Triple negative	15 (25%)	16 (26%)

Table 8. Performance of the 5 miRNA signatures for the prediction of breast tumors and breast cancer molecular subtypes.

Classifier	Dataset	Accur. (%)	Sensitiv. (%)	Specific. (%)	miRs (n)	miRNA signature
Breast tumors	Training 61 BT 7 NT	98	100	83	25	miR-125b-5p*,miR-3613-3p*, miR-4668-5p*,miR-3656*,miR-5704*,miR-3676-3p,miR-3196, miR-3941*,miR-585,miR-1264*, miR-200a-3p*,miR-1273g-3p, miR-5581-3p,miR-877-5p*,miR-96-5p*,miR-744-3p*,miR-2276*,miR-342-5p*,miR-760*,miR203*, miRPlus-A1086,miR-185-5p*,miR-20b-5p,miR-4521*,miR-4692
	Test 61 BT 4 NT	100	100	100		
Triple negative tumors	Training 15 TN 46 RT	98	95	99	8	miR-125b-5p,miR-126-3p*****,miR-214-3p*****,miR-29c-3p,miR-4290, miR-3149,miR-10b-5p*****,miR-5193*****
	Test 16 TN 45 RT	98	94	100		
Her2 tumors	Training 13 H2 48 RT	98	93	99	7	miR-5704,miR-3676-3p,miR-1264,miR-3195,miR-205-5p, miR-4536-3p,miR-4692
	Test 13 H2 48 RT	100	100	100		
Luminal B tumors	Training 17 LB 44 RT	97	93	99	16	miR-205-5p,miR-320a,miR-10a-5p,miR-130a-3p,miR-1273g-3p, miR-15b-5p*****,miR-4667-5p,miR-4633-5p,miR-4800-5p,miR-708-5p,miR-1321,miR-125b-1-3p, miR-374a-5p,miR-3912,miR-221-3p, miR-3161
	Test 16 LB 45 RT	95	94	95		
Luminal A tumors	Training 16 LA 45 RT	98	97	99	3	miR-3676-3p,miR-532-3p,miR-22-5p
	Test 15 LA 46 RT	96	93	97		

*miRNAs detected as specifically deregulated (in the case of molecular subtypes) or commonly deregulated (in the case of breast tumors) in the previous analysis. BT: breast tumors, NT: normal breast tissues, TN: triple negative, H2: Her2, LB: Luminal B, LA: Luminal A, RT: Rest of tumors.

In order to validate the discriminatory potential of the 5 miRNA signatures, we classified an independent series of 65 samples (test set) in a blind approach. Patients in the test set were correctly identified with 100% sensitivity and 100% specificity using the breast tumor miRNA signature, 94% sensitivity and 100% specificity using the triple negative miRNA signature, 100% sensitivity and 100% specificity using the Her2 miRNA signature, 94% sensitivity and 95% specificity using the luminal B miRNA signature, and 93% sensitivity and 97% specificity using the luminal A miRNA signature (**Table 8**).

1.4. Pathway enrichment analysis

Given the fact that a single miRNA can target a large number of mRNA transcripts, aberrant expression of a set of miRNAs could have significant effect on cellular function by affecting multiple signaling pathways. To get more insight into the biological relevance of deregulated miRNA expression in breast cancer molecular subtypes, we used Diana miRPath web-based computational tool and investigated biological processes that are predicted to be targeted collectively by each of the 5 miRNA signatures identified. KEGG pathway enrichment analysis has revealed that the miRNAs that compose each signature are expected to regulate multiple pathways that are known to be relevant for cancer development and progression, such as PI3K-Akt, MAPK, Wnt, mTOR, p53, Notch, ErbB, VEGF, TGF-beta and HIF-1 signaling pathways. Top 10 statistically significant biological functions enriched in each miRNA signature are shown in **Table 9**. Interestingly, ErbB signaling pathway is one of the most significantly overrepresented pathways in the Her2 miRNA signature, indicating that the deregulation of miR-1264, miR-205-5p, miR-4536-3p and miR-4692 may contribute to the alteration of the ErbB signaling pathway in this group of tumors. Of note, some of the biological functions enriched in each miRNA signature are common to other cancers such as chronic myeloid leukemia, thyroid, prostate, endometrial, bladder or small cell lung cancer, suggesting potential similarities in the molecular mechanisms that operate in particular breast cancer molecular subtypes and different types of cancer.

Table 9. Top 10 significantly enriched signaling pathways associated with the different miRNA sigantures.

MiRNA signature	KEGG pathway	p-value	#genes	#miRNAs
Breast tumors	MAPK signaling pathway	2.56E-38	131	21
25-miR signature	Regulation of actin cytoskeleton	8.01E-31	108	19
	Ubiquitin mediated proteolysis	2.40E-29	82	16
	Axon guidance	1.58E-27	72	18
	Wnt signaling pathway	2.22E-25	80	21
	Neurotrophin signaling pathway	5.47E-25	69	18
	Gap junction	1.25E-21	55	18
	Endocytosis	7.46E-21	99	17
	ErbB signaling pathway	7.84E-21	53	17
	PI3K-Akt signaling pathway	9.03E-21	151	18
Triple negative	PI3K-Akt signaling pathway	9.54E-17	93	8
tumors 8-miR	Protein digestion and absorption	1.18E-14	33	6
signature	ECM-receptor interaction	1.18E-14	29	8
	Focal adhesion	2.88E-14	60	8
	Small cell lung cancer	1.14E-11	30	7
	MAPK signaling pathway	2.58E-08	66	7
	Insulin signaling pathway	2.58E-08	39	8
	Lysine degradation	7.43E-08	17	6
	mTOR signaling pathway	9.06E-08	22	8
	Prostate cancer	1.05E-07	27	6
HER2 tumors	Adherens junction	5.06E-12	22	3
7-miR signature	Thyroid cancer	1.80E-11	12	3
	Prostate cancer	1.01E-10	22	3
	Endometrial cancer	1.97E-09	15	3
	Bladder cancer	1.56E-06	12	3
	Wnt signaling pathway	1.71E-06	28	5
	Glioma	2.44E-06	16	4
	ErbB signaling pathway	6.28E-06	16	4
	Endocrine and other factor-regulated calcium reabsorption	1.92E-05	14	3
	Aldosterone-regulated sodium reabsorption	3.99E-05	10	3
Luminal B tumors	TGF-beta signaling pathway	4.08E-38	43	12
16-miR signature	Prostate cancer	8.62E-25	42	12
	mTOR signaling pathway	5.26E-24	33	12
	Pathways in cancer	5.43E-19	115	15
	Chronic myeloid leukemia	7.96E-18	34	11
	Endocytosis	5.76E-16	75	13
	Wnt signaling pathway	4.51E-14	60	13
	PI3K-Akt signaling pathway	4.99E-14	108	13
	Focal adhesion	7.26E-13	70	14
	ABC transporters	1.26E-12	21	9

Luminal A tumors 3-miR signature	Mucin type O-Glycan biosynthesis	0.009179	4	2
	Valine, leucine and isoleucine biosynthesis	0.019153	1	1
	Fanconi anemia pathway	0.019153	6	2
	MAPK signaling pathway	0.019153	19	2
	Endocytosis	0.019153	16	2
	Focal adhesion	0.019153	16	2
	Fc gamma R-mediated phagocytosis	0.019153	9	2
	Pentose phosphate pathway	0.032454	3	2
	PI3K-Akt signaling pathway	0.032454	22	2
	Fc epsilon RI signaling pathway	0.032454	7	2

2. CIRCULATING MICRORNAs IN EARLY BREAST CANCER DETECTION

Circulating miRNAs are attracting the attention of researchers as they are highly stable, resistant to degradation and can be easily obtained by non-invasive procedures (Chen et al., 2008). Although it is unclear how these extracellular miRNAs are liberated into plasma, there is evidence that some miRNAs are selectively released from malignant mammary epithelial cells while retained by non-malignant cells (Pigati et al., 2010). Taking into account the potential clinical relevance of circulating miRNAs in breast cancer, the second objective of this thesis was to study in plasma the status of the most deregulated miRNAs identified in breast tumors and to analyze their utility as non-invasive biomarkers for early breast cancer detection.

2.1. Validation of the most relevant miRNAs in breast tumors by qRT-PCR

We selected miRNAs with the smallest FDR and highest FC when comparing breast tumors and normal breast tissues and tried to validate them by qRT-PCR in 44 tumors from the test set (11 from each molecular subtype) and 12 normal breast tissues. A total of 19 miRNAs were selected: 10 from the breast tumor classifier and 9 from the differential expression analysis. Unfortunately, 8 were discarded due to unspecific amplification or no amplification (miR-3613-3p, miR-4668-5p, miR-5704, miR-1264, miR-5581-3p, miR-548as-3p, miR-3686, miR-4419b), and therefore 11 could be analyzed: 5 from the breast tumor classifier and 6 from the differential expression analysis. MiR-103a-3p was used for data normalization as it is a widely-used endogenous control for miRNA qRT-PCR and was stably expressed among our samples. Statistical analysis led to the validation of all of them (p-value<0.05) except miR-1273g-3p. Hence miR-183-3p, miR-96-5p, miR-142-3p, miR-141-3p, miR-21-5p and miR-200a-3p were confirmed to be significantly upregulated in breast tumors whereas miR-125b-5p, miR-3656, miR-638 and miR-505-5p were confirmed to be downregulated (**Figure 10**).

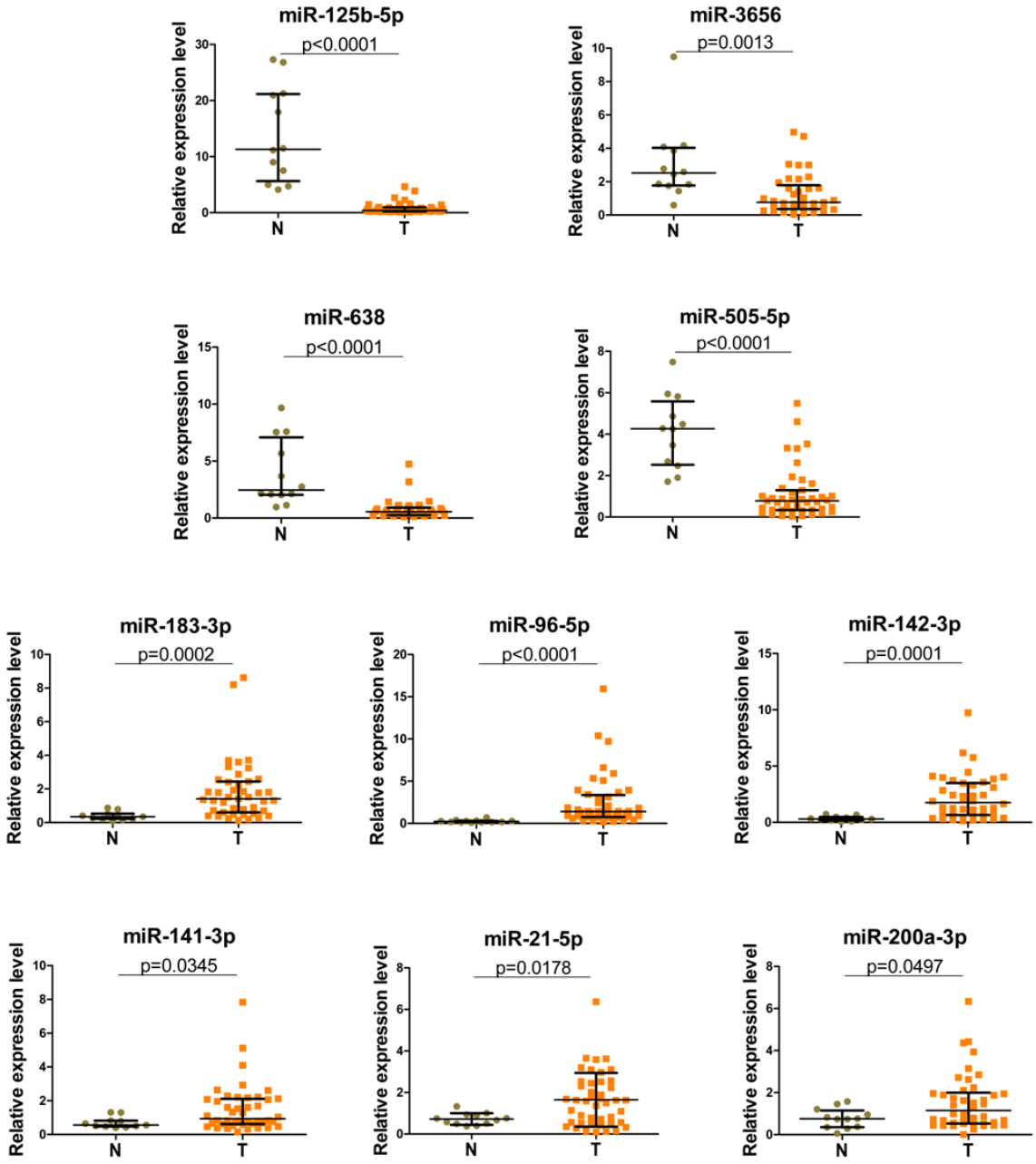


Figure 10. Validation of most relevant miRNAs by qRT-PCR in breast tumors and normal breast tissues. Scatter plots show relative expression levels of miR-183-3p, miR-96-5p, miR-142-3p, miR-141-3p, miR-21-5p, miR-200a-3p, miR-125b-5p, miR-3656, miR-638 and miR-505-5p in 44 breast tumors (T) and 12 normal breast tissues (N).

2.2. From tumor to plasma: analysis of circulating miRNAs

In order to evaluate the expression of our selected miRNAs in plasma, amplification by qRT-PCR was performed in 26 healthy individuals, 36 pretreated and 47 postreated breast cancer patients. Clinicopathologic details of patients included in each series are shown in **Table 10**. Of note, most of the patients had an early-stage breast cancer. Among 10 miRNAs analyzed, 9 had detectable levels of expression in plasma and only miR-183-3p could not be detected. Comparison of the expression levels in pretreated breast cancer patients and normal individuals led to the identification of 5 differentially expressed miRNAs in plasma: miR-3656, miR-505-5p, miR-125b-5p, miR-21-5p and miR-142-3p. In addition, although not significant, miR-96-5p showed a trend (p-value<0.1). Interestingly, some of the miRNAs analyzed were deregulated in opposite directions when compared with tumors. That is, miR-21-5p, miR-142-3p and miR-96-5p were both overexpressed in breast tumors and plasma from breast cancer patients, whereas miR-3656, miR-505-5p and miR-125b-5p were downregulated in breast tumors but upregulated in plasma from breast cancer patients (**Figure 11**). A selective release of certain miRNAs from tumors to plasma might be the explanation for this discordance.

Next we compared the expression levels of the 5 significant miRNAs and the one with p-value<0.1 in pretreated breast cancer patients and patients who underwent surgery and/or systemic therapy. While miR-125b-5p, miR-142-3p and miR-96-5p did not show a significant change in their levels, the expression of miR-3656, miR-21-5p and miR-505-5p was significantly reduced in plasma after treatment, suggesting that these miRNAs may be sensitive to changes in tumor mass (**Figure 11**).

Table 10. Patients' clinicopathologic characteristics at the time of breast cancer diagnosis.

Parameter	Plasma Pretretaed - Discovery (n=36)	Plasma Postreated - Discovery (n=47)	Plasma Pretreated - Validation (n=114)
Age, years	n=35	n=46	n=114
Mean	57	50	57
Range	33-82	28-74	28-85
Tumor size	n=35	n=33	n=99
T1 (≤2cm)	26 (74%)	20 (61%)	69 (70%)
T2 (2.1-5cm)	8 (23%)	13 (39%)	30 (30%)
T3/T4 (>5cm)	1 (3%)	0	0
Grade	n=33	n=34	n=102
I	7 (22%)	4 (12%)	16 (16%)
II	13 (39%)	10 (29%)	48 (47%)
III	13 (39%)	20 (59%)	38 (37%)
Nodes	n=34	n=42	n=107
negative	20 (59%)	19 (45%)	63 (59%)
positive	14 (41%)	23 (55%)	44 (41%)
Stage	n=34	n=32	n=96
I	18 (53%)	10 (32%)	36 (38%)
II	12 (35%)	19 (59%)	48 (50%)
III	3 (9%)	3 (9%)	12 (12%)
IV	1 (3%)	0	0
ER	n=36	n=42	n=113
negative	10 (28%)	16 (38%)	28 (25%)
positive	26 (72%)	26 (62%)	85 (75%)
PR	n=34	n=42	n=113
negative	9 (26%)	19 (45%)	35 (31%)
positive	25 (74%)	23 (55%)	78 (69%)
KI-67	n=34	n=35	n=113
low (≤14%)	10 (29%)	9 (26%)	31 (27%)
intermediate-high (>14%)	24 (71%)	26 (74%)	82 (73%)
Her2	n=34	n=39	n=112
negative	26 (76%)	30 (77%)	78 (70%)
positive	8 (24%)	9 (23%)	34 (30%)
Subtype	n=36	n=36	n=113
Luminal A	10 (28%)	9 (25%)	33 (29%)
Luminal B	16 (44%)	14 (39%)	52 (46%)
Her2	4 (11%)	5 (14%)	9 (8%)
Triple negative	6 (17%)	8 (22%)	19 (17%)

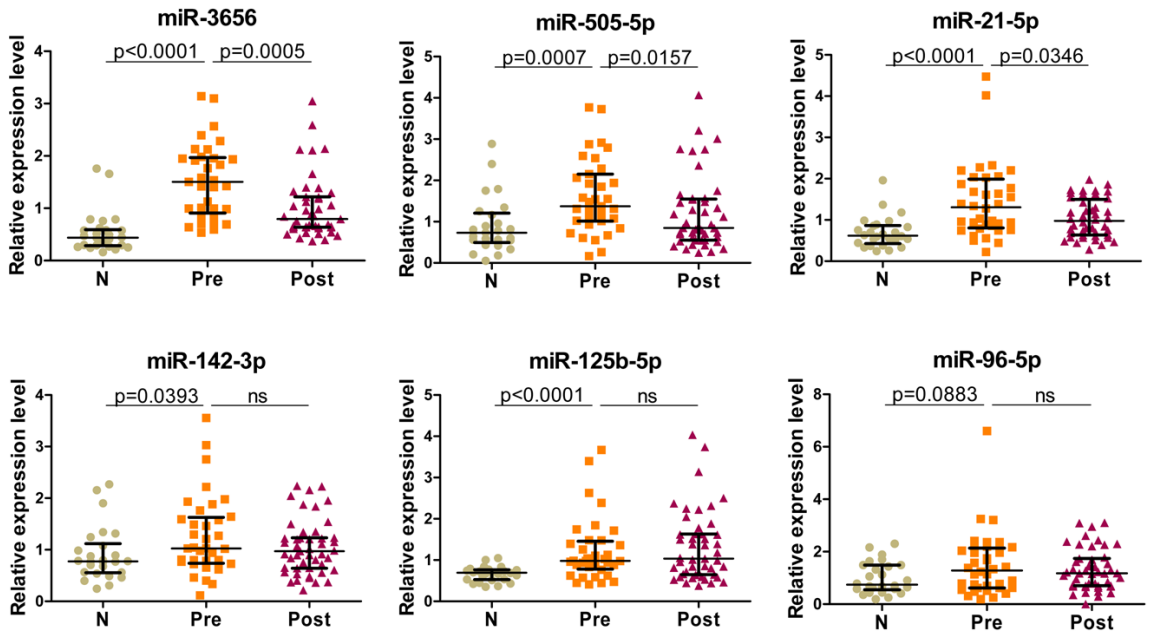


Figure 11. Relative expression levels of circulating miRNAs deregulated in the plasma of 36 pretreated breast cancer patients (Pre) in comparison with 26 healthy individuals (N). Although not significant, miR-96-5p showed a trend (p-value<0.1). In addition, miRNA plasma concentration was quantified in 47 postreated breast cancer patients (Post).

2.3. Marker validation

The 5 significant miRNAs identified were then subjected to validation in a second set of plasma from 114 pretreated breast cancer patients and 116 healthy women. Since miR-96-5p showed some evidence in the previous analysis, we decided to include this miRNA in the validation stage as well. **Table 10** summarizes the clinicopathologic features of the patients, showing again a high proportion of early-stage breast cancers. The expression levels of miR-505-5p, miR-125b-5p, miR-21-5p and miR-96-5p were confirmed to be significantly overexpressed in the plasma of breast cancer patients (**Figure 12A**). ROC curve analysis was performed to evaluate the diagnostic utility of these miRNAs. The resultant curves showed that miR-505-5p and miR-96-5p were the most valuable biomarkers for discriminating patients from healthy individuals, with AUC of 0.7213 (95% CI: 0.6558 to 0.7867, p<0.0001) and 0.7167 (95% CI: 0.6507 to 0.7827, p<0.0001), respectively, and sensitivity and specificity at the optimal cutoff of 75% and 60% for miR-505 and 73% and 66% for miR-96-5p, respectively. MiR-125b-5p and miR-21-5p showed AUC of 0.6368 (95% CI: 0.5642 to 0.7093, p<0.03699) and 0.6070 (95% CI: 0.5336 to 0.6803, p<0.03742), respectively (**Figure 12B**).

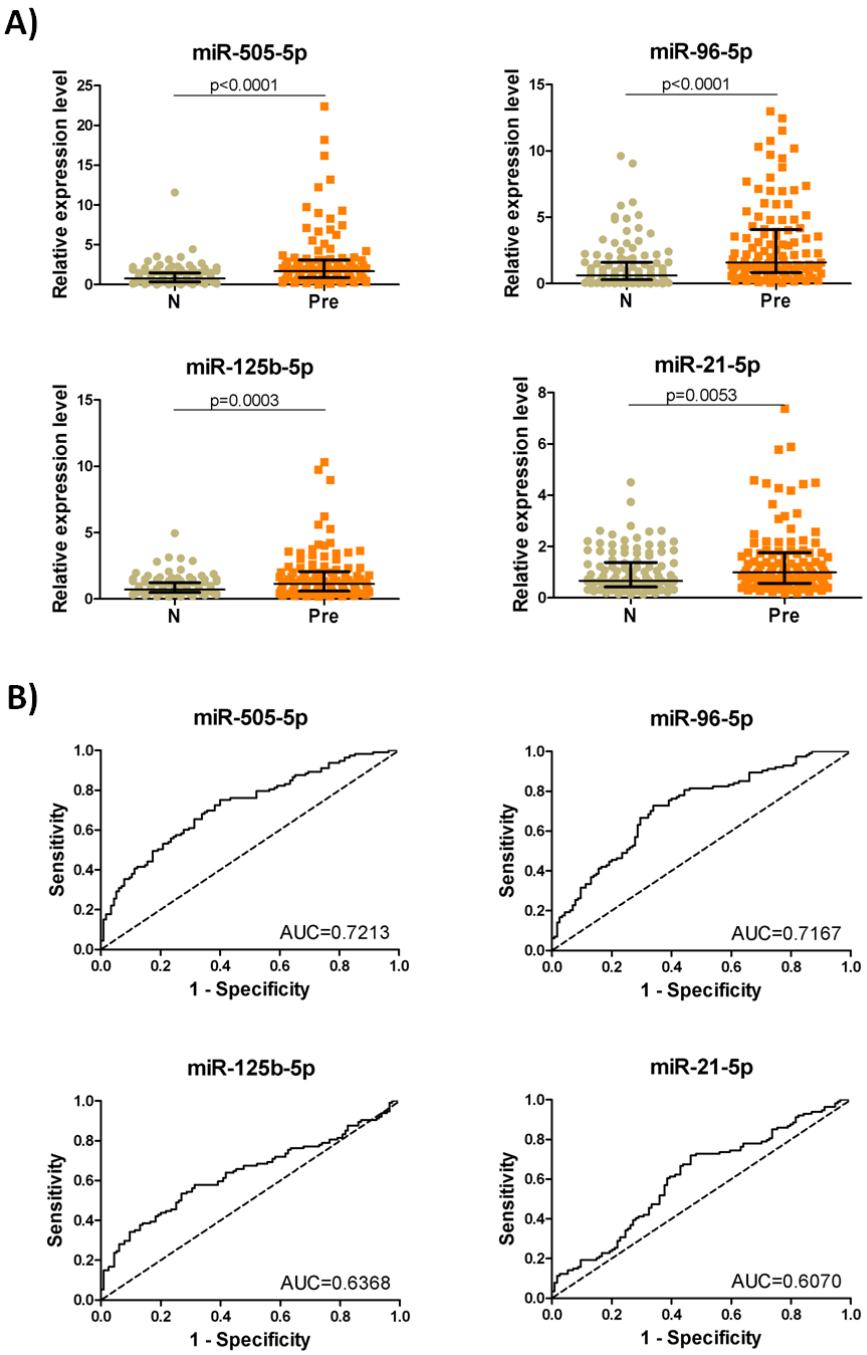


Figure 12. Circulating miRNAs validated as being overexpressed in the plasma of 114 breast cancer patients (Pre) when compared with 116 healthy women (N). **A)** Relative expression levels of miR-505-5p, miR-96-5p, miR-125b-5p and miR-21-5p with their corresponding p-values. **B)** ROC curves for each miRNA and the resultant area under the curve (AUC).

2.4. Association between circulating miRNA expression and clinicopathologic characteristics

We further compared the expression of circulating miR-505-5p, miR-125b-5p, miR-21-5p and miR-96-5p with clinicopathologic characteristics in the 114 pretreated breast cancer patients. The variables evaluated were: age of the patient at diagnosis, tumor size, histologic grade, lymph node, ER, PR and Her2 status, ki-67 levels and molecular subtype. Non-parametric Mann–Whitney test (for two groups' comparison) and Kruskal Wallis test (for multiple groups' comparison) were applied. Significant higher levels of miR-505-5p were observed in older patients, lower grade and ER positive tumors. Significant differences were also observed among molecular subtypes, with luminal A tumors with the highest levels of miR-505-5p (**Figure 13**). No significant associations between circulating miR-125b-5p, miR-21-5p and miR-96-5p and clinicopathologic features were detected.

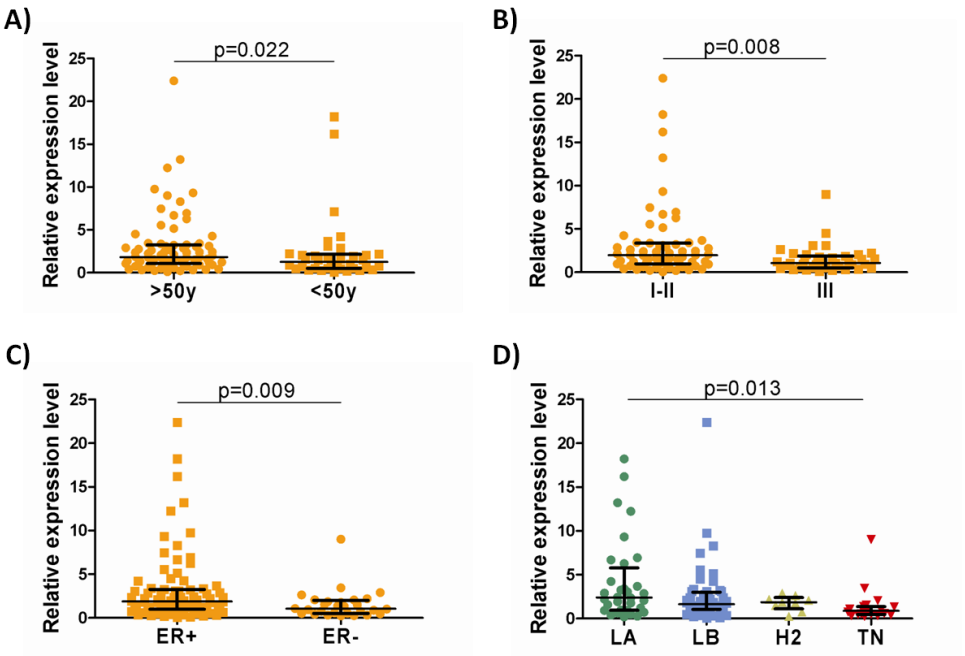


Figure 13. Association between circulating miR-505-5p expression and different clinicopathologic characteristics in 114 pretreated breast cancer patients. **A)** age at diagnosis, **B)** histologic grade, **C)** ER status, and **D)** molecular subtype.

3. PROGNOSTIC MICRORNAs IN TRIPLE NEGATIVE BREAST CANCER PATIENTS

Triple negative breast cancer (TNBC) is a very aggressive form of breast cancer with higher recurrence rates and greater likelihood of death compared to other breast cancer subtypes (Dent et al., 2007). Moreover, the lack of response to endocrine and anti-Her2 therapies makes its treatment an extremely challenging process. Taking into account the absence of prognostic and predictive markers in this subtype of breast cancer, the third objective of this thesis was to identify miRNAs that can distinguish at the time of diagnosis between high and low risk TNBC patients. This subclassification would permit the administration of different treatments to patients with different clinical outcomes.

3.1. miRNAs diferentially expressed in node-positive triple negative patients with different outcome

Since nodal status is known to be an independent prognosis factor for breast cancer, we decided to explore if this factor was also associated with survival in our series of 21 triple negative tumors with available follow-up information. As expected, nodal status divided the patients in two groups: those with good prognosis (node-negative, 0.8 cumulative proportion of surviving) and those with worse prognosis (node-positive, 0.4 cumulative proportion of surviving) (**Figure 14A**). Since node-positive triple negative tumors were associated with a more variable outcome, we decided to focus on this group of patients to identify miRNAs linked to prognosis.

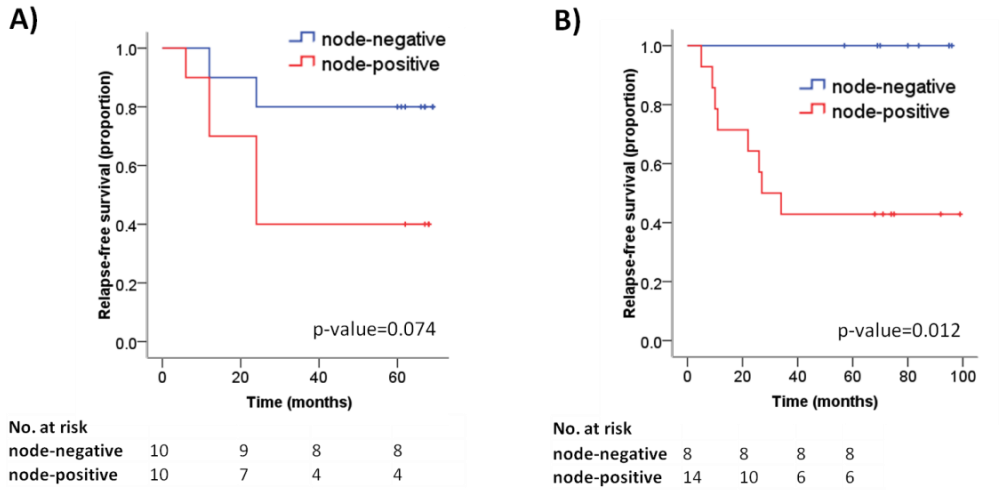


Figure 14. Kaplan-Meier survival curves of two different cohorts of triple negative patients based on nodal status: **A)** 20 patients used for microarray and qRT-PCR expression analysis, and **B)** 22 patients used for qRT-PCR expression analysis. Patients at risk at the indicated time intervals are shown at the bottom of the graph.

Follow-up studies have demonstrated that the prognosis of TNBC patients is highly time dependent, with some patients experiencing disease recurrence in the first 3-5 years following diagnosis and other patients having excellent long term survival (Liedtke et al., 2008; Mulligan et al., 2008). In order to detect differentially expressed miRNAs between node-positive triple negative patients with RFS shorter than 5 years and patients with longer RFS, differential expression analysis was performed with the 698 miRNAs obtained after microarray data filtering. A total of 17 miRNAs showed significant differential expression (FDR<0.05) and all of them were downregulated in triple negative patients with worse prognosis (**Figure 15**).

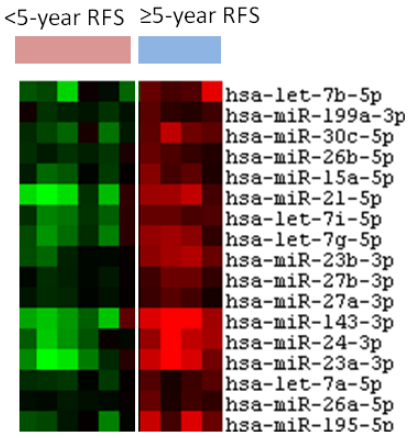


Figure 15. miRNAs differentially expressed (FDR<0.05) between 6 node-positive triple negative patients with RFS shorter than 5 years and 4 node-positive triple negative patients with longer RFS.

3.2. miRNAs associated with recurrence in node-positive triple negative patients

We performed survival analysis of node-positive triple negative patients to investigate associations between the expression levels of the 17 differentially expressed miRNAs and RFS. By using univariate Cox proportional hazards regression analysis we identified 4 miRNAs significantly related to outcome: let-7b-5p, miR-195-5p, miR-24-3p and miR-30c-5p (hazard ratio (HR): 8.957; 95% confidence interval (CI): 1.024-78.310, p=0.048) (**Figure 16**).

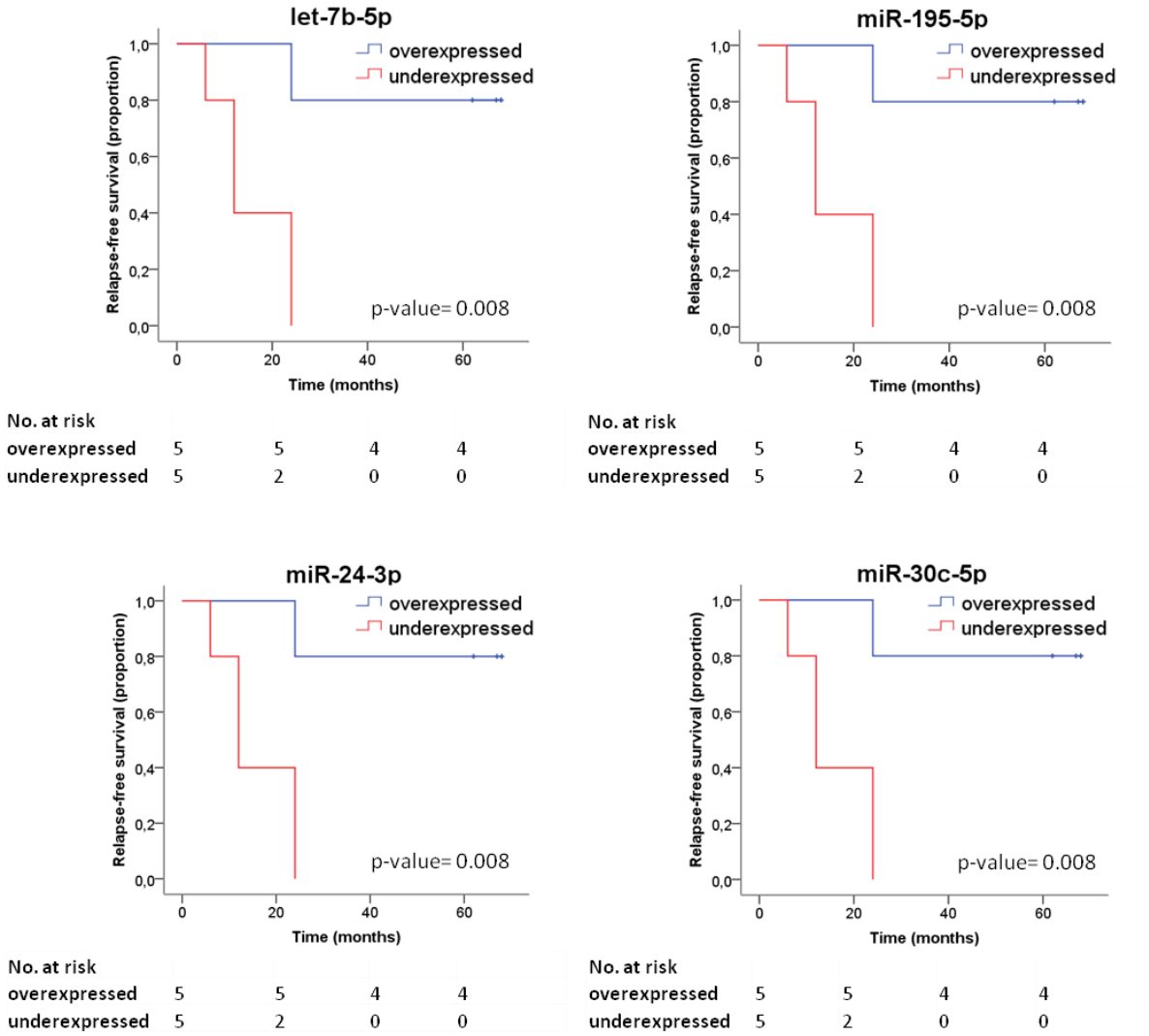


Figure 16. Kaplan-Meier survival curves of 10 node-positive triple negative patients based on the expression of let-7b-5p, miR-195-5p, miR-24-3p and miR-30c-5p. Patients at risk at the indicated time intervals are shown at the bottom of each graph.

3.3. Validation of prognostic miRNAs in triple negative patients

To validate the prognostic value of let-7b-5p, miR-195-5p, miR-24-3p and miR-30c-5p, we analyzed their expression levels by qRT-PCR in the same 21 triple negative patients plus a second cohort of 22 triple negative patients. In this second cohort we could also observe a strong relationship between nodal status and clinical outcome, with all node-negative patients surviving after 5 years of diagnosis and node-positive patients having 40% probability of relapse-free survival (**Figure 14B**).

Univariate Cox proportional hazards regression analysis confirmed that miR-30c-5p was significantly associated with RFS in both node-positive (HR: 4.101; 95% CI: 1.264-13.299, p=0.019) and all triple negative patients (HR: 4.157; 95% CI: 1.350-12.800, p=0.013) (**Figure 17A,B**). Likewise, miR-195-5p was significantly related to outcome in all triple negative patients (HR: 3.338; 95% CI: 1.171-9.516, p=0.024) and almost significant in node-positive triple negative patients (HR: 2.717; 95% CI: 0.898-8.217, p=0.077) (**Figure 17C,D**). In addition, the combination of miR-30c-5p and miR-195-5p showed a significant association with RFS, with patients having both miRNAs overexpressed experiencing a significantly better outcome than patients having both of them underexpressed (**Figure 17E,F**).

Analysis of histopathological variables (tumors size, age at diagnosis, Ki-67 expression levels and nodal status) confirmed the association of tumor size (HR: 1.379; 95% CI: 1.154-1.648, p<0.001) and nodal status (HR: 4.700; 95% CI: 1.347-16.399, p=0.015) with RFS. Multivariate Cox proportional hazards regression analysis including tumor size and nodal status revealed an independent association of miR-195-5p with RFS in both node-positive and all triple negative patients (**Table 11**). However, after adjustment for these cofactors, miR-30c-5p lost its association in node-positive patients and showed weak evidence in all triple negative patients, indicating that expression of this miRNA may be correlated with tumor size. In fact, we found that triple negative tumors expressing low levels of miR-30c-5p were significantly bigger than tumors expressing high levels (**Supplementary Figure S1**).

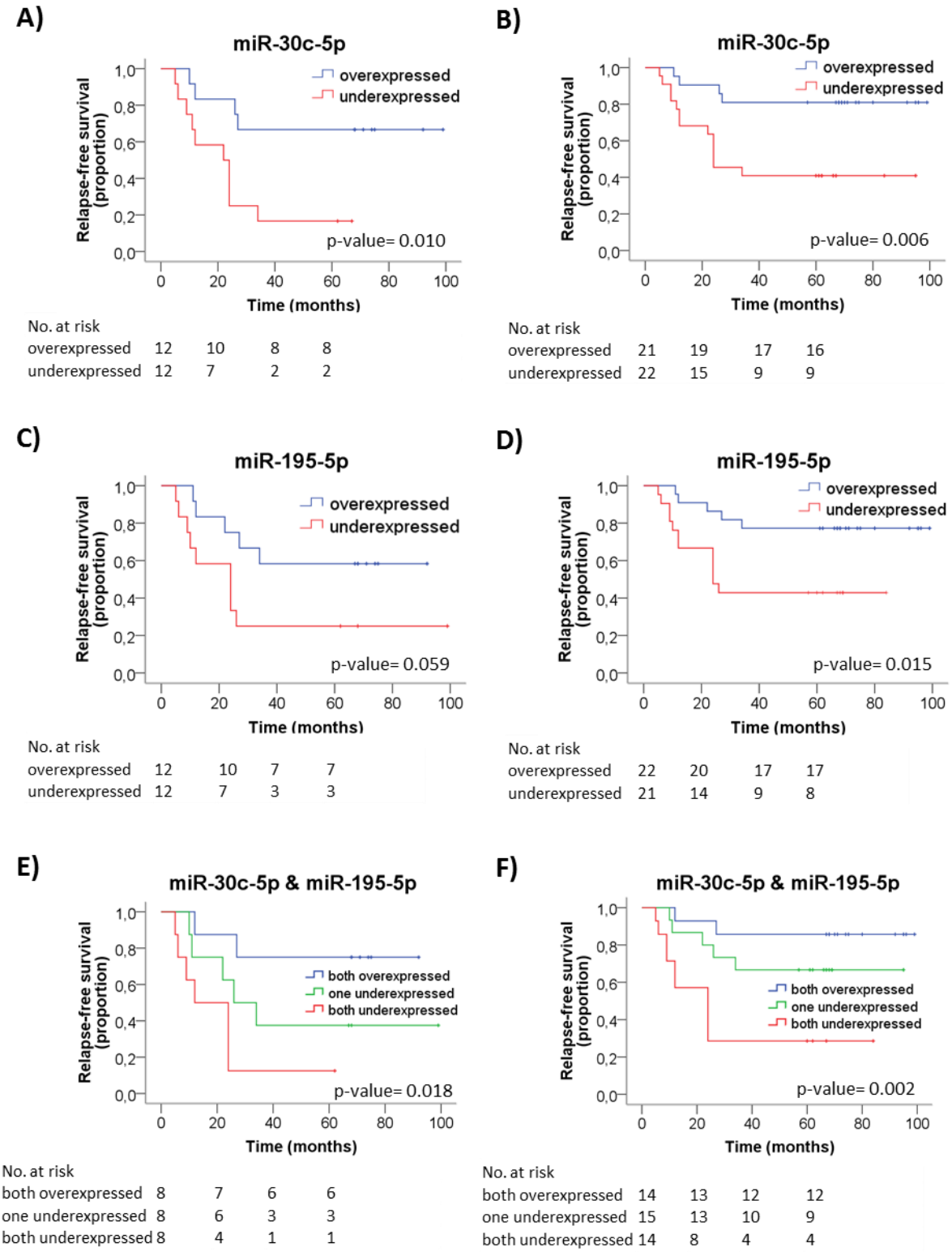


Figure 17. Kaplan-Meier survival curves of 24 node-positive triple negative patients (**A, C, E**) and the whole cohort of 43 triple negative patients (**B, D, F**) based on the expression of miR-30c-5p, miR-195-5p and the combination of both. Patients at risk at the indicated time intervals are shown at the bottom of each graph.

Table 11. Multivariate Cox regression analysis of prognostic factors for RFS.

Patients	Comparison	p-value	HR (95% CI)	p(adj)
MiR-30c-5p				
Node-positive TN	MiRNA overexpression vs underexpression	0.010	2.649 (0.720-9.749)	0.143
	Tumor size		1.251 (1.013-1.545)	0.037
All TN	MiRNA overexpression vs underexpression	0.006	3.322 (0.964-11.447)	0.057
	Tumor size		1.210 (0.990-1.479)	0.063
	Nodal status		6.466 (1.438-29.076)	0.015
MiR-195-5p				
Node-positive TN	MiRNA overexpression vs underexpression	0.059	7.924 (2.149-29.223)	0.002
	Tumor size		1.523 (1.201-1.931)	0.001
All TN	MiRNA overexpression vs underexpression	0.015	6.788 (2.101-21.931)	0.001
	Tumor size		1.436 (1.165-1.770)	0.001
	Nodal status		10.459 (2.178-50.228)	0.003

P-values in the univariate analysis calculated with log-rank test; p(adj) as calculated in the multivariate analysis by using Cox regression model in node-positive triple negative patients and the whole series of triple negative patients with RFS as endpoint; TN, triple negative; HR, hazard ratio; CI, confidence interval.

3.4. Pathway enrichment analysis

To get more insight into the role of miR-30c-5p and miR-195-5p in TNBC, we used Diana miRPath web-based computational tool and investigated biological processes that are predicted to be targeted collectively by both miRNAs. KEGG pathway enrichment analysis has revealed that miR-30c-5p and miR-195-5p target different effectors of pathways involved in cell cycle, proliferation, angiogenesis, apoptosis and cell survival (**Table 12**). Hence, downregulation of miR-30c-5p and miR-195-5p in triple negative breast tumors could confer them enhanced proliferative, angiogenic and invasive potentials.

Table 12. Top 10 significantly enriched signaling pathways associated with miR-30c-5p and miR-195-5p.

KEGG pathway	p-value	Genes
Ubiquitin mediated proteolysis	4,78E-14	UBE2R2,BTRC,RFWD2,FBXW7,TRIM37,WWP1,CUL2, SMURF2,NEDD4L,UBE4A,HERC3,UBE2J1,UBE2I,SOCS3, CBLB,SKP2,SOCS1,UBE3C,BIRC6,SMURF1,UBE2Q1, RCHY1,UBE2K,SIAH1,NEDD4,UBE2F,UBE2D1,CDC27, UBE2D2,UBE2G1,CUL3,PPIL2,UBE4B
B cell receptor signaling pathway	3,24E-09	SOS2,RAF1,BCL10,KRAS,IKBKB,PPP3CA,NFAT5,PIK3CD, PPP3CB,NFATC2,PIK3R1,SOS1,DAPP1,AKT3,MAP2K1, RASGRP3,VAV3,NFATC3,LYN
Prostate cancer	1,05E-08	SOS2,RAF1,BCL2,IGF1R,KRAS,IKBKB,AR,PIK3CD,CCND1, CCNE2,E2F3,PIK3R1,SOS1,AKT3,PDGFC,CCNE1,MAP2K1, FGFR1,FOXO1,PDGFA
PI3K-Akt signaling pathway	1,08E-08	PRLR,YWHAH,MYB,SOS2,ITGA9,ITGA8,COL24A1,CCND2, RAF1,BCL2,EFNA3,PPP2R2B,IGF1R,FGF20,PPP2R5C, KRAS,CDK6,IFNAR2,GHR,IKBKB,GNG10,DDIT4,PIK3CD, CCND1,EIF4E,CCNE2,PPP2R1A,EIF4B,PIK3R1,SOS1, IL2RA,YWHAZ,IRS1,INSR,FGF2,FGF18,AKT3,PDGFC, CCNE1,FOXO3,MAP2K1,ITGA4,ITGA6,VEGFA,FGFR1, PPP2R1B,FGF7,KDR,CSF1,BCL2L11,CCND3,PDGFA
Neurotrophin signaling pathway	1,08E-08	CAMK2D,SOS2,SH2B3,CAMK4,NTRK3,FRS2,RAF1,BCL2, KRAS,IKBKB,PIK3CD,MAPK8,KIDINS220,ARHGDIA, PIK3R1,SOS1,IRS1,RPS6KA3,AKT3,FOXO3,MAP2K1, IRAK2,PRDM4,RAP1B,ABL1,MAP3K5
Long-term depression	1,14E-07	GNA12,GUCY1A3,CRHR1,GNA13,RAF1,IGF1R,GNAI3, KRAS,GRIA2,NOS1,PPP2R1A,GNAQ,MAP2K1,PRKG1, PPP2R1B,LYN
Melanoma	1,92E-07	RAF1,IGF1R,FGF20,KRAS,CDK6,PIK3CD,CCND1,E2F3, PIK3R1,FGF2,FGF18,AKT3,PDGFC,MAP2K1,FGFR1,FGF7, PDGFA
Non-small cell lung cancer	2,61E-07	SOS2,RAF1,RARB,KRAS,CDK6,PIK3CD,CCND1,E2F3, PIK3R1,SOS1,AKT3,FOXO3,MAP2K1
p53 signaling pathway	7,30E-07	ZMAT3,RFWD2,CCND2,CDK6,CHEK1,CASP3,CCND1, CCNE2,SHISA5,RCHY1,SESN1,SIAH1,TNFRSF10B,CCNE1, PPM1D,CCND3
Cell cycle	1,25E-06	ESPL1,YWHAH,SMAD2,CDC14A,ORC2,CCND2,ORC4, DBF4,STAG2,WEE1,CDK6,CHEK1,CCNA1,CCND1,CCNE2, SKP2,E2F3,CDC14B,YWHAZ,TFDP1,CCNE1,CDC27,ABL1, CCND3,CDC25A

4. MICRORNA REGULATION OF *BRCA1* GENE EXPRESSION IN SPORADIC TRIPLE NEGATIVE BREAST CANCER

Sporadic triple negative tumors share many characteristics with *BRCA1*-mutated breast tumors, which reveals a possible role of BRCA1 dysfunction in the pathogenesis of sporadic TNBC. Despite being *BRCA1*$^{+/+}$, most sporadic triple negative tumors have a reduced expression of the *BRCA1* gene (Mueller and Roskelley, 2003; Turner et al., 2007), suggesting that other mechanisms might be involved in *BRCA1* somatic inactivation. Considering the central role that miRNAs play in gene expression regulation, the last objective of this thesis was to investigate the involvement of miRNAs in *BRCA1* regulation in sporadic triple negative breast cancer.

4.1. Triple negative-specifc miRNAs predicted to target *BRCA1*

In an attempt to identify miRNAs regulating the *BRCA1* gene, five different target prediction algorithms (Miranda, Pita, TargetScan, Microtar and RNAhybrid) were used with the 78 miRNAs specifically upregulated in sporadic triple negative tumors identified in the first part of this thesis (**Supplementary Table S1**). We selected miR-498 and miR-187-5p since they were predicted to bind to the 3'UTR of the *BRCA1* gene with high scores by at least two prediction methods (**Table 13, Figure 18A**). Overexpression of these two miRNAs in sporadic triple negative tumors could lead to reduced levels of *BRCA1* expression.

Table 13. miRNAs specifically overexpressed in triple negative breast tumors that are predicted to target the 3'UTR of *BRCA1* by at least two prediction methods.

miRNA mature	Gene	Agreement	Miranda	Pita	Microtar	RNAHybrid	Targetscan	Mean value
hsa-miR-498	*BRCA1*	2	0.9006	-	-	-	1	0.9503
hsa-miR-187-5p	*BRCA1*	3	0.9123	-	0.0022	0.0900	-	0.3348

4.2. *BRCA1* is a target of miR-498 and miR-187-5p

We investigated whether the 3'UTR of *BRCA1* is a functional target of miR-498 and miR-187-5p by using a reporter vector into which the entire 3'UTR of *BRCA1* was inserted downstream of the firefly luciferase reporter gene. This reporter vector was transiently transfected into 293T cells together with pre-miR-498, pre-miR-187-5p, non-targeting control, positive control (pre-miR-146a-5p) or no miRNA precursor. Each experiment was performed in quadruplicate. An average of 35% and 50% reduction of reporter activity as compared to the mock transfection control was observed for miR-498 and miR-187-5p, respectively (**Figure 18B**), indicating that these miRNAs target *BRCA1* 3'UTR. Similar to previous studies, the degree of luciferase inhibition with miR-146a-5p reached 60%.

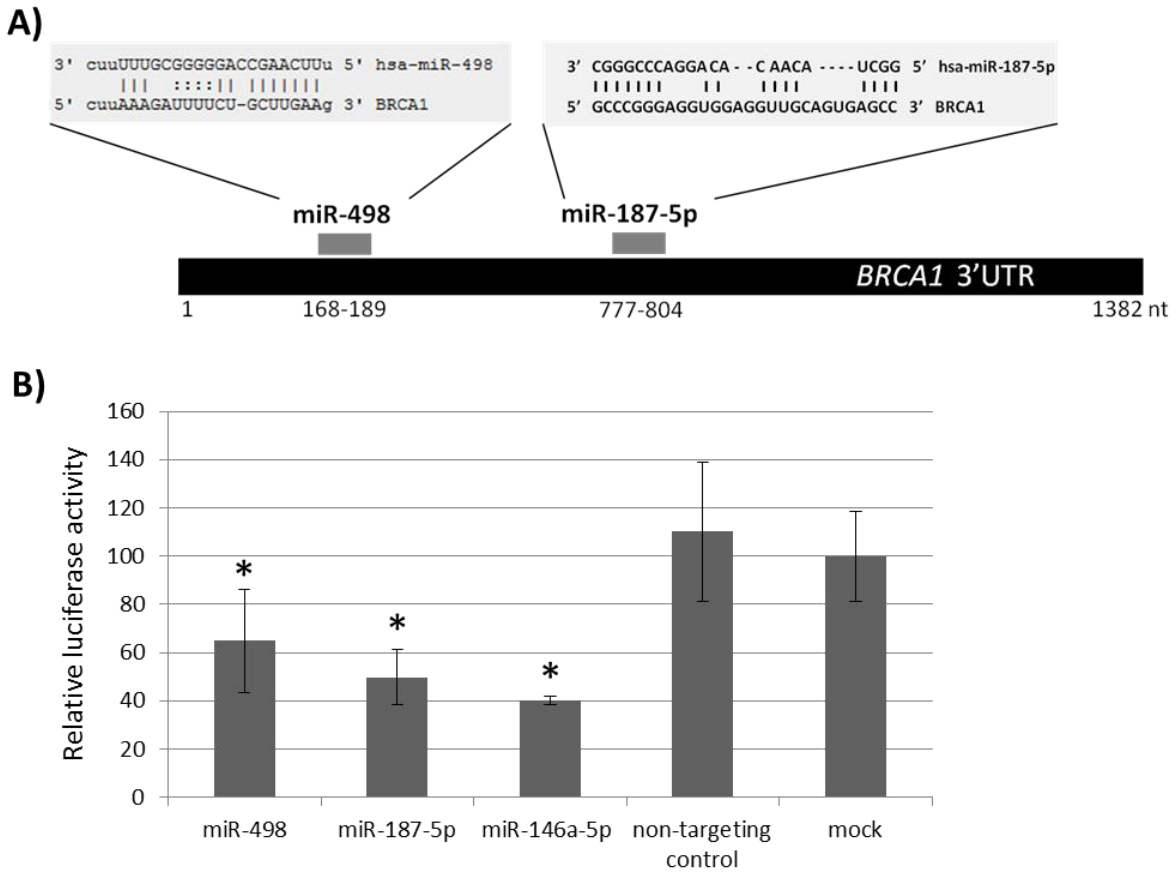


Figure 18. Negative regulation of *BRCA1* expression by miR-498 and miR-187-5p. **A)** Schematic representation of miRNA binding sites within the *BRCA1* 3'UTR. **B)** Relative luciferase activity of a reporter vector carrying the *BRCA1* 3'UTR downstream of the firefly luciferase gene. The vector was co-transfected with each of the indicated miRNA precursors or with no miRNA precursor (mock) into 293T cells. Error bars represent standard deviation for four replicates of one representative experiment. Data were normalized versus the luciferase levels generated by the mock transfection. * $p<0.05$.

4.3. MiR-498 and miR-187-5p expression in breast tumors and breast cancer cell lines of different subtypes

In the first part of this thesis, we found that miR-498 and miR-187-5p expression was increased in sporadic triple negative breast tumors but not in other subtypes when compared with normal breast tissues (**Figure 19A**). We next analyzed the expression of miR-498 and miR-187-5p in 5 different breast cancer cell lines, including two triple negative (MDA-MB-231 and HS578T), one Her2 (SKBR3), one luminal B (BT474) and one luminal A (MCF7) cell line. We found that miR-498 was expressed at high levels in a triple negative cell line (HS578T) while miR-187-5p was highly expressed in a luminal cell line (BT474). Regarding *BRCA1* expression, lower levels were found in the triple negative and the Her2 cell lines when compared with the luminal cell lines (**Figure 19B**). Since we were interested in miRNAs with increased expression levels in triple negative cell lines and negatively correlated with *BRCA1* expression levels, we decided to focus on miR-498 for following experiments.

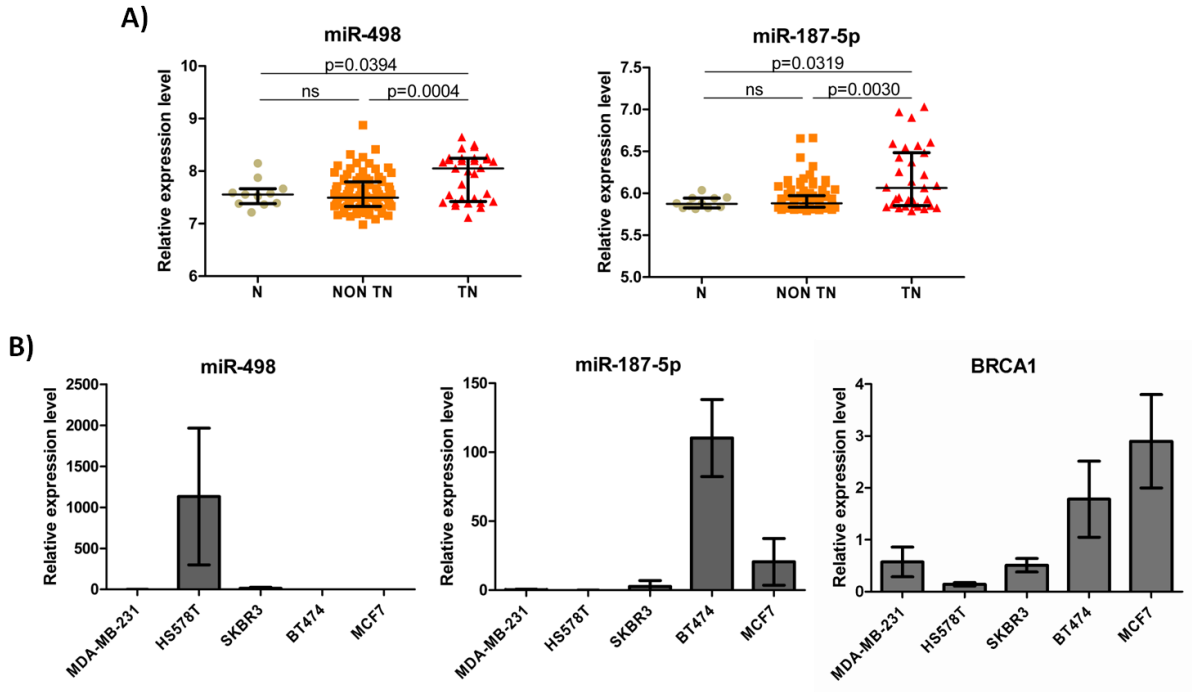


Figure 19. miR-498 and miR-187-5p expression levels in breast tumors and breast cancer cell lines of different subtypes. **A)** Relative expression of miR-498 and miR-187-5p in 11 normal breast tissues (N), 91 non triple negative tumors (non TN) and 31 triple negative tumors (TN). **B)** Relative expression of miR-498, miR-187-5p and *BRCA1* in two triple negative (MDA-MB-231 and HS578T), one Her2 (SKBR3), one luminal B (BT474) and one luminal A (MCF7) breast cancer cell lines. Error bars represent standard deviation for triplicates of one representative experiment.

4.4. MiR-498-mediated regulation of *BRCA1* in breast cancer cell lines

Since miR-498 was expressed at high levels in HS578T cells and at low levels in MCF7 cells, we next investigated the consequences of miR-498 inhibition in HS578T cell line and of miR-498 overexpression in MCF7 cell line. As expected, miR-498 inhibition led to an increase in the amount of BRCA1 (480% increase at the mRNA level and 164% increase at the protein level) while its overexpression produced a reduction of BRCA1 (38% decrease at the mRNA level), as compared with mock transfection (**Figure 20**). These results demonstrate that miR-498 regulates the expression of *BRCA1* in breast cancer cells.

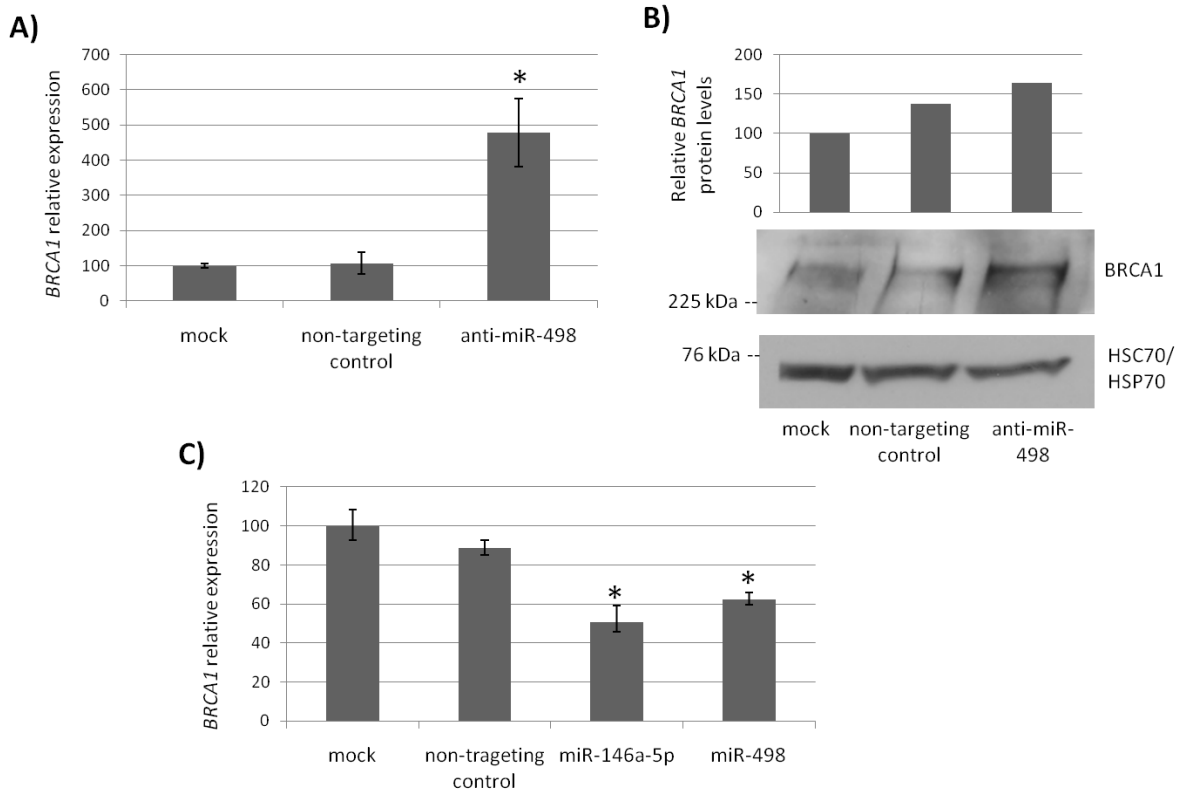


Figure 20. *BRCA1* expression levels after miR-498 inhibition or overexpression. **A)** Relative mRNA levels of *BRCA1* after transfection of HS578T cells with anti-miR-498, non-targeting control or no miRNA inhibitor. **B)** Western blot analysis of BRCA1 expression in HS578T cells after transfection with anti-miR-498, non-targeting control or no miRNA inhibitor. Full-length BRCA1 was detected using a monoclonal anti-BRCA1 antibody (Calbiochem, #OP92) and HSC70/HSP70 served as a loading control. **C)** Relative mRNA levels of *BRCA1* after transfection of MCF7 cells with pre-miR-498, pre-miR-146a-5p, non-targeting control or no miRNA precursor. Error bars represent standard deviation for triplicates of one representative experiment. * $p<0.05$.

4.5. Inhibition of miR-498 reduces proliferation in triple negative breast cancer cells

To gain more insight into the biological effect of miR-498 on breast tumorigenesis and given that BRCA1 is presumed to have a growth suppressor function (Holt et al., 1996; Thompson et al., 1995), we transfected HS578T cells, which previously showed elevated levels of miR-498, with anti-miR-498 or mock and analyzed the cell growth by WST-1 cell viability assay. **Figure 21** shows that inhibition of miR-498 resulted in reduced proliferation in comparison to mock transfected cells. These results indicate that down-regulation of *BRCA1* by miR-498 can promote proliferation and contribute to tumorigenesis.

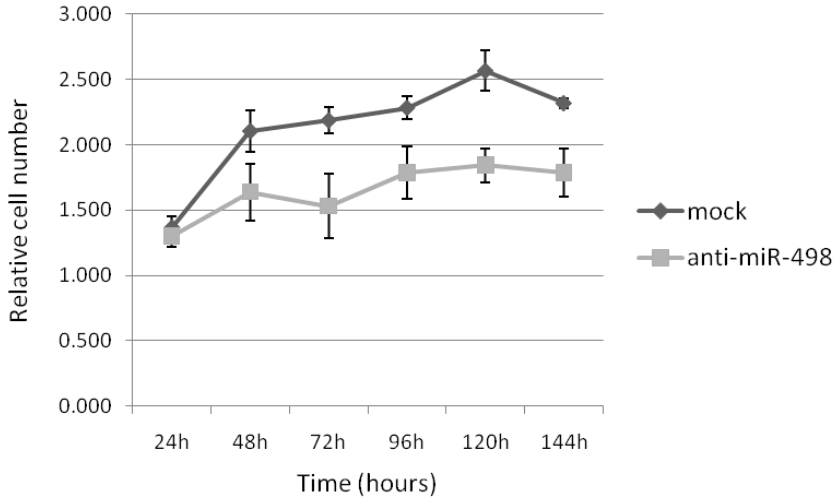
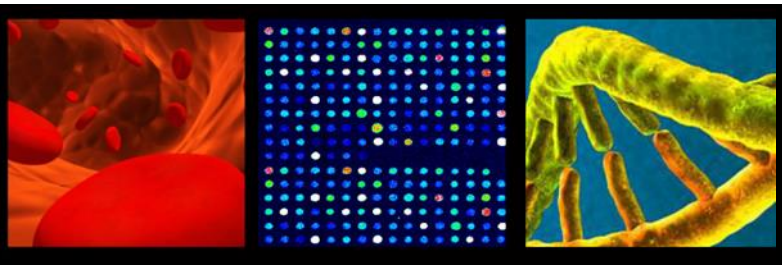


Figure 21. Effect of miR-498 inhibition on proliferation of HS578T cells. WST-1 cell viability assay was performed at 24, 48, 72, 96, 120 and 144 hours after transfection of HS578T cells with anti-miR-498 or mock transfected.



Discussion

1. MICRORNA DEREGULATION IN BREAST CANCER MOLECULAR SUBTYPES

Breast cancer is a highly heterogeneous disease. In the past few years, gene expression profiling has identified at least four major subtypes (luminal A, luminal B, Her2 and triple negative) with distinct biological features, clinical outcomes and responses to therapies (Perou et al., 2000; Sorlie et al., 2001; Sorlie et al., 2003). As a consequence of this subclassification, a number of therapies that target specific molecules involved in cancer progression have been developed, improving survival of patients. Nevertheless, although breast cancer-related genes have been extensively investigated, little is known about the role of miRNAs in breast cancer molecular subtypes. In a preliminary study, Blenkiron and colleagues (Blenkiron et al., 2007) showed that miRNAs might contribute to the stratification of breast tumors into the intrinsic subtypes. Since miRNAs are key regulators of many cellular processes, identification of subtype-specific miRNAs would provide better understanding of the biology of these tumors, especially in the case of triple negative cancers, which are associated with the most aggressive clinical behavior and do not respond to current targeted therapies. By analyzing the expression levels of 1919 human miRNAs in a large series of breast tumors and normal breast tissues, we aimed to find miRNAs associated with breast cancer and identify those miRNAs specifically deregulated in breast cancer molecular subtypes.

1.1. MiRNA expression profiling differentiates breast tumors from normal breast tissues although stratification of molecular subtypes is imperfect

We first performed unsupervised hierarchical clustering of our samples and observed a clear separation between breast tumors and normal breast tissues, thus confirming previous studies that suggest that miRNA expression profiling can be used to classify breast tissues (Iorio et al., 2005; Lu et al., 2005). We detected common expression profiles in tumors having the same molecular subtype, but also clusters of samples belonging to different subtypes (**Figure 8**). One possible explanation could be that miRNAs are regulating multiple processes in the cell and our samples could be grouping not exclusively according to their molecular subtype but to other tumor characteristics. Interestingly, Rothé et al. came to the

same conclusion after comparing clustering of breast tumors based on miRNA and mRNA expression profiles (Rothe et al., 2011). For instance, it has been suggested that miRNA expression profiles may discriminate tumors of breast cancer patients with different prognosis (Perez-Rivas et al., 2014) and could predict BRCA mutation status in hereditary tumors (Tanic et al., 2015). Therefore, miRNAs may offer additional information on breast tumors stratification. Another explanation to our imperfect unsupervised cluster could be the stratification of breast tumors into new molecular entities. The integration of DNA copy number alterations, DNA methylation, exome sequencing and mRNA, miRNA and protein expression has led to the identification of novel subtypes with characteristic alterations (Curtis et al., 2012b; Network, 2012; Prat et al., 2010). New subtypes could be also present in our samples, making it difficult the stratification into the four classical subtypes. Further studies are needed in order to obtain a comprehensive picture of breast cancer heterogeneity that would allow improvements in the clinical management of the patients.

1.2. MiRNAs commonly and specifically deregulated in breast cancer molecular subtypes

In order to identify miRNAs associated with breast tumors and the intrinsic subtypes, we performed a supervised analysis. The large amount of deregulated miRNAs identified in our set of breast tumors highlights the important role that miRNAs play in breast tumorigenesis. Among the most deregulated miRNAs, miR-125b-5p and miR-21-5p have been repeatedly associated with breast cancer (Iorio et al., 2005; Volinia et al., 2006). MiR-21-5p is known to function as an oncogene by targeting tumor suppressor genes including tropomyosin 1 (*TPM1*), programmed cell death 4 (*PDCD4*) and phosphatase and tensin homolog (*PTEN*), leading to cell proliferation and inhibition of apoptosis and regulating cancer invasion and metastasis in breast cancer (Frankel et al., 2008; Huang et al., 2009; Zhu et al., 2007). MiR-125b-5p is upregulated in many cancers but downregulated in others such as breast cancer, and controls many different cellular processes by targeting numerous transcription factors such as ETS1, E2F3 and BCL3 (Guan et al., 2011; Huang et al., 2011; Zhang et al., 2011). We found that both miR-21 and miR-125b were deregulated through all the molecular subtypes, underlying their essential role in breast cancer. In addition and due to the large number of miRNAs analyzed in the present study, we have identified new

miRNAs that have not been associated with breast cancer before, expanding the knowledge on miRNA deregulation in breast cancer. Some of the most significant ones are deregulated through all the molecular subtypes, like miR-3613-3p, miR-4668-5p, miR-4516, miR-548as-3p, miR-4488, miR-3656 and miR-5704, making them ideal candidates for breast cancer detection.

Even though several studies have focused on the identification of miRNAs associated with breast cancer, little is known about deregulated miRNAs in breast cancer molecular subtypes. By analyzing the expression of 309 human miRNAs, Blenkiron and colleagues detected some miRNAs differentially expressed among the intrinsic subtypes (Blenkiron et al., 2007). In the present study, we analyzed the expression of 1919 human miRNAs and consequently identified a larger number of miRNAs associated with tumor subtypes (**Figure 9**). Interestingly, comparison of our results with the ones obtained by Blenkiron et al. revealed similar patterns of expression for several key miRNAs (**Figure 22**). Of note, all these miRNAs seem to be associated with ER status. For example, miR-150, miR-155 and miR-187 are upregulated in triple negative and Her2 tumors when compared with luminal samples, while miR-145, miR-199a, miR-30a, let-7a, let-7b, let-7c, let-7f and miR-342 are downregulated in triple negative and Her2 samples when compared with luminal tumors. These results are consistent with previous studies that suggest that the expression levels of miR-155 are inversely correlated with ER (Lu et al., 2012) while the expression of miR-342 is higher in ER positive tumors compared with triple negative tumors (Lowery et al., 2009). In addition, overexpression of miR-150 and miR-155 and downregulation of miR-145, miR-30a and members of the let-7 family has been associated with breast cancer progression, tumor aggressiveness, self-renewal and loss of differentiation (Bussing et al., 2008; Chen et al., 2012; Cheng et al., 2012; Huang et al., 2013; Zou et al., 2012), thus explaining the differences observed between luminal and non-luminal tumors.

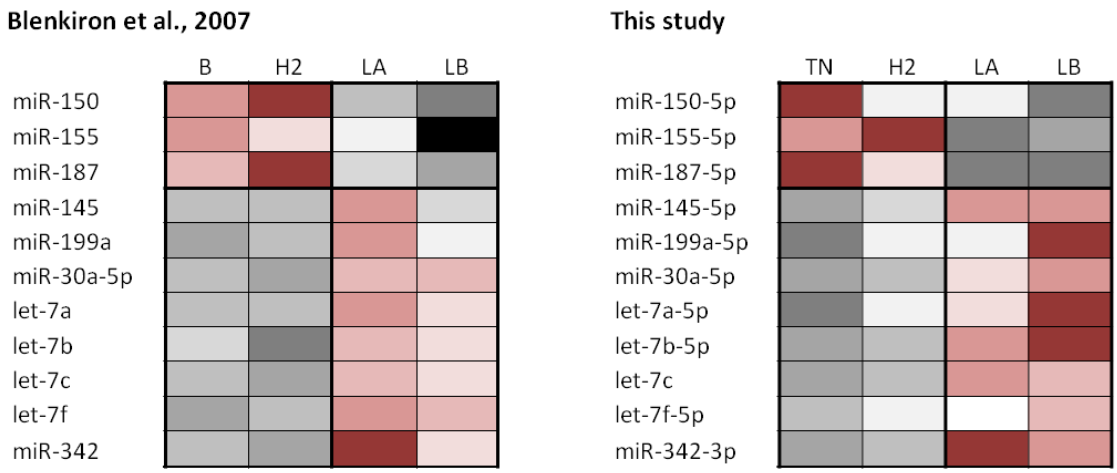


Figure 22. miRNAs associated with breast cancer molecular subtypes in the study carried out by Blenkiron et al. that have a similar pattern of expression in the present study. Colors represent the average expression value for each group of interest: dark red indicates high expression and dark grey, low expression.

To our knowledge, this is the first report that takes into account the miRNA expression profile of normal tissues to obtain miRNAs associated with breast cancer intrinsic subtypes. We believe that this approach might be more appropriate for the identification of specifically deregulated miRNAs. For example, we found that miR-342 is downregulated in triple negative and Her2 when compared with luminal tumors, as previous studies have suggested (Blenkiron et al., 2007; Lowery et al., 2009). However, all the subtypes showed overexpression of this miRNA when compared with the normal tissues, suggesting that the upregulation of this miRNA might play a general role in breast tumorigenesis and could not be specific to ER positive tumors.

Interestingly, we could only identify one specific miRNA for Her2 tumors (**Figure 9B**). These results are in accordance with the ones obtained by van Schooneveld et al. (van Schooneveld et al., 2012). After comparing miRNA expression profiles of tumor subtypes, they could not find any specific miRNA for Her2 tumors while identified triple negative subtype as the subgroup with more specific miRNAs. These findings suggest that Her2 subtype might enclose tumors with diverse miRNA profiles, making it difficult the identification of exclusive miRNAs. In fact, when compared with the normal tissues, Her2 was the subtype with less differentially expressed miRNAs, highlighting its heterogeneous miRNA profile. On the other hand, a great number of specifically deregulated miRNAs were

identified for triple negative tumors, including several members of the let-7 family (let-7d-5p, let-7i-5p, let-7a-5p, let-7d-3p). Let-7 is a family of miRNAs highly conserved across species and is often cited as the archetypal tumor-suppressing miRNA family. It has been shown that downregulation of let-7 promotes self-renewal and leads to a less differentiated cellular state in human and murine breast cells (Ibarra et al., 2007; Yu et al., 2007). Thus, misregulation of member of this family in triple negative breast tumors could explain at least in part why this subtype tends to grow and spread more quickly than other types of breast cancer and why triple negative cancer cells are often poorly differentiated.

1.3. miRNA signatures predict breast cancer molecular subtypes

Machine learning is a promising tool in disease diagnosis since it allows the recognition of expression patterns in groups of samples and the use of these patterns for the classification of new samples. In fact, several prognostic classifiers are currently used as treatment decision tools. The 70-gene predictor Mammaprint (Glas et al., 2006) and the 21-gene signature Oncotype (Paik et al., 2004) predict risk of recurrence in early-stage breast cancer patients and inform the utility of chemotherapy as part of the treatment plan. By using support vector machines, we have generated five microarray classifiers that discriminate breast tumors and breast cancer molecular subtypes with high sensitivity and specificity (**Table 8**). The accuracy for our classifiers is in the range 0.97-0.98, suggesting excellent classification ability. These classifiers have been validated in an independent set of samples with similar sensitivity and specificity, confirming the potential of these miRNAs to stratify breast tumors. Since inmunohistochemical (IHC) tests can be limited by their need of well-preserved tissues and the subjective interpretation of stain intensity by pathologists (Dunstan et al., 2011), we consider that the microarray classifiers reported here could be useful tools to complement IHC tests for breast cancer classification. MiRNAs have been shown to be unusually well-preserved in a range of specimen types and are ideal substrates for the molecular characterization of FFPE tissues due to their small size and resistance to degradation (Liu et al., 2009). Nevertheless, since reproducibility is a major criticism about microarray technology, validation in a prospective multicenter trial is required before any translational application.

The signatures generated here include miRNAs that have previously been associated with breast cancer and other cancers, and that are implicated in the regulation of cellular processes such as proliferation, migration, invasion and apoptosis. The pathways predicted to be significantly enriched for each signature are closely related to cancer development and progression, suggesting that the identified miRNAs are biologically relevant and their choice is not arbitrary. The Cancer Genome Atlas Network has recently shown that somatic mutations in *PIK3CA, TP53, MAP3K1* and *MAP2K4* genes occur at high incidence across breast cancer molecular subtypes (Cancer Genome Atlas, 2012). Interestingly, PI3K-AKT, P53 and MAPK signaling pathways are predicted to be associated with the identified signatures with high statistical significance, which suggests that deregulation of these miRNAs might contribute to the alteration of these pathways in breast cancer.

Some of the miRNAs that compose each signature have previously been associated with the corresponding intrinsic subtype, thus confirming the utility of these signatures in the prediction of breast cancer molecular subtypes. For example, reduced levels of miR-29c has been found in basal-like cancers when compared with other subtypes (Sandhu et al., 2014); a lower expression of miR-205 is associated with HER2 breast tumors (Mattie et al., 2006) and interestingly, miR-205 has been reported to regulate Her3 in human breast cancer (Iorio et al., 2009) and, in turn, to be regulated by Her2 (Adachi et al., 2011); miR-221 increases proliferation in ER-positive cells (Di Leva et al., 2010); and miR-22 is a suppressor of ER alpha and is downregulated in ER-positive breast cancer cells and clinical samples (Xiong et al., 2010). Consequently, we hypothesize that the identified miRNA signatures could be informative for breast cancer diagnosis and might assist in defining specific targets for future therapy.

2. CIRCULATING MICRORNAs AS EARLY DETECTION MARKERS FOR BREAST CANCER

Despite improvements in screening techniques and treatment strategies, breast cancer is still one of the leading causes of cancer death among women (Ferlay et al., 2015), mainly due to late stage at initial diagnosis. Thus, efficient diagnostic tools are urgently needed to improve early breast cancer detection and consequently patient outcome. Even though mammography is the most reliable way to detect breast cancer, it has some limitations including low sensitivity in dense breasts, severe pain caused to some women or radiation risk, especially for women below age 35 (Kolb et al., 2002; Law and Faulkner, 2001; Sharp et al., 2003). Biopsy is the method used to establish a definitive diagnosis but it is an invasive procedure. Serum tumor markers such as CEA or CA 15-3, although being promising at the time of their identification, are not recommended by the ASCO and other expert panels for screening or diagnosis of breast cancer due to their low sensitivity in early stages of the disease (Harris et al., 2007). Consequently, there is an urgent need for the identification of sensitive, specific and non-invasive markers for early breast cancer detection. Given the important role that microRNAs play in tumorigenesis and their remarkable stability in body fluids, we have evaluated their potential as novel non-invasive breast cancer biomarkers by analyzing two independent series of plasma.

2.1. Comparison of miRNA expression between tumor and plasma

In the first part of this thesis, we have identified a large number of miRNAs deregulated in breast tumors when compared with normal breast tissues. Since these miRNAs might be relevant for breast cancer detection, we decided to investigate the expression of the most relevant ones in plasma. First, we validated their deregulation in tumors by qRT-PCR. A strong correspondence between microarray expression and qRT-PCR was observed, as 91% of the explored miRNAs were validated ($p\text{-value}<0.05$). These results are in line with those reported by Git et al., where a high correlation (0.82-0.92) between Exiqon platform and qPCR was described (Git et al., 2010). However, the validation of significant tumoral miRNAs in plasma was more controversial. Among 10 miRNAs analyzed in the discovery set, 5 were found differentially expressed in the plasma of breast cancer patients when compared with

healthy women and one showed a trend. However, only half of these six miRNAs were deregulated in the same direction as in tumors (miR-21-5, miR-142-3p and miR-96-5p). Dissimilar patterns of miRNA expression between tumor and plasma have been reported recently. Chan et al. performed miRNA profiling of tumors and sera from breast cancer patients and healthy individuals and observed 73 miRNAs deregulated in breast tumors and 85 in plasma. However, only 21 were in common in both tissues, and 13 of them were deregulated in opposite directions (Chan et al., 2013). Similarly, Pigati et al. studied the liberation of miRNAs from malignant and non-malignant mammary cells into body fluids and suggested that miRNAs are released from breast cancer cells in a selective manner and therefore, extracellular and cellular miRNA profiles are different (Pigati et al., 2010).

2.2. miR-505-5p, miR-96-5p, miR-125b-5p and miR-21-5p are deregulated in tumors and plasma of breast cancer patients

Validation in a second series of plasma led to the confirmation that miR-505-5p, miR-125b-5p, miR-21-5p and miR-96-5p are overexpressed in the plasma of breast cancer patients. Since these miRNAs are significantly deregulated both in tumor and plasma from pretreated breast cancer patients, we hypothesize that miR-505-5p, miR-125b-5p, miR-21-5p and miR-96-5p might be candidates for non-invasive breast cancer detection. Overexpression of circulating miR-21-5p has been described in breast cancer patients (Asaga et al., 2011; Mar-Aguilar et al., 2013; Ng et al., 2013; Si et al., 2013) but also in other cancers such as esophageal, gastric, colorectal and lung (Du et al., 2014; Wang and Zhang, 2012; Zheng et al., 2011). These findings show the potential utility of circulating miR-21-5p as a broad-spectrum biomarker for the detection of various cancers and not specifically for breast cancer diagnosis. Mir-125b-5p has also been reported to be upregulated in the serum of breast cancer patients (Mar-Aguilar et al., 2013; Wang et al., 2012) and has been associated with chemotherapeutic resistance, with non-responsive patients having higher expression levels (Wang et al., 2012). To the best of our knowledge, this is the first report of circulating miR-505-5p and miR-96-5p being associated with breast cancer. Nevertheless, downregulation of miR-505 and upregulation of miR-96-5p have been reported in breast tumors and have been related to increased cell proliferation (Li et al., 2014; Yamamoto et al., 2011).

ROC curve analysis showed that the discrimination potential of these four miRNAs was acceptable, with AUC ranging from 0.6070 to 0.7213, being miR-505 and miR-96-5p the most valuable biomarkers for discriminating breast cancer patients from healthy individuals. In addition, the levels of miR-505-5p and miR-21-5p were significantly reduced after surgery/treatment, suggesting that the expression of these two miRNAs is dependent on tumor dynamics. Although quantification of levels of miR-505-5p and miR-21-5p in the same cohort of patients is required, the reduction of expression in the postreated group indicates the potential utility of these miRNAs to monitor treatment response and highlights their clinical value for breast cancer detection and surveillance.

Moreover, we found significant higher levels of circulating miR-505 in patients with luminal low-grade tumors, suggesting that this miRNA could be used not only in the detection and surveillance of breast cancer but also in the recognition of luminal subtypes. Similar to other studies (Asaga et al., 2011; Wang and Zhang, 2012), circulating miR-21 did not show any association with age of the patient, tumor size, grade, lymph node, ER, PR and Her2 status, ki-67 levels and molecular subtype. Contrarily, other reports have shown an association with larger tumor size and lymph node metastasis (Si et al., 2013). Although higher expression levels of circulating miR-125b have been related to higher tumor grade and lymph node metastasis (Wang et al., 2012), we did not observed any association with these factors. In order to use circulating miRNAs as a liquid biopsy, further studies with larger cohorts of patients are required to elucidate the relationship between miRNA expression levels and clinicopathologic features of breast cancer patients.

In conclusion, we have reported and validated the overexpression of miR-505-5p, miR-125b-5p, miR-21-5p and miR-96-5p in the plasma of breast cancer patients and demonstrated the potential utility of these miRNAs as non-invasive biomarkers for breast cancer screening. A great advantage of our study is that most of the patients had an early stage breast cancer at the time of blood sample collection, which highlights the relevance of the identified miRNAs in early breast cancer detection. Although promising, prospective studies on larger cohorts of patients are required to confirm the diagnostic value of these miRNAs.

3. MICRORNAs ASSOCIATED WITH RECURRENCE-FREE SURVIVAL IN TRIPLE NEGATIVE BREAST CANCER

Triple negative tumors are associated with the most aggressive clinical behavior and poorest prognosis in breast cancer (Dent et al., 2007). Due to the lack of therapeutic targets, chemotherapy is the only possibility for triple negative patients, but although some patients have an excellent response, others experience early disease recurrence in the first 3-5 years following diagnosis (Liedtke et al., 2008; Mulligan et al., 2008). During the last decade, several gene expression signatures for outcome prediction have been described and validated in breast cancer with consistent results (Glas et al., 2006; Paik et al., 2004). However, these assays are not useful in ER-negative disease (Fan et al., 2011) and therefore, identification of robust prognostic and predictive markers in triple negative patients is urgently needed. In this context, we used miRNA expression profiling to find miRNAs that can distinguish groups of triple negative patients with different clinical outcomes.

3.1. Deregulated miRNAs in node-positive triple negative patients with different outcome

While women with node-negative triple negative tumors generally have an excellent five-year disease-free survival when treated, the presence of any lymph node metastases at the time of diagnosis is a negative prognostic indicator (Hernandez-Aya et al.; Rakha et al., 2007). In accordance to these data, we found that most node-negative women in our two cohorts of triple negative patients survived after 5 years of diagnosis but only 40% of node-positive women had no disease recurrence at the end of this period. In order to identify miRNAs that can distinguish node-positive patients with different outcome, the expression of 1919 human miRNAs was analyzed by microarray technology in 10 node-positive triple negative patients. Among the 17 miRNAs identified, only two (let-7i-5p and let-7a-5p) were specific of triple negative subtype, suggesting that most of the miRNAs involved in the aggressiveness of triple negative tumors are not subtype specific. The miRNAs identified showed significant lower expression in patients with shorter RFS, indicating that the expression of these miRNAs might be lost in highly aggressive triple negative tumors. In fact, most of these miRNAs have been reported to act as cancer suppressor genes by inhibiting

cell migration, invasion and proliferation. Similarly, Avery-Kiejda et al. suggested that miRNA expression profiles tend to be downregulated in node-positive triple negative patients while are overexpressed in node-negative disease (Avery-Kiejda et al., 2014). Interestingly, many of the identified miRNAs belong to a miRNA cluster, such as miR-23b/27b/24-1 cluster in chromosome 9, miR-23a/24-2/27a cluster in chromosome 19 and let-7a/7b cluster in chromosome 22. Since it has been shown that clustered miRNAs have a tendency to coordinately regulate target genes (Grun et al., 2005; Hausser and Zavolan, 2014), deregulation of these miRNAs might have an additive effect in the same molecular pathway. Surprisingly and contrarily to what has been published before in breast cancer (Yan et al., 2008), miR-21 was found underexpressed in patients with worse prognosis. Although further investigation is required, a different role of miR-21 in triple negative disease could be the explanation for this finding.

3.2. MiR-30c-5p and miR-195-5p are associated with recurrence in triple negative breast cancer

We found that decreased expression levels of let-7b-5p, miR-195-5p, miR-24-3p and miR-30c-5p were significantly associated with increased risk of recurrence of node-positive TNBC patients. However, validation by qRT-PCR in a larger cohort of patients confirmed the prognostic value of only miR-30c-5p and miR-195-5p in both node-positive and the whole group of triple negative patients. Moreover, the stratification of the patients according to the combination of both miRNAs resulted in a great separation of high- and low-risk groups. Interestingly, the pathway enrichment analysis suggested that these miRNAs seem to collectively target a broad range of signaling pathways related to proliferation, invasion and cell cycle regulation.

It has been suggested that miR-195-5p acts as a tumor suppressor gene, the expression of which is downregulated in breast cancer (Li et al., 2011). More importantly, its overexpression in breast cancer cells inhibits cell proliferation, reduces cell colony formation, suppresses cell migration and promotes apoptosis through inhibition of *RAF-1* and cyclins E1 (*CCNE1*) and D1 (*CCND1*) (Li et al., 2011; Luo et al., 2014; Yang et al., 2013). Furthermore, the expression of miR-195-5p has recently been associated with sensitivity to the anthracycline drug Adriamycin (also known as Doxorubicin), and it has been shown that

multidrug-resistant breast cancer tissues have low levels of this miRNA (Yang et al., 2013). Likewise, miR-30c-5p has been reported to regulate invasion and proliferation in breast cancer cells by targeting the oncogene *KRAS* and the cytoskeleton network genes encoding twinfilin 1 (TWF1) and vimentin (VIM) (Bockhorn et al., 2013; Tanic et al., 2012). Moreover, miR-30c-5p has lately been shown to be downregulated in Doxorubicin-resistant breast cancer cell lines and its overexpression sensitivizes tumor cells to Doxorubicin by inhibiting the anti-apoptotic gene *YWHAZ* (Fang et al., 2014).

Of note, miR-195-5p exhibited association with poor prognosis even after adjustment for relevant clinical variables, indicating that this miRNA might be an independent prognostic marker in TNBC. Although preliminary, the association of miR-195-5p with disease recurrence could potentially serve to define a group of triple negative patients who may benefit from a more aggressive therapy. On the other hand, we found that the expression of miR-30c-5p seems to be associated with tumor size in triple negative patients. Similarly, Tanic et al. showed that the overexpression of miR-30c-5p in the TNBC cell line MDA-MB-436 reduces KRAS levels and inhibits proliferation (Tanic et al., 2012). Hence, tumors that express low levels of miR-30c-5p might proliferate more and increase in size. Furthermore and taking into consideration the association of both miRNAs with Doxorubicin sensitivity (Fang et al., 2014; Yang et al., 2013), therapeutic delivery of miR-195-5p and miR-30c-5p could improve chemotherapy response in triple negative patients with low levels of these miRNAs. Although our results may require further external validation in a larger cohort, the prognostic and predictive value of these two miRNAs in TNBC is promising.

4. MiR-498 REGULATES *BRCA1* EXPRESSION IN SPORADIC TRIPLE NEGATIVE BREAST CANCER

In early 1990s, family-based linkage analysis and positional cloning led to the identification of high-penetrance genes *BRCA1* (Miki et al., 1994) and *BRCA2* (Wooster et al., 1994), two tumor-suppressor genes that are frequently mutated in hereditary breast cancers. During the past two decades, BRCA1 has been found to play a critical role in various cellular processes, including DNA repair by distinct pathways, cell cycle checkpoints control, centrosome amplification, transcriptional activation of target genes, and ubiquitin ligation (Drost and Jonkers, 2014; Narod and Foulkes, 2004). While the specific functions of BRCA1 are still being elucidated, it is clear that functional BRCA1 protein is required to prevent breast transformation (Xu et al., 1999). Although sporadic triple negative tumors share many characteristics with *BRCA1*-germline mutated breast tumors, they usually do not present somatic mutations in the *BRCA1* gene. However, several studies have shown that most sporadic triple negative tumors have a reduced expression of the *BRCA1* gene (Mueller and Roskelley, 2003; Turner et al., 2007), which suggests a possible role of BRCA1 dysfunction in the pathogenesis of sporadic triple negative breast cancer. Taking into consideration that miRNAs function as negative regulators of gene expression, we investigated their possible involvement in the inactivation of *BRCA1* in sporadic triple negative tumors.

4.1. *BRCA1* is a target of miR-498 and miR-187-5p

In the first part of this thesis, we identified 78 miRNAs that were overexpressed in sporadic triple negative tumors but not in other breast cancer subtypes when compared with normal breast tissues. Since reduced expression of *BRCA1* in triple negative tumors could be produced by high levels of a miRNA targeting this gene, we investigated if these miRNAs have binding sites in the 3'UTR of *BRCA1*. At least two bioinformatics algorithms predicted with high scores that miR-498 and miR-187-5p bind to the 3'UTR of the *BRCA1* gene, and we functionally validated these results by luciferase reporter assay. In addition, we have confirmed with similar levels of repression, previous studies that report that miR-146a targets *BRCA1* 3'UTR (Garcia et al., 2011; Shen et al., 2008).

The role of miR-498 in cancer development has not been well documented. While it seems to be dowregulated in some cancers such as colon and ovarian cancer (Gopalan et al., 2015; Kasiappan et al., 2012), its overexpression has been reported in metastatic medullary thyroid carcinoma and retinoblastoma (Santarpia et al., 2013; Zhao et al., 2009). Regarding its targets, miR-498 has been shown to bind to the 3'UTR of *HER2* and *TERT* (Kasiappan et al., 2012; Leivonen et al., 2014), but reports showing targeting of *BRCA1* have not been described so far. Similarly, high levels of miR-187-5p have been associated with ovarian cancer (Chao et al., 2012) but its downregulation has been reported in clear cell renal cell carcinoma and prostate cancer (Fuse et al., 2012; Zhao et al., 2013). Interestingly, its overexpression has been associated with poor outcome in breast cancer, leading to a more aggressive phenotype (Mulrane et al., 2012). These findings suggest that miR-498 and miR-187-5p might act as oncogenes or tumor suppressors depending on the cellular context, as it happens with other miRNAs such as miR-125b-5p.

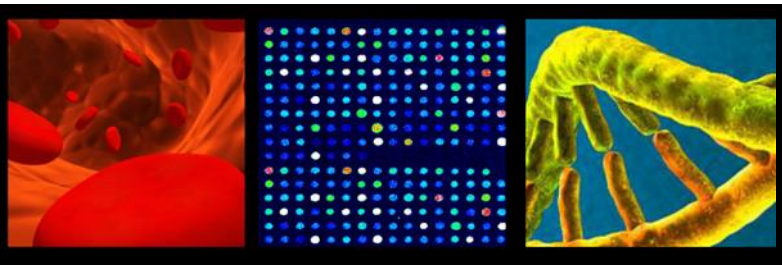
4.2. MiR-498 regulates *BRCA1* expression in breast cancer cell lines

After analyzing the expression of miR-498 and miR-187-5p in breast cancer cell lines of different subtypes, we found that miR-498 was overexpressed in the triple negative cell line HS578T while miR-187-5p seemed to be overexpressed in the luminal cell lines BT474 and MCF7. Since we were interested in miRNAs with increased expression levels in triple negative cell lines and negatively correlated with *BRCA1* expression levels, we decided to focus on miR-498 for following experiments. We functionally demonstrated the interaction between miR-498 and *BRCA1* in breast cancer cell lines: inhibition of miR-498 in HS578T cell line increased BRCA1 levels and its overexpression in MCF7 cell line reduced BRCA1 expression. These results suggest that miR-498 regulates *BRCA1* expression in breast cancer and its overexpression could contribute to the pathogenesis of sporadic TNBC via *BRCA1* downregulation. These findings confirm previous studies that suggest that miRNA deregulation might be involved in the inactivation of *BRCA1* in sporadic breast cancer (Garcia et al., 2011; He et al., 2014; Moskwa et al., 2011; Tan et al., 2014).

4.3. MiR-498 plays a role in triple negative breast cancer cell proliferation

In addition, we have demonstrated that miR-498 inhibition leads to reduced proliferation of triple negative breast cancer cells. Our findings are consistent with a role of miR-498 in the regulation of *BRCA1*, since induction of *BRCA1* expression has been shown to inhibit growth in breast tumors and cell lines (Holt et al., 1996). Hence, our data support that miR-498 promotes cell proliferation in triple negative breast cancer through direct regulation of *BRCA1* expression. Although the effect of miR-498 deregulation on DNA repair needs to be investigated, the findings reported here have potential clinical implications. The most relevant one is that tumors with high levels of miR-498 might be more sensitive to PARP inhibitors and DNA damaging chemotherapeutic agents. Therefore, the monitoring of miR-498 expression could serve to identify a group of breast cancer patients that may benefit from these therapies.

In conclusion, this study sheds light on the mechanisms behind the decreased expression of *BRCA1* in sporadic TNBC. Determination of these mechanisms is essential to increase our understanding of triple negative breast cancer etiology and to permit better therapeutic approaches.



Conclusions / Conclusiones

CONCLUSIONS

1. We have defined five miRNA signatures that discriminate breast tumors and the main breast cancer molecular subtypes with high sensitivity and specificity. Pathway enrichment analysis indicates that these miRNAs might regulate different biological processes related to cancer development and progression.

2. Comparison of miRNA expression between tumor and plasma showed dissimilar patterns of expression for some miRNAs, which could indicate a selective release from breast cancer cells into the blood. We have reported and validated the overexpression of miR-505-5p, miR-125b-5p, miR-21-5p and miR-96-5p in the plasma of breast cancer patients when compared with healthy women and demonstrated the potential utility of these miRNAs as non-invasive biomarkers for early breast cancer detection.

3. We have identified a set of 17 miRNAs that are downregulated in breast tumors of node-positive triple-negative patients with poor outcome. Moreover, we found that miR-30c-5p and miR-195-5p are associated with recurrence in triple negative breast cancer, and that miR-195-5p might be an independent prognostic marker that could serve to define a group of triple negative patients who may benefit from a more aggressive therapy.

4. Two triple-negative specific miRNAs, miR-498 and miR-187-5p, were found to target *BRCA1* 3'UTR. We demonstrated that miR-498 regulates *BRCA1* expression in breast cancer cell lines and its inhibition leads to reduced proliferation in triple negative breast cancer cells. Our results shed light on the mechanisms behind the decreased expression of *BRCA1* in sporadic triple negative breast cancer.

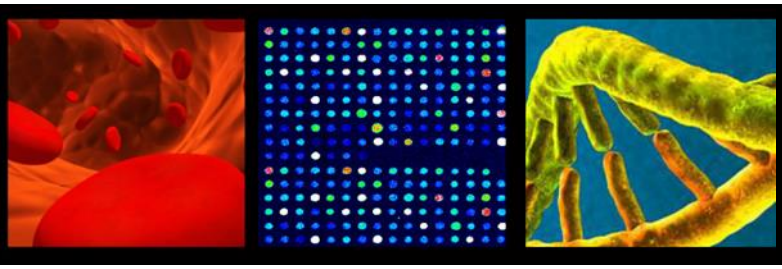
CONCLUSIONES

1. Hemos definido cinco firmas de miRNAs que discriminan los tumores de mama y los principales subtipos moleculares de cáncer de mama con gran sensibilidad y especificidad. El análisis de enriquecimiento funcional indica que estos miRNAs podrían estar regulando distintos procesos biológicos relacionados con el desarrollo y la progresión del cáncer.

2. La comparación de la expresión de ciertos miRNAs en tumores y plasma mostró patrones de expresión diferentes para algunos miRNAs, lo que podría indicar que existe una liberación selectiva por parte de las células cancerosas a la sangre. Hemos validado la sobreexpresión de los miR-505-5p, miR-125b-5p, miR-21-5p y miR-96-5p en el plasma de pacientes con cáncer de mama en comparación con mujeres sanas, y hemos demostrado la posible utilidad de estos miRNAs como biomarcadores no invasivos para la detección temprana del cáncer de mama.

3. Hemos identificado un grupo de 17 miRNAs que están infraexpresados en tumores de mama de pacientes triple negativas con ganglios positivos y mal pronóstico. Además, hemos encontrado que los miR-30c-5p y miR-195-5p están asociados con recurrencia y que el miR-195-5p podría ser un marcador de pronóstico independiente en cáncer de mama triple negativo. El análisis de la expresión del miR-195-5p en tumores podría servir para definir un grupo de pacientes triple negativas que podrían beneficiarse de una terapia más agresiva.

4. Hemos encontrado que dos miRNAs específicos de tumores triple negativos, miR-498 y miR-187-5p, tienen como diana la región 3'UTR de *BRCA1*. Además, hemos demostrado que el miR-498 regula la expresión de *BRCA1* en líneas celulares de cáncer de mama, y que su inhibición da lugar a una reducción en la proliferación de las células triple negativas. Estos resultados podrían explicar la disminución en la expresión de *BRCA1* que se observa en cáncer de mama esporádico triple negativo.



Bibliography

Adachi, R., Horiuchi, S., Sakurazawa, Y., Hasegawa, T., Sato, K., Sakamaki, T., 2011. ErbB2 down-regulates microRNA-205 in breast cancer. Biochemical and biophysical research communications 411, 804-808.

Ali, H.R., Rueda, O.M., Chin, S.F., Curtis, C., Dunning, M.J., Aparicio, S.A., Caldas, C., 2014. Genome-driven integrated classification of breast cancer validated in over 7,500 samples. Genome biology 15, 431.

Ali, S., Coombes, R.C., 2002. Endocrine-responsive breast cancer and strategies for combating resistance. Nat Rev Cancer 2, 101-112.

Arroyo, J.D., Chevillet, J.R., Kroh, E.M., Ruf, I.K., Pritchard, C.C., Gibson, D.F., Mitchell, P.S., Bennett, C.F., Pogosova-Agadjanyan, E.L., Stirewalt, D.L., Tait, J.F., Tewari, M., 2011. Argonaute2 complexes carry a population of circulating microRNAs independent of vesicles in human plasma. Proceedings of the National Academy of Sciences of the United States of America 108, 5003-5008.

Asaga, S., Kuo, C., Nguyen, T., Terpenning, M., Giuliano, A.E., Hoon, D.S., 2011. Direct serum assay for microRNA-21 concentrations in early and advanced breast cancer. Clin Chem 57, 84-91.

Avery-Kiejda, K.A., Braye, S.G., Mathe, A., Forbes, J.F., Scott, R.J., 2014. Decreased expression of key tumour suppressor microRNAs is associated with lymph node metastases in triple negative breast cancer. BMC cancer 14, 51.

Badve, S., Dabbs, D.J., Schnitt, S.J., Baehner, F.L., Decker, T., Eusebi, V., Fox, S.B., Ichihara, S., Jacquemier, J., Lakhani, S.R., Palacios, J., Rakha, E.A., Richardson, A.L., Schmitt, F.C., Tan, P.H., Tse, G.M., Weigelt, B., Ellis, I.O., Reis-Filho, J.S., 2011. Basal-like and triple-negative breast cancers: a critical review with an emphasis on the implications for pathologists and oncologists. Modern pathology : an official journal of the United States and Canadian Academy of Pathology, Inc 24, 157-167.

Banerji, S., Cibulskis, K., Rangel-Escareno, C., Brown, K.K., Carter, S.L., Frederick, A.M., Lawrence, M.S., Sivachenko, A.Y., Sougnez, C., Zou, L., Cortes, M.L., Fernandez-Lopez, J.C., Peng, S., Ardlie, K.G., Auclair, D., Bautista-Pina, V., Duke, F., Francis, J., Jung, J., Maffuz-Aziz, A., Onofrio, R.C., Parkin, M., Pho, N.H., Quintanar-Jurado, V., Ramos, A.H., Rebollar-Vega, R., Rodriguez-Cuevas, S., Romero-Cordoba, S.L., Schumacher, S.E., Stransky, N., Thompson, K.M., Uribe-Figueroa, L., Baselga, J., Beroukhim, R., Polyak, K., Sgroi, D.C., Richardson, A.L., Jimenez-Sanchez, G., Lander, E.S., Gabriel, S.B., Garraway, L.A., Golub, T.R., Melendez-Zajgla, J., Toker, A., Getz, G., Hidalgo-Miranda, A., Meyerson, M., 2012. Sequence analysis of mutations and translocations across breast cancer subtypes. Nature 486, 405-409.

Bartel, D.P., 2009. MicroRNAs: target recognition and regulatory functions. Cell 136, 215-233.

Benjamini, Y., Drai, D., Elmer, G., Kafkafi, N., Golani, I., 2001. Controlling the false discovery rate in behavior genetics research. Behav Brain Res 125, 279-284.

Bergh, J., Holmquist, M., 2001. Who should not receive adjuvant chemotherapy? International databases. Journal of the National Cancer Institute. Monographs, 103-108.

Blenkiron, C., Goldstein, L.D., Thorne, N.P., Spiteri, I., Chin, S.F., Dunning, M.J., Barbosa-Morais, N.L., Teschendorff, A.E., Green, A.R., Ellis, I.O., Tavare, S., Caldas, C., Miska, E.A., 2007. MicroRNA expression profiling of human breast cancer identifies new markers of tumor subtype. Genome biology 8, R214.

Bloom, H.J., Richardson, W.W., 1957. Histological grading and prognosis in breast cancer; a study of 1409 cases of which 359 have been followed for 15 years. British journal of cancer 11, 359-377.

Bockhorn, J., Yee, K., Chang, Y.F., Prat, A., Huo, D., Nwachukwu, C., Dalton, R., Huang, S., Swanson, K.E., Perou, C.M., Olopade, O.I., Clarke, M.F., Greene, G.L., Liu, H., 2013. MicroRNA-30c targets cytoskeleton genes involved in breast cancer cell invasion. Breast cancer research and treatment 137, 373-382.

Bussing, I., Slack, F.J., Grosshans, H., 2008. let-7 microRNAs in development, stem cells and cancer. Trends in molecular medicine 14, 400-409.

Calin, G.A., Dumitru, C.D., Shimizu, M., Bichi, R., Zupo, S., Noch, E., Aldler, H., Rattan, S., Keating, M., Rai, K., Rassenti, L., Kipps, T., Negrini, M., Bullrich, F., Croce, C.M., 2002. 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.

Calin, G.A., Sevignani, C., Dumitru, C.D., Hyslop, T., Noch, E., Yendamuri, S., Shimizu, M., Rattan, S., Bullrich, F., Negrini, M., Croce, C.M., 2004. Human microRNA genes are frequently located at fragile sites and genomic regions involved in cancers. Proceedings of the National Academy of Sciences of the United States of America 101, 2999-3004.

Cancer Genome Atlas, N., 2012. Comprehensive molecular portraits of human breast tumours. Nature 490, 61-70.

Cardoso, F., Van't Veer, L., Rutgers, E., Loi, S., Mook, S., Piccart-Gebhart, M.J., 2008. Clinical application of the 70-gene profile: the MINDACT trial. Journal of clinical oncology : official journal of the American Society of Clinical Oncology 26, 729-735.

Carey, L., Winer, E., Viale, G., Cameron, D., Gianni, L., 2010. Triple-negative breast cancer: disease entity or title of convenience? Nature reviews. Clinical oncology 7, 683-692.

Cimmino, A., Calin, G.A., Fabbri, M., Iorio, M.V., Ferracin, M., Shimizu, M., Wojcik, S.E., Aqeilan, R.I., Zupo, S., Dono, M., Rassenti, L., Alder, H., Volinia, S., Liu, C.G., Kipps, T.J., Negrini, M., Croce, C.M., 2005. miR-15 and miR-16 induce apoptosis by targeting BCL2. Proceedings of the National Academy of Sciences of the United States of America 102, 13944-13949.

Colleoni, M., Minchella, I., Mazzarol, G., Nole, F., Peruzzotti, G., Rocca, A., Viale, G., Orlando, L., Ferretti, G., Curigliano, G., Veronesi, P., Intra, M., Goldhirsch, A., 2000. Response to primary chemotherapy in breast cancer patients with tumors not expressing estrogen and progesterone receptors. Annals of oncology : official journal of the European Society for Medical Oncology / ESMO 11, 1057-1059.

Cuk, K., Zucknick, M., Heil, J., Madhavan, D., Schott, S., Turchinovich, A., Arlt, D., Rath, M., Sohn, C., Benner, A., Junkermann, H., Schneeweiss, A., Burwinkel, B., 2013. Circulating microRNAs in plasma as early detection markers for breast cancer. International journal of cancer. Journal international du cancer 132, 1602-1612.

Curtis, C., Shah, S.P., Chin, S.F., Turashvili, G., Rueda, O.M., Dunning, M.J., Speed, D., Lynch, A.G., Samarajiwa, S., Yuan, Y., Graf, S., Ha, G., Haffari, G., Bashashati, A., Russell, R., McKinney, S., Group, M., Langerod, A., Green, A., Provenzano, E., Wishart, G., Pinder, S., Watson, P., Markowetz, F., Murphy, L., Ellis, I., Purushotham, A., Borresen-Dale, A.L., Brenton, J.D., Tavare, S., Caldas, C., Aparicio, S., 2012a. The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. Nature 486, 346-352.

Curtis, C., Shah, S.P., Chin, S.F., Turashvili, G., Rueda, O.M., Dunning, M.J., Speed, D., Lynch, A.G., Samarajiwa, S., Yuan, Y., Graf, S., Ha, G., Haffari, G., Bashashati, A., Russell, R., McKinney, S., Langerod, A., Green, A., Provenzano, E., Wishart, G., Pinder, S., Watson, P., Markowetz, F., Murphy, L., Ellis, I., Purushotham, A., Borresen-Dale, A.L., Brenton, J.D., Tavare, S., Caldas, C., Aparicio, S., 2012b. The genomic and transcriptomic architecture of 2,000 breast tumours reveals novel subgroups. Nature 486, 346-352.

Cuzick, J., 2008. Assessing risk for breast cancer. Breast cancer research : BCR 10 Suppl 4, S13.

Chan, M., Liaw, C.S., Ji, S.M., Tan, H.H., Wong, C.Y., Thike, A.A., Tan, P.H., Ho, G.H., Lee, A.S., 2013. Identification of circulating microRNA signatures for breast cancer detection. Clin Cancer Res 19, 4477-4487.

Chao, A., Lin, C.Y., Lee, Y.S., Tsai, C.L., Wei, P.C., Hsueh, S., Wu, T.I., Tsai, C.N., Wang, C.J., Chao, A.S., Wang, T.H., Lai, C.H., 2012. Regulation of ovarian cancer progression by microRNA-187 through targeting Disabled homolog-2. Oncogene 31, 764-775.

Chen, J., Wang, B.C., Tang, J.H., 2012. Clinical significance of microRNA-155 expression in human breast cancer. Journal of surgical oncology 106, 260-266.

Chen, X., Ba, Y., Ma, L., Cai, X., Yin, Y., Wang, K., Guo, J., Zhang, Y., Chen, J., Guo, X., Li, Q., Li, X., Wang, W., Wang, J., Jiang, X., Xiang, Y., Xu, C., Zheng, P., Zhang, J., Li, R., Zhang, H., Shang, X., Gong, T., Ning, G., Zen, K., Zhang, C.Y., 2008. Characterization of microRNAs in serum: a novel class of biomarkers for diagnosis of cancer and other diseases. Cell Res 18, 997-1006.

Cheng, C.W., Wang, H.W., Chang, C.W., Chu, H.W., Chen, C.Y., Yu, J.C., Chao, J.I., Liu, H.F., Ding, S.L., Shen, C.Y., 2012. MicroRNA-30a inhibits cell migration and invasion by downregulating vimentin expression and is a potential prognostic marker in breast cancer. Breast cancer research and treatment 134, 1081-1093.

Dent, R., Trudeau, M., Pritchard, K.I., Hanna, W.M., Kahn, H.K., Sawka, C.A., Lickley, L.A., Rawlinson, E., Sun, P., Narod, S.A., 2007. Triple-negative breast cancer: clinical features and patterns of recurrence. Clinical cancer research : an official journal of the American Association for Cancer Research 13, 4429-4434.

Di Leva, G., Gasparini, P., Piovan, C., Ngankeu, A., Garofalo, M., Taccioli, C., Iorio, M.V., Li, M., Volinia, S., Alder, H., Nakamura, T., Nuovo, G., Liu, Y., Nephew, K.P., Croce, C.M., 2010. MicroRNA cluster 221-222 and estrogen receptor alpha interactions in breast cancer. Journal of the National Cancer Institute 102, 706-721.

Diederichs, S., Haber, D.A., 2006. Sequence variations of microRNAs in human cancer: alterations in predicted secondary structure do not affect processing. Cancer research 66, 6097-6104.

Drost, R., Jonkers, J., 2014. Opportunities and hurdles in the treatment of BRCA1-related breast cancer. Oncogene 33, 3753-3763.

Du, M., Liu, S., Gu, D., Wang, Q., Zhu, L., Kang, M., Shi, D., Chu, H., Tong, N., Chen, J., Adams, T.S., Zhang, Z., Wang, M., 2014. Clinical potential role of circulating microRNAs in early diagnosis of colorectal cancer patients. Carcinogenesis 35, 2723-2730.

Dunstan, R.W., Wharton, K.A., Jr., Quigley, C., Lowe, A., 2011. The use of immunohistochemistry for biomarker assessment--can it compete with other technologies? Toxicol Pathol 39, 988-1002.

Elston, C.W., Ellis, I.O., 1991. Pathological prognostic factors in breast cancer. I. The value of histological grade in breast cancer: experience from a large study with long-term follow-up. Histopathology 19, 403-410.

Ellis, M.J., Suman, V.J., Hoog, J., Lin, L., Snider, J., Prat, A., Parker, J.S., Luo, J., DeSchryver, K., Allred, D.C., Esserman, L.J., Unzeitig, G.W., Margenthaler, J., Babiera, G.V., Marcom, P.K., Guenther, J.M., Watson, M.A., Leitch, M., Hunt, K., Olson, J.A., 2011. Randomized phase II neoadjuvant comparison between letrozole, anastrozole, and exemestane for postmenopausal women with estrogen receptor-rich stage 2 to 3 breast cancer: clinical and biomarker outcomes and predictive value of the baseline PAM50-based intrinsic subtype--ACOSOG Z1031. Journal of clinical oncology : official journal of the American Society of Clinical Oncology 29, 2342-2349.

Esquela-Kerscher, A., Slack, F.J., 2006. Oncomirs - microRNAs with a role in cancer. Nature reviews. Cancer 6, 259-269.

Fan, C., Prat, A., Parker, J.S., Liu, Y., Carey, L.A., Troester, M.A., Perou, C.M., 2011. Building prognostic models for breast cancer patients using clinical variables and hundreds of gene expression signatures. BMC medical genomics 4, 3.

Fang, Y., Shen, H., Cao, Y., Li, H., Qin, R., Chen, Q., Long, L., Zhu, X.L., Xie, C.J., Xu, W.L., 2014. Involvement of miR-30c in resistance to doxorubicin by regulating YWHAZ in breast cancer cells. Brazilian journal of medical and biological research = Revista brasileira de pesquisas medicas e biologicas / Sociedade Brasileira de Biofisica ... [et al.] 47, 60-69.

Farazi, T.A., Horlings, H.M., Ten Hoeve, J.J., Mihailovic, A., Halfwerk, H., Morozov, P., Brown, M., Hafner, M., Reyal, F., van Kouwenhove, M., Kreike, B., Sie, D., Hovestadt, V., Wessels, L.F., van de Vijver, M.J., Tuschl, T., 2011. MicroRNA sequence and expression analysis in breast tumors by deep sequencing. Cancer research 71, 4443-4453.

Ferlay, J., Soerjomataram, I., Dikshit, R., Eser, S., Mathers, C., Rebelo, M., Parkin, D.M., Forman, D., Bray, F., 2014. Cancer incidence and mortality worldwide: Sources, methods and major patterns in GLOBOCAN 2012. Int J Cancer.

Ferlay, J., Soerjomataram, I., Dikshit, R., Eser, S., Mathers, C., Rebelo, M., Parkin, D.M., Forman, D., Bray, F., 2015. Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. International journal of cancer. Journal international du cancer 136, E359-386.

Frankel, L.B., Christoffersen, N.R., Jacobsen, A., Lindow, M., Krogh, A., Lund, A.H., 2008. Programmed cell death 4 (PDCD4) is an important functional target of the microRNA miR-21 in breast cancer cells. J Biol Chem 283, 1026-1033.

Fuse, M., Kojima, S., Enokida, H., Chiyomaru, T., Yoshino, H., Nohata, N., Kinoshita, T., Sakamoto, S., Naya, Y., Nakagawa, M., Ichikawa, T., Seki, N., 2012. Tumor suppressive microRNAs (miR-222 and miR-31) regulate molecular pathways based on microRNA expression signature in prostate cancer. J Hum Genet 57, 691-699.

Galea, M.H., Blamey, R.W., Elston, C.E., Ellis, I.O., 1992. The Nottingham Prognostic Index in primary breast cancer. Breast cancer research and treatment 22, 207-219.

Garcia, A.I., Buisson, M., Bertrand, P., Rimokh, R., Rouleau, E., Lopez, B.S., Lidereau, R., Mikaelian, I., Mazoyer, S., 2011. Down-regulation of BRCA1 expression by miR-146a and miR-146b-5p in triple negative sporadic breast cancers. EMBO molecular medicine 3, 279-290.

Geiss, G.K., Bumgarner, R.E., Birditt, B., Dahl, T., Dowidar, N., Dunaway, D.L., Fell, H.P., Ferree, S., George, R.D., Grogan, T., James, J.J., Maysuria, M., Mitton, J.D., Oliveri, P., Osborn, J.L., Peng, T., Ratcliffe, A.L., Webster, P.J., Davidson, E.H., Hood, L., Dimitrov, K., 2008. Direct multiplexed measurement of gene expression with color-coded probe pairs. Nature biotechnology 26, 317-325.

Ghoussaini, M., Pharoah, P.D., Easton, D.F., 2013. Inherited genetic susceptibility to breast cancer: the beginning of the end or the end of the beginning? The American journal of pathology 183, 1038-1051.

Git, A., Dvinge, H., Salmon-Divon, M., Osborne, M., Kutter, C., Hadfield, J., Bertone, P., Caldas, C., 2010. Systematic comparison of microarray profiling, real-time PCR, and next-generation sequencing technologies for measuring differential microRNA expression. RNA 16, 991-1006.

Glas, A.M., Floore, A., Delahaye, L.J., Witteveen, A.T., Pover, R.C., Bakx, N., Lahti-Domenici, J.S., Bruinsma, T.J., Warmoes, M.O., Bernards, R., Wessels, L.F., Van't Veer, L.J., 2006. Converting a breast cancer microarray signature into a high-throughput diagnostic test. BMC genomics 7, 278.

Goldhirsch, A., Winer, E.P., Coates, A.S., Gelber, R.D., Piccart-Gebhart, M., Thurlimann, B., Senn, H.J., Panel, m., 2013. Personalizing the treatment of women with early breast cancer: highlights of the St Gallen International Expert Consensus on the Primary Therapy of Early Breast Cancer 2013. Annals of oncology : official journal of the European Society for Medical Oncology / ESMO 24, 2206-2223.

Goldhirsch, A., Wood, W.C., Coates, A.S., Gelber, R.D., Thurlimann, B., Senn, H.J., 2011. Strategies for subtypes--dealing with the diversity of breast cancer: highlights of the St. Gallen International Expert Consensus on the Primary Therapy of Early Breast Cancer 2011. Ann Oncol 22, 1736-1747.

Gopalan, V., Smith, R.A., Lam, A.K., 2015. Downregulation of microRNA-498 in colorectal cancers and its cellular effects. Exp Cell Res 330, 423-428.

Grun, D., Wang, Y.L., Langenberger, D., Gunsalus, K.C., Rajewsky, N., 2005. microRNA target predictions across seven Drosophila species and comparison to mammalian targets. PLoS computational biology 1, e13.

Guan, Y., Yao, H., Zheng, Z., Qiu, G., Sun, K., 2011. MiR-125b targets BCL3 and suppresses ovarian cancer proliferation. Int J Cancer 128, 2274-2283.

Harris, L., Fritsche, H., Mennel, R., Norton, L., Ravdin, P., Taube, S., Somerfield, M.R., Hayes, D.F., Bast, R.C., Jr., 2007. American Society of Clinical Oncology 2007 update of recommendations for the use of tumor markers in breast cancer. J Clin Oncol 25, 5287-5312.

Hasemeier, B., Christgen, M., Kreipe, H., Lehmann, U., 2008. Reliable microRNA profiling in routinely processed formalin-fixed paraffin-embedded breast cancer specimens using fluorescence labelled bead technology. BMC Biotechnol 8, 90.

Hassiotou, F., Geddes, D., 2013. Anatomy of the human mammary gland: Current status of knowledge. Clin Anat 26, 29-48.

Hausser, J., Zavolan, M., 2014. Identification and consequences of miRNA-target interactions--beyond repression of gene expression. Nature reviews. Genetics 15, 599-612.

He, X., Xiao, X., Dong, L., Wan, N., Zhou, Z., Deng, H., Zhang, X., 2014. MiR-218 regulates cisplatin chemosensitivity in breast cancer by targeting BRCA1. Tumour biology : the journal of the International Society for Oncodevelopmental Biology and Medicine.

Hernandez-Aya, L.F., Chavez-Macgregor, M., Lei, X., Meric-Bernstam, F., Buchholz, T.A., Hsu, L., Sahin, A.A., Do, K.A., Valero, V., Hortobagyi, G.N., Gonzalez-Angulo, A.M., Nodal status and clinical outcomes in a large cohort of patients with triple-negative breast cancer. J Clin Oncol 29, 2628-2634.

Herschkowitz, J.I., Simin, K., Weigman, V.J., Mikaelian, I., Usary, J., Hu, Z., Rasmussen, K.E., Jones, L.P., Assefnia, S., Chandrasekharan, S., Backlund, M.G., Yin, Y., Khramtsov, A.I., Bastein, R., Quackenbush, J., Glazer, R.I., Brown, P.H., Green, J.E., Kopelovich, L., Furth, P.A., Palazzo, J.P., Olopade, O.I., Bernard, P.S., Churchill, G.A., Van Dyke, T., Perou, C.M., 2007. Identification of conserved gene expression features between murine mammary carcinoma models and human breast tumors. Genome biology 8, R76.

Holt, J.T., Thompson, M.E., Szabo, C., Robinson-Benion, C., Arteaga, C.L., King, M.C., Jensen, R.A., 1996. Growth retardation and tumour inhibition by BRCA1. Nature genetics 12, 298-302.

Hu, Z., Fan, C., Oh, D.S., Marron, J.S., He, X., Qaqish, B.F., Livasy, C., Carey, L.A., Reynolds, E., Dressler, L., Nobel, A., Parker, J., Ewend, M.G., Sawyer, L.R., Wu, J., Liu, Y., Nanda, R., Tretiakova, M., Ruiz Orrico, A., Dreher, D., Palazzo, J.P., Perreard, L., Nelson, E., Mone, M., Hansen, H., Mullins, M., Quackenbush, J.F., Ellis, M.J., Olopade, O.I., Bernard, P.S., Perou, C.M., 2006. The molecular portraits of breast tumors are conserved across microarray platforms. BMC genomics 7, 96.

Huang, G.L., Zhang, X.H., Guo, G.L., Huang, K.T., Yang, K.Y., Shen, X., You, J., Hu, X.Q., 2009. Clinical significance of miR-21 expression in breast cancer: SYBR-Green I-based real-time RT-PCR study of invasive ductal carcinoma. Oncol Rep 21, 673-679.

Huang, L., Luo, J., Cai, Q., Pan, Q., Zeng, H., Guo, Z., Dong, W., Huang, J., Lin, T., 2011. MicroRNA-125b suppresses the development of bladder cancer by targeting E2F3. Int J Cancer 128, 1758-1769.

Huang, S., Chen, Y., Wu, W., Ouyang, N., Chen, J., Li, H., Liu, X., Su, F., Lin, L., Yao, Y., 2013. miR-150 promotes human breast cancer growth and malignant behavior by targeting the pro-apoptotic purinergic P2X7 receptor. PloS one 8, e80707.

Hudis, C.A., Gianni, L., 2011. Triple-negative breast cancer: an unmet medical need. The oncologist 16 Suppl 1, 1-11.

Ibarra, I., Erlich, Y., Muthuswamy, S.K., Sachidanandam, R., Hannon, G.J., 2007. A role for microRNAs in maintenance of mouse mammary epithelial progenitor cells. Genes & development 21, 3238-3243.

Iorio, M.V., Casalini, P., Piovan, C., Di Leva, G., Merlo, A., Triulzi, T., Menard, S., Croce, C.M., Tagliabue, E., 2009. microRNA-205 regulates HER3 in human breast cancer. Cancer research 69, 2195-2200.

Iorio, M.V., Ferracin, M., Liu, C.G., Veronese, A., Spizzo, R., Sabbioni, S., Magri, E., Pedriali, M., Fabbri, M., Campiglio, M., Menard, S., Palazzo, J.P., Rosenberg, A., Musiani, P., Volinia, S., Nenci, I., Calin, G.A., Querzoli, P., Negrini, M., Croce, C.M., 2005. MicroRNA gene expression deregulation in human breast cancer. Cancer Res 65, 7065-7070.

Kasiappan, R., Shen, Z., Tse, A.K., Jinwal, U., Tang, J., Lungchukiet, P., Sun, Y., Kruk, P., Nicosia, S.V., Zhang, X., Bai, W., 2012. 1,25-Dihydroxyvitamin D3 suppresses telomerase expression and human cancer growth through microRNA-498. J Biol Chem 287, 41297-41309.

Kolb, T.M., Lichy, J., Newhouse, J.H., 2002. Comparison of the performance of screening mammography, physical examination, and breast US and evaluation of factors that influence them: an analysis of 27,825 patient evaluations. Radiology 225, 165-175.

Lakhani, S.R., Ellis. I.O., Schnitt, S.J., Tan, P.H., van de Vijver, M.J., 2012. WHO Classification of Tumours of the Breast, Fourth Edition**.**

Law, J., Faulkner, K., 2001. Cancers detected and induced, and associated risk and benefit, in a breast screening programme. Br J Radiol 74, 1121-1127.

Lawrie, C.H., Gal, S., Dunlop, H.M., Pushkaran, B., Liggins, A.P., Pulford, K., Banham, A.H., Pezzella, F., Boultwood, J., Wainscoat, J.S., Hatton, C.S., Harris, A.L., 2008. Detection of elevated levels of tumour-associated microRNAs in serum of patients with diffuse large B-cell lymphoma. British journal of haematology 141, 672-675.

Lee, R.C., Feinbaum, R.L., Ambros, V., 1993. The C. elegans heterochronic gene lin-4 encodes small RNAs with antisense complementarity to lin-14. Cell 75, 843-854.

Leivonen, S.K., Sahlberg, K.K., Makela, R., Due, E.U., Kallioniemi, O., Borresen-Dale, A.L., Perala, M., 2014. High-throughput screens identify microRNAs essential for HER2 positive breast cancer cell growth. Mol Oncol 8, 93-104.

Li, D., Zhao, Y., Liu, C., Chen, X., Qi, Y., Jiang, Y., Zou, C., Zhang, X., Liu, S., Wang, X., Zhao, D., Sun, Q., Zeng, Z., Dress, A., Lin, M.C., Kung, H.F., Rui, H., Liu, L.Z., Mao, F., Jiang, B.H., Lai, L., 2011. Analysis of MiR-195 and MiR-497 expression, regulation and role in breast cancer. Clinical cancer research : an official journal of the American Association for Cancer Research 17, 1722-1730.

Li, P., Sheng, C., Huang, L., Zhang, H., Huang, L., Cheng, Z., Zhu, Q., 2014. MiR-183/-96/-182 cluster is up-regulated in most breast cancers and increases cell proliferation and migration. Breast cancer research : BCR 16, 473.

Liedtke, C., Mazouni, C., Hess, K.R., Andre, F., Tordai, A., Mejia, J.A., Symmans, W.F., Gonzalez-Angulo, A.M., Hennessy, B., Green, M., Cristofanilli, M., Hortobagyi, G.N., Pusztai, L., 2008. Response to neoadjuvant therapy and long-term survival in patients with triple-negative breast cancer. Journal of clinical oncology : official journal of the American Society of Clinical Oncology 26, 1275-1281.

Liu, A., Tetzlaff, M.T., Vanbelle, P., Elder, D., Feldman, M., Tobias, J.W., Sepulveda, A.R., Xu, X., 2009. MicroRNA expression profiling outperforms mRNA expression profiling in formalin-fixed paraffin-embedded tissues. International journal of clinical and experimental pathology 2, 519-527.

Lowery, A.J., Miller, N., Devaney, A., McNeill, R.E., Davoren, P.A., Lemetre, C., Benes, V., Schmidt, S., Blake, J., Ball, G., Kerin, M.J., 2009. MicroRNA signatures predict oestrogen receptor, progesterone receptor and HER2/neu receptor status in breast cancer. Breast cancer research : BCR 11, R27.

Lu, J., Getz, G., Miska, E.A., Alvarez-Saavedra, E., Lamb, J., Peck, D., Sweet-Cordero, A., Ebert, B.L., Mak, R.H., Ferrando, A.A., Downing, J.R., Jacks, T., Horvitz, H.R., Golub, T.R., 2005. MicroRNA expression profiles classify human cancers. Nature 435, 834-838.

Lu, Z., Ye, Y., Jiao, D., Qiao, J., Cui, S., Liu, Z., 2012. miR-155 and miR-31 are differentially expressed in breast cancer patients and are correlated with the estrogen receptor and progesterone receptor status. Oncology letters 4, 1027-1032.

Luo, Q., Wei, C., Li, X., Li, J., Chen, L., Huang, Y., Song, H., Li, D., Fang, L., 2014. MicroRNA-195-5p is a potential diagnostic and therapeutic target for breast cancer. Oncology reports 31, 1096-1102.

Mar-Aguilar, F., Mendoza-Ramirez, J.A., Malagon-Santiago, I., Espino-Silva, P.K., Santuario-Facio, S.K., Ruiz-Flores, P., Rodriguez-Padilla, C., Resendez-Perez, D., 2013. Serum circulating microRNA profiling for identification of potential breast cancer biomarkers. Dis Markers 34, 163-169.

Mattie, M.D., Benz, C.C., Bowers, J., Sensinger, K., Wong, L., Scott, G.K., Fedele, V., Ginzinger, D., Getts, R., Haqq, C., 2006. Optimized high-throughput microRNA expression profiling provides novel biomarker assessment of clinical prostate and breast cancer biopsies. Molecular cancer 5, 24.

McPherson, K., Steel, C.M., Dixon, J.M., 2000. ABC of breast diseases. Breast cancer-epidemiology, risk factors, and genetics. Bmj 321, 624-628.

Melchor, L., Benitez, J., 2013. The complex genetic landscape of familial breast cancer. Human genetics 132, 845-863.

Michailidou, K., Beesley, J., Lindstrom, S., Canisius, S., Dennis, J., Lush, M.J., Maranian, M.J., Bolla, M.K., Wang, Q., Shah, M., Perkins, B.J., Czene, K., Eriksson, M., Darabi, H., Brand, J.S., Bojesen, S.E., Nordestgaard, B.G., Flyger, H., Nielsen, S.F., Rahman, N., Turnbull, C., Bocs, Fletcher, O., Peto, J., Gibson, L., Dos-Santos-Silva, I., Chang-Claude, J., Flesch-Janys, D., Rudolph, A., Eilber, U., Behrens, S., Nevanlinna, H., Muranen, T.A., Aittomaki, K., Blomqvist, C., Khan, S., Aaltonen, K., Ahsan, H., Kibriya, M.G., Whittemore, A.S., John, E.M., Malone, K.E., Gammon, M.D., Santella, R.M., Ursin, G., Makalic, E., Schmidt, D.F., Casey, G., Hunter, D.J., Gapstur, S.M., Gaudet, M.M., Diver, W.R., Haiman, C.A., Schumacher, F., Henderson, B.E., Le Marchand, L., Berg, C.D., Chanock, S.J., Figueroa, J., Hoover, R.N., Lambrechts, D., Neven, P., Wildiers, H., van Limbergen, E., Schmidt, M.K., Broeks, A., Verhoef, S., Cornelissen, S., Couch, F.J., Olson, J.E., Hallberg, E., Vachon, C., Waisfisz, Q., Meijers-Heijboer, H., Adank, M.A., van der Luijt, R.B., Li, J., Liu, J., Humphreys, K., Kang, D., Choi, J.Y., Park, S.K., Yoo, K.Y., Matsuo, K., Ito, H., Iwata, H., Tajima, K., Guenel, P., Truong, T., Mulot, C., Sanchez, M., Burwinkel, B., Marme, F., Surowy, H., Sohn, C., Wu, A.H., Tseng, C.C., Van Den Berg, D., Stram, D.O., Gonzalez-Neira, A., Benitez, J., Zamora, M.P., Perez, J.I., Shu, X.O., Lu, W., Gao, Y.T., Cai, H., Cox, A., Cross, S.S., Reed, M.W., Andrulis, I.L., Knight, J.A., Glendon, G., Mulligan, A.M., Sawyer, E.J., Tomlinson, I., Kerin, M.J., Miller, N., kConFab, I., Group, A., Lindblom, A., Margolin, S., Teo, S.H., Yip, C.H., Taib, N.A., Tan, G.H., Hooning, M.J., Hollestelle, A., Martens, J.W., Collee, J.M., Blot, W., Signorello, L.B., Cai, Q., Hopper, J.L., Southey, M.C., Tsimiklis, H., Apicella, C., Shen, C.Y., Hsiung, C.N., Wu, P.E., Hou, M.F., Kristensen, V.N., Nord, S., Alnaes, G.I., Nbcs, Giles, G.G., Milne, R.L., McLean, C., Canzian, F., Trichopoulos, D., Peeters, P., Lund, E., Sund, M., Khaw, K.T., Gunter, M.J., Palli, D., Mortensen, L.M., Dossus, L., Huerta, J.M., Meindl, A., Schmutzler, R.K., Sutter, C., Yang, R., Muir, K., Lophatananon, A., Stewart-Brown, S., Siriwanarangsan, P., Hartman, M., Miao, H., Chia, K.S., Chan, C.W., Fasching, P.A., Hein, A., Beckmann, M.W., Haeberle, L., Brenner, H., Dieffenbach, A.K., Arndt, V., Stegmaier, C., Ashworth, A., Orr, N., Schoemaker, M.J., Swerdlow, A.J., Brinton, L., Garcia-Closas, M., Zheng, W., Halverson, S.L., Shrubsole, M., Long, J., Goldberg, M.S., Labreche, F., Dumont, M., Winqvist, R., Pylkas, K., Jukkola-Vuorinen, A., Grip, M., Brauch, H., Hamann, U., Bruning, T., Network, G., Radice, P., Peterlongo, P., Manoukian, S., Bernard, L., Bogdanova, N.V., Dork, T., Mannermaa, A., Kataja, V., Kosma, V.M., Hartikainen, J.M., Devilee, P., Tollenaar, R.A., Seynaeve, C., Van Asperen, C.J., Jakubowska, A., Lubinski, J., Jaworska, K., Huzarski, T., Sangrajrang, S., Gaborieau, V., Brennan, P., McKay, J., Slager, S., Toland, A.E., Ambrosone, C.B., Yannoukakos, D., Kabisch, M., Torres, D., Neuhausen, S.L., Anton-Culver, H., Luccarini, C., Baynes, C., Ahmed, S., Healey, C.S., Tessier, D.C., Vincent, D., Bacot, F., Pita, G., Alonso, M.R., Alvarez, N., Herrero, D., Simard, J., Pharoah, P.P., Kraft, P., Dunning, A.M., Chenevix-Trench, G., Hall, P., Easton, D.F., 2015. Genome-wide association analysis of more than 120,000 individuals identifies 15 new susceptibility loci for breast cancer. Nature genetics.

Miki, Y., Swensen, J., Shattuck-Eidens, D., Futreal, P.A., Harshman, K., Tavtigian, S., Liu, Q., Cochran, C., Bennett, L.M., Ding, W., et al., 1994. A strong candidate for the breast and ovarian cancer susceptibility gene BRCA1. Science 266, 66-71.

Moskwa, P., Buffa, F.M., Pan, Y., Panchakshari, R., Gottipati, P., Muschel, R.J., Beech, J., Kulshrestha, R., Abdelmohsen, K., Weinstock, D.M., Gorospe, M., Harris, A.L., Helleday, T., Chowdhury, D.,

2011. miR-182-mediated downregulation of BRCA1 impacts DNA repair and sensitivity to PARP inhibitors. Molecular cell 41, 210-220.

Mueller, C.R., Roskelley, C.D., 2003. Regulation of BRCA1 expression and its relationship to sporadic breast cancer. Breast cancer research : BCR 5, 45-52.

Mulrane, L., Madden, S.F., Brennan, D.J., Gremel, G., McGee, S.F., McNally, S., Martin, F., Crown, J.P., Jirstrom, K., Higgins, D.G., Gallagher, W.M., O'Connor, D.P., 2012. miR-187 is an independent prognostic factor in breast cancer and confers increased invasive potential in vitro. Clin Cancer Res 18, 6702-6713.

Mulligan, A.M., Pinnaduwage, D., Bull, S.B., O'Malley, F.P., Andrulis, I.L., 2008. Prognostic effect of basal-like breast cancers is time dependent: evidence from tissue microarray studies on a lymph node-negative cohort. Clin Cancer Res 14, 4168-4174.

Narod, S.A., Foulkes, W.D., 2004. BRCA1 and BRCA2: 1994 and beyond. Nature reviews. Cancer 4, 665-676.

National Institutes of Health Consensus Development, P., 2001. National Institutes of Health Consensus Development Conference statement: adjuvant therapy for breast cancer, November 1-3, 2000. Journal of the National Cancer Institute. Monographs, 5-15.

Network, T.C.G.A., 2012. Comprehensive molecular portraits of human breast tumours. Nature 490, 61-70.

Ng, E.K., Li, R., Shin, V.Y., Jin, H.C., Leung, C.P., Ma, E.S., Pang, R., Chua, D., Chu, K.M., Law, W.L., Law, S.Y., Poon, R.T., Kwong, A., 2013. Circulating microRNAs as specific biomarkers for breast cancer detection. PLoS One 8, e53141.

Nielsen, T.O., Parker, J.S., Leung, S., Voduc, D., Ebbert, M., Vickery, T., Davies, S.R., Snider, J., Stijleman, I.J., Reed, J., Cheang, M.C., Mardis, E.R., Perou, C.M., Bernard, P.S., Ellis, M.J., 2010. 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 16, 5222-5232.

Paik, S., Shak, S., Tang, G., Kim, C., Baker, J., Cronin, M., Baehner, F.L., Walker, M.G., Watson, D., Park, T., Hiller, W., Fisher, E.R., Wickerham, D.L., Bryant, J., Wolmark, N., 2004. A multigene assay to predict recurrence of tamoxifen-treated, node-negative breast cancer. The New England journal of medicine 351, 2817-2826.

Parker, J.S., Mullins, M., Cheang, M.C., Leung, S., Voduc, D., Vickery, T., Davies, S., Fauron, C., He, X., Hu, Z., Quackenbush, J.F., Stijleman, I.J., Palazzo, J., Marron, J.S., Nobel, A.B., Mardis, E., Nielsen, T.O., Ellis, M.J., Perou, C.M., Bernard, P.S., 2009. Supervised risk predictor of breast cancer based on intrinsic subtypes. Journal of clinical oncology : official journal of the American Society of Clinical Oncology 27, 1160-1167.

Perez-Rivas, L.G., Jerez, J.M., Carmona, R., de Luque, V., Vicioso, L., Claros, M.G., Viguera, E., Pajares, B., Sanchez, A., Ribelles, N., Alba, E., Lozano, J., 2014. A microRNA signature associated with early recurrence in breast cancer. PloS one 9, e91884.

Perou, C.M., Sorlie, T., Eisen, M.B., van de Rijn, M., Jeffrey, S.S., Rees, C.A., Pollack, J.R., Ross, D.T., Johnsen, H., Akslen, L.A., Fluge, O., Pergamenschikov, A., Williams, C., Zhu, S.X., Lonning, P.E., Borresen-Dale, A.L., Brown, P.O., Botstein, D., 2000. Molecular portraits of human breast tumours. Nature 406, 747-752.

Persson, H., Kvist, A., Rego, N., Staaf, J., Vallon-Christersson, J., Luts, L., Loman, N., Jonsson, G., Naya, H., Hoglund, M., Borg, A., Rovira, C., 2011. Identification of new microRNAs in paired normal and tumor breast tissue suggests a dual role for the ERBB2/Her2 gene. Cancer research 71, 78-86.

Pigati, L., Yaddanapudi, S.C., Iyengar, R., Kim, D.J., Hearn, S.A., Danforth, D., Hastings, M.L., Duelli, D.M., 2010. Selective release of microRNA species from normal and malignant mammary epithelial cells. PLoS One 5, e13515.

Prat, A., Cheang, M.C., Martin, M., Parker, J.S., Carrasco, E., Caballero, R., Tyldesley, S., Gelmon, K., Bernard, P.S., Nielsen, T.O., Perou, C.M., 2013. Prognostic significance of progesterone

receptor-positive tumor cells within immunohistochemically defined luminal A breast cancer. Journal of clinical oncology : official journal of the American Society of Clinical Oncology 31, 203-209.

Prat, A., Parker, J.S., Karginova, O., Fan, C., Livasy, C., Herschkowitz, J.I., He, X., Perou, C.M., 2010. Phenotypic and molecular characterization of the claudin-low intrinsic subtype of breast cancer. Breast Cancer Res 12, R68.

Rakha, E.A., El-Sayed, M.E., Green, A.R., Lee, A.H., Robertson, J.F., Ellis, I.O., 2007. Prognostic markers in triple-negative breast cancer. Cancer 109, 25-32.

Reis, P.P., Waldron, L., Goswami, R.S., Xu, W., Xuan, Y., Perez-Ordonez, B., Gullane, P., Irish, J., Jurisica, I., Kamel-Reid, S., 2011. mRNA transcript quantification in archival samples using multiplexed, color-coded probes. BMC biotechnology 11, 46.

Romualdi, C., Campanaro, S., Campagna, D., Celegato, B., Cannata, N., Toppo, S., Valle, G., Lanfranchi, G., 2003. Pattern recognition in gene expression profiling using DNA array: a comparative study of different statistical methods applied to cancer classification. Hum Mol Genet 12, 823-836.

Roth, C., Rack, B., Muller, V., Janni, W., Pantel, K., Schwarzenbach, H., 2010. Circulating microRNAs as blood-based markers for patients with primary and metastatic breast cancer. Breast Cancer Res 12, R90.

Rothe, F., Ignatiadis, M., Chaboteaux, C., Haibe-Kains, B., Kheddoumi, N., Majjaj, S., Badran, B., Fayyad-Kazan, H., Desmedt, C., Harris, A.L., Piccart, M., Sotiriou, C., 2011. Global microRNA expression profiling identifies MiR-210 associated with tumor proliferation, invasion and poor clinical outcome in breast cancer. PloS one 6, e20980.

Sandhu, R., Rivenbark, A.G., Mackler, R.M., Livasy, C.A., Coleman, W.B., 2014. Dysregulation of microRNA expression drives aberrant DNA hypermethylation in basal-like breast cancer. International journal of oncology 44, 563-572.

Santarpia, L., Calin, G.A., Adam, L., Ye, L., Fusco, A., Giunti, S., Thaller, C., Paladini, L., Zhang, X., Jimenez, C., Trimarchi, F., El-Naggar, A.K., Gagel, R.F., 2013. A miRNA signature associated with human metastatic medullary thyroid carcinoma. Endocr Relat Cancer 20, 809-823.

Schnitt, S.J., 2010. Classification and prognosis of invasive breast cancer: from morphology to molecular taxonomy. Modern pathology : an official journal of the United States and Canadian Academy of Pathology, Inc 23 Suppl 2, S60-64.

Sempere, L.F., Christensen, M., Silahtaroglu, A., Bak, M., Heath, C.V., Schwartz, G., Wells, W., Kauppinen, S., Cole, C.N., 2007. Altered MicroRNA expression confined to specific epithelial cell subpopulations in breast cancer. Cancer research 67, 11612-11620.

Servick, K., 2014. Breast cancer. Breast cancer: a world of differences. Science 343, 1452-1453.

Sharp, P.C., Michielutte, R., Freimanis, R., Cunningham, L., Spangler, J., Burnette, V., 2003. Reported pain following mammography screening. Arch Intern Med 163, 833-836.

Shen, J., Ambrosone, C.B., DiCioccio, R.A., Odunsi, K., Lele, S.B., Zhao, H., 2008. A functional polymorphism in the miR-146a gene and age of familial breast/ovarian cancer diagnosis. Carcinogenesis 29, 1963-1966.

Si, H., Sun, X., Chen, Y., Cao, Y., Chen, S., Wang, H., Hu, C., 2013. Circulating microRNA-92a and microRNA-21 as novel minimally invasive biomarkers for primary breast cancer. J Cancer Res Clin Oncol 139, 223-229.

Skog, J., Wurdinger, T., van Rijn, S., Meijer, D.H., Gainche, L., Sena-Esteves, M., Curry, W.T., Jr., Carter, B.S., Krichevsky, A.M., Breakefield, X.O., 2008. Glioblastoma microvesicles transport RNA and proteins that promote tumour growth and provide diagnostic biomarkers. Nature cell biology 10, 1470-1476.

Sorlie, T., Perou, C.M., Tibshirani, R., Aas, T., Geisler, S., Johnsen, H., Hastie, T., Eisen, M.B., van de Rijn, M., Jeffrey, S.S., Thorsen, T., Quist, H., Matese, J.C., Brown, P.O., Botstein, D., Lonning, P.E., Borresen-Dale, A.L., 2001. Gene expression patterns of breast carcinomas distinguish

tumor subclasses with clinical implications. Proceedings of the National Academy of Sciences of the United States of America 98, 10869-10874.

Sorlie, T., Tibshirani, R., Parker, J., Hastie, T., Marron, J.S., Nobel, A., Deng, S., Johnsen, H., Pesich, R., Geisler, S., Demeter, J., Perou, C.M., Lonning, P.E., Brown, P.O., Borresen-Dale, A.L., Botstein, D., 2003. Repeated observation of breast tumor subtypes in independent gene expression data sets. Proceedings of the National Academy of Sciences of the United States of America 100, 8418-8423.

Sparano, J.A., Paik, S., 2008. Development of the 21-gene assay and its application in clinical practice and clinical trials. Journal of clinical oncology : official journal of the American Society of Clinical Oncology 26, 721-728.

Stephens, P.J., Tarpey, P.S., Davies, H., Van Loo, P., Greenman, C., Wedge, D.C., Nik-Zainal, S., Martin, S., Varela, I., Bignell, G.R., Yates, L.R., Papaemmanuil, E., Beare, D., Butler, A., Cheverton, A., Gamble, J., Hinton, J., Jia, M., Jayakumar, A., Jones, D., Latimer, C., Lau, K.W., McLaren, S., McBride, D.J., Menzies, A., Mudie, L., Raine, K., Rad, R., Chapman, M.S., Teague, J., Easton, D., Langerod, A., Oslo Breast Cancer, C., Lee, M.T., Shen, C.Y., Tee, B.T., Huimin, B.W., Broeks, A., Vargas, A.C., Turashvili, G., Martens, J., Fatima, A., Miron, P., Chin, S.F., Thomas, G., Boyault, S., Mariani, O., Lakhani, S.R., van de Vijver, M., van 't Veer, L., Foekens, J., Desmedt, C., Sotiriou, C., Tutt, A., Caldas, C., Reis-Filho, J.S., Aparicio, S.A., Salomon, A.V., Borresen-Dale, A.L., Richardson, A.L., Campbell, P.J., Futreal, P.A., Stratton, M.R., 2012. The landscape of cancer genes and mutational processes in breast cancer. Nature 486, 400-404.

Tan, X., Peng, J., Fu, Y., An, S., Rezaei, K., Tabbara, S., Teal, C.B., Man, Y.G., Brem, R.F., Fu, S.W., 2014. miR-638 mediated regulation of BRCA1 affects DNA repair and sensitivity to UV and cisplatin in triple-negative breast cancer. Breast cancer research : BCR 16, 435.

Tanic, M., Yanowski, K., Gomez-Lopez, G., Rodriguez-Pinilla, M.S., Marquez-Rodas, I., Osorio, A., Pisano, D.G., Martinez-Delgado, B., Benitez, J., 2015. MicroRNA expression signatures for the prediction of BRCA1/2 mutation-associated hereditary breast cancer in paraffin-embedded formalin-fixed breast tumors. International journal of cancer. Journal international du cancer 136, 593-602.

Tanic, M., Yanowsky, K., Rodriguez-Antona, C., Andres, R., Marquez-Rodas, I., Osorio, A., Benitez, J., Martinez-Delgado, B., 2012. Deregulated miRNAs in hereditary breast cancer revealed a role for miR-30c in regulating KRAS oncogene. PloS one 7, e38847.

The Cancer Genome Atlas, N., 2012. Comprehensive molecular portraits of human breast tumours. Nature 490, 61-70.

Thompson, M.E., Jensen, R.A., Obermiller, P.S., Page, D.L., Holt, J.T., 1995. Decreased expression of BRCA1 accelerates growth and is often present during sporadic breast cancer progression. Nature genetics 9, 444-450.

Turner, N.C., Reis-Filho, J.S., 2006. Basal-like breast cancer and the BRCA1 phenotype. Oncogene 25, 5846-5853.

Turner, N.C., Reis-Filho, J.S., Russell, A.M., Springall, R.J., Ryder, K., Steele, D., Savage, K., Gillett, C.E., Schmitt, F.C., Ashworth, A., Tutt, A.N., 2007. BRCA1 dysfunction in sporadic basal-like breast cancer. Oncogene 26, 2126-2132.

Valadi, H., Ekstrom, K., Bossios, A., Sjostrand, M., Lee, J.J., Lotvall, J.O., 2007. Exosome-mediated transfer of mRNAs and microRNAs is a novel mechanism of genetic exchange between cells. Nature cell biology 9, 654-659.

van Schooneveld, E., Wouters, M.C., Van der Auwera, I., Peeters, D.J., Wildiers, H., Van Dam, P.A., Vergote, I., Vermeulen, P.B., Dirix, L.Y., Van Laere, S.J., 2012. Expression profiling of cancerous and normal breast tissues identifies microRNAs that are differentially expressed in serum from patients with (metastatic) breast cancer and healthy volunteers. Breast Cancer Res 14, R34.

Visvader, J.E., 2009. Keeping abreast of the mammary epithelial hierarchy and breast tumorigenesis. Genes Dev 23, 2563-2577.

Volinia, S., Calin, G.A., Liu, C.G., Ambs, S., Cimmino, A., Petrocca, F., Visone, R., Iorio, M., Roldo, C., Ferracin, M., Prueitt, R.L., Yanaihara, N., Lanza, G., Scarpa, A., Vecchione, A., Negrini, M., Harris, C.C., Croce, C.M., 2006. A microRNA expression signature of human solid tumors defines cancer gene targets. Proceedings of the National Academy of Sciences of the United States of America 103, 2257-2261.

Volinia, S., Galasso, M., Sana, M.E., Wise, T.F., Palatini, J., Huebner, K., Croce, C.M., 2012. Breast cancer signatures for invasiveness and prognosis defined by deep sequencing of microRNA. Proceedings of the National Academy of Sciences of the United States of America 109, 3024-3029.

Wang, B., Zhang, Q., 2012. The expression and clinical significance of circulating microRNA-21 in serum of five solid tumors. J Cancer Res Clin Oncol 138, 1659-1666.

Wang, H., Tan, G., Dong, L., Cheng, L., Li, K., Wang, Z., Luo, H., 2012. Circulating MiR-125b as a marker predicting chemoresistance in breast cancer. PloS one 7, e34210.

Wessels, L.F., Reinders, M.J., Hart, A.A., Veenman, C.J., Dai, H., He, Y.D., van't Veer, L.J., 2005. A protocol for building and evaluating predictors of disease state based on microarray data. Bioinformatics 21, 3755-3762.

Wooster, R., Neuhausen, S.L., Mangion, J., Quirk, Y., Ford, D., Collins, N., Nguyen, K., Seal, S., Tran, T., Averill, D., et al., 1994. Localization of a breast cancer susceptibility gene, BRCA2, to chromosome 13q12-13. Science 265, 2088-2090.

Xiong, J., Yu, D., Wei, N., Fu, H., Cai, T., Huang, Y., Wu, C., Zheng, X., Du, Q., Lin, D., Liang, Z., 2010. An estrogen receptor alpha suppressor, microRNA-22, is downregulated in estrogen receptor alpha-positive human breast cancer cell lines and clinical samples. The FEBS journal 277, 1684-1694.

Xu, X., Wagner, K.U., Larson, D., Weaver, Z., Li, C., Ried, T., Hennighausen, L., Wynshaw-Boris, A., Deng, C.X., 1999. Conditional mutation of Brca1 in mammary epithelial cells results in blunted ductal morphogenesis and tumour formation. Nature genetics 22, 37-43.

Yamamoto, Y., Yoshioka, Y., Minoura, K., Takahashi, R.U., Takeshita, F., Taya, T., Horii, R., Fukuoka, Y., Kato, T., Kosaka, N., Ochiya, T., 2011. An integrative genomic analysis revealed the relevance of microRNA and gene expression for drug-resistance in human breast cancer cells. Molecular cancer 10, 135.

Yan, L.X., Huang, X.F., Shao, Q., Huang, M.Y., Deng, L., Wu, Q.L., Zeng, Y.X., Shao, J.Y., 2008. MicroRNA miR-21 overexpression in human breast cancer is associated with advanced clinical stage, lymph node metastasis and patient poor prognosis. Rna 14, 2348-2360.

Yang, G., Wu, D., Zhu, J., Jiang, O., Shi, Q., Tian, J., Weng, Y., 2013. Upregulation of miR-195 increases the sensitivity of breast cancer cells to Adriamycin treatment through inhibition of Raf-1. Oncology reports 30, 877-889.

Yu, F., Yao, H., Zhu, P., Zhang, X., Pan, Q., Gong, C., Huang, Y., Hu, X., Su, F., Lieberman, J., Song, E., 2007. let-7 regulates self renewal and tumorigenicity of breast cancer cells. Cell 131, 1109-1123.

Zhang, Y., Yan, L.X., Wu, Q.N., Du, Z.M., Chen, J., Liao, D.Z., Huang, M.Y., Hou, J.H., Wu, Q.L., Zeng, M.S., Huang, W.L., Zeng, Y.X., Shao, J.Y., 2011. miR-125b is methylated and functions as a tumor suppressor by regulating the ETS1 proto-oncogene in human invasive breast cancer. Cancer Res 71, 3552-3562.

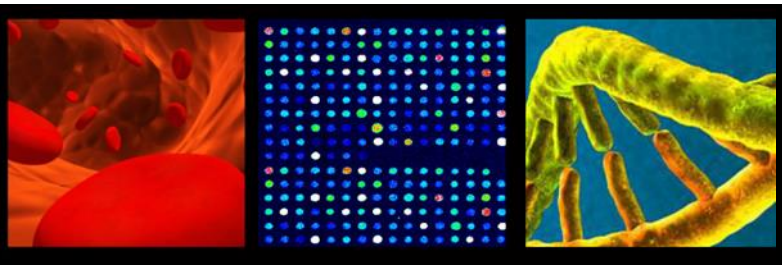
Zhao, J., Lei, T., Xu, C., Li, H., Ma, W., Yang, Y., Fan, S., Liu, Y., 2013. MicroRNA-187, down-regulated in clear cell renal cell carcinoma and associated with lower survival, inhibits cell growth and migration though targeting B7-H3. Biochem Biophys Res Commun 438, 439-444.

Zhao, J.J., Yang, J., Lin, J., Yao, N., Zhu, Y., Zheng, J., Xu, J., Cheng, J.Q., Lin, J.Y., Ma, X., 2009. Identification of miRNAs associated with tumorigenesis of retinoblastoma by miRNA microarray analysis. Childs Nerv Syst 25, 13-20.

Zheng, Y., Cui, L., Sun, W., Zhou, H., Yuan, X., Huo, M., Chen, J., Lou, Y., Guo, J., 2011. MicroRNA-21 is a new marker of circulating tumor cells in gastric cancer patients. Cancer biomarkers : section A of Disease markers 10, 71-77.

Zhu, S., Si, M.L., Wu, H., Mo, Y.Y., 2007. MicroRNA-21 targets the tumor suppressor gene tropomyosin 1 (TPM1). J Biol Chem 282, 14328-14336.

Zou, C., Xu, Q., Mao, F., Li, D., Bian, C., Liu, L.Z., Jiang, Y., Chen, X., Qi, Y., Zhang, X., Wang, X., Sun, Q., Kung, H.F., Lin, M.C., Dress, A., Wardle, F., Jiang, B.H., Lai, L., 2012. MiR-145 inhibits tumor angiogenesis and growth by N-RAS and VEGF. Cell cycle 11, 2137-2145.



Appendix I.
Supplementary materials

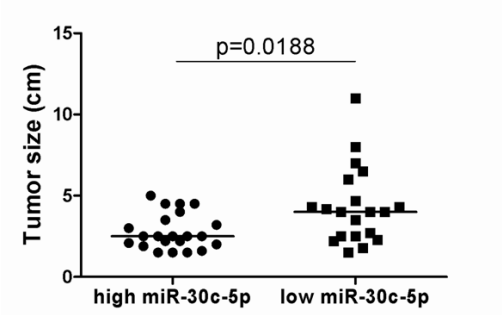
Supplementary Table S1. Specifically up or downregulated miRNAs in breast cancer molecular subtypes after comparison with normal breast tissues.

Molecular subtype	Deregulation	Number	miRNAs
Triple negative	Up	78	miR-642b-5p,miR-4795-3p,miR-4653-3p,miR-3124-3p,miR-4501,miR-4698,miR-1973,miR-3976,miR-4417,miR-634,miR-4723-5p,miR-150-5p,miR-4639-3p,miR-659-5p,miR-4633-5p,miR-5584-3p,miR-4449,miR-921,miR-5000-3p,miR-4707-3p,miR-513a-5p,miR-519e-5p,miR-552,miR-4431,miR-4677-3p,miR-4329,miR-1273e,miR-711,miR-4503,miR-3687,miR-492,miR-3591-5p,miR-5588-3p,miR-874,miR-181a-2-3p,miR-4636,miR-498,miR-4782-5p,miR-490-5p,miR-548ao-3p,miR-5193,miR-3648,miR-4511,miR-1827,miR-187-5p,miR-550b-2-5p,miR-4674,miR-3912,let-7d-3p,miR-2113,miR-4264,miR-675-3p,miR-5689,miR-4784,miR-1285-5p,miR-4694-5p,miR-5089,miR-5187-5p,miR-650,miR-506-5p,miR-891a,miR-1265,miR-1197,miR-3944-5p,miR-4683,miR-4536-3p,miR-3618,miR-4535,miR-4778-5p,miR-3925-5p,miR-1321,miR-4436b-5p,miR-3161,miR-5006-3p,miR-3606,miR-146a-5p,miR-5580-5p
	Down	27	miR-554,miR-4444,miR-629-5p,miR-10a-5p,miR-140-3p,miR-374a-5p,miR-124-5p,let-7d-5p,miR-4791,miR-548o-3p,miR-1202,miR-4328,miR-448,miR-382-3p,miR-29b-2-5p,miR-590-3p,miR-10b-5p,let-7i-5p,miR-4301,miR-126-3p,miR-199a-5p,miR-3607-3p,miR-214-3p,miR-491-3p,miR-4285,miR-5701,let-7a-5p
Her2	Up	0	-
	Down	1	miR-574-3p
Luminal B	Up	23	miR-16-5p,let-7g-5p,miR-23a-3p,miR-200c-3p,miR-34a-5p,miR-101-3p,miR-26b-5p,miR-193a-3p,miR-30b-5p,miR-20a-5p,miR-29a-3p,miR-15a-5p,miR-27b-3p,miR-93-5p,miR-4714-5p,miR-1280,miR-29b-3p,miR-15b-5p,miR-26a-5p,miR-374b-5p,miR-429,miR-107
	Down	16	miR-184,miR-4462,miR-4649-5p,miR-1299,miR-4433-3p,miR-4472,miR-302a-3p,miR-4507,miR-3646,miR-371b-5p,miR-4497,miR-4787-5p,miR-548ap-5p,miR-548j,miR-3940-5p,miR-4505,miR-1275
Luminal A	Up	5	miR-4421,miR-3667-5p,miR-5196-3p,miR-548k,miR-331-3p
	Down	12	miR-486-5p,miR-4492,miR-664-3p,miR-124-3p,miR-4446-5p,miR-3664-5p,miR-513b,miR-3620,miR-4769-3p,miR-4646-3p,miR-3182,miR-4723-3p
Common	Up	25	miR-21-5p,miR-3613-3p,miR-4668-5p,miR-106b-5p,miR-200a-3p,miR-5704,miR-1278,miR-182-5p,miR-1264,miR-96-5p,miR-190b,miR-3611,miR-185-5p,miR-339-5p,miR-7-5p,miR-340-5p,miR-203,miR-342-5p,miR-1244,miR-600,miR-760,miR-92a-2-5p,miR-219-2-3p,miR-553,miR-4521
	Down	27	miR-889,miR-192-3p,miR-432-5p,miR-1224-5p,miR-1205,miR-1228-5p,miR-571,miR-1229,miR-4253,miR-2276,miR-577,miR-548s,miR-744-3p,miR-1207-5p,miR-1270,miR-877-5p,miR-499a-5p,miR-574-5p,miR-718,miR-4695-3p,miR-3164,miR-3148,miR-3941,miR-4488,miR-4516,miR-3656,miR-125b-5p

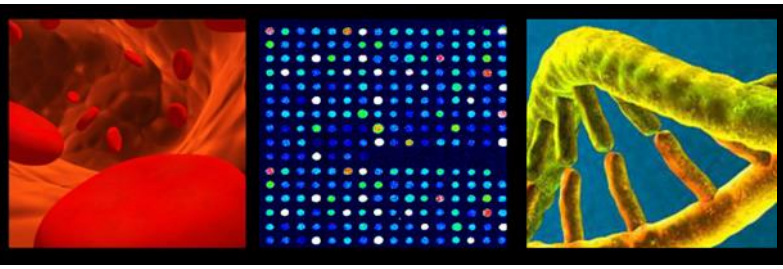
Supplementary Table S2. Comparison of the performance of different algorithms used for the classification of breast tumors and the different molecular subtypes.

Classifier	Algorithm	Accuracy	MCC	RMSE	AUC	miRNAs
Breast tumors vs normal breast tissues	**KNN**	0.95	0.74	0.18	0.91	25
	SVM	0.98	0.90	0.06	0.92	
	RF	0.92	0.49	0.22	0.91	
Triple negative vs rest of tumors	**KNN**	0.78	0.31	0.37	0.82	8
	SVM	0.98	0.95	0.05	0.97	
	RF	0.80	0.42	0.36	0.84	
Her2 vs rest of tumors	**KNN**	0.69	0.15	0.44	0.60	7
	SVM	0.98	0.94	0.05	0.96	
	RF	0.78	0.00	0.41	0.50	
Luminal B vs rest of tumors	**KNN**	0.76	0.38	0.41	0.72	16
	SVM	0.97	0.93	0.07	0.96	
	RF	0.80	0.45	0.41	0.67	
Luminal A vs rest of tumors	**KNN**	0.70	0.02	0.44	0.61	3
	SVM	0.98	0.97	0.04	0.98	
	RF	0.71	-0.05	0.44	0.57	

KNN: K Nearest Neighbor, SVM: Support Vector Machine, RF: Random Forest, MCC: Mathews Correlation Coefficient, RMSE: Root Median Square Error, AUC: Area Under the Curve.



Supplementary Figure S1. Differences in size of triple negative tumors expressing high or low levels of miR-30c-5p.



Appendix II.
Publications and posters

PUBLICATIONS

- **Matamala N**, Vargas MT, González-Cámpora R, Miñambres R, Arias JI, Menéndez P, Andrés-León E, Gómez-López G, Yanowsky K, Calvete-Candenas J, Inglada-Pérez L, Martínez-Delgado B, Benítez J. Tumor microRNA expression profiling identifies circulating microRNAs for earlier breast cancer detection. (under review in Clinical Chemistry).

POSTERS IN CONFERENCES

- **Matamala N**, Vargas MT, González-Cámpora R, Arias JI, Menéndez P, Andrés E, Yanowsky K, Miñambres R, Martínez-Delgado B, Benítez J (2014). Tumor miRNA expression profiling to identify circulating miRNAs for breast cancer detection. European Association for Cancer Research, Munich.
- Carlos Rey, Irene Vázquez, Laura Pérez, Miljana Tanic, Eduardo Andrés, **Nerea Matamala**, José Silva, Javier Benítez, Beatriz Martínez-Delgado (2014). *BRCA1* regulating microRNAs. Familial Cancer Conference, Madrid.
- **Matamala N**, Vargas MT, González-Cámpora R, Arias JI, Menéndez P, Andrés E, Yanowsky K, Calvete-Candenas J, Miñambres R, Martínez-Delgado B, Benítez J (2014). Identification of circulating miRNAs as potential biomarkers for breast cancer detection. Familial Cancer Conference, Madrid.
- **Matamala Nerea**, Vargas de los Monteros MT, González-Campora R, Calvete-Candenas J, Gil-Borja MT, Ballester de Matías A, Miñambres Herráiz R, Santillán Garzón S, Tanic M, Yanowsky Ruiz K, Martínez-Delgado B, Benítez Ortiz J (2013). MicroRNA signatures in breast cancer molecular subtypes. CNIOSA Lab Day, Madrid.
- K Yanowsky, **N Matamala**, M Tanic, I Marquez-Rodas, A Barroso, V Fernandez, J Benitez, B Martínez-Delgado (2013). Identification of tumor microRNAs in plasma samples from hereditary breast cancer patients. European Society of Human Genetics, Paris.
- **Matamala Nerea**, Vargas de los Monteros MT, Gonzalez-Campora R, Gil-Borja MT, Ballester de Matías A, Miñambres Herráiz R, Santillan Garzón S, Yanowsky Ruiz K, Martinez-Delgado B, Benitez Ortiz J (2013). Desafío integral el cáncer de mama: hacia

una medicina personalizada basada en el análisis de miRNAs. IX Simposio Internacional de GEICAM, Valencia.

- B Martínez-Delgado, **N Matamala**, K Yanowsky, M Tanic, I Marquez-Rodas, A Barroso, M González-Neira, V Fernandez, J Benitez (2013). Identificación de microRNAs tumorales en muestras de plasma de pacientes con cáncer de mama hereditario. Asociación Española de Genética Humana, Madrid.