Bruno Filipe da Silva Santos

Degree in Biomedical Sciences in Histocellular Pathology

Lisbon School of Health Technology - Polytechnic Institute of Lisbon

Immunohistochemistry detection of putative miR-200c and miR-203 Targets in Breast Cancer Patients

Dissertation for Master's degree in Molecular Genetics and Biomedicine

Supervisor: António Sebastião Rodrigues, PhD, Assistant Professor, CIGMH/FCM-UNL

Co- Supervisor: Manuela Martins, MD, Pathologist, CHLC-HSJ

Co- Supervisor: Bruno Costa Gomes, MSc, CIGMH/FCM-UNL

Jury:

President: Margarida Casal Ribeiro Castro Caldas Braga, PhD

Arguer: Maria Alexandra Núncio de Carvalho Ramos Fernandes, PhD

Member: António Sebastião Rodrigues, PhD

FACULDADE DE CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE NOVA DE LISBOA

November 2014

Bruno Filipe da Silva Santos

Degree in Biomedical Sciences in Histocellular Pathology
Lisbon School of Health Technology - Polytechnic Institute of Lisbon

Immunohistochemistry detection of putative miR-200c and miR-203 Targets in Breast Cancer Patients

Dissertation for Master's degree in Molecular Genetics and Biomedicine

Supervisor: António Sebastião Rodrigues, PhD, Assistant Professor, CIGMH/FCM-UNL
Co- Supervisor: Manuela Martins, MD, Pathologist, CHLC-HSJ
Co- Supervisor: Bruno Costa Gomes, MSc, CIGMH/FCM-UNL

Jury:
President: Margarida Casal Ribeiro Castro Caldas Braga, PhD
Arguer: Maria Alexandra Núncio de Carvalho Ramos Fernandes, PhD
Member: António Sebastião Rodrigues, PhD

November 2014

FACULDADE DE CIÊNCIAS E TECNOLOGIA
UNIVERSIDADE NOVA DE LISBOA

Immunohistochemistry detection of putative miR-200c and miR-203 Targets in Breast Cancer Patients

“Your work is going to fill a large part of your life, and the only way to be truly satisfied is to do what you believe is great work. And the only way to do great work is to love what you do.”

Steve Jobs

Part of this thesis will be published in the following:

Gomes, B.C., Santos, B. 2015. Methods for studying microRNAs expression and their targets in formalin-fixed, paraffin-embedded (FFPE) breast cancer tissues. *In* Methods in Molecular Biology: Cancer Drug Resistance (Rueff, J. & Rodrigues, A.S. eds), Springer Protocols.

Acknowledgment

Most people whom I intend to demonstrate my appreciation primarily speak Portuguese thus I chose to write my acknowledgment in Portuguese.

Em primeiro lugar, gostaria de demonstrar o meu apreço ao Centro de Investigação em Genética Molecular Humana, Departamento de Genética – FCM, UNL, por me ter acolhido no âmbito da realização do meu estágio de preparação de dissertação, sendo de mencionar a sua importância no que diz respeito à cedência de espaços, materiais, reagentes, e equipamentos, indispensáveis à execução do mesmo. Não obstante, estendo o meu apreço à Unidade de Anatomia Patológica do Hospital de S. José que connosco colaborou, prestando sempre que possível o apoio necessário e ao Serviço de Anatomia Patológica do Hospital de Egas Moniz que tão hospitaleiramente me acolheu e ajudou a ultrapassar obstáculos inerentes ao decorrer da tese.

Deixo um profundo agradecimento aos meus três orientadores – Prof. Sebastião Rodrigues, Drª Manuela Martins e Dr. Bruno Gomes – que sempre estiveram ao meu lado e me apoiaram no decorrer deste trabalho e que sempre acreditaram nas minhas capacidades, motivando-me para ultrapassar todas as dificuldades encontradas. Sem eles não seria possível a entrega desta dissertação para a qual muito contribuíram com os seus conhecimentos e experiência.

Sem detrimento dos restantes não poderia deixar de realçar o apoio que o Bruno Gomes me deu durante este ano de estágio, não só do ponto de vista profissional como também pessoal, pelo qual eu lhe estou inteiramente grato.

À Inês Silva, minha parceira de laboratório, um imenso obrigado por ajudares a tornar todos os momentos mais divertidos e fáceis de encarar.

À Mariana Santos e Maria Jardim agradeço todos os momentos de descontração e conversas sem fim que ajudaram a relativizar os problemas e a seguir sempre com otimismo.

Em último lugar, não poderia deixar de agradecer aos meus familiares e amigos e que me apoiaram e suportaram ao longo deste processo, e sem os quais não teria sido possível chegar ao fim. Em particular agradeço ao Paulo Ferro por ter estado sempre ao meu lado a aturar-me mesmo nos meus piores momentos e por se demonstrar sempre prestável aos meus pedidos.

Abstract

The morphological variability and clinical evolution of breast cancer have prompted researchers to find a strategy to classify the disease and to possibly define supportive prognostic and predictive indicators. Recently, some studies have focused on the putative utility of miRNA as a novel class of cancer markers. Since potential targets of miRNAs are often provided only by bioinformatic tools there is a gap in this area of study. Therefore, parallel to the quantification of the expression of miR-203 and miR-200c in tumor tissue from patients from Central Lisbon Hospital with breast cancer, the aim of this dissertation was to analyze the expression of their putative targets – ATM, BMI1, SIX1 and SOX2 – by immunohistochemistry. 45 samples were analyzed corresponding to 43 patients whose mean age at diagnosis was 62 years. The most common type was invasive carcinoma NOS (71,1%) followed by invasive lobular carcinoma (8,9%). 86,4% of samples were ER positive, 79,1% PR positive, 13,6% HER2 positive and 45,5% high ki67. miR-200c was downregulated in 12,8% of the samples and upregulated in 23,1%. Comparatively, 20,5% of tumors presented miR-203 downregulation and 30,8% upregulation. Anti-ATM and anti-BMI1 antibodies didn't perform properly thus, they were not assessed. Regarding to SIX1 and SOX2, only 13.3% and 8.9% of tumors were positive, respectively. Furthermore, a statistically significant association between the expression of both proteins and various clinicopathological parameters was not found, except for the number of pregnancies that seems to be associated with SIX1 positivity ($p = 0.034$). Regarding the relationship between levels of miRNAs and expression of their putative targets, there was no statistically association. In the future a bigger sample size should be used to increase the robustness of results and patient's follow-up would allow evaluating the association between SIX1 and SOX2 with therapeutic outcome.

Keywords

ATM, BMI1, Breast Cancer, miR-200c, miR-203, SIX1, SOX2

Resumo

A variabilidade morfológica e clínica do cancro da mama tem levado diversos investigadores a procurar uma estratégia para melhor classificar a doença e, desejavelmente, definir indicadores prognósticos e preditivos de resposta à terapêutica. Recentemente, alguns estudos têm-se centrado na possível utilidade dos miRNAs como uma nova classe de marcadores de cancro. Potenciais alvos dos miRNAs são frequentemente identificados apenas por ferramentas bioinformáticas existindo uma lacuna nesta área de estudo. Por isso, paralelamente à quantificação da expressão de miR-203 e miR-200c no tecido tumoral de pacientes do Centro Hospitalar de Lisboa Central com cancro de mama, o objetivo deste trabalho foi analisar a expressão dos seus alvos putativos - ATM, BMI1, SIX1 e SOX2 - por imunohistoquímica. 45 amostras foram analisadas, correspondendo a 43 pacientes com idade média, no momento do diagnóstico, de 62 anos. O tipo mais frequente foi o carcinoma invasivo SOE (71,1%), seguido pelo carcinoma lobular invasivo (8,9%). 86,4% das amostras eram ER positivas, 79,1% PR, 13,6% HER2 positivas e 45,5% com Ki67 elevado. miR-200c estava sub-expresso em 12,8% das amostras e sobre-expresso em 23,1%. Relativamente ao miR-203 20,5% das amostras apresentaram sub-expressão e 30,8% sobre-expressão. Como não marcaram corretamente, os anticorpos anti-ATM e anti-BMI1 não foram avaliados. Relativamente ao SIX1 e SOX2, apenas 13,3% e 8,9% dos tumores foram positivos, respectivamente. Uma associação estatisticamente significativa entre a expressão de ambas as proteínas e os vários parâmetros clínicopatológicos também não foi encontrada, com exceção do número de gestações que parece estar associado com a positividade do SIX1 ($p = 0,034$). Quanto à relação entre os níveis de miRNAs e expressão dos seus potenciais alvos, não houve associação. No futuro, uma maior casuística deve ser usada para aumentar a robustez dos resultados e o follow-up da paciente permitiria avaliar a associação entre SIX1 e SOX2 e o resultado terapêutico.

Termos-chave

ATM, BMI1, Cancro da Mama, miR-200c, miR-203, SIX1, SOX2

Table of Contents

Index of Figures

Index of Tables

List of abbreviations, acronyms and symbols

-	Negative
%	Percentage
+	Positive
<	Lower than
>	Higher than
CIGMH	Centro de Investigação em Genética Molecular Humana
EMT	Epithelial-mesenchymal transition
ER	Estrogen Receptor
HER2	Human epidermal growth factor receptor 2
IHC	Immunohistochemistry
miRNA	microRNA
NOS	Not otherwise specified
PR	Progesterone Receptor
TNM	Tumor, (lymph) Node and Metastasis

1 Introduction

This dissertation, titled *"Immunohistochemistry detection of putative miR-200c and miR-203 Targets in Breast Cancer Patients"*, was held under the Master's Degree in Molecular Genetics and Biomedicine, FCT-UNL. It is inserted in a project that resulted from a partnership between the Centro de Investigação em Genética Molecular Humana (CIGMH), under the responsibility of professor Sebastião Rodrigues and Dr. Bruno Gomes, and the Surgical and Pathology Departments of the Breast Pathology Unit of Central Lisbon Hospital, under the responsibility of Dr.ª Paulina Lopes and Dr.ª Manuela Martins, respectively. The project involving the collection and analysis of human breast cancer samples, as well as surveys of patients, was approved by the Board of Directors of Central Lisbon Hospital and by the Ethics Committee thereof.

1.1 Anatomophysiology of the Breast

Breasts develop as downgrowths from the epidermis along the milk line which runs obliquely from the axilla toward the groin on each side. The nipple and its simple system of ducts are present at birth, but full development does not occur until puberty, and then usually only in females. Normally the male breast remains a rudimentary system of simple nipple ducts with a small amount of surrounding fibrocollagenous tissue. Continued estrogen secretion after onset of puberty leads to progressive enlargement and complexity of the breast (Fig. 1.1), due initially to an increase in adipose tissue, and then the ductular system of the nipple becomes more complex, with branches extending into the adipose tissue (Stevens & Lowe, 2005).

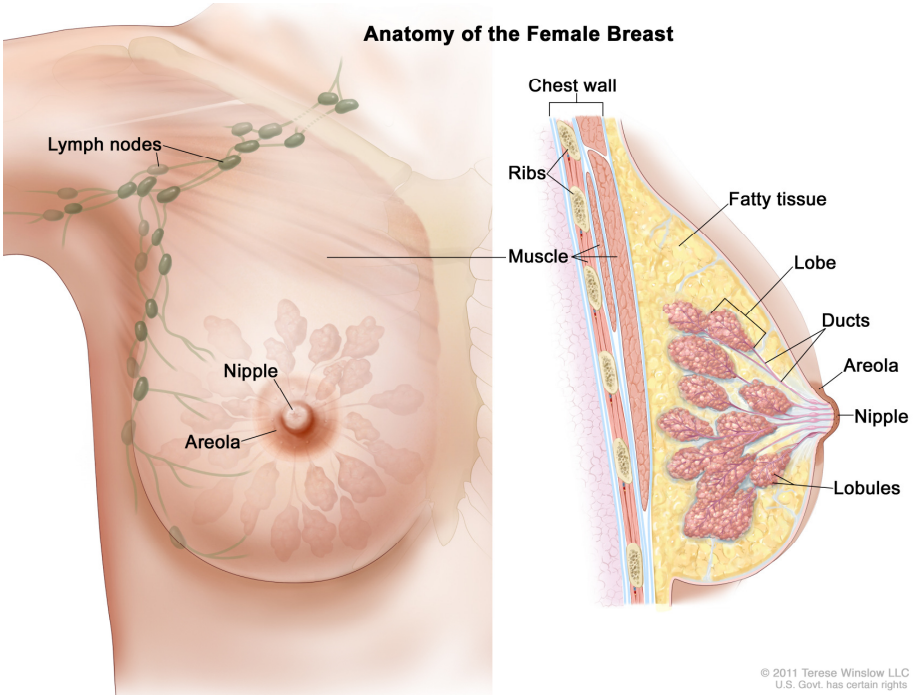


Fig. 1.1 Anatomy of the Female Breast.
The nipple and areola are shown on the outside of the breast. The lymph nodes, lobes, lobules, ducts, and other parts of the inside of the breast are also shown. Source: https://saintfranciscare.com/saintfrancisdoctors/cancercenter/nci/CancerSummary.aspx?id=CDR257995.xml.

Inside the breasts are located the mammary glands, modified sweat glands responsible for the production and secretion of milk, which are composed of 12-20 distinct lobes, with its own separate opening at the nipple, and embedded in adipose tissue. Each lobe is a system of ever-branching ducts that penetrate deep into the fibroadipose tissue of the breast. The branching duct system ends in a cluster of blind-ending terminal ductules, each cluster and its feeding duct comprising a mammary lobule (Fig. 1.2). The terminal ducts and lobules are embedded in a loose fibrous support tissue, which is rich in capillaries and also contains a few lymphocytes, macrophages and mast cells (Seeley et al., 2007; Stevens & Lowe, 2005). The ducts are lined by cuboidal epithelium (luminal side) with an outer discontinuous layer of myoepithelial cells (basal surface) (Fig. 1.3). This polarity of cellular organization allows transport through the mammary epithelium in a single direction: secretion into the lumen. Contractile myoepithelial cells have the ability to generate a flow of milk through the ducts to the nipple (Hinck & Näthke, 2014).

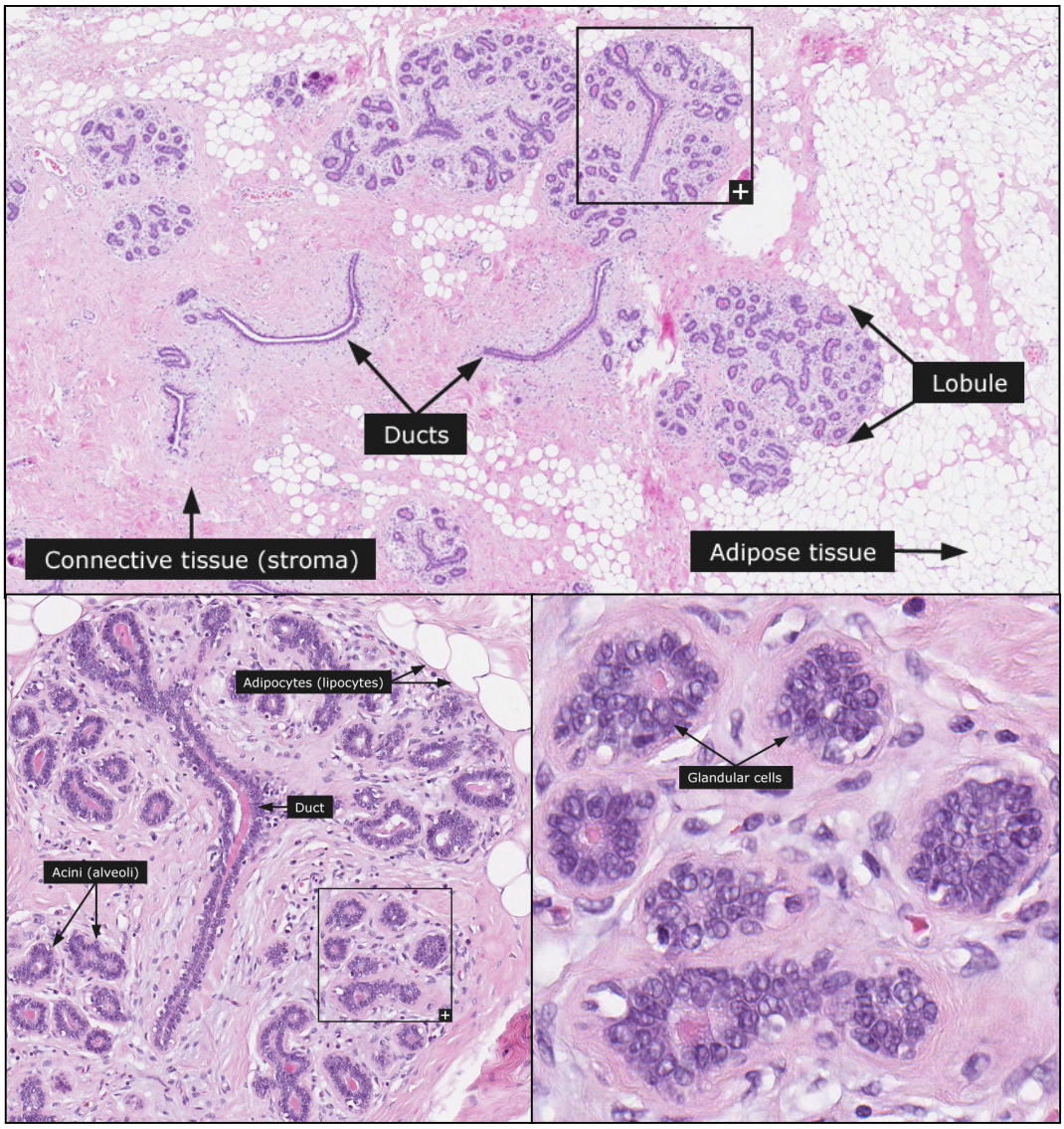


Fig. 1.2 Normal Breast histology.
The glandular epithelium is composed of two distinct types of cells, the secretory or luminal cells and the myoepithelial cells. In the collecting ducts the lining cells are usually columnar whereas in the *acini* they are usually cuboidal. Adapted from http://www.proteinatlas.org/dictionary/normal/breast+1.

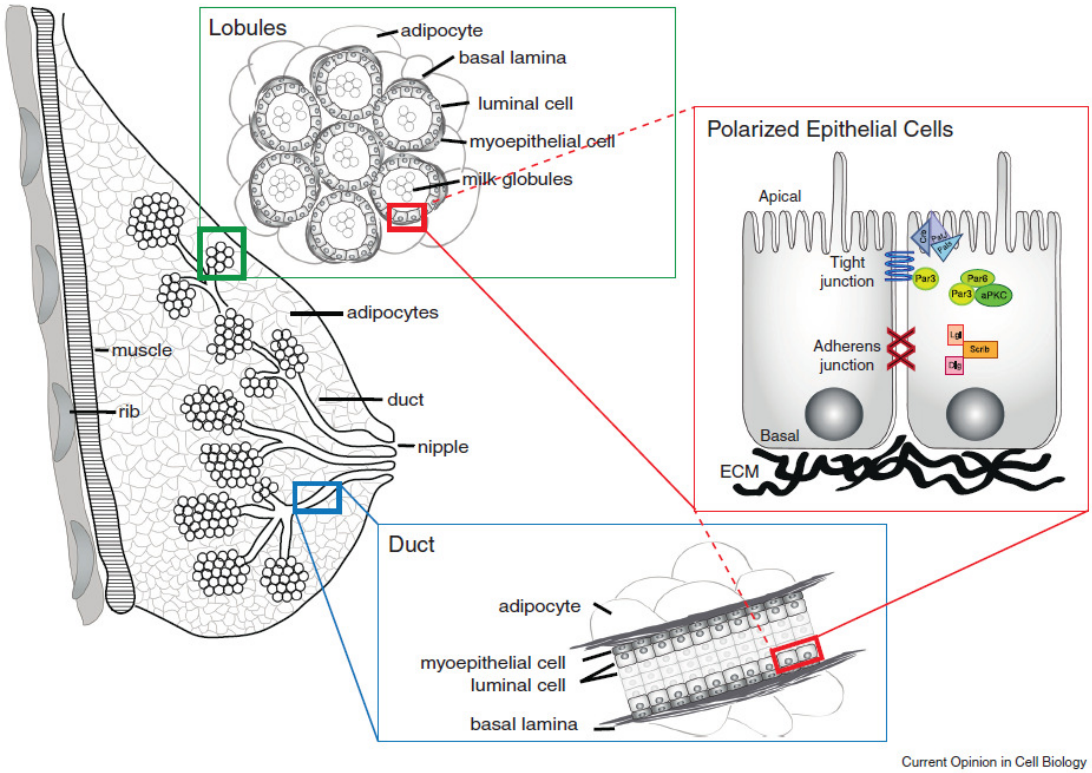


Fig. 1.3 Schematized view of the tissue organization in breast.
ECM: Extracellular matrix. Source: Hinck & Näthke, 2014.

Breast's full functional activity is only reached under the influence of pituitary (prolactin) and ovarian hormones (estrogen and progesterone), which are secreted in high concentrations during pregnancy and throughout breastfeeding. During the follicular phase of the menstrual cycle, cell proliferation is low and does not increase with the pre-ovulatory peak of estrogen. Following ovulation, progesterone stimulates up to three times the proliferation of the epithelial cells in the terminal ductules that become enlarged and begin to show evidence of early secretory activity. If fertilization does not occur, progesterone levels fall dramatically at the end of the menstrual cycle and the structure of the breast lobule reverts to normal along with some cell death. However, if fertilization occurs, increasing amounts of progesterone stimulate the continued proliferation and secretory activity in the terminal ductules of the lobule, in order to produce milk. When breastfeeding ceases, the breast returns to its normal state by gradual involution together with massive apoptosis leading to the loss of alveolar structures (Brisken, 2013; Stevens & Lowe, 2005; Vogel, 2000). Throughout a woman's lifetime the breast goes through *ca.* 450 cycles of growth and involution in response to hormones produced during the menstrual cycle and pregnancy (Hinck & Näthke, 2014). As women age the amount of fibrocollagenous tissue in the breast increases, replacing some of the adipose tissue, and the mammary lobules become enclosed in dense collagen (Stevens & Lowe, 2005).

Knowledge of normal histological appearances is essential to recognize abnormal structures, and to understand how the altered biochemical and physiological processes result in disease.

1.2 Breast Cancer

Because of the innumerous cycles of growth and involution in response to hormones, hormonal disturbances are most likely responsible for several breast pathologies, both benign and malignant. In most cases the development of invasive cancer appears to be preceded by carcinoma *in situ*, in which the malignant cells proliferate within the mammary ducts or lobules but do not breach the basement membrane. The most frequent breast invasive cancers are those that arise from the terminal ductolobular unit (Fig. 1.4) (Stevens et al., 2002).

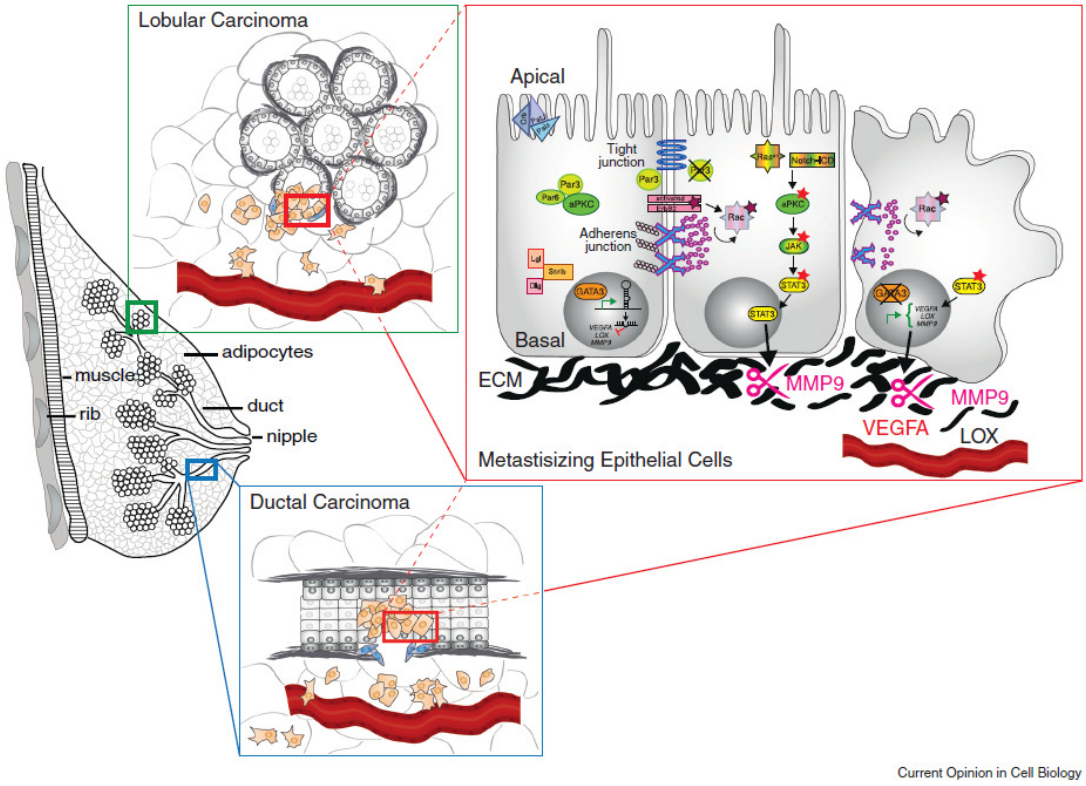
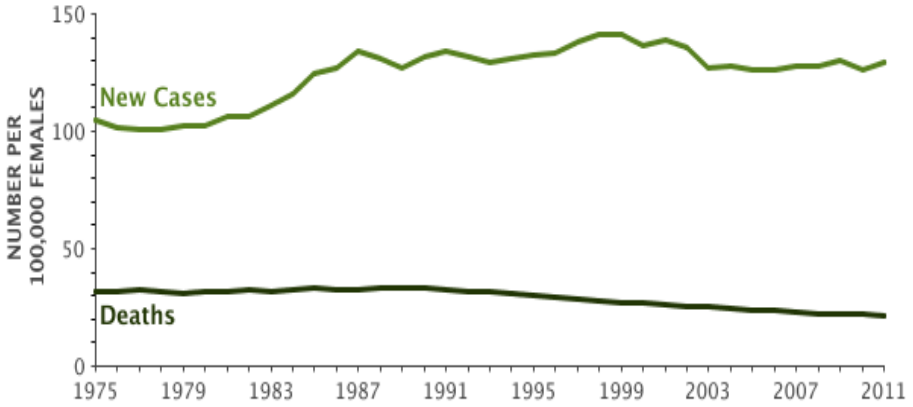


Fig. 1.4 Relationship between cellular and tissue changes in breast cancer.
The loss of polarity is a characteristic of cancers of epithelial origin and occurs in the early stages of progression in the breast tissue as well as stromal remodeling due to signals from the cancer cells that control the activity of leukocytes, fibroblasts, endothelial and other cells, contributing to tumor progression, particularly for metastasis. ECM: Extracellular matrix. Source: Hinck & Näthke, 2014.

1.2.1 Epidemiology

According to GLOBOCAN 2012, breast cancer is the second most frequent cause of cancer death worldwide, accounting for 11.9% (1.7 million) of the total cancer cases and 6.4% (521,817) of the total cancer deaths worldwide in 2012. For 2014, The American Cancer Society's estimates that in the United States alone 232,670 new cases of invasive breast cancer will be diagnosed in women and 40,000 women will die from breast cancer (Fig. 1.5).



New Cases, Deaths and

Estimated New Cases in 2014	232,670
% of All New Cancer Cases	14.0%
Estimated Deaths in 2014	40,000
% of All Cancer Deaths	6.8%

Year	1975	1980	1985	1990	1994	1998	2002	2006
5-Year Relative Survival	75.2%	74.9%	78.4%	84.6%	86.5%	89.5%	90.3%	90.6%

Fig. 1.5 New cases, deaths and 5-year relative survival.
Adapted from http://seer.cancer.gov/statfacts/html/breast.html

After increasing for more than 2 decades, female breast cancer incidence rates began decreasing in 2000. This is thought to be due to the decline in use of hormone therapy after menopause that occurred after the results of the Women's Health Initiative were published in 2002 that linked the use of hormone therapy to an increased risk of breast cancer and heart diseases. Incidence rates have been stable in recent years. Death rates from breast cancer have been declining since about 1989, with larger decreases in women younger than 50. These are believed to be the results of earlier detection through screening and increased awareness, as well as improved treatment (DeSantis et al., 2014).

In Portugal, according to the Portuguese Cancer League (Liga Portuguesa Contra o Cancro), about 4,500 new cases of breast cancer are detected annually and 1,500 women die from this disease.

1.2.2 Risk Factors

Simply being a woman is the main risk factor for developing breast cancer. Men can develop breast cancer too, but according to the American Cancer Society this disease is about 100 times more common among women than men.

Breast cancer often shows familial clustering with 5% to 10% of cases thought to be hereditary. Two high penetrance genes have been identified, *BRCA1 and BRCA2*. In families with *BRCA1* mutations the average lifetime risk of breast cancer seems to be in the range of 55- 65%, and for BRCA2 mutations the risk is lower, around 45%. Low or moderate penetrance genes are also known and include *CHEK2*, *PTEN*, *TP53*, *ATM*, *STK11*/*LKB1*, *CDH1*, *BRIP1* and *PALB2* (Bradbury & Olopade, 2007; Campeau, et al.,2008; WHO, 2012). However, the etiology of breast cancer is multifactorial and also involves diet, reproductive factors, and related hormonal imbalances.

Current age, age of menarche, age at first birth (or nulliparity), number of breast biopsies, atypical hyperplasia and number of first-degree relatives with breast cancer (ie. Mother, sisters, daughters) are considered risk factors in breast cancer. Compared with women who experience

menarche at age 16, girls who experience menarche two to five years earlier have a 10% to 30% greater risk of developing breast cancer later in life. If women who experience menopause between the ages of 45 and 55 years are used as referent group, women who experience menopause at age 55 or older have a 50% higher risk of subsequently developing breast cancer. Furthermore, women who cease menstruating at age 45 or younger have a 30% lower risk of subsequently developing breast cancer (Vogel, 2000; WHO, 2012). Thus, risk increases with the number of menstrual cycles a woman experiences during her lifetime and hence the exposure time of the mammary epithelium to ovarian hormones (Brisken, 2013).

Additionally, although current data do not support specific dietary guidelines for reducing breast cancer risk, the American Cancer Society still recommends that women maintain a healthy weight and limit intake of high-fat foods, particularly those from animal sources, as part of a healthy lifestyle. Obesity has been associated with an increased risk of breast cancer in post-menopausal women, which may occur because fat stores provide an important source of hormone substrates in postmenopausal women. It has not been shown clinically that reducing body weight can lower breast cancer risk, but this topic deserves further study. Smoking affects overall health and may increase risk for breast cancer, but no controlled trials have thus far established a definite link between smoking and breast cancer. Alcohol consumption has, however, been linked to higher serum estrogen levels; moderate or high amounts of alcohol consumption have been associated with increased breast cancer incidence. On the other hand, exercise enhances immune function, is associated with lower body fat, and affects hormonal levels, all of which may influence the incidence of breast cancer nevertheless, confounding factors make it difficult to assess this relationship, because women who exercise regularly are also likely to smoke less, drink less, have different menstrual and reproductive patterns, and consume different diets than sedentary women (Vogel, 2000; Yager & Davidson, 2006).

Breast cancer risk is also decreased when reproductive hormone levels are altered through surgery. Risks associated with the use of oral contraceptives are not well defined, though recent studies found that current users of oral contraceptives had a higher relative risk of breast cancer. Nonetheless, use of hormonal replacement therapy is associated with an increased risk of invasive breast cancer. Also, relatively low doses of radiation have been associated with an increased incidence of solid tumors such as breast cancer although very low doses of radiation associated with screening mammography do not increase the risk appreciably (Vogel, 2000).

1.2.3 Diagnosis

Most clinically significant breast disorders present as a lump and the major imperative is to identify those which are malignant tumors so that the patient may be treated quickly. Currently, several national screening programs use radiological techniques (mammography) and/or ultrasound to identify early suspicious breast lesions, including abnormal calcifications. A tissue diagnosis is then made by fine needle aspiration biopsy, core biopsy, vacuum assisted biopsy or excision biopsy before definitive treatment is undertaken (Stevens et al., 2002).

1.2.3.1 Histological classification

Histopathologic classification is primarily based on the histological appearance of the tumor seen upon light microscopy of routine haematoxylin and eosin stained sections. The typing of invasive breast cancer and its histological variants are well established.

Invasive carcinoma not otherwise specified (NOS) comprises the largest group of invasive breast cancers, comprising between 40% and 75% in published series (Fig. 1.6). It is a heterogeneous group of tumors that fail to exhibit sufficient characteristics to achieve classification as a specific histological type, such as lobular or tubular carcinoma. Invasive lobular carcinoma is the second most common type, representing approximately 5-15% of invasive breast cancers (Fig. 1.7). The less common types include mucinous, cribriform, micropapillary, papillary, tubular, medullary, metaplastic, and inflammatory carcinomas (Sandhu et al., 2010; WHO, 2012).

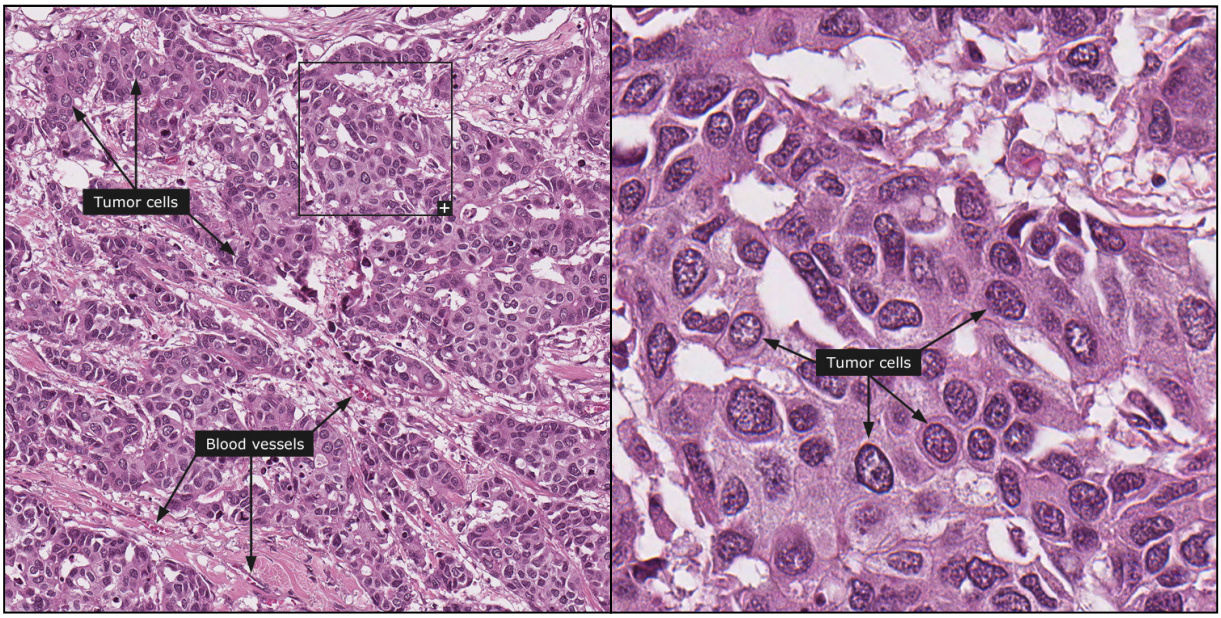


Fig. 1.6 Invasive Carcinoma NOS.
Invading malignant epithelial cells form small ductal structures. Adapted from http://www.proteinatlas.org/dictionary/cancer/breast+cancer+3

1.2.3.2 Grade

Invasive ductal carcinomas and all other invasive tumors are routinely graded based on an assessment of tubule formation as an expression of glandular differentiation, nuclear pleomorphism and mitotic counts (total number of mitoses per 10 high power fields). A numerical scoring system of 1-3 is used to ensure that each factor is assessed individually. The three values are added together to produce scores of 3 to 9, to which the grade is assigned as follows: Grade 1 - well differentiated: 3-5 points; Grade 2 - moderately differentiated: 6-7 points; Grade 3 - poorly differentiated: 8-9 points. Many studies have demonstrated a significant association between histological grade and survival in invasive breast carcinoma. Elston and Ellis are the most recent modifications to the grading method used (WHO, 2012).

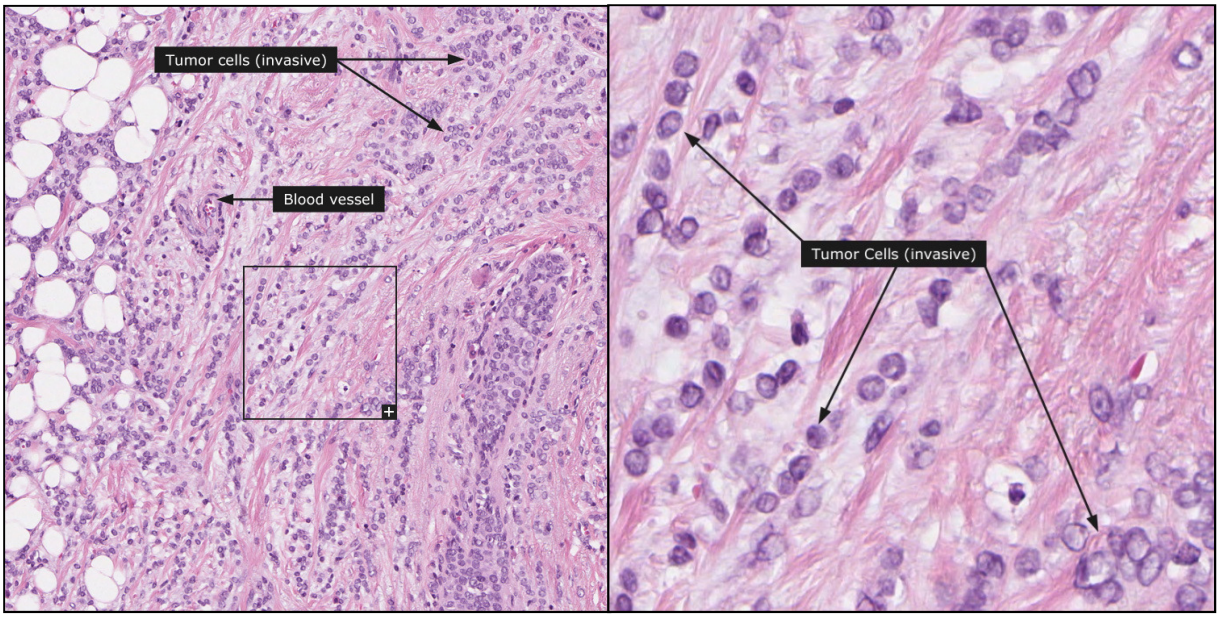


Fig. 1.7 Invasive Lobular Carcinoma.
Malignant cells infiltrate the stroma in rows of cells which do not form ducts. Adapted from http://www.proteinatlas.org/dictionary/cancer/breast+cancer+4.

1.2.3.3 Stage

The staging system currently in most widespread use is the TNM Classification. This system assesses cancer in 3 ways: the size and extension of the primary tumor (pT), regional lymph node involvement (pN), and the presence of distant metastases (pM) (Siegel et al., 2013; WHO, 2012).

1.2.3.4 Molecular Subtypes

Tumors show great multidimensional variation in gene expression, with many different sets of genes showing independent patterns of variation. These sets of genes relate to biological processes such as proliferation or cell signaling. Despite this variation, there are striking similarities between tumors, providing new opportunities for tumor classification. Using microarray technology, which details the expression level of thousands of genes simultaneously, several studies have shown that breast tumors can be grouped into at least four molecular subtypes: Luminal A, Luminal B, HER2 and Basal-like. Routinely, approach to defining tumors molecular subtypes is to use standard immunohistochemical markers, including ER (estrogen receptor), PR (progesterone receptor), HER2 (Human epidermal growth factor receptor 2), and the proliferation marker Ki-67 (Fig. 1.8).

Approximately 70% of invasive breast cancers are ER positive. These are collectively classified as luminal cancers which are subclassified into luminal A and luminal B subtypes based on their proliferation rate (Ki-67 expression) and HER2. The majority of ER positive tumors also express PR. The ER negative breast cancers are subclassified as HER2+ (~15% of all the breast cancer) and as triple-negative, based on the HER2 over-expression/gene amplification. Basal-like breast cancers (~15% of all breast carcinomas) are distinguished from other kind of triple-negative breast cancers by expression of CK5/6, CK14, CK17, CK34BE12 or P63. More recently, additional subtypes, such as the claudin-low, have been identified, while the existence of the normal like subtype is still debatable as it

could be an artifact of gene profiling due to a disproportionally high content of normal cells. These breast cancer subtypes have been linked with specific molecular alterations and, equally important, they seem to correlate with specific incidence, baseline prognosis and response to therapy (Fumagalli et al., 2012; Ignatiadis & Sotiriou, 2013; Sandhu et al., 2010; Singh & Mo, 2013; Schnitt, 2010; WHO, 2012).

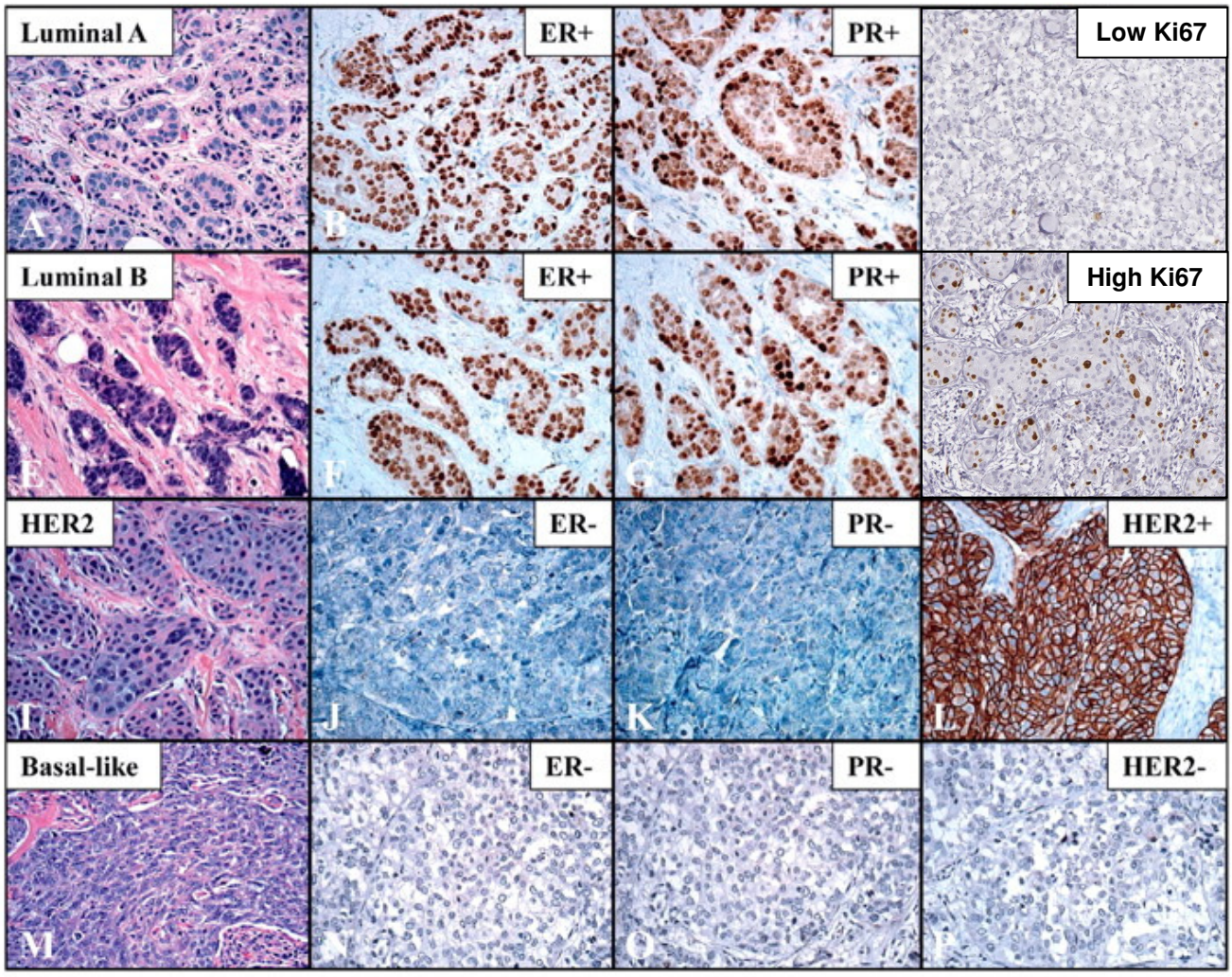


Fig. 1.8 Major molecular subtypes of breast cancer determined by immunohistochemistry.
Luminal A: ER+ and/or PR+, HER2−, and low Ki67 (<14%); Luminal B: ER+ and/or PR+ and HER2+ (luminal-HER2 group), or ER+ and/or PR+, HER2–, and high Ki67 (>14%); HER2: ER–, PR–, and HER2+; Basal-like: ER–, PR–, HER2–, and CK5/6 and/or EGFR+ (Adapted from Sandhu et al., 2010).

The traditional pathological factors of lymph node status, tumor size, histological type, and histological grade are the most useful prognostic factors in breast cancer patients; however, they are now challenged by gene expression profiling. In general, the ER+ breast cancer subtypes (luminal A and luminal B) exhibit a good prognosis and excellent long-term survival (approximately 80%–85% 5-year survival). The ability of patients with ER+ breast cancers to survive their disease reflects the availability of effective targeted therapy in the form of anti-estrogen treatment such as tamoxifen and aromatase inhibitors, both of which target the ER signaling. The low-grade luminal A tumors seem indeed indolent and may be treated only with anti-estrogens, whereas the high proliferative luminal B

tumors often have lower expression levels of ER, lower or no PR expression, and are considered to have lower sensitivity to endocrine treatment and higher sensitivity to chemotherapy. In contrast, the ER- subtypes (HER2-positive and basal-like) are difficult to treat and are associated with poor prognosis (approximately 50%–60% 5-year survival). Although HER2-positive tumors have an aggressive progression, the survival rate has improved in the last decade due to the onset of target therapies using trastuzumab, an antibody against HER2, which have been shown to be effective in 20% of patients. Basal-like tumors, despite being more aggressive than other tumor types, can be especially sensitive to chemotherapy but promising strategies are being developed to treat these type of cancer, such as poly-ADP ribose polymerase-1 inhibitors. Regarding claudin-low breast cancer, this subtype has a poor long-term prognosis (Ignatiadis & Sotiriou, 2013; Fumagalli et al., 2012; Sandhu et al., 2010; Singh & Mo, 2013; WHO, 2012).

The purpose of breast cancer classification, subtyping and risk assessment is to, ultimately, select the best course of treatment for each patient.

1.3 Novel Cancer Markers

As previously referred, breast cancer is not a single disease; rather it represents a diverse spectrum of diseases including several distinct biological entities and subtypes. The complex and varied presentation and clinical evolution of breast cancer have long prompted professionals in the field to find a strategy to classify the disease and to possibly define supportive prognostic and predictive indicators. In a landmark report in 2000, Perou and coworkers provided a molecular classification of breast cancer based on gene expression profiles, which has been used as a standard scheme for many basic and clinical studies in the field. The molecular signatures of the five subtypes (luminal A, luminal B, HER2 enriched, basal-like and normal breast-like), based on differential immunohistochemical staining for ER, PR, HER2, and cytokeratins, among others, were shown to differ in terms of a number of characteristics including biology, relapse rate and response to therapy. The differential expression of these protein biomarkers is used as an immunohistochemical surrogate for gene expression analysis to determine molecular subtype (Hudson, 2013; Sandhu et al., 2010). If from one side it has been argued that the definition of breast cancer subtypes using immunohistochemistry (IHC) is not able to recapitulate the information provided by the gene expression intrinsic classification, from the other side it has been posited that the additional clinical value of the molecular classification is limited by its close correspondence to ER and HER2 status and proliferation markers defined by IHC (Colombo et al. 2011; Fumagalli et al., 2012).

Nonetheless, these prognostic indicators have shown limited ability to predict individual patient outcomes since patients with the same clinico-pathological parameters can have largely different clinical courses. One plausible explanation for this phenomenon is the presence of molecular differences due to distinct cellular evolution in the tumor. After the initial transformation events occur, neoplastic cells undergo a series of changes due to the interactions between the altered cells and external signals. As a consequence, the incipient cancer cells arising from similar initiating events may diverge, resulting in neoplasms that are quite dissimilar from one another. These differences may be molecular (reflecting gene expression patterns), but otherwise they are not easily discernible (at the level of cellular morphology) and may represent the mechanisms accounting for biological subsets

(Fumagalli et al., 2012; Sandhu et al., 2010). Hence, further efforts should be directed at the earliest detection of cancer cells, which would depend on novel molecular biomarkers.

Furthermore, the hormone receptors (ER and PR) and HER2, the only two validated predictive biomarkers used in the clinic, do not provide information about which chemotherapy regimen should be selected for adjuvant therapy (Fumagalli et al., 2012). Women receive adjuvant chemotherapy following the removal of detectable breast cancer tissue, to treat additional microscopic disease and reduce the risk of relapse and cancer-related death. Although the efficacy of this practice has been demonstrated, many patients currently treated with adjuvant therapy are actually overtreated. This underlines the importance of good prediction strategies to tailor treatment for each individual patient (Hudson, 2013; van der Vegt et al., 2009). It then follows that detailed molecular analysis of breast cancer could yield diagnostic tests that might be more accurate than existing clinical-pathological prediction models, or at least be complementary to them.

Gene expression profiling is one of many new and powerful tools that have become available for the purpose of dissecting the biological complexity of breast cancer and improving its clinical course. Given that the phenotype is determined by gene expression, the clinical behavior of subgroups of breast cancer will be associated with specific gene expression patterns, potentially leading to identification of therapeutic targets or biomarkers for progression (Sandhu et al., 2010). One example of a microarray-based gene expression profiling currently employed in the clinical assessment of breast cancer is MammaPrint (Table 1.1), the 70-gene prognosis profile, which was approved by the Food and Drug Administration (FDA) in 2007. This test differentiates between patients at low risk or at high risk for metastasis on the basis of a score yielded by the assay. High-risk patients are recommended for more aggressive chemotherapy compared to patients with a low score. Another good example is the Oncotype Dx Recurrence Score, which is a 21-gene prognostic and predictor assay based on a continuous variable algorithm used to predict the likelihood of relapse, stratifying the patients that need to be aggressively treated versus those where conservative treatment will suffice (Colombo et al., 2011; Sandhu et al., 2010).

Table 1.1 Prognostic multitude signatures in breast cancer commercially available or in commercial development.
Adapted from: Colombo et al., 2011.

Signature	MammaPrint	Oncotype DX	Theros/MGI	MapQuant DX/ simplified	Veridex 76-gene
Commercially available/Provider	Yes/Agendia BV (Amsterdam, The Netherlands)	Yes/Genomic Health (Redwood City, CA, USA)	Yes/ bioTheranostics, Inc. (San Diego, CA, USA)	Yes/Ipsogen Inc. (Stamford, CT, USA)	No/Johnson & Johnson (New Brunswick, NJ, USA)
Study population	ER+ and ER–, N0, <5 cm diameter, age <55 years	ER+, N0, TAM treated	ER+, N0	ER+ and ER–, N0 and N+	ER+ and ER–
Assay	70-gene signature	21-gene Recurrence Score	2-gene HOXB13:IL17R/ molecular-grade index	97-gene signature/8-gene PCR	76-gene signature
Platform	Microarray (Agilent Technologies, Inc., Santa Clara, CA, USA)	RT-PCR	RT-PCR	Microarray (Affymetrix, Santa Clara, CA, USA)/ RT-PCR	Microarray (Affymetrix)
Tissue type	Frozen or stabilized mRNA	FFPE	FFPE	Frozen/FFPE	Frozen
Prognostic value in other populations	Age 55-70 years, 1-3 N+, N0 and N+, HER2+	ER+ and 1-3 N+, ER+ postmenopausal receiving aromatase inhibitors	-	ER+ receiving aromatase inhibitors	
Predictive value	Neoadjuvant and adjuvant CT (poor signature)	Neoadjuvant and adjuvant CT [71] (high-RS), response to TAM (low-RS)	Resistance to TAM (high-ratio)	Response to neoadjuvant CT (high-risk)	Response to TAM (high-risk patients)
Indication	Prognostic in N0, <5 cm diameter, stage I/II BC, age <61 years	Prediction of recurrence risk in ER+ and N0 BC treated with TAM	Prognostic in ER+ BC, prediction of response to TAM	Molecular grading, for ER+, histological grade II BC	Prognostic in ER+ BC
Level of evidence	III	II	III	III	III
FDA approval	Yes	No	No	No	No
Randomized trial	MINDACT	TAILORx	-	-	-
Availability	Europe and USA	Europe and USA	USA	Europe	-

BC, breast cancer; CT, chemotherapy; ER; estrogen receptor status (+ or –); FDA, US Food and Drug Administration; FFPE, formalin-fixed paraffin-embedded; HER2, human epidermal growth factor receptor 2; HOXB13, homeobox 13; IL-17BR, interleukin-17B receptor; MGI, molecular grade index; MINDACT, Microarray In Node-negative and 1-3 positive lymph-node Disease may Avoid ChemoTherapy; N+, lymph node-positive; N0, lymph node-negative; PCR, polymerase chain reaction; RS, recurrence score; RT-PCR, real time – polymerase chain reaction ; TAILORx, Trial Assigning IndividuaLized Options for Treatment Rx; TAM, tamoxifen.

The multiplicity of tests that have been developed has created difficulties for clinicians who are unsure of which tests to use (Hudson, 2013). Moreover, skepticism has arisen regarding the reliability and the reproducibility of the microarray technology and the computational approaches used to develop and evaluate predictive models because different tests may contain completely different gene sets. Nonetheless, many genetic profiles can examine the same molecular roadmaps, given that so many genes are responsible for controlling the many biochemical pathways that are expressed by the tumor. Thus, the biological subgroups predicted by those genes are equal (Fumagalli et al., 2012; van der Vegt et al., 2009; De Snoo et al., 2009). The power of these signatures is their high accuracy in identifying low-risk patients who could be spared aggressive adjuvant chemotherapy, in contrast to their ability to identify high-risk patients, which could still be improved. Given these observations, it is justified to wonder if standard pathological biomarkers could provide similar information (Fumagalli et al., 2012; Rakha et al., 2008).

Recently, some studies have focused on the putative utility of non-coding RNAs profiling, such as miRNA, as a novel class of cancer markers (Bartels & Tsongaliset 2009; Sandoval & Esteller, 2012; Serpico et al., 2014).

1.4 Biogenesis of microRNAs

miRNAs were first discovered in 1993 when researchers found components of the genome, until then considered nonfunctional, with gene regulatory capacity. The lin-4 gene in *C. elegans* did not code for a protein but the coded RNA was involved in an antisense regulatory mechanism (Lee et al.,

1993). miRNAs have since been found in plants, viruses, and more deeply branching animals (Axtell et al, 2011; Bartel, 2009).

Currently, there are over 32,000 miRNA-related publications indexed in pubmed and the latest version of the miRBase Sequence database (mirbase.org) lists over 2,588 mature human miRNAs. miRNAs are now recognized as major players in almost every biological process such as cell proliferation, apoptosis, differentiation, and organogenesis, by regulating key processes of gene activation and suppression. They account for ~1% of all predicted genes in *Homo sapiens* and bioinformatics studies suggest that they control up to one third of all human genes (Tétreault & De Guire, 2013). Essentially, miRNAs are single-stranded RNAs, broadly conserved across species, of 18~23 nucleotides derived from endogenous hairpin-shaped transcripts that regulate gene expression through a posttranscriptional mechanism (Singh & Mo, 2013).

miRNAs (Fig. 1.9) are transcribed as long primary transcripts (up to several thousands of nucleotides) characterized by hairpin structures (pri-miRNAs). Approximately 50% of miRNAs have their own promoter and the other half can be found in intronic or exonic regions of coding or non-coding transcription units. After transcription, pri-miRNAs are processed by RNAse III Drosha and DGCR8. Briefly, pri-miRNAs are cleaved into ~70 nucleotide pre-miRNAs, creating an imperfect stem loop structure. Then, Exportin 5, a RAN-GTP dependent nucleo/cytoplasmatic cargo transporter, mediates the export of the originated precursor molecules to the cytoplasm. This change in cellular localization allows an additional step mediated by the RNAse III Dicer which acts in complex with the transactivating response RNAbinding protein (TRBP) to generate a small double stranded RNA duplex, approximately 22 nucleotides long, that contains both the mature miRNA strand and its complementary strand. Dicer recognizes the double strand region of the pre-miRNA in association with different proteins. Completed the processing steps, TRBP then recruits Argonaute 2 (Ago2), which is the major component of the RISC complex (RNA Induced Silencing Complex). The role of the RISC complex is to select and recruit the RNA strand that has the lowest thermodynamic stability at its 5'-end, termed the mature miRNA-3p guide strand (Iorio & Croce, 2012; Tétreault & De Guire, 2013).

The mature single stranded miRNA product is then incorporated in the RISC complex. Guided by the base pairing between the noncoding RNA and the target mRNA, miRNA-RISC-mediated gene inhibition can be split into three processes: site-specific cleavage, enhanced mRNA degradation, and translational inhibition. The initial process, commonly defined as RNAi and restricted to miRNAs with a perfect or near-perfect match to the target RNA, is a very rare event in mammals, exclusively Ago2 dependent. Instead, the other two processes are more commonly associated with mismatched miRNA/target sequences that are the most likely scenario in mammals. The combination of these two processes is commonly defined as a non-cleavage repression (Bartels & Tsongaliset 2009; Iorio & Croce, 2012).

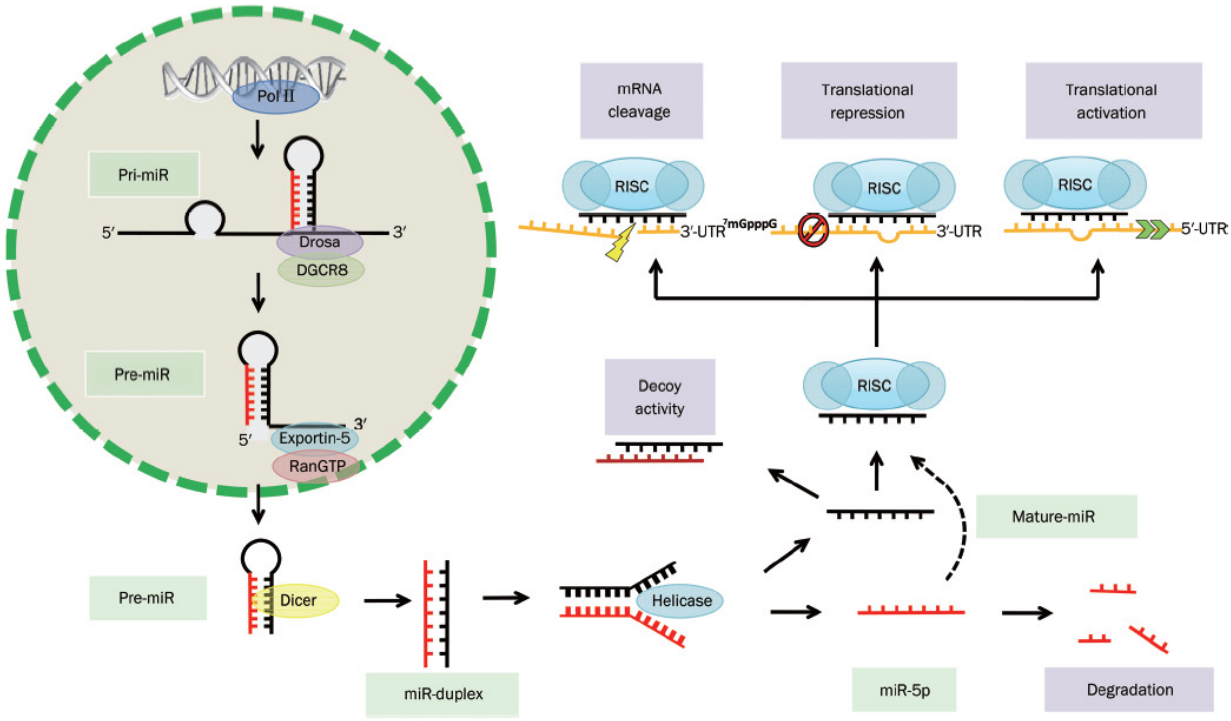


Fig. 1.9 The mechanism of microRNA biogenesis and regulation of gene expression.
Source: Li & Yang, 2013.

Most of the time, miRNAs mainly bind their targets via a sequence between their 2nd and 8th nucleotides of their 5' extremity: the seed sequence. Although interaction between the miRNA sequence of the 9th to 20th nucleotide of small RNAs is required, a perfect complementary between a miRNA and the 3'-untranslated regions (UTRs) sequence of their target mRNAs is not necessary. The limited complementary was first considered as a weak-point; however, this property holds intrinsic advantages for the regulation of gene expression by miRNAs. Moderate affinity of one miRNA for a given target allows a single miRNA to inhibit simultaneously the expression of hundreds of different mRNAs. Since many of these miRNA targets are involved in various signaling pathways, their impact on gene expression can be significantly amplified. Moreover, bioinformatics analysis predicts that 3'-UTRs of single genes are often targeted by several different miRNAs. Although less frequently, miRNA can also bind to the 5'-UTR or the ORF and, even more surprisingly, they can upregulate translation upon growth arrest conditions (Iorio & Croce, 2012; Singh & Mo, 2013; Tétreault & De Guire, 2013).

Apart from the cells in which they are produced, miRNAs can be found in blood, urine, saliva and many other biological fluids. Not only detectable in blood cells, an important fraction of blood circulating miRNAs is found in a free state that may include transportation in exosomes and lipoproteins, as well as bound to proteins. For that reason, some consider miRNAs as hormones because, when secreted into circulation by a mother cell, miRNAs can regulate gene expression of a distant cell following endocytosis (De Guire et al., 2013; Tétreault & De Guire, 2013).

1.5 miRNA and Cancer

Cancer has traditionally been viewed as a set of diseases that are driven by the accumulation of genetic mutations that have been considered the major causes of neoplasia. However, this paradigm has now been expanded, since tumorigenesis is a multistep process, including initiation, promotion and progression, and a multifactorial pathology characterized by the accumulation of a multitude of alterations including genetic, cytogenetic, and epigenetic changes. The epigenetic pathway to cancer is not simple and is determined by chromatin structure remodeling including DNA methylation, histone variants modifications, nucleosome remodeling as well as small non-coding regulatory RNAs. Importantly, alterations in epigenetic mechanisms can lead to genetic mutations, and genetic mutations in epigenetics regulators lead to an altered epigenome (You & Jones, 2012; Sandoval & Esteller, 2012).

The link between miRNA dysregulation and human disease has been reported in almost all medical fields (Tétreault & De Guire, 2013). To date, according to the Human MicroRNA Disease Database (HMDD), research has demonstrated that miRNAs are associated to 378 human diseases (Li et al., 2013a). The first link of the involvement of miRNA in human cancer derived from studies on chronic lymphocytic leukemia by Dr. Croce's group who showed the first evidence that miRNAs could be involved in the pathogenesis of human cancer as the chromosome deletion caused the loss of miR-15a and miR-16-1 (Calin et al., 2008).

The widespread differential expression of miRNA genes between malignant and normal cells is a complex phenomenon and can be exerted through several mechanisms, all acting in concert for abnormal expression levels of miRNAs: chromosomal abnormalities, mutations, polymorphisms, defects in miRNA biogenesis machinery, epigenetic changes, or transcription activity dysregulation of a transcription factor at the promoter (Iorio & Croce, 2012; Nicoloso et al., 2009; Serpico et al., 2014).
It is now well characterized that the genomic sequences of 50% of miRNAs are located at cancer-associated genomic regions which are fragile sites that are either deleted or amplified in cancer (Croce, 2009; De Guire et al., 2013; Nicoloso et al., 2009). Indeed, chromosomal regions encompassing miRNAs involved in the negative regulation of a transcript encoding a known tumor suppressor gene can be amplified in cancer development. This amplification would result in the increased expression of the miRNA and consequent silencing of the tumor suppressor gene, being, therefore, designated as oncogenic miRNAs (oncomiRs). Conversely, tumor suppressor miRNAs (tsmiRs) repress oncogenes and are often located in fragile loci, where deletions or mutations can occur and result in reduced microRNA levels and overexpression of the target oncogene (Fig. 1.10). It is important to note that miRNAs may act in a tissue-specific way such that a single miRNA type can be either an oncomiR or a tsmiR in different types of tumors (Iorio & Croce, 2012; Li & Yang, 2013; Serpico et al., 2014). Either as tsmiRs or oncomiRs, there are essential features of cancer progression in which miRNAs may act: self-sufficiency in growth signals, insensitivity to anti-growth signals, apoptosis evasion, limitless replicative potential, sustained angiogenesis, tissue invasion and metastasis (Iorio & Croce, 2012).

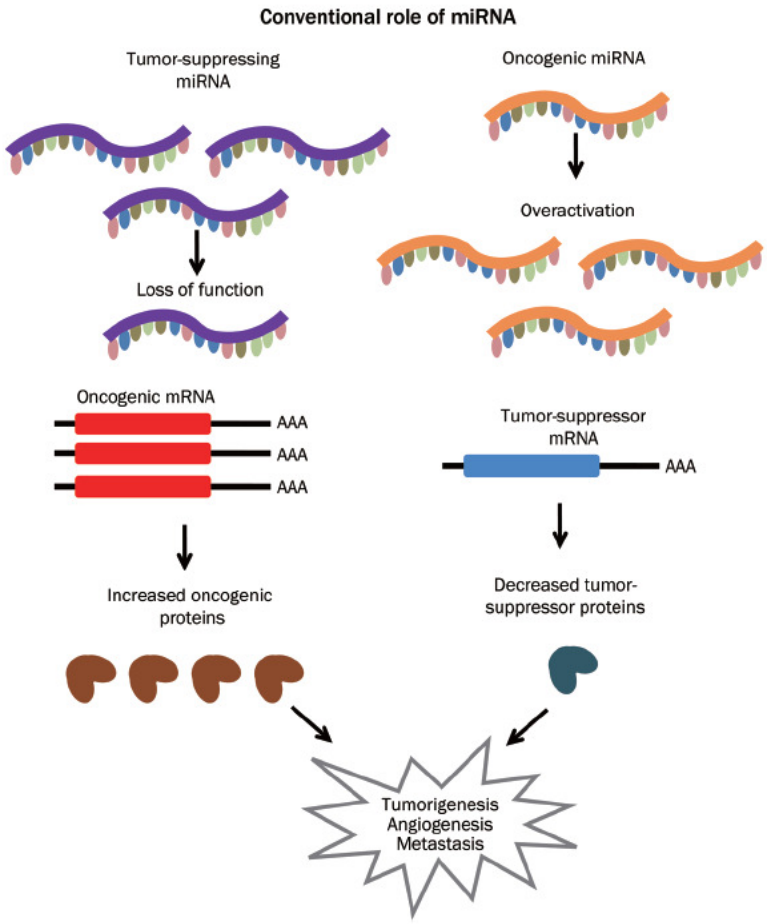


Fig. 1.10 The role of microRNA in cancer.
Source: Li &Yang, 2013.

Since microRNAs have multiple targets, their function in tumorigenesis could be due to their regulation of very few specific targets, possibly even one, or many targets. A future challenge will be to identify all the targets of the microRNAs involved in cancer and establish their contribution to malignant transformation in vivo. An additional challenge will be the identification of all the microRNAs that are dysregulated by pathways that are consistently dysregulated in various types of human cancers. This point is of particular potential importance, since instead of focusing on the specific alterations in protein coding oncogenes or tumor suppressor genes that are components of the pathways dysregulated in cancer, that may be difficult to be addressed therapeutically, we may focus on their downstream microRNA targets (Croce, 2009).

Of considerable interest, a unique miRNA expression pattern can be associated with certain breast cancer subtypes (Table 1.2). For example, miR-21, miR-210 and miR-221 are significantly overexpressed in triple-negative breast cancer, whereas miR-10b, miR-145, miR-205 and miR-122a are significantly underexpressed in these cancer types. Furthermore, miR-21, miR-210 and miR-221 correlate with worse patient disease-free and overall survival, and hence they play a significant role in triple-negative primary breast cancers (Singh & Mo, 2013).

Table 1.2 miRNAs and breast cancer subtypes.
Source: Serpico et al,, 2014.

Signature*	Cancer subtype
Up: miR-150, -142-3p, -142-5p, 148a, -106a/b, -18a, -93, -155, -25, -187, -135b	Basal-like
Up: miR-150, -142-3p, -142-5p, -148a, -106b, -93, -155, -25, -187, -375 Down: miR-125a and b	HER2 positive
Up: miR-191, -26, -126, -136, -100, -99a, -145, -146b, -10a, -199a/b, -130a, -30a-3p,-30a-5p, -224, -214, let-7a/b/c/f, -342 Down: miR-206, -15b, -107, -103	Luminal A
Up: miR-191, -26, -106a/b, -93, -25, -10a, -30a-3p, -30a-5p, -224, let-7b/c/f and -342, -15b, -107, -103 Down: miR-206, -100, -99a, -130, -126, -136, -146b	Luminal B
Up: miR-142-5p, -135b, -126, -136, -100, -99a, -145, -10a, -199a/b, -130a, -30a-3p, -214, -7a/c	Normal-like

* As compared with normal tissue or with parental cell lines in case of preclinical data.

In addition, miRNAs can specifically classify estrogen receptor (ER), progesterone receptor (PR) and HER2 receptor status (Iorio & Croce, 2012). ER, which serves as the target of endocrine therapeutic agents such as tamoxifen and raloxifen, is regulated by let-7, miR-206, and miR-221 in breast cancer. Interestingly, miR-206 and miR-221 are believed to be responsible for tamoxifen insensitivity, while induction of tamoxifen sensitivity by let-7 could be due to a different binding region (Li & Yang, 2013). On the other hand, miR-21 upregulation has been associated to Trastuzumab resistance in HER2-positive breast cancers (Serpico et al., 2014). Besides hormonal therapy, miRNAs are closely related to chemoresistance (Table 1.3). For instances, the ABC transporter MDR1 gene has been shown to be directly regulated by miR-200c that induces chemosensitivity to doxorubicin. Alternatively, SOCS3 has been shown to be regulated by miR-203, inducing chemoresistance to Cisplatin (Singh & Mo, 2013). These resistance/sensitivity can be mediated by miRNAs through regulating drug resistance-related proteins, altering drug targets, changing drug concentration, influencing therapeutic induced cell death, promoting angiogenesis, influencing tumor stem cell, and promoting epithelial-mesenchymal transition (EMT) (Li & Yang, 2013; Zhou, 2013a).

Table 1.3 miRNAs and multigdrug resistance.
Source: Tian et al., 2013.

Mechanism	miRNA	Alteration	Targets	Resistant to
ABC transporters	miR-200c	Down-regulated	Pgp/MDR1	Doxorubicin
	miR-451	Down-regulated	Pgp/MDR1	Doxorubicin
	miR-345,-7	Down-regulated	MRP-1/ABCC1	Cisplatin
	miR-326	Down-regulated	MRP-1/ABCC1	VP-16
	miR-328,-519c	Down-regulated	BCRP/ABCG2	Mitoxantrone
	miR-19	Up-regulated	MDR-1	Taxol
			MRP-1	VP-16
			BCRP	Mitoxantrone
Anti-apoptotic proteins	miR-125b	Up-regulated	Bak1	Taxol
	miR-19	Up-regulated	PTEN	Taxol
				VP-16
				Mitoxantrone
	miR-21	Up-regulated	PTEN	Doxorubicin
	miR-21	Up-regulated	PTEN	Trastuzumab
EMT	miR-200	Down-regulated	E-cadherin	Doxorubicin
			ZEB1/ZEB2	
	miR-221/222			
	miR-203			
CSCs	miR-128	Down-regulated	Bmi-1	
			ABCC5	

While numerous miRNAs have been shown to control EMT during breast tumor progression, recently miRNAs have been identified that govern multiple aspects of metastasis, including microenvironmental remodeling, in addition to epithelial plasticity (Fig. 1.11) (Hinck & Näthke, 2014; Nicoloso et al., 2009; Wang & Wang, 2012). miR-148b, for instance, is downregulated in aggressive breast tumors and has been found to be a major coordinator of malignancy by regulating over 130 genes involved in epithelial cell motility and stromal cell proliferation (Hinck & Näthke, 2014). The well-characterized miR-200 family and let-7 family, beyond the regulatory role in normal development and tumorigenesis, also inhibit invasion, metastasis, and EMT of breast cancer (Liu, 2012). In addition, it was recently showed that cancer cell-derived miRNAs could bind to Toll like receptor of immune cells and act as agonist to induce a pro-metastatic inflammatory response (De Guire et al., 2013).

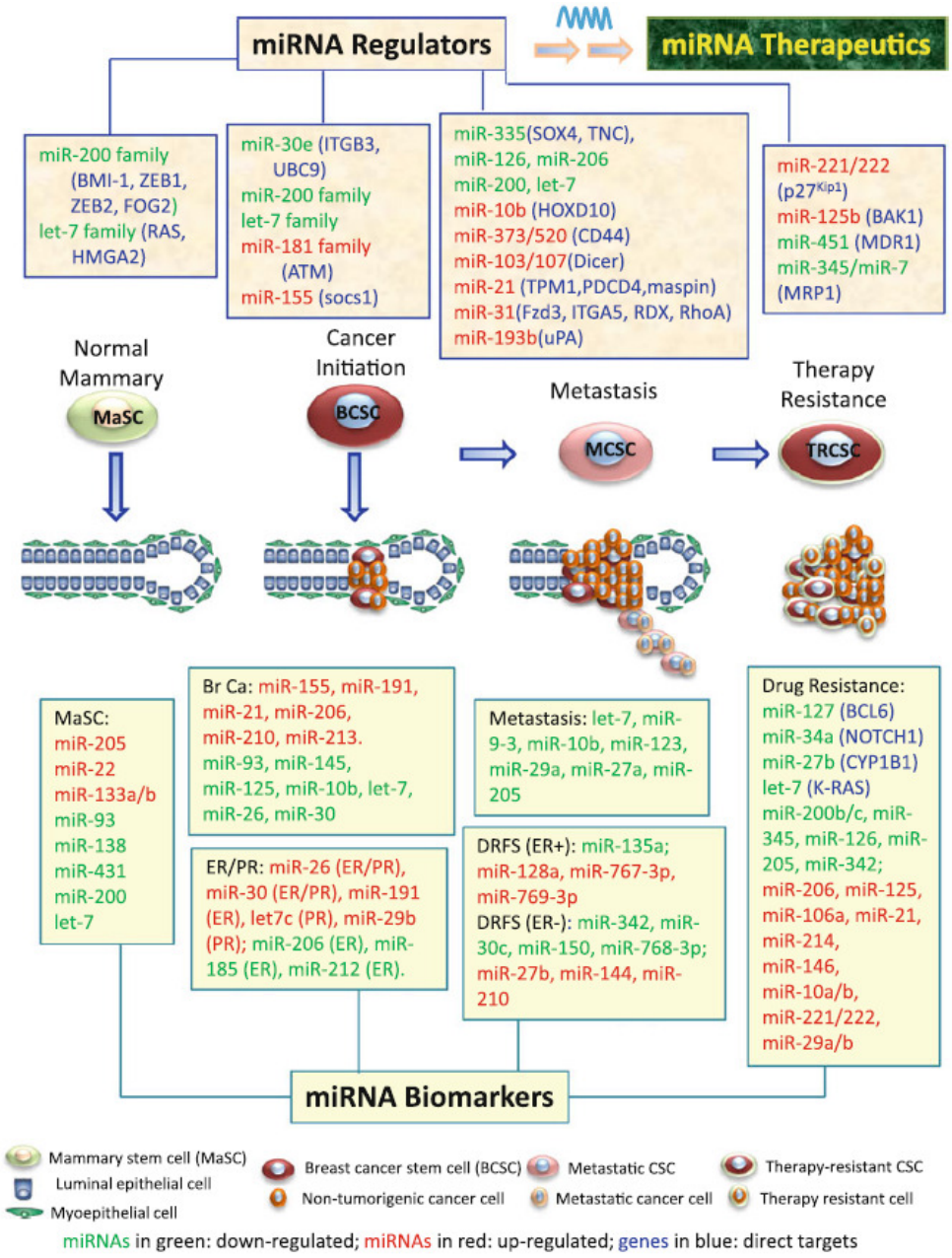


Fig. 1.11 Summary of miRNA regulators and biomarkers in the development of the normal mammary gland, breast cancer initiation, metastasis, and therapy resistance.
MaSC: mammary stem cells; Br Ca: breast cancer; ER: Estrogen receptor; PR; Progesterone Receptor; DRFS: Distant relapse-free survival. Source: Liu, 2012.

Besides the use as predictive biomarkers, the correlation between microRNA expression and response to specific therapies has also suggested their promising potential as therapeutic adjuvant (Iorio & Croce, 2012; Liu, 2012). However, because miRNAs target a broad range of target genes in a context-dependent manner, they may also elicit unexpected effects in certain cells or tissues (Liu, 2012). For that reason, the miRNAs and antagomiRs agents are still in preclinical studies and *in vitro* toxicity studies are underway (Nicoloso et al., 2009). To date, 17 trials are found in ClinicalTrials.gov for the keywords "miRNA" + "breast cancer" and they all refer to miRNA profiling.

Overall, there is still a long way to go before miRNA-targeted therapies as effective approaches for the sequence-specific inhibition of miRNAs in breast and other tumors become a reality, and using miRNAs as diagnostic and prognostic markers remains a scientific and clinical challenge (Wang & Wang, 2012).

1.6 miRNA as a cancer marker

The use of genome-wide approaches has enabled the production of miRNAs fingerprints in a range of tumors and its normal counterpart (Bartels & Tsongaliset 2009; Sandoval & Esteller, 2012; Serpico et al., 2014). As a result, miRNA expression signatures (miRNome) allowed different types of cancer to be discriminated with high accuracy and the tissue of origin of poorly differentiated tumors to be identified. For example, Rosetta Genomics® developed an algorithm based on 48 miRNAs that identified the origin of cancers of unknown primary with an accuracy of 90%. By contrast, mRNA profiles are highly inaccurate indicators of tissue or cancer type. Given that cancers of undefined origin account for approximately 4% of all malignancies and are associated with poor prognosis, the continued development of miRNA classifiers has foreseeable benefits in aiding clinical diagnosis and subsequent treatment (De Guire et al., 2013; Iorio & Croce, 2012).

In addition, information contained in the RNA of the tumor cells is degraded when the tissue is processed and embedded in paraffin, the most common procedure for tumor storage (De Snoo et al., 2009; van der Vegt et al., 2009). miRNAs are certainly more stable due to their small size as compared to long mRNAs, allowing expression profiling from fixed tissues or other biological material, and thus supporting their possible use as novel, minimally invasive and robust biomarkers (Iorio & Croce, 2012; Liu, 2012).

Moreover, there's a specific signature of miRNAs for a wide range of disease in blood, urine, saliva and many other biological fluids and the profile of circulating miRNAs has been shown to reflect the pattern observed in the tumor tissues. For instances, Henegan et al. identified breast cancer patients with a sensitivity and specificity of 87.7% and 91% by measuring the circulating levels of miR-195. Even more impressively, overexpression of miR-205 and miR-21 in ductal adenocarcinoma has been reported to precede phenotypic changes in the ducts. This suggests the attractive possibility of using circulating miRNAs as easily detectable tumor biomarkers, especially for early diagnosis (De Guire et al, 2013; Iorio & Croce, 2012; Tétreault & De Guire, 2013).

In a preliminary study in our lab, an expression plate for 95 miRNAs (QuantiMir™ Cancer MicroRNA qPCR Array) was performed on the breast cell lines MCF-10A (non tumoral line), MCF-7 (tumor line) and MDA-MB-231 (metastatic tumor line). Some miRNAs with dysregulated expression were identified, including miR-203a and miR-200c. The dysregulated expression was confirmed with

miRCURY LNA™, a miRNA-specific, LNA™-based system designed for sensitive and accurate detection of microRNA by quantitative real-time PCR using SYBR® Green. Briefly, our data indicate that miR-203a was overexpressed on MCF-7, when compared to all other breast cell lines, and miR-200c was underexpressed in tumor cell lines when compared to the non tumoral line (data not published). Hence, miR-203a may be associated with tumorigenesis and possibly ER status, since of these three cell lines MCF7 is the only to express ERα, and miR-200c could be associated with a better prognosis, given that its expression decreases with the tumor aggression.

1.7 miRNAs Targets

Regarding the identification of miRNA targets, some computational tools are available that can take advantage of algorithms developed for miRNA target prediction such as “Targetscan”, “EMBL”, “PicTar”, “EIMMo”, “Miranda”, “miRBase Targets”, “PITA”, “mirWIP”, “RNA22”, “Tarbase”, “microRNA”, “Diana-microT”, “miRecords” and “Starbase”. TargetScan is the one with the most robust discrimination rankings (Bartel, 2009; Iorio & Croce, 2012).

This identification has been difficult because most of the times only the seed sequence (about 6–8 bases) of the approximately 22 nucleotides aligns perfectly with the target mRNA’s 3’-UTR. Therefore, bioinformatics approaches can only identify putative targets for particular miRNAs through analysis of the miRNA seed sequences. As a result, these miRNAs need to be assayed *in vitro* or *in vivo* to determine if they truly affect the proposed mRNA because in many cases the predicted targets are not real targets. However, since the number of predicted targets is large, this approach is rather laborious (Bartels & Tsongaliset 2009; Croce, 2009).

The vast number of predicted targets, often with quite disparate functions, presents researchers with the challenge of choosing which is worthy of experimental follow-up. In some cases, known properties of a predicted target will suggest that the biological process of interest might be particularly sensitive to changes in its expression, making it especially promising for follow-up. Another way to choose targets to investigate is to assume that those messages with multiple conserved sites and particularly favorable sites might be among the most responsive to the miRNA (Bartel, 2009).

Therefore, in order to select targets that could be of interest in the context of breast cancer, according to some publications and to their function at the cellular level, we used TargetScan to find transcripts with conserved sites for miR-203 and miR-200c, the miRNAs that we previously indentify in breast cancer cell lines. As targets we selected ATM (target of miR-203), BMI1 (target of miR-203 and miR-200c), SIX1 (target of miR-200c), and SOX2 (target of miR-200c).

1.7.1 ATM

Cellular sensitivity to DNA damaging agents such as radiotherapy and radiomimetic drugs are regulated by a cascade of DNA damage response proteins. Among them is ATM (ataxia-telangiectasia mutated). *ATM* gene is mutated in the autosomal recessive disorder Ataxia-Telangiectasia, manifested by progressive neuronal degeneration, cancer predisposition, immunodeficiency and hypersensitivity to radiotherapy (Guo et al., 2013; Hesse et al., 2013). This gene encodes a serine/threonine kinase, member of the phosphoinositide 3-kinase (PI3-kinase)-like family (PIKK). ATM is a nuclear phosphoprotein activated in response to double strand DNA breaks damage, with a more general role

in signal transduction and in maintaining the stability of the genome by phosphorylation and activation of a series of downstream targets, such as CHK1, CHK2, BRCA1, 53BP1, and MDC1 which signal to DNA repair coordinators such as BRCA2, PALB2 and to cell cycle checkpoints and the apoptotic machinery. There is significant cross-talk between the various pathways of damage response depending on the nature and severity of the DNA damage (Caldo, 2014; Cuatrecasas et al., 2006; Guo et al., 2013).

ATM mRNA is downregulated in breast cancers when compared with normal tissue by competitive RT-PCR. Low ATM expression might thus have effects on crucial cell cycle checkpoints, allowing cells that harbor DNA damage to divide and acquire genetic alterations leading to increased tumor grade (Cuatrecasas et al., 2006).

Additionally, about half of unselected breast cancer patients have been reported to be heterozygotes for *ATM* mutations in some series. While heterozygous carriers do not suffer from ataxia telangiectasia syndrome, they have an increased risk of developing heart disease, diabetes, and cancers, specifically breast cancer, compared to individuals with normal ATM expression levels. Hence, ATM has been proposed as a candidate tumor suppressor with a potential pathogenic function in breast carcinomas (Cuatrecasas et al., 2006; Hesse et al., 2013). Nonetheless, *ATM* hyperactivation has been observed in many stages of tumor tissues. In the early stage of tumorigenesis is oncogene driven and represents the antitumor function of the kinase. However, in late stages contributes to breast cancer metastasis. Moreover, since radiation induces ATM autophosphorylation increasing its protein kinase activity, the use of specific ATM inhibitors might achieve more clinical benefits as this might specifically increase tumor sensitivities to many of the chemotherapeutic drugs as well as radiotherapy (Guo et al., 2013).

ERα downregulates transcription of *ATM* via the activation of miR-18a and miR-106a, which interferes with the induction of cell cycle checkpoints so that cells continue to progress through the cell cycle after DNA damage, and DNA repair is delayed or not engaged. This could explain why ATM levels are higher in ER negative breast cancers and its high expression correlated with recurrence in breast cancer (Caldo, 2014; Guo et al., 2013). Moreover, *ATM* can also be regulated by directly binding of miR-203 to the *ATM* 3'-UTR site, as demonstrated in colorectal cancer (Zhou et al., 2013b).

1.7.2 BMI1

BMI1 (BMI1 polycomb ring finger oncogene) which encodes a polycomb ring finger protein, was identified as a proto-oncogene cooperating with c-Myc during the initiation of lymphomas. It has subsequently been identified as a transcriptional repressor belonging to the polycomb group (PcG) proteins, and is also a key factor in the polycomb repressor complex 1 (PRC1), which serves as an important epigenetic regulatory complex for modulation of chromatin remodeling. To date, many PRC1 target genes have been identified including homeobox (HOX) genes and p16INK4a, whose promoters contain interactive elements which bind directly to BMI1 (Wang et al., 2014; Yin et al., 2013).

A striking finding in recent studies is that the activity of BMI1 is indispensable for cell survival and self-renewal of stem cells or cancer stem cells. Over-expression of *BMI1* has been found in a large number of human cancers, including breast cancer, which indicates that BMI1 might play important roles in cancer initiation and progression. Some molecular mechanisms underlying the role

of BMI1 in cancer have been proposed, such as inhibition of the tumor suppressors p16INK4a and p14ARF, and PTEN to promote EMT and malignancy. In addition, investigation has shown that BMI1 expression is positively correlated with ER status in breast cancer, since it is induced mainly by ERα. However if there is no expression of ERα, BMI1 may be induced by other factors such as E2F1 and MYC, both of which are usually expressed in cancers (Parvathi et al.,2013; Wang et al., 2014; Yin et al., 2013).

BMI1 might be used as a diagnostic and prognostic marker of human cancer. In particular, in patients suspected of having breast cancer, *BMI1* mRNA can be detected with highly sensitive RT-PCR and may be a good marker to support diagnosis even when very few cells are obtained (Parvathi et al., 2013).

Furthermore, functional studies support a pivotal role of miR-200c and miR-203 in regulating self-renewal of mammary stem cells through a direct targeting of *BMI1* 3'-UTR. On the contrary, *BMI1* can also inhibited miR-200c expression, but not miR-203 (Liu, 2012; Yin et al., 2013).

1.7.3 SIX1

In vertebrates, the Six family is classified into three subgroups, SIX1/SIX2 (So), SIX3/SIX6 (Optix), and SIX4/SIX5 (Dsix4), which are characterized by a Six-type homeodomain and Six-domain. The SIX family members are known to play an important role in the expansion of precursor populations prior to differentiation. SIX homeobox 1 (SIX1), the most extensively investigated, is involved in the development of many tissues and organs, such as muscle, kidney, auditory system, sensory organs and craniofacial structures. SIX1 is broadly expressed in various tissues and organs, although levels vary, and throughout developmental from embryo to adulthood, suggesting that there may be a basic promoter sequence to maintain its basic expression in these various contexts. In addition, much attention has been paid to the role of SIX1 in tumorigenesis, including breast cancer (Jin et al., 2014; Wu et al., 2014). *SIX1* mRNA is overexpressed in 50% of primary breast cancers, and in a much larger 90% percent of metastatic lesions, suggesting that it may be involved in more than just tumor initiation (Iwanaga et al., 2012).

SIX1 overexpression can activate breast cancer cell proliferation and stimulate tumorigenesis by directly activating cyclin A1 transcription, a critical factor in cell proliferation, survival, DNA repair and angiogenesis. Other pro-tumorigenesis genes are also regulated by SIX1 (e.g., cyclin D1, c-Myc, and Ezrin). In addition to the role that SIX1 plays in proliferation and survival, its overexpression also leads to the induction of an epithelial to EMT via upregulation of TGF-b signaling and may also increase the cancer stem cell or tumor initiating cell population via its ability to activate both the TGF-b signaling and MEK/ERK signaling pathways. SIX1 can induce as well lymphangiogenesis and distant metastasis by up-regulating VEGF-C in breast cancer (Iwanaga et al., 2012; Li et al., 2013b; Wu et al., 2014).

Moreover, elevated expression of SIX1 is associated with relapse of breast cancer treated with paclitaxel-containing chemotherapy. Hence, overexpression of SIX1 correlates significantly with worse survival (Iwanaga et al., 2012; Li et al., 2013b).

Surprisingly, a recent study found that microRNA-185 can inhibit SIX1 expression, resulting in the suppression of tumor growth. Because a single gene may be targeted by one or multiple miRNAs

SIX1 is likely to be regulated by other miRNAs (Wu et al., 2014). In fact, according to TargetScan, *SIX1* mRNA is targeted by several miRNAs including miR-200c.

1.7.4 SOX2

The recent interest in stem cell research and its clinical implications have brought into focus transcription factors of the sex-determining region (SRY)-related high mobility group (HMG)-box family (SOX family). SOX2 is widely regarded as a key founding member of the cohort of core transcription regulators that controls pluripotency and self-renewal in embryonic stem cells by physically interacts with OCT4 and NANOG forming a protein complex that binds the promoters of numerous stem cell differentiation factors, suppressing their expression. Thus, SOX2 plays critical roles in the embryonic development of several tissues and in organogenesis. It is expressed in stem cells and precursor cells during development, displaying a restricted spatial-temporal pattern, and therefore it is likely to be involved in self-renewal and precursor differentiation. Due to its role in maintenance of embryonic stem cells and tumorigenesis, SOX2 has been implicated as a marker for cancer stem cells (Huang et al., 2014; Lengerke et al., 2001; Tam & Ng, 2014; Zhang et al., 2012).

Consistent with their role in sustaining stemness of embryonic stem cells, pluripotency-related factors have been suggested to be expressed with higher frequency in tumors displaying lower degrees of differentiation (Lengerke et al., 2001). However, to date, little is known about the role of SOX2 during breast tissue development, and no significant SOX2 expression has been detected in healthy human breast tissue. There is, however, increasing evidence showing that SOX2 is expressed in human breast cancers. These suggest that SOX2 plays a role in breast cancer tumorigenesis (Huang et al., 2014; Zhang et al., 2012).

SOX2 appears to be preferably expressed in basal-like breast carcers. This observation indicated that SOX2 expression decreases upon tumor cell differentiation, raising the possibility that SOX2 may be important in maintaining the ‘low differentiation’ status, tumor progression and lymph node metastasis of breast cancer (Huang et al., 2014; Lengerke et al., 2001).

Interestingly, it was found that miR-200c regulates SOX2 expression through a feedback loop and is associated with colorectal carcinoma stemness, growth, and metastasis (Lu et al. 2014).

1.8 Goals

Since potential targets of miRNAs are often provided only by bioinformatic tools there is a gap in this area of study, which leads to the topic of this dissertation. Therefore, parallel to the quantification of the expression of these miRNAs in tumor tissue from patients with breast cancer, it is necessary to observe the expression of these miRNAs targets described bioinformatically, in order to infer if *in vivo* when a miRNA is overexpressed its targets are downregulated and subsequently the respective protein is underexpressed, and vice-versa.

Hence, the aim of this dissertation is to analyze the expression of miR-203 and miR-200c putative targets – ATM, BMI1, SIX1 and SOX2 – in breast cancer samples by immunohistochemistry.

As specific objectives we delineated the following:

- To determine the expression of ATM, BMI1, SIX1 and SOX2 in breast cancer samples by immunohistochemistry;

- To compare immunohistochemistry profile of ATM, BMI1, SIX1 and SOX2 with histologic characteristics of breast cancer samples and;
- To compare immunohistochemistry profile of ATM, BMI1, SIX1 and SOX2 with miR-203 and miR-200c expression of breast cancer samples

2 Materials and Methods

2.1 Breast Cancer Samples

Blood samples, fresh tumor samples preserved in RNALater and formalin fixed and paraffin embedded (FFPE) tumor and normal counterpart tissue were collected from patients diagnosed and subject to lumpectomies between 2013 and 2014. Collection of biological samples and clinical / personal information was carried out by the Surgical and Pathology Departments of the Breast Pathology Unit of Central Lisbon Hospital, in accordance with the Declaration of Helsinki; the confidentiality of patients was fully safeguarded. To date, the biobank is composed of 121 diagnosed and characterized samples from a total of 115 patients. Samples were diagnosed, graded according to Elston-Ellis, and staged under the competence of pathologist Dr.ª Manuela Martins, and were further characterized for ER, PR, HER2 and Ki67 expression by immunohistochemistry. The majority of tumors were invasive carcinomas NOS, accounting for 66.4% of all tumors collected. The remaining were invasive lobular carcinomas (12.6%), carcinomas in situ (6.7%), mucinous carcinoma (3.4%), invasive papillary carcinoma (2.5%), invasive micropapillary carcinoma (2.5%), invasive cribriform carcinoma (0.8%), medullary carcinoma (0.8%), neuroendocrine tumor (0.8%), intraductal carcinoma (0.8%), sebaceous carcinoma (0.8%), mixed type carcinoma (0.8%), or glycogen-rich clear cell carcinoma (0.8%). 88.8% of the samples were estrogen positive (>1% stained cells), 82.3% progesterone positive (>1% stained cells), 12.3% HER2 positive (3+ by immunohistochemistry or gene amplification by FISH) and 46.5% high ki67 (>14% stained cells). Women's mean age at diagnosis was 63 years (range 35-86). 91 of these women were postmenopausal and 1 was perimenopausal. 35.7% were classified as lymph node positive.

2.2 miRNA Quantification

miRNAs quantification was performed at CIGMH by Dr. Bruno Gomes. Total RNA was extracted from the same FFPE tissue samples of breast cancer used for diagnosis, and normal counterpart. From tumor specimens ten sections of 10 µm were collected whereas from normal counterpart a punch was made in the sample to guarantee that only nontumoral tissue was collected. RNA was isolated following the manufacturer's protocol (RecoverAll™ Total Nucleic Acid Isolation Kit - ambion™ Total Nucleic Acid Isolation Kit – ambion® by life technologies™). Briefly, samples were deparaffinized with xylene at 50ºC, washed with ethanol and then air dried. Next, protein digestion was achieve with a protease incubation for 15 min at 50ºC followed by 15 min at 80ºC. After dilution in an isolation additive/ethanol mixture the samples were filtered for nucleic acid isolation. For purification, the samples were treated with a DNase for 30 min at room temperature and then filtered again. Elution was made with RNase-free water and the total RNA concentration measured using NanoDrop 1000 (Thermo Scientific). Samples were stored at -80ºC.Template RNA was then adjusted to 5ng/µL and, using miRCURY LNA™ Universal RT microRNA PCR (Exiqon), incubated with enzyme mix for 60 min at 42ºC followed by heat-inactivation of the reverse transcriptase for 5 min at 95ºC. Samples were immediately cool to 4ºC. No reverse transcription enzyme and no sample negative controls were included. For real-time PCR amplification the cDNA was diluted 80x, merged with PCR Master mix

and PCR primer mix for miR-203 or miR-200c and inserted in the cycler for a 40 amplification cycles run (10sec at 95ºC and 1 min at 60º). Threshold cycles were defined manually for the first cases and used for the remaining samples.

A fold-change expression was calculated to the tumor samples relatively to the normal counterpart and a cut-off established as follows:

- If fold-change ≤ -2.5, the miRNA was considered underexpressed in the tumor relatively to the normal counterpart;
- If fold-change ≥ 2.5, the miRNA was considered overexpressed in the tumor relatively to the normal counterpart;
- If fold-change > -2.5 and < 2.5, it was considered that the miRNA levels do not vary between tumor and normal tissue.

2.3 Immunohistochemistry

2.3.1 Optimization

Immunohistochemistry (IHC) technique was implemented from scratch. Moreover, as the work conditions of the primary antibodies may not correspond to the ones recommended by the manufacturer there was also the need to previously identify, for each primary antibody, the optimal dilution that satisfied the requirements of this study. For each target, we purchased 2 antibodies that were subject to optimization prior to application in the patients' samples:

- Anti-ATM, rabbit, monoclonal, clone Y170, Millipore, lot 2445819, ref. 04-200;
- Anti-ATM, rabbit, monoclonal, clone Y170, Millipore, lot 2472976, ref. 04-200;
- Anti-BMI1, mouse, monoclonal, clone 229F6, Nordic-MUbio, lot 1362, ref. MUB2004S;
- Anti-BMI1, mouse, monoclonal, clone 10H8, Cell Applications, lot 101, ref. CB16351;
- Anti-SIX1, rabbit, polyclonal, Sigma-Aldrich, lot B40182, ref. HPA001893;
- Anti-SIX1, mouse, monoclonal, clone CL0185, Sigma-Aldrich, lote 02582, ref. AMAb90544;
- Anti-SOX2, rabbit, polyclonal, Sigma-Aldrich, lote S9072, ref. S9072;
- Anti-SOX2, mouse, monoclonal, clone 10F10, Sigma-Aldrich, lot PM12110682, ref. SAB5300177.

The tissue sections used as positive controls – normal breast tissue for ATM and BMI1, normal cervix for SIX1, and normal tonsil for SOX2 – were cut at the Pathology Department of Central Lisbon Hospital. Immunohistochemistry to assess the optimal conditions was performed manually at CIGMH and the main steps that were adjusted were:

- Antigen retrieval – duration (10min-20min), microwave potency (50-100%) and buffer (tris-EDTA pH9 or citrate pH6), and;
- Primary antibody incubation – duration (30min-60min), and dilution (1/50-1/2000).

Optimal dilution was considered the one which allowed more specific staining intensity in all expected structures of the positive control sample with no background or nonspecific staining. Although all antibodies have been tested , only two were optimized.

2.3.2 Technique

IHC technique (Table 2.1) was performed manually (Fig. 2.1) on the same FFPE breast cancer samples from which total RNA was extracted for miRNA quantification. So far, 45 samples of the biobank were analyzed, corresponding to 43 patients. Merely tumor samples were used since normal mammary gland was also present in all of them, avoiding the need to use another sample only to observe the protein expression in the normal counterpart.

Fig. 2.1 Immunohistochemistry technique being performed manually.

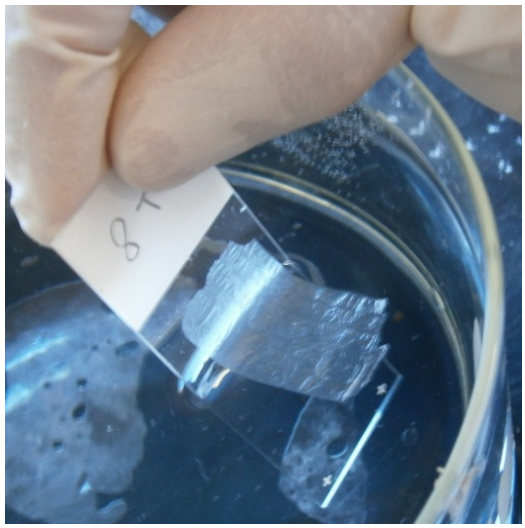


Fig. 2.2 Tumor section of ~2 µm after being cut on a micrtotome.

Serial 2-3µm paraffin sections were cut onto Superfrost® Plus slides (Thermo Scientific, ref. J1800AMNZ) at the Pathology Department of Egas Moniz Hospital (Fig. 2.2). The subsequent steps were performed at CIGMH. For proper adhesion, sections were left overnight at 37ºC and then put in microwave for 2 minutes at full potency. Then, they were deparaffinized in Xylene, and rehydrated. Antigen retrieval was achieved in MW with Tris-EDTA Buffer (10mM Tris Base, 1mM EDTA Solution with 0.05% Tween 20, pH 9.0), and endogenous peroxidase activity blocked with 3% Hydrogen Peroxide. Protein block was performed with ready-to-use (2.5%) normal horse blocking solution (Vector Laboratories, ref. MP-7402), and then sections were incubated (Fig. 2.3 - A) at room temperature with:

- Anti-SIX1 antibody (mouse, clone CL0185, Sigma-Aldrich), dilution 1:100, for 1 hour or;
- Anti-SOX2 antibody (mouse, clone 10F10, Sigma-Aldrich), dilution 1:500, for 30 minutes.

Antibodies were diluted in Phosphate buffer saline (PBS) (137mM NaCl, 2,7mM KCl, 10mM Na_2HPO_4, 1,8mM KH_2PO_4, with 0.05% Tween 20, pH 7.4). Bound antibody was visualized using ready-to-use ImmPRESS™ anti-mouse Ig detection kit (Vector Laboratories, ref. MP-7402) (Fig. 2.3 -

B) based on horseradish peroxidase (HRP) and 3,3'-diaminobenzidine (DAB) (Invitrogen, ref. 88-2014) (Fig. 2.3 - C). Slides were counterstained with Mayer's Hematoxylin (Sigma-Aldrich, ref. MHS80) (Fig. 2.3 - D). Subsequently, sections were washed, dehydrated in a graded alcohol series, clarified in Xylene and mounted with Entellan® (Merck, ref. 1.07961.0500).

In order to validate the results of samples that eventually did not express the studied protein, we included, per run, a slide containing a section of a tissue with known expression that served as control, the same kind that was used during optimization step. That way we could guarantee that the technique performed appropriately.

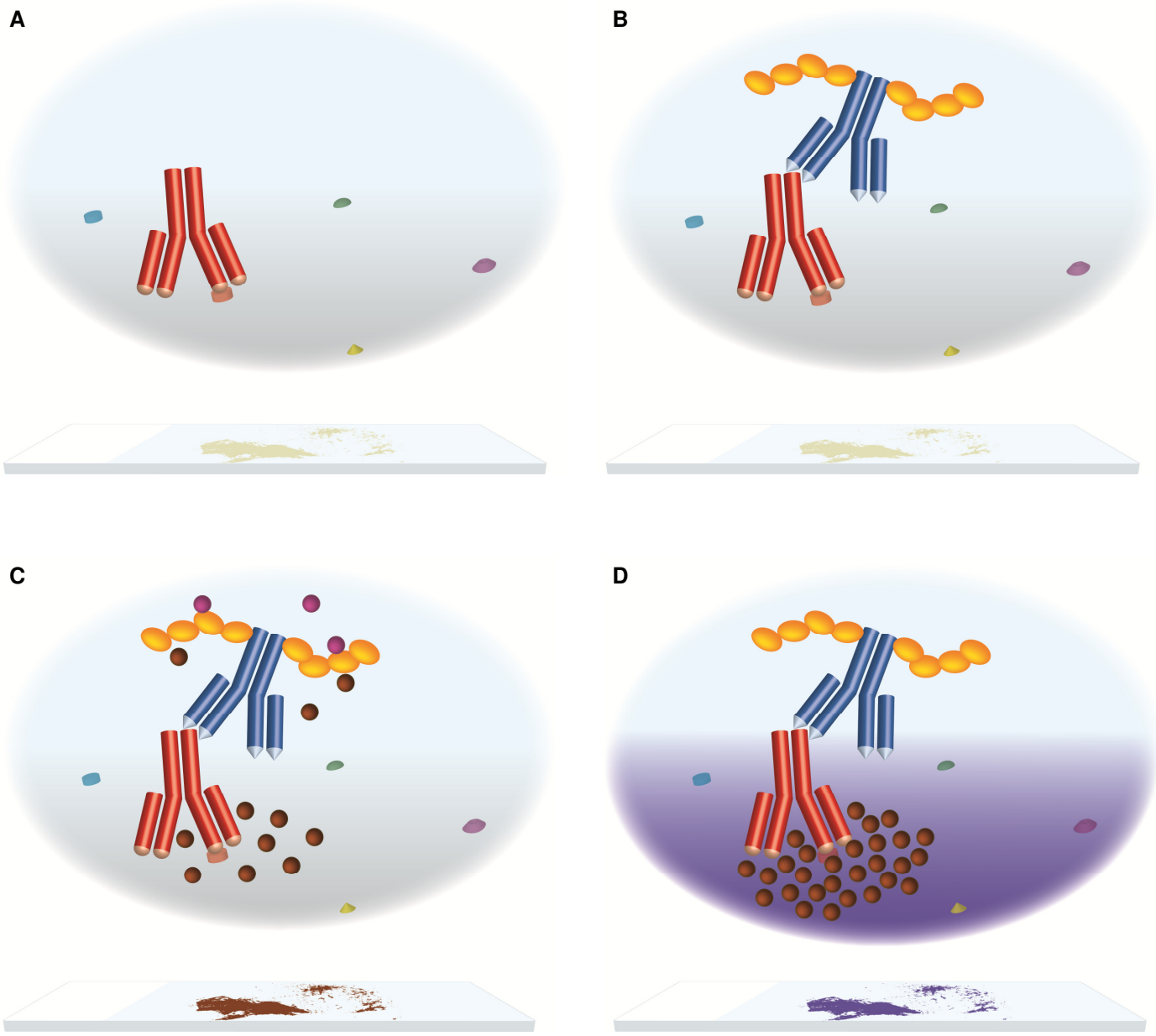


Fig. 2.3 Schematic representation of what happens during immunohistochemistry of the tumor section, macroscopically and at molecular level.
A) Mouse primary antibody (red) specifically recognize its epitope; B) ImmPRESS™ anti-mouse Ig (blue) recognizes the primary antibody FC region; C) DAB (purple) is converted by HRP enzymes (orange) of the detection system in a insoluble precipitate (brown). D) Hematoxylin stains de nuclei giving contrast to a tissue previously translucent.

Table 2.1 Immunohistochemistry Protocol.
All the steps were performed at room temperature except for antigen retrieval (step 4). In this step, the buffer, in which the slides are dipped, was boiled in a microwave. Same amount of slides were always used to ensure that the temperature cycles were uniform.

Step	Time	
	SIX1	SOX2
1. Deparaffinization in Xylene	2 x 10min	
2. Rehydrate through decreasing concentrations of Ethanol	-	
3. Wash in bidistilled water	-	
4. Microwave the slides in 600mL of Tris-EDTA Buffer (pH 9.0) at 80% potency	20min	
5. Let the solution cool at room temperature	5min	
6. Wash the slides in running water	15min	
7. Block endogenous peroxidase activity with Hydrogen Peroxide at 3%	10min	
8. Wash in running water	-	
9. Apply hydrophobic pen	-	
10. Wash with PBS with Tween 20 at 0.05%	-	
11. Incubate with blocking solution	20min (kit)	
12. Incubate with Primary Antibody (dilution)	60min (1/100)	30min (1/500)
13. Wash in PBS with Tween 20 at 0.05%	2 x 5min	
14. Incubate with Secondary Antibody	30min (kit)	
15. Wash in PBS with Tween 20 at 0.05%	2 x 5min	
16. Incubate with DAB	5min	
17. Wash in running water	2min	
18. Counterstain in Mayer's Hematoxylin	2min	
19. Wash in running water	5min	
20. Dehydrate through increasing concentrations of Ethanol and Xylene	-	
21. Mount with Entellan®	-	

2.3.3 Slide evaluation

All slides were further evaluated by 3 observers that used the same criteria, previously validated by a pathologist. Sections were evaluated in its totality and positivity was considered when 10% or more of tumor cells stained specifically with a moderate to strong intensity. After an initial independent observation, all results were reviewed in group and a final result assigned to each case by consensus. Slide evaluation was blind to tumor characteristics and other clinical data.

Afterward, images were captured at the Pathology Department of Egas Moniz Hospital with a Leica DFC320 digital camera coupled to a Leica DM1000 microscope using Leica IM50 software.

2.4 Data Treatment

Statistical analyzes were carried out with IBM® SPSS® Statistics 20 in order to properly interpreted all collected data. Fisher's exact test was employed for categorical data analysis and Fisher-Freeman-Halton test when contingency tables were larger than 2x2; t-test was applied to interval variables and the Mann-Whitney test for non-normal distributions. Statistical significance was established at $P < 0.05$.

3 Results

3.1 Sample characterization

45 samples were analyzed corresponding to 43 patients. Women's mean age at diagnosis was 62 years (range 42-84), mean age of menarche 13 years (range 9-15) and the mean age of menopause 49 years (range 36-59). On average women had 2 pregnancies (range 0-8) and a mean body mass index (BMI) of 26.56 (range 19.23-38.58). 37.8% of them have used oral contraceptives and 15.0% had hormone replacement treatment. Relatively to the tumors, 71.1% were invasive carcinomas NOS (Fig. 3.1), 8.9% were invasive lobular carcinomas (Fig. 3.2), and the remaining were carcinomas in situ (6.7%), mucinous carcinomas (6.7%), invasive papillary carcinomas (2.2%), invasive micropapillary carcinoma (2.2%) and a medullary carcinoma (2.2%). 86.4% of the samples were estrogen positive, 79.1% progesterone positive, 13.6% HER2 positive and 45.5% high ki67. Accordingly, 42.2% of the tumors were Luminal A, 31.1% Luminal B (HER2-), 11.1% Luminal B (HER2+), 6.7% Basal-Like, and 6.7% match in situ carcinomas and therefore weren't categorized. The most abundant grade was II representing 65% of all tumors, and the stage was IA, representing 30.2% of all tumors. miRNA quantification reveal that miR-200c was downregulated in 12.8% of the samples and upregulated in 23.1%. As for miR-203, 20.5% samples presented downregulation and 30.8% upregulation. Tumor classification and clinical information are summarized in Table 3.1 and Table 3.2.

Additional characteristics like ethnicity and alcohol consumption, among others, were not evaluated due to the substantial lack of data for all patients.

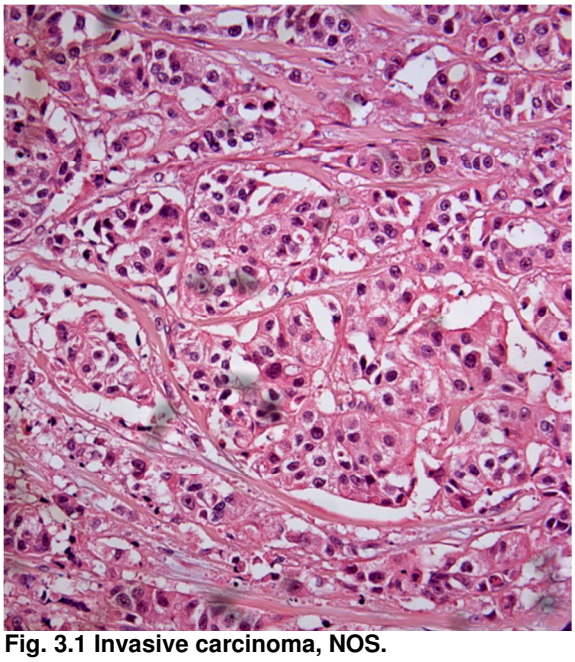


Fig. 3.1 Invasive carcinoma, NOS.
Malignant cells form nests with small ductal structures. The stroma is very fibrotic. Hematoxylin and Eosin stain (20x).

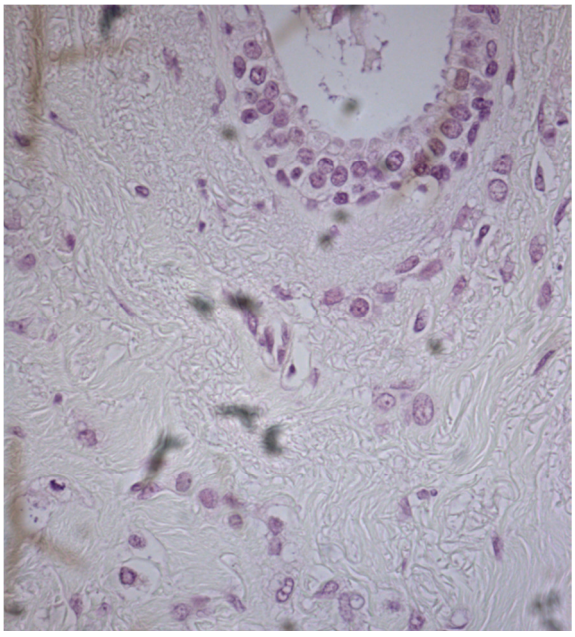


Fig. 3.2 Invasive lobular carcinoma.
Malignant cells infiltrate the stroma next to a normal duct. Hematoxylin and Eosin stain (40x).

Table 3.1 Frequency analysis of clinical parameters.

Characteristics	Value
Age at Diagnosis	
Mean	62
Range	42-84
N	44
BMI [a]	
Mean	26,56
Range	19.23-38.58
N	40
Age of Menarche	
Mean	13
Range	9-15
N	41
Age of Menopause	
Mean	49
Range	36-59
N	32
Number of Pregnancies	
Mean	2
Range	0-8
N	41
Oral Contraceptives (%)	
No	22 (56.4%)
Yes	17 (37.8%)
Total	39
Hormone Replacement Treatment (%)	
No	34 (85.0%)
Yes	6 (15.0%)
Total	40

[a] BMI: Body Mass Index

Table 3.2 Frequency analysis of pathological parameters.

Characteristics	Value
Histopathological Type (%)	
Carcinoma *In Situ*	3 (6.7%)
Invasive carcinoma, NOS	32 (71.1%)
Invasive lobular carcinoma	4 (8.9%)
Mucinous carcinoma	3 (6.7%)
Medullary carcinoma	1 (2.2%)
Invasive papillary carcinoma	1 (2.2%)
Invasive micropapillary carcinoma	1 (2.2%)
Total	45
Grade (%)	
I	7 (17.5%)
II	26 (65.0%)
III	7 (17.5%)
Total	40
Estrogen receptor (%)	
Negative	6 (13.6%)
Positive	38 (86.4%)
Total	44
Progesterone receptor (%)	
Negative	9 (20.9%)
Positive	34 (79.1%)
Total	43
HER2 (%)	
Negative	38 (86.4%)
Positive	6 (13.6%)
Total	44
Ki67 (%)	
Negative	24 (54.5%)
Positive	20 (45.5%)
Total	44
Molecular Subtype (%)	
Luminal A	19 (43.2%)
Luminal B (HER2-)	14 (31.8%)
Luminal B (HER2+)	5 (11.4%)
Basal-Like	3 (6.8%)
In Situ	3 (6.8%)
Total	44
Stage (%)	
0	1 (2.3%)
IA	13 (30.2%)
IB	8 (18.6%)
IIA	10 (23.3%)
IIB	9 (20.9%)
IIIA	1 (2.3%)
IIIC	1 (2.3%)
Total	43
miR-200c (%)	
Without variation	25 (64.1%)
Downregulated [a]	5 (12.8%)
Upregulated [b]	9 (23.1%)
Total	39
miR-203 (%)	
Without variation	19 (48.7%)
Downregulated [a]	8 (20.5%)
Upregulated [b]	12 (30.8%)
Total	39

[a] Downregulated: fold change ≤ -2.5
[b] Upregulated: foldchange ≥ 2.5.

3.2 Immunohistochemistry Optimization

The optimal staining was considered the one which allowed specific intense staining in all expected structures of the sample used as control, with no background or nonspecific staining. Following this simple criterion, each variation to the protocol was evaluated during optimization of all the antibodies.

Table 3.3 summarizes the outcome of all the variations tested per antibody. Only Anti-SIX1, clone CL0185 (Fig. 3.3), and Anti-SOX2, clone 10F10 (Fig. 3.4), immunostain were validated, all the other antibodies didn't perform properly; for this reason ATM (Fig. 3.5) and BMI1 (Fig. 3.6) could not be assessed in the tumor samples.

Table 3.3 Outcome of immunohistochemistry technique optimization.

Antibody	Specifications	Slides	Outcome
ATM	rabbit, monoclonal, clone Y170, Millipore, lot 2445819, ref. 04-200	5	Wrong structures stained
	rabbit, monoclonal, clone Y170, Millipore, lot 2472976, ref. 04-200	3	Wrong structures stained
BMI1	mouse, monoclonal, clone 229F6, Nordic-MUbio, lot 1362, ref. MUB2004S	17	Lack of staining
	mouse, monoclonal, clone 10H8, Cell Applications, lot 101, ref. CB16351	12	Overstaining
SIX1	rabbit, polyclonal, Sigma-Aldrich, lot B40182, ref. HPA001893	34	Unspecific staining
	mouse, monoclonal, clone CL0185, Sigma-Aldrich, lot 02582, ref. AMAb90544	18	Optimal staining
SOX2	rabbit, polyclonal, Sigma-Aldrich, lot 061M0604, ref. S9072	23	Overstaining
	mouse, monoclonal, clone 10F10, Sigma-Aldrich, lot PM12110682, ref. SAB5300177	15	Optimal staining

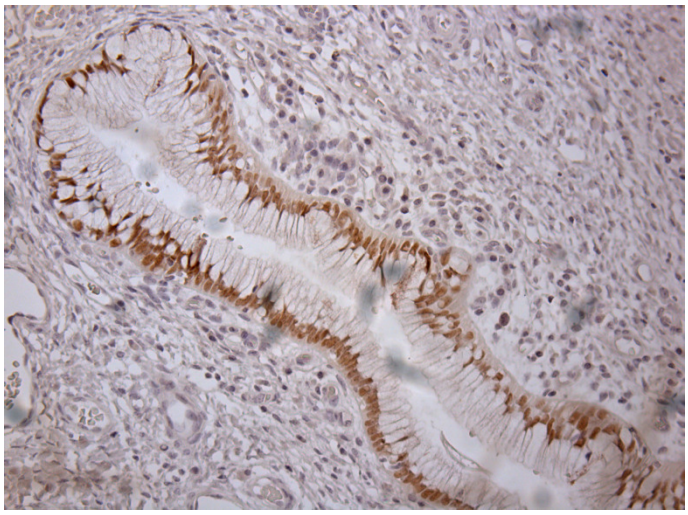


Fig. 3.3 Normal cervix stained for SIX1.
Only the nuclei of glandular epithelium express the protein, as expected (20x).

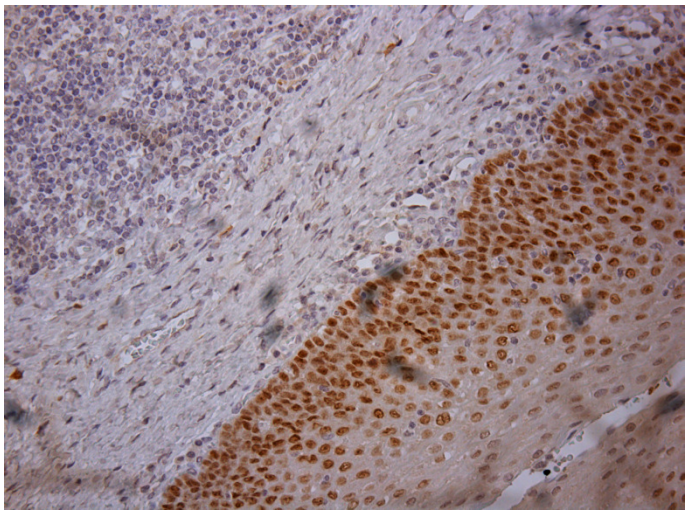


Fig. 3.4 Normal tonsil stained for SOX2.
Only the nuclei of stratified epithelium express the protein, as expected (20x).

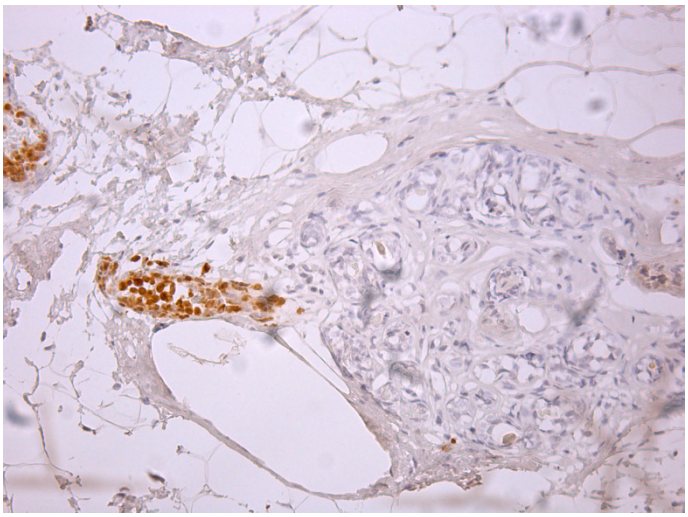


Fig. 3.5 Normal breast stained for ATM.
Nuclei, that should be expressing the protein, appear blue. Only neutrophils appear to be stained, cytoplasmically (20x).

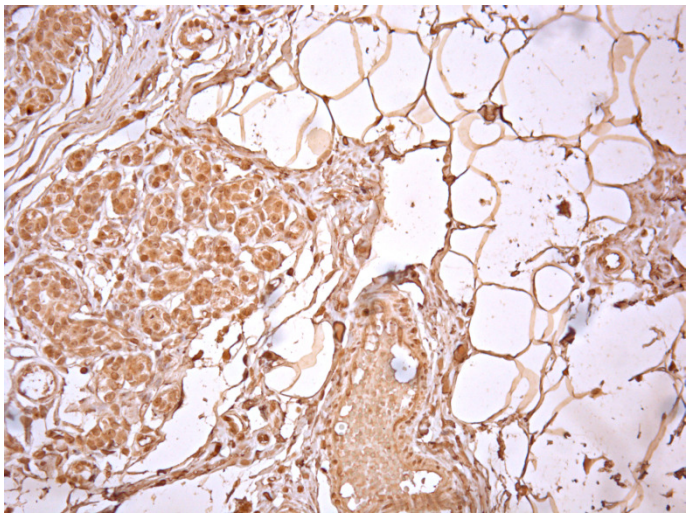


Fig. 3.6 Normal breast stained for BMI1.
Every structure appears to be stained, even at an antibody dilution of 1/1500, clearly indicating its unspecificity (20x).

3.3 Immunohistochemistry Results

The tumor sections stained for SIX1 and SOX2 were evaluated in its totality by all 3 observers. After an initial independent observation, all results were reviewed in group and a final result assigned to each case by consensus. Positivity was considered when 10% or more of tumor cells stained with a moderate to strong intensity (Fig. 3.7). In sum, 13.3% of tumors were SIX1-positive (Fig. 3.8, Fig. 3.9 and Fig. 3.10) and 8.9% were SOX2-positive (Fig. 3.11, Fig. 3.12 and Fig. 3.13). Only 2 tumors were positive for both. All positive cases exhibited normal mammary gland cells negatively stained. In each run, the control tissue performed accordingly. Most of the tumor sections had poorly preserved morphology that occurred during the antigen retrieval step that was too harsh to the tissues due to their insufficient paraffin impregnation (Fig. 3.14). However, morphology did not jeopardize evaluation since there were tumor cells present in all cases.

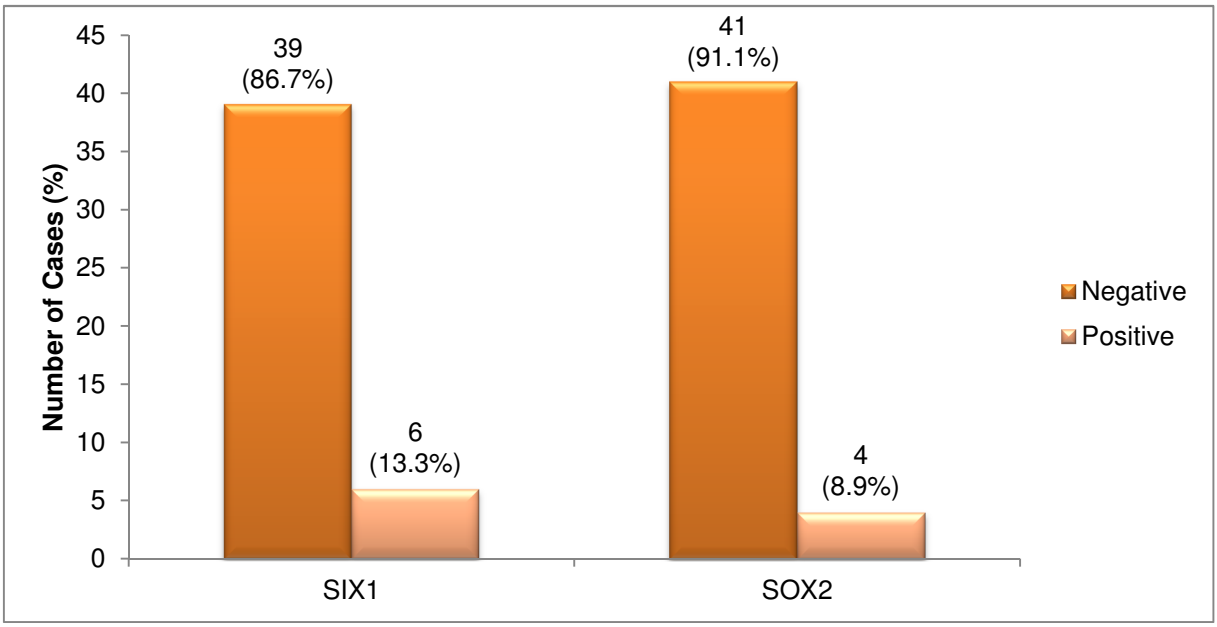


Fig. 3.7 Results of slide evaluation of SIX1 and SOX2 immunostain.

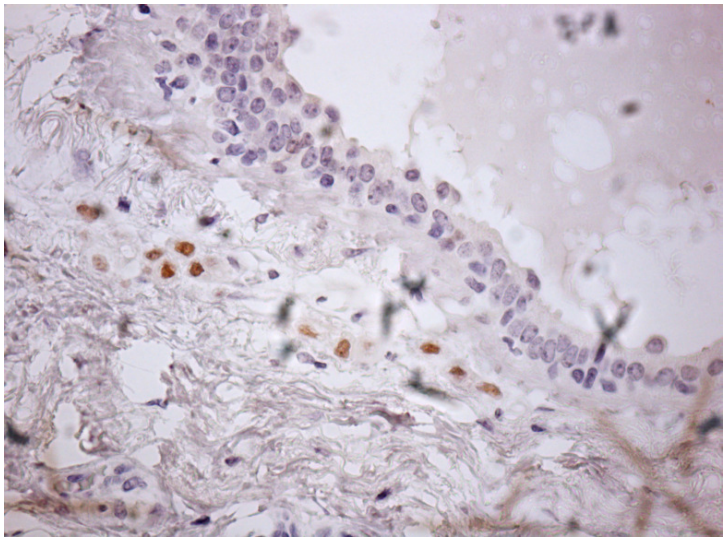


Fig. 3.8 SIX1 positive invasive lobular carcinoma.
Same case as Fig. 3.2 (40x).

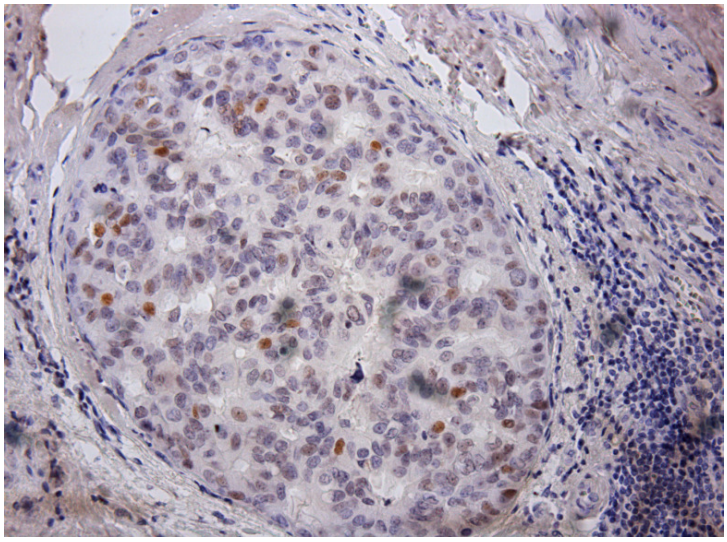


Fig. 3.9 SIX1 positive invasive carcinoma NOS.
(20x).

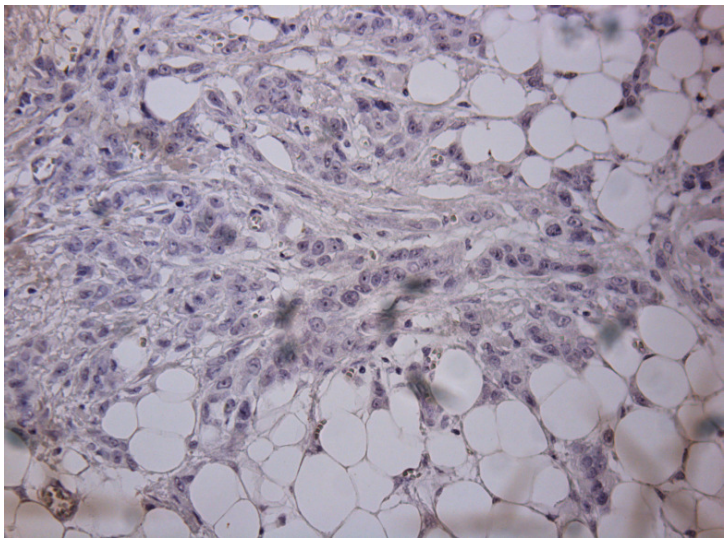


Fig. 3.10 SIX1 negative invasive carcinoma NOS.
(20x).

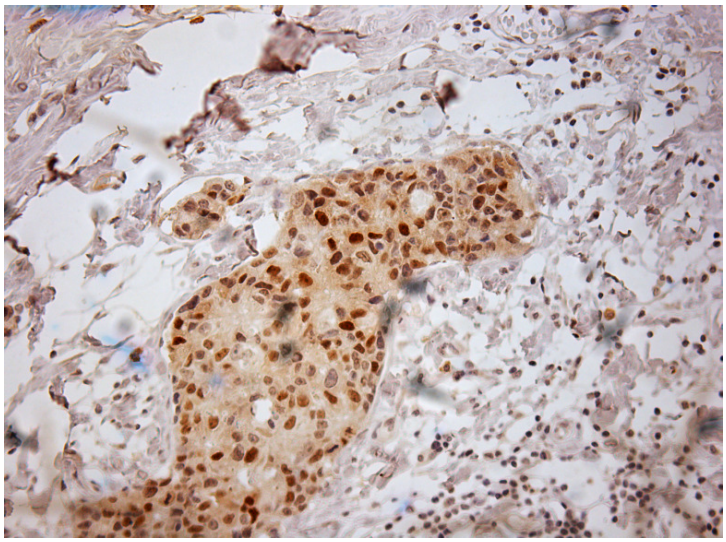


Fig. 3.11 SOX2 positive invasive carcinoma NOS.
Same case as Fig. 3.1 (20x).

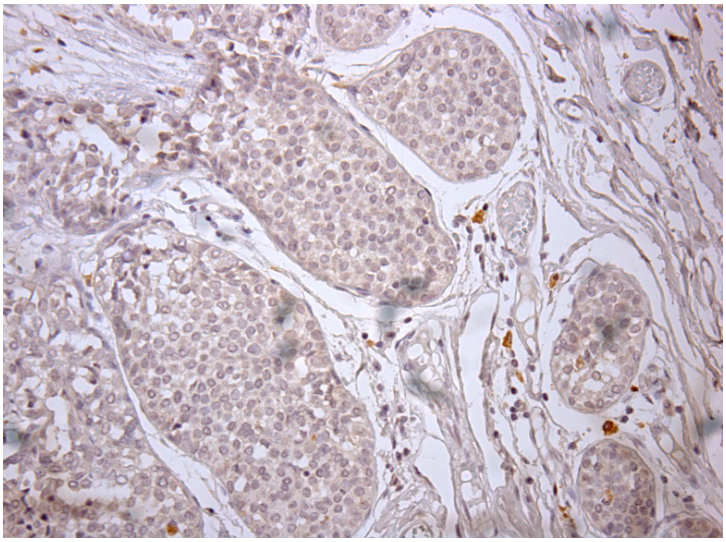


Fig. 3.12 SOX2 negative invasive carcinoma NOS.
(20x).

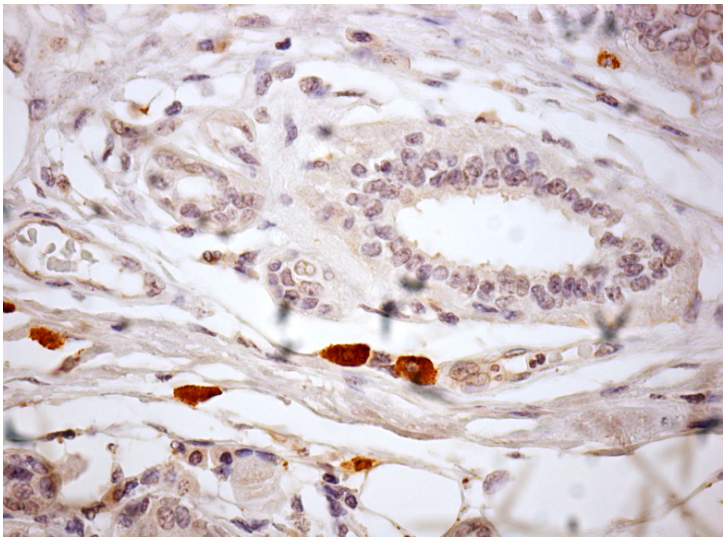


Fig. 3.13 Mast cells stained with anti-SOX2 antibody.
Same case as Fig. 3.12 where it can be seen mast cells exhibiting completely normal unspecific cytoplasmic granular staining (40x).

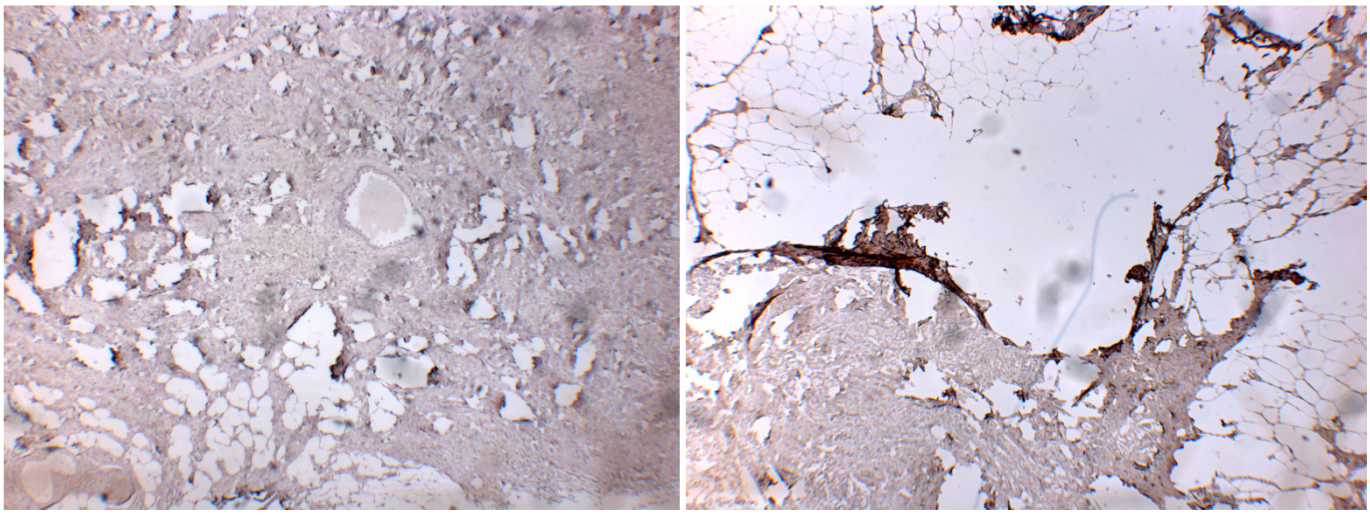


Fig. 3.14 Poor tissue preservation.
Example of poor tissue preservation found in an invasive lobular carcinoma and in an invasive carcinoma, NOS. Note the tattered aspect of these sections and the presence of folded regions (4x).

3.4 Statistical analysis

The results of statistical analysis are summarized in Table 3.4, Table 3.5, Table 3.6, Table 3.7, and Table 3.8. None of the characteristics analyzed exhibited relation with the immunohistochemical expression of both SIX1 and SOX2 except for the number of pregnancies. SIX1 expression appears to be statistically associated with the number of pregnancies (p-value=0.034), being its positivity related with multiple pregnancies. SOX2 does not show the same pattern. Although not statistically relevant, SOX2 displays a certain trend relatively to age at diagnosis (p-value=0.108), BMI (p-value=0.123) and ER (p-value=1.000) since all the positive cases are above 62 years of age at diagnosis, above a BMI of 25 (overweight) and all ER positive. Furthermore, SIX1 expression was not associated with SOX2 expression (p-value=0.08).

The association between miR-200c and miR-203 levels with clinicopathological features was not a purpose of this study, however, their statistical analysis was performed and did not reveal any statistically significant association. Accordingly, it was decided not to include this data.

Table 3.4 Association between clinical parameters and immunohistochemical expression of SIX1 and SOX2.

		SIX1				SOX2			
		-	+	Total	P-value[a]	-	+	Total	P-value[a]
Age at Diagnosis	< 62	20	2	22	0.664	22	0	22	0.108
	≥ 62	18	4	22		18	4	22	
	Total	38	6	44		40	4	44	
BMI	< 24.99 (normal weight)	15	2	17	1.000	17	0	17	0.123
	≥ 25 (overweight)	20	3	23		19	4	23	
	Total	35	5	40		36	4	40	
Age of Menarche	< 13	16	3	19	0.649	17	2	19	1.000
	≥ 13	20	2	22		20	2	22	
	Total	36	5	41		37	4	41	
Age of Menopause	< 49	10	1	11	1.000	8	3	11	0.106
	≥ 49	18	3	21		20	1	21	
	Total	28	4	32		28	4	32	
Oral Contraceptives	Yes	16	1	17	0.363	16	1	17	1.000
	No	18	4	22		20	2	22	
	Total	34	5	39		36	3	39	
Hormone Replacement Treatment	Yes	5	1	6	1.000	5	1	6	0.493
	No	30	4	34		31	3	34	
	Total	35	5	40		36	4	40	

(+) Positive; (-) Negative; BMI: Body Mass Index.

[a] Fisher's Exact Test.

Table 3.5 Association between number of pregnancies and immunohistochemical expression of SIX1 and SOX2.

		SIX1			SOX2		
		-	+	P-value[a]	-	+	P-value[a]
Number of pregnancies[b]	N	36	5	**0.034**	37	4	0.212
	Mean	1.83	3.80		1.95	3.25	
	Std. Deviation	1.81	2.39		1.78	3.40	
	Std. Error Mean	0.30	1.07		0.29	1.70	

(+) Positive; (-) Negative.

[a] t-test.

[b] Kolmogorov-Smirnov test indicates that this variable has Normal distribution (p-value=0.170) and Levene's Test demonstrates that equal variances are assumed (p-value=0.600 for SIX1 and p-value=0.120 for SOX2).

Table 3.6 Association between tumor characteristics and immunohistochemical expression of SIX1 and SOX2.

		SIX1				SOX2			
		-	+	Total	P-value[a]	-	+	Total	P-value[a]
Histopathological Type	Carcinoma In Situ	3	0	3	0.889	3	0	3	0.156
	Invasive carcinoma, NOS	27	5	32		30	2	32	
	Invasive lobular carcinoma	3	1	4		4	0	4	
	Medullary carcinoma	1	0	1		1	0	1	
	Mucinous carcinoma	3	0	3		2	1	3	
	Invasive papillary carcinoma	1	0	1		0	1	1	
	Invasive micropapillary carcinoma	1	0	1		1	0	1	
	Total	39	6	45		41	4	45	
Grade	I	6	1	7	1.000	6	1	7	0.602
	II	22	4	26		24	2	26	
	III	6	1	7		6	1	7	
	Total	34	6	40		36	4	40	
Estrogen Receptor	Negative	6	0	6	0.573	6	0	6	1.000
	Positive	32	6	38		34	4	38	
	Total	38	6	44		40	4	44	
Progesterone Receptor	Negative	6	3	9	0.095	8	1	9	1.000
	Positive	31	3	34		31	3	34	
	Total	37	6	43		39	4	43	
HER2	Negative	32	6	38	0.573	34	4	38	1.000
	Positive	6	0	6		6	0	6	
	Total	38	6	44		40	4	44	
Ki67	Negative	20	4	24	0.673	22	2	24	1.000
	Positive	18	2	20		18	2	20	
	Total	38	6	44		40	4	44	
Molecular Subtype	Luminal A	16	3	19	0.938	18	1	19	0.528
	Luminal B (HER2-)	11	3	14		11	3	14	
	Luminal B (HER2+)	5	0	5		5	0	5	
	Basal-Like	3	0	3		3	0	3	
	In Situ	3	0	3		3	0	3	
	Total	38	6	44		40	4	44	
Stage	0	1	0	1	1.000	1	0	1	0.421
	IA	11	2	13		13	0	13	
	IB	7	1	8		6	2	8	
	IIA	8	2	10		9	1	10	
	IIB	8	1	9		8	1	9	
	IIIA	1	0	1		1	0	1	
	IIIC	1	0	1		1	0	1	
	Total	37	6	43		38	4	43	
miR-200c Regulation	Without Variation	21	4	25	1.000	22	3	25	1.000
	Down Regulation	5	0	5		5	0	5	
	Up Regulation	8	1	9		8	1	9	
	Total	34	5	39		35	4	39	
miR-203 Regulation	Without Variation	14	5	19	0.082	16	3	19	0.800
	Down Regulation	8	0	8		8	0	8	
	Up Regulation	12	0	12		11	1	12	
	Total	34	5	39		35	4	39	

(+) Positive; (-) Negative.

[a] Fisher's Exact Test and Fisher-Freeman-Halton Test

Table 3.7 Association between miRNAs fold-change and immunohistochemical expression of SIX1 and SOX2.

		SIX1			SOX2		
		-	+	P-value[a]	-	+	P-value[a]
miR-200c Fold[b]	N	34	5	0.356	35	4	0.876
miR-203 Fold[b]	N	34	5	0.610	35	4	0.911

(+) Positive; (-) Negative.

[a] Mann-Whitney's test.

[b] Kolmogorov-Smirnov test indicates that these variables don't follow a Normal distribution (p-value=0.000 for both SIX1 and SOX2).

Table 3.8 Association between immunohistochemical expression of SIX1 with SOX2.

		SIX1			P-value[a]
		-	+	Total	
SOX2	-	37	4	41	0.080
	+	2	2	4	
	Total	39	6	45	

(+) Positive; (-) Negative.

[a] Fisher's Exact Test.

4 Discussion

miR-200 family consists of five members that includes miR-200c. They cooperatively down-regulate the E-cadherin transcriptional repressors zinc finger E-box binding homeobox 1 (ZEB1, also known as dEF1) and Smad interacting protein 1 (SIP1, also known as ZEB2) implicated in EMT and tumor metastasis (Serpico, 2013; Singh & Mo, 2013). As regards to miR-203, this miRNA is a key molecule involved in the induction of apoptosis or cell cycle arrest, and inhibition of EMT, migration and invasion. Taken together with the fact that miR-203 is upregulated in nonmetastatic breast cancer cell lines (MCF-7) when compared to non-tumorigenic cells (MCF-10A) but absent in metastatic cell lines (MDA-MB-231), it could be speculated that miR-203 is upregulated in primary breast cancers to restrain metastatic behavior, and downregulation of miR-203 may be a mechanism by which cancer cells can acquire an aggressive, pro-metastatic phenotype (DeCastro et al., 2013; Liu, 2012; Zhang et al., 2011). Therefore, the study of such miRNAs may impact the understanding of the functioning of the tumor cells and this information can be used to predict tumor behavior and, ultimately halting its proliferation. The therapeutic use of miRNAs is far from being a reality however, it can perhaps be used as tumor markers to better characterize them. Since the effect of miRNA is closely associated with the numerous regulated targets, its true potential as a marker depend on the expression of these same targets. Hence the relevance of this work to try to understand if the levels of miR-200c and miR-203 can have an impact as cancer markers through the expression of putative targets involved in tumorigenesis and tumor progression.

The 45 studied samples were randomly selected from the biobank therefore; a major concern was if these samples represented properly the diversity of cases in the biobank. In the selected samples, 71,1% were invasive carcinoma NOS and 8.9% invasive lobular carcinomas, which portrays the 66.4% of invasive carcinoma NOS and 12.6% of invasive lobular carcinomas of the biobank. Additionally, these values reflect published data as invasive carcinoma NOS represents approximately 40% to 70% of tumors, and invasive lobular carcinomas 5% to 15% (Sandhu et al., 2010; WHO, 2012). Regarding molecular patterns, the 45 samples reveal ER positivity in 86.4% of the cases, PR positivity in 79.1%, HER2 positivity in 13.6% and high ki67 in 45.5%. This values follow the ones found in the totality of the biobank: 88.8% of ER positive samples, 82.3% PR positive, 12.3% HER2 positive and 46.5% high ki67. These results could permit a generalization of the data obtained with the 45 samples to the rest of the biobank.

Regarding immunohistochemistry data, ATM nuclear staining of normal breast epithelium, myoepithelial cells, and fibroblasts can serve as an internal positive control (Fig. 4.1). For that reason the absence of nuclear staining in normal breast tissue, using the antibodies purchased, could not be validated. Conversely, BMI1 expression was excluded from this study since it's expected to stain just the nuclei of normal breast (Fig. 4.2) and not the whole tissue. Anti-SIX1 and Anti-SOX2 antibodies performed well and as expected, which allowed validating the immunostaining and, subsequently, apply them to the tumor samples. In the breast cancer samples the technique did not perform as anticipated, since there was extensive loss of architectural detail within loose connective and adipose tissue that may reflect inadequate fixation or faults in tissue processing, such as too short a

processing cycle and/or use of exhausted reagents. Because these samples were handed to us after being previously diagnosed and quantified for miRNAs, we could only try to potentiate tissue adhesion but not revert their poorly processed state. Nonetheless, due to adhesion slides and thinner sections there were always tumor cells to evaluate for the expression of SIX1 and SOX2.

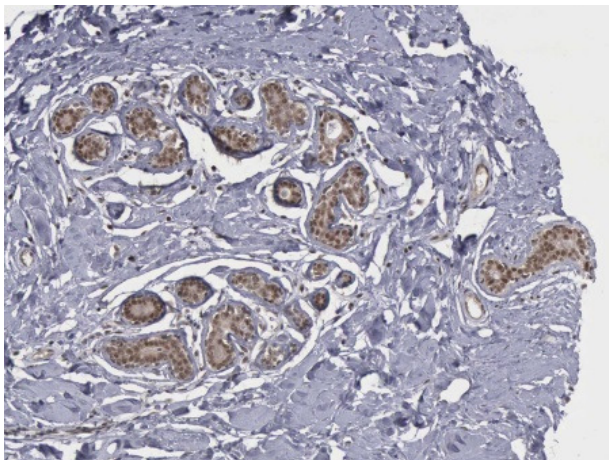


Fig. 4.1 ATM expression by immunohistochemistry in normal breast tissue.
Source:
http://www.proteinatlas.org/ENSG00000149311/tissue/breast

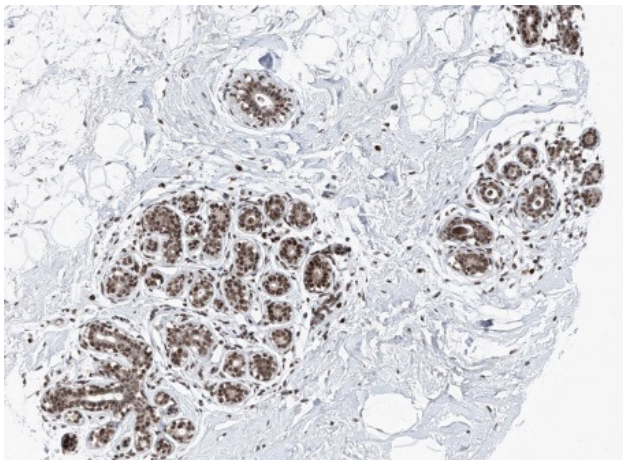


Fig. 4.2 BMI1 expression by immunohistochemistry in normal breast tissue.
Source:
http://www.proteinatlas.org/ENSG00000168283/tissue/breast

SIX1 is involved in the development of many tissues and organs, and their levels vary, throughout developmental stages from embryo to adulthood, suggesting that there may be a basic promoter sequence to maintain its basic expression in these various contexts (Jin et al., 2014; Wu et al., 2014). On the other hand, SOX2 is widely regarded as a key founding member of the cohort of core transcription regulators that controls pluripotency and self-renewal in embryonic stem cells by physically interacting with OCT4 and NANOG forming a protein complex that binds the promoters of numerous stem cell differentiation factors, suppressing their expression. Thus, SOX2 plays critical roles in the embryonic development of several tissues and in organogenesis, and therefore it is likely to be involved in self-renewal and precursor differentiation (Huang et al., 2014; Lengerke et al., 2001; Tam & Ng, 2014; Zhang et al., 2012). Due to their role as general transcription factors promoting cell proliferation, SIX1 and SOX2 have been implicated as potential cancer markers. However, to date, little is known about their role in breast cancer tumorigenesis.

In this study, 13.3% of the samples were SIX1 positive. These results do not follow the trend found in a study conducted in China in 2014 by Jin and colleagues on 314 patients, where IHC analysis revealed a significantly higher strongly positive rate of SIX1 protein in breast cancer (61.8%) and carcinoma *in situ* (23.1%) compared with adjacent normal breast tissues (6.7%). They also concluded that SIX1 protein expression was significantly correlated with clinical stage, lymph node metastasis and Her2 expression status, suggesting that SIX1 may be a useful marker for prognostic evaluation of breast cancer. On the contrary, our results showed no association with any of the clinicopathological characteristics analyzed, except for the number of pregnancies where positivity is associated with multiple pregnancies. This may be counter intuitive since women who have had no children or who had their first pregnancy after age 30 are the ones who have a slightly higher breast

cancer risk (Vogel, 2000; ; WHO, 2012; Yager & Davidson, 2006). Additionally, there seems to be a variation about the accurate location of SIX1 using different antibodies since the protein has been seen in both cytoplasm and nucleus. Hence, further studies are required to elucidate the mechanism that regulates SIX1 homeoprotein localization in order to properly understand its expression and influence in a cellular context. Moreover, according to TargetScan, *SIX1* mRNA is targeted by several miRNAs including miR-200c but, in our study it was not possibly to conclude that miR-200c levels could be related to SIX1 protein levels. Even more, none of the positive cases had miR-200c downregulated, as it should be expected.

Relatively to SOX2 expression, 8.9% of analyzed samples were positive. These results do not fit with those of other studies such as the one of Lengerke and colleagues performed in Germany in 2011 with 95 patients, wherein 27.9% of analyzed samples of invasive breast carcinoma were SOX2 positive such as 44.4% of carcinomas *in situ*. In another study of Huang and colleagues, carried out in 2014 in China with 609 samples, SOX2 was detected in 19.0% of invasive breast carcinomas and 12.3% carcinomas *in situ*. This study also showed that SOX2 expression was associated closely with high histological grade, large tumor size, molecular subtypes with adverse outcome (preferably expressed in basal-like breast cancers), negative hormone receptors status and high proliferation index, in other words, SOX2 contributes to a less differentiated state, tumor progression and lymph node metastasis in breast cancer. However, in our study SOX2 expression could not be associated with any of those characteristics, neither with the other characteristics assessed. There seems to be a trend, although not statistically relevant, of SOX2 with higher body mass index and higher age at diagnosis which could indicate a possibly expression of SOX2 in women with more percentage of body fat, which is known to be associated with higher estrogen levels. Curiously, all the SOX2 positive cases are also ER positive, even though this association is not statically relevant. Nonetheless, the number of positive cases probably did not allow us to draw a conclusion. Furthermore, although *SOX2* mRNA is, according to Targetscan, a target of miR-200c and that it is described *in vivo* in colorectal carcinoma that they are involved in a feedback loop regulation (Lu et al. 2014), we did not find any association between the two of them in our study and again, no positive cases were found with miR-200c downregulation.

There are few studies regarding the expression of SIX1 and SOX2 in human patients with breast cancer and as far as we know, there are no studies that characterize the Portuguese population for expression of both proteins. As such, a conclusion we can draw from our study is that, in fact, in the Portuguese population breast cancers we observe relatively low levels of SIX1 and SOX2, 13.3% and 8.9%, respectively. Furthermore, we could not find a statistically significant association between the expression of both proteins and various clinicopathological parameters, except for the number of pregnancies that seems to be associated with SIX1positivity (p-value = 0.034). Regarding the relationship between levels of miR-200c and expression of their putative targets, SIX1 and SOX2, it was also not possible to find a statistically significant association. This may be due to the rather small sample size used in this study, which restricts us from drawing well established conclusions.

Therefore, we should continue to analyze the rest of samples that already exist in the biobank (76 samples) in order to increase the sample size and thus increase the robustness of the results.

Additionally, because SIX1 and SOX2 are actually transcription factors generally acting in transcriptional complexes, the mere expression of one these proteins can have no relevance in understanding the behavior of the tumor cell but rather a set of markers should be evaluated in order to correctly predict their behavior. As such several interacting proteins could have a greater impact on the understanding of the final results of dysregulated pathways instead of just one. This would be a wiser approach in the future. Also, as a complement, it would be interesting to assess in the same samples the levels of SIX1 and SOX2 mRNA to determine whether they may indeed be a link between miRNAs and these proteins expression and in cases where protein was not detected if expression of their respective gene indeed occurs. A clinical aspect which was not possible to evaluate with this study was the therapeutic outcome. In this sense, it would be enriching to follow-up these patients assessing after five years the relationship between SIX1 and SOX2 with the effectiveness of therapy, recurrence, and survival. Regarding the technical procedure, as ATM and BMI1 are important cancer markers candidates one should test other antibodies that would function as expected. Moreover, in order to minimize the poor tissue morphology observed, immunohistochemistry should be executed on automated staining instruments which are less aggressive to the tissue sections than the technique performed manually and allow more reproducible results.

5 References

- Axtell, M. J., Westholm, J. O. & Lai, E. C. 2011. Vive la différence: biogenesis and evolution of microRNAs in plants and animals. Genome Biol 12(4):221.
- Bartel, D. P. 2009. MicroRNAs: target recognition and regulatory functions. Cell 136(2):215–33.
- Bartels, C.L. & Tsongaliset, G.J. 2009. MicroRNAs: Novel Biomarkers for Human Cancer. Clin Chem 55(4):623–31.
- Berber, U., Yilmaz, I., Narli, G., Haholu, A., Kucukodaci, Z. & Demirel, D. 2014. miR-205 and miR-200c: Predictive Micro RNAs for Lymph Node Metastasis in Triple Negative Breast Cancer. J Breast Cancer 17(2): 143–8.
- Bradbury, A. R. & Olopade, O. I. 2007. Genetic susceptibility to breast cancer. Rev Endocr Metab Disord 8:255–267.
- Brisken, C. 2013. Progesterone signalling in breast cancer: a neglected hormone coming into the limelight. Nat Rev Cancer 13(6): 385–96.
- Caldon, C.E. 2014. Estrogen signaling and the DNA damage response in hormone dependent breast cancers. Front Oncol 4: 106.
- Calin, G. A., Cimmino, A., Fabbri, M., Ferracin, M., Wojcik, S. E., Shimizu, M., Taccioli, C., Zanesi, N., Garzon, R., Aqeilan, R. I., Alder, H., Volinia, S., Rassenti, L., Liu, X, Liu, C., Kipps, T. J., Negrini, M. & Croce, C. M. 2008. MiR-15a and miR-16-1 cluster functions in human leukemia. PNAS 105(13): 5166–5171.
- Campeau, P. M., Foulkes, W. D. & Tischkowitz, M. D. 2008. Hereditary breast cancer: new genetic developments, new therapeutic avenues. Hum Genet 124:31–42.
- Colombo, P., Milanezi, F., Weigelt, B. & Reis-Filho, J. S. 2011. Microarrays in the 2010s: the contribution of microarray-based gene expression profiling to breast cancer classification, prognostication and prediction. Colombo et al. Breast Cancer Research 13:212.
- Croce, C. M. 2009. Causes and consequences of microRNA dysregulation in cancer. Nat Rev Genet 10(10):704–14
- Cuatrecasas, M., Santamaria, G., Velasco, M., Camacho, E., Hernandez, L., Sanchez, M., Orrit, C., Murcia, C., Cardesa, A., Campo, E. & Fernandez, P.L. 2006. ATM gene expression is associated with differentiation and angiogenesis in infiltrating breast carcinomas. Histol Histopathol 21(2): 149–56.
- De Guire, V., Robitaille, R., Tétreault, N., Guérin, R., Ménard, C., Bambace, N. & Sapieha, P. 2013. Circulating miRNAs as sensitive and specific biomarkers for the diagnosis and monitoring of human diseases: Promises and challenges. Clin Biochem 46(10-11): 846–860.
- De Snoo, F., Bender, R., Glas, A. & Rutgers, E. 2009. Gene expression profiling: Decoding breast cancer. Surg Oncol 18(4):366-78.
- DeCastro, A.J., Dunphy, K.A., Hutchinson, J., Balboni, A.L., Cherukuri, P., Jerry, D.J. & DiRenzo, J. 2013. MiR203 mediates subversion of stem cell properties during mammary epithelial differentiation via repression of ΔNP63α and promotes mesenchymal-to-epithelial transition. Cell Death Dis 4: e514.
- DeSantis, C., Ma, J., Bryan, L. & Jemal, A. 2014. Breast cancer statistics, 2013. CA Cancer J Clin 64: 52-62.
- Fumagalli, D., Andre, F., Piccart-Gebhart, M. J., Sotitiou, C. & Desmedt, C. 2012. Molecular biology in breast cancer: should molecular classifiers beassessed by conventional

tools or by gene expression arrays?. Crit Rev Oncol Hematol 84(1), e58–e69.

- Guo, X., Yang, C., Qian, X., Lei, T., Li, Y., Shen, H., Fu, L. & Xu, B. 2013. Estrogen receptor α regulates ATM Expression through miRNAs in breast cancer. Clin Cancer Res 19(18): 4994–5002.
- Hesse, J.E., Liu, L., Innes, C.L., Cui, Y., Palii, S.S. & Paules, R.S. 2013. Genome-wide small RNA sequencing and gene expression analysis reveals a microRNA profile of cancer susceptibility in ATM-deficient human mammary epithelial cells. PloS One 8(5): e64779.
- Hinck, L. & Näthke, I. 2014. Changes in cell and tissue organization in cancer of the breast and colon. Curr Opin Cell Biol 26: 87-95.
- Huang, Y.H., Luo, M.H., Ni, Y.B., Tsang, J.Y., Chan, S.K., Lui, P.C., Yu, A.M., Tan, P.H. & Tse, G.M. 2014. Increased SOX2 expression in less differentiated breast carcinomas and their lymph node metástases. Histopathology 64: 494–503.
- Hudson, T.J. 2013. Genome variation and personalized cancer medicine. J Intern Med 274: 440–50.
- IARC (International Afency for Research on Cancer). GLOBOCAN 2012: Estimated Cancer Incidence, Mortality and Prevalence Worldwide in 2012. http://globocan.iarc.fr/Default.aspx
- Ignatiadis, M. & Sotiriou, C. 2013. Luminal breast cancer: from biology to treatment. Nat Rev Clin Oncol 10(9): 494–506.
- Iorio, M. V. & Croce, C. M. 2012. MicroRNA dysregulation in cancer: diagnostics, monitoring and therapeutics. A comprehensive review. EMBO Mol Med 4(3): 143–59.
- Iwanaga, R., Wang, C.A., Micalizzi, D.S., Harrell, J.C., Jedlicka, P., Sartorius, C.A., Kabos, P., Farabaugh, S.M., Bradford, A.P. & Ford, H.L. 2012. Expression of SIX1 in luminal breast cancers predicts poor prognosis and promotes increases in tumor initiating cells by activation of extracellular signal-regulated kinase and transforming growth factor-beta signaling pathways. Breast Cancer Res 14(4): R100.
- Jin, H., Cui, M., Kong, J., Cui, X., Lin, Z., Wu, Q. & Liu, S. 2014. Sineoculis homeobox homolog 1 protein is associated with breast cancer progression and survival outcome. Exp Mol Pathol 97(2): 247–252.
- Lengerke, C., Fehm, T., Kurth, R., Neubauer, H., Scheble, V., Müller, F., Schneider, F., Petersen, K., Wallwiener, D., Kanz, L., Fend, F., Perner, S., Bareiss, P.M. & Staebler, A. 2011. Expression of the embryonic stem cell marker SOX2 in early-stage breast carcinoma. BMC Cancer 11: 42.
- Li, H. & Yang, B. B. 2013. Friend or foe: the role of microRNA in chemotherapy resistance. Acta Pharmacol Sin 34(7):870-9
- Li, Y., Qiu, C., Tu, J., Geng, B., Yang, J., Jiang, T., & Cui, Q. 2013a. HMDD v2.0: a database for experimentally supported human microRNA and disease associations. Nucl. Acids Res 42 (D1): D1070-D1074.
- Li, Z., Tian, T., Hu, X., Zhang, X., Nan, F., Chang, Y.,Lv, F. & Zhang, M. 2013b. SIX1 mediates resistance to paclitaxel in breast cancer cells. Biochem Biophys Res Commun 441(3): 538–43.
- Liu, H. 2012. MicroRNAs in breast cancer initiation and progression. Cell Mol Life Sci 69(21): 3587–99.
- Lu, Y.X., Yuan, L., Xue, X.L., Zhou, M., Liu, Y., Zhang, C., Li, J.P., Zheng, L., Hong, M. & Li, X.N. 2014. Regulation of colorectal carcinoma stemness, growth, and metastasis by an miR-200c-Sox2-negative feedback loop mechanism. Clin Cancer Res 20(10): 2631-42.
- Moes, M., Le Béchec, A., Crespo, I., Laurini, C., Halavatyi, A., Vetter, G., Del Sol, A. &

Friederich, E. 2012. A novel network integrating a miRNA-203/SNAI1 feedback loop which regulates epithelial to mesenchymal transition. PloS One 7(4): e35440.

- Nicoloso, M.S., Spizzo, R., Shimizu, M., Rossi S. & Calin, G.A. 2009. MicroRNAs – the micro steering wheel of tumous metastases. Nat Rev Cancer 9(4): 293–302.
- Parvathi, M. V., Murthy, P. B., Vennila, M. & Suresh, B. V. 2013. Regulation of BMI1 Polycomb gene expression in histological grades of invasive ductal breast carcinomas and its correlation with hormone receptor status. Tumour Biol 36(6): 3807-15
- Rakha, E. A., El-Sayed, M. E., Reis-Filho, J. S. & Ellis, I. O. 2008. Expression profiling technology: its contribution to our understanding of breast cancer. Histopathology 52, 67–81.
- Sandhu, R., Parker, J. S., Jones, W. D., Livasy, C. A. & Coleman, W. B. 2010. Microarray-Based Gene Expression Profiling for Molecular Classification of Breast Cancer and Identification of New Targets for Therapy. LabMedicine 41, 364–372.
- Sandoval, J. & Esteller. M. 2012. Cancer epigenomics: beyond genomics. Curr Opin Genet Dev 22(1): 50–5.
- Seeley, R. R., Stephens, T. D. & Tate, P. 2007. Anatomia & Fisiologia, 6ª ed., Lusociência, Loures.
- Serpico, D., Molino, L. & Di Cosimo, S. 2014. microRNAs in breast cancer development and treatment. Cancer Treat Rev 40(5):595-604.
- Siegel, R., DeSantis, C., Virgo, K., Stein, K., Mariotto, A. Smith, T., Cooper ,D., Gansler, T., Lerro, C., Fedewa, S., Lin, C., Leach, C., Cannady, R.S., Cho, H., Scoppa, S., Hachey, M., Kirch, R., Jemal ,A. & Ward, E. 2013. Cancer Treatment and Survivorship Statistics, 2012. CA Cancer J Clin 62(4):220-41.
- Siegel, R., Ma, J., Zou, Z. & Jemal, A. 2014. Cancer Statistics, 2014. CA Cancer J Clin 64: 9-29.
- Singh, R. & Mo, Y.Y. 2013. Role of microRNAs in breast cancer. Cancer Biol Ther 14(3): 201–12.
- Stevens, A. & Lowe, J. S. 2005. Human Histology, 3rd ed., Mosby, Spain.
- Stevens, A., Lowe, J. S. & Young, B. 2002. Wheater's Basic Histopathology: A Colour Atlas and Text, 4th ed., Churchill Livingstone.
- Tam, W. L. & Ng, H. H. 2014. Sox2: Masterminding the Root of Cancer. Cancer Cell 26(1): 3–5.
- Tétreault, N. & De Guire, V. 2013. miRNAs: Their discovery, biogenesis and mechanism of action. Clin Biochem 46(10-11): 842–5.
- The Human Protein Atlas. Version 5 December 2013. http://www.proteinatlas.org/
- Tian, W., Chen, J., He, H. & Deng, Y. 2013. MicroRNAs and drug resistance of breast cancer: basic evidence and clinical applications. Clin transl oncol 15(5): 335–42.
- van der Vegt, B. de Bock, G.H., Hollema, H. & Wesseling J. 2009. Microarray methods to identify factors determining breast cancer progression: Potentials, limitations, and challenges. Crit Rev Oncol Hematol 70(1): 1–11.
- Vogel, V. G. 2000. Breast Cancer Prevention : A Review of Current Evidence. CA Cancer J Clin 50(3): 156-170.
- Wang, H., Liu, H., Li, X., Zhao, J., Zhang, H., Mao, J., Zou, Y., Zhang, H., Zhang, S., Hou, W., Hou, L., McNutt, M.A. & Zhang, B. 2014. Estrogen receptor α-coupled BMI1 regulation pathway in breast cancer and its clinical implications. BMC Cancer 14(1): 122.

- Wang, L. & Wang, J. 2012. MicroRNA-mediated breast cancer metastasis: from primary site to distant organs. Oncogene 31(20): 2499–511.
- WHO (World Health Organization). 2012. Classification of Tumours of the Breast, IARC Press, Lyon
- Wu, W., Ren, Z., Li, P., Yu, D., Chen, J., Huang, R. & Liu, H. 2014. SIX1: A critical transcription factor in tumorigenesis. Int J Cancer.
- Yager, J. D. & Davidson, N. E. 2006. Estrogen carcinogenesis in breast cancer. N Engl J Med 354(3): 270–82.
- Yin, J., Zheng, G., Jia, X., Zhang, Z., Zhang, W., Song, Y. & He, Z. 2013. A Bmi1-miRNAs Cross-Talk Modulates Chemotherapy Response to 5-Fluorouracil in Breast Cancer Cells. PloS One 8(9): e73268.
- You, J.S. & Jones. P. A. 2012. Cancer Genetics and Epigenetics: Two Sides of the Same Coin?. Cancer Cell 22(1): 9–20.
- Zhang, Y., Eades, G., Yao, Y., Li, Q. & Zhou, Q. 2012. Estrogen Receptor α Signaling Regulates Breast Tumor-initiating Cells by Down-regulating miR-140 Which Targets the Transcription Factor SOX2. J Biol Chem 287(49): 41514–22.
- Zhang, Z., Zhang, B., Li, W., Fu, L., Fu, L., Zhu, Z. & Dong, J.T. 2011. Epigenetic Silencing of miR-203 Upregulates SNAI2 and Contributes to the Invasiveness of Malignant Breast Cancer Cells. Genes Cancer 2(8): 782–91.
- Zhou, J., Teng, R., Wang, Q., Xu, C., Guo, J., Yuan, C., Shen, J., Hu, W., Wang, L. & Xie, S. 2013a. Endocrine resistance in breast cancer: Current status and a perspective on the roles of miRNAs (Review). Oncol lett 6(2): 295–305.
- Zhou, Y., Wan, G., Spizzo, R., Ivan, C., Mathur, R., Hu, X., Ye, X., Lu, J., Fan, F., Xia, L., Calin, G.A., Ellis, L.M. & Lu, X. 2013b. miR-203 induces oxaliplatin resistance in colorectal cancer cells by negatively regulating ATM kinase. Mol Oncol 8(1):8.-92.