UNIVERSITÀ DEGLI STUDI DI MILANO

Facoltà di Medicina e Chirurgia

Scuola di Dottorato in Scienze Biomediche Cliniche e Sperimentali

Dottorato di Ricerca in Biotecnologie Applicate alle Scienze Mediche

Ciclo XXV

Settore Disciplinare BIO/13

BIOLOGICALLY MEANINGFUL AND CLINICALLY RELEVANT GENE EXPRESSION PROFILE FOR OPTIMAL TREATMENT PLANNING IN BREAST CANCER

Coordinatore: Prof. Alessandro Gianni

Relatore: Prof. Carmelo Carlo-Stella

Tutor: Dott.ssa Vera Cappelletti

Dottorando: Valeria Musella

Matricola: R08628

A.A. 2011-2012

This thesis work was carried out at the Fondazione IRCCS

Istituto Nazionale dei Tumori of Milan,

in Biomarkers Unit of

the Department of Experimental Oncology and Molecular Medicine

directed by Dr. Maria Grazia Daidone.

1. TABLE OF CONTESTS

ABSTRACT

Breast cancer is a highly heterogeneous disease from the molecular and clinical point of view. Optimal treatment with the currently available drugs depends therefore on the ability to individually predict treatment. Presently available outcome prediction models are however suboptimal, single predictive variables have limited accuracy, and the actual clinical outcomes remain heterogeneous in any given prognostic group. ER status and HER-2 status are helpful in identifying patients who are not eligible for endocrine or trastuzumab therapies by virtue of their high negative predictive values (NPVs) and high sensitivities. However, only a minority of ER-positive or HER-2-positive patients respond to receptor-targeted therapy. The positive predictive values (PPVs) of these tests are <50%. Moreover, currently there are no accepted molecular predictors of response to various chemotherapeutic drugs. These limitations have driven biomarker research to develop more accurate molecular predictors of clinical outcome.

In the present thesis work we report data on an innovative strategy to predict treatment response to conventional therapies by combining in a hierarchical way genomic predictors related to prognosis and to treatment sensitivity and resistance. The strategy takes into account the intrinsic molecular heterogeneity of breast tumors by distinctly developing genomic predictors for each of the three main tumor subtypes defines as: ER+/Her2-, Her2+ and ER-/Her2-. The considered genomic predictors are built based on literature data and results previously obtained at the INT (as in the case of the prognostic role of ISG genes) and are treated as metagenes mainly to facilitate cross-platform comparisons and to stick to pathways with s clear biological role. An important part of the development of the prediction strategy deals with analytical approaches which were optimized to gain more accurate information from FFPE samples.

Gene expression profiles used in such approach were obtained from public database, but also from expression studies carried out at INT on two types of samples; fresh frozen and formalin fixed samples. A detailed comparative analysis of technical solutions (Illumina HT 12, Illumina Ref8, Illumina DASL and Affymetrix HG Plus2.0 chips) for optimizing gene expression profiles in FFPE samples with heavily degraded and chemically-modified RNA is reported. A new robust protocol was developed based on linear amplification of RNA under conditions minimizing rRNA amplification, and on use of the Affymetrix HG Plus 2.0 chips. The protocol was tested in a pilot study on 60 samples to estimate the actual percentage of archived FFPE clinical samples which could yield technically acceptable gene expression profiles and to evaluate the biological reliability

of gene expression profiles obtained from fixed samples. Reliability was tested by comparing ER-status classifiers developed using FF-derived expression data and testing the classifier on predicting ER status in FFPE and FF dataset. Prediction accuracy between the two types of samples was comparable (FF Cohen's k 0,92, FFPE Cohen's k 0.89).

With the various tools developed as described above, it was possible to identify *a priori* in the ER+/Her2- a subset of patients who were predicted (and confirmed by external validation) to have 95% 5-year disease free survival. Such accurate (and validated) prediction was achieved by combining optimized metagenes with clear biological roles in proliferation, ER signaling and immunity and separately analyzing prognostic and predictive information.
Work is still in progress for the other molecular subtypes (Her2+ and ER-/Her2-).

The role of immune genes was particularly interesting as it definitely added important information to refine prognosis in our strategy, but at the same way raised some questions due to its paradoxical role encompassing both tumor promotion and tumor inhibition. In our gene expression profiling data in node negative breast cancer patients, ISG expression was associated to likelihood to develop distant metastases in patients with ER+/Her2-tumors, but upon validation in larger data set it clearly emerged that the ISG prognostic role was strongly dependent on the molecular subtypes of the tumor (protective in patients with Her2+ tumors, risk-associated in patients with ER+/Her2- tumors, and neutral in those with ER-/Her2- tumors). Furthermore in the clinical setting a high ISG expression did not reflect stronger lymphocyte infiltration, and upon correction for genuine T cell associated genes ISG genes were found to be associated to proliferation. Interestingly in our clinical samples the ISG genes were demonstrated to be expressed by the cancer cells rather than the stroma.
To further search for biological mechanism justifying the prognostic role of ISG expression we switched to *in vitro* co-cultures of epithelial cells with normal fibroblasts of different origin or CAFs. Many different experimental setting either 2D or 3D were adopted and described in detail. The experimental setting played a major role on the experimental results. On the average basal-like cells were more sensitive to fibroblasts promoted effects in terms of cytokine secretion in the culture medium and up-regulation of ISG genes in the epithelial cells. Those cells were characterized by high intrinsic migratory and invasive ability which did not change much upon co-cultures.

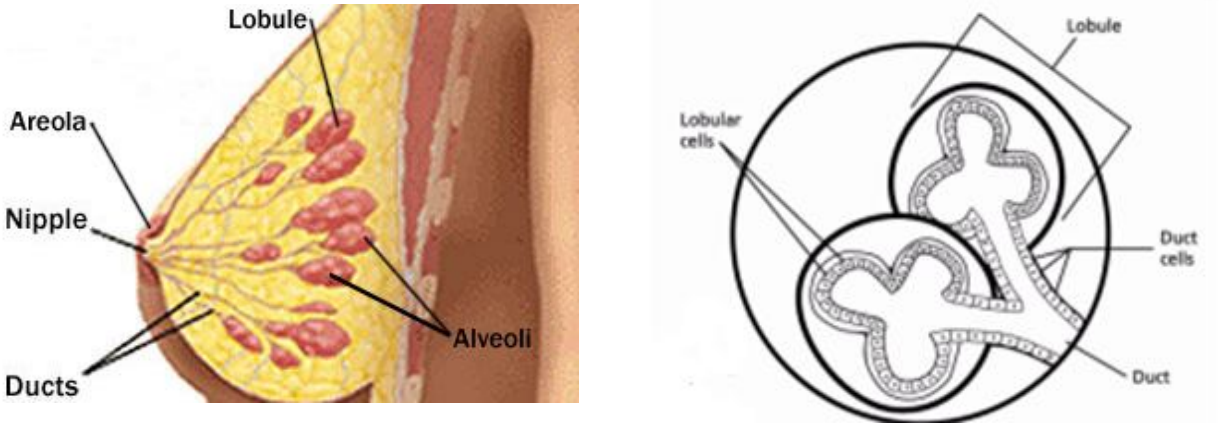
On the contrary luminal cells were stimulated to growth and gained invasive and migratory abilities after stimulation with fibroblasts and CAFs.
Our experiments showed a complicated and sometimes controversial interplay between fibroblasts and epithelial cells which fully justifies the paradoxical role of the immune system in clinical tumors.

In conclusion FFPE samples derived from clinical trials are non more an inaccessible source of biological and clinical information, and can be exploited to build true genomic predictor allowing a true personalized cancer treatment. While our prediction model proofed to be effective the role of immune genes and of stroma, although recognized as very important, is still elusive despite the development of different types of heterotypic cells co-culture models.

1. INTRODUCTION

1.1 THE NORMAL BREAST

To understand breast cancer, it helps to have some basic knowledge about its the normal structure.
The structure of the female breast is complex, including fat and connective tissue, as well as lobes, lobules, ducts and lymph nodes.
Each breast is made up mainly of lobules, that are milk-producing glands, ducts, tiny tubes that carry the milk from the lobules to the nipple, and stroma composed of fatty tissue and connective tissue surrounding the ducts and lobules, blood vessels, and lymphatic vessels (Fig 1a and 1b).



Figures 1: **a.** mammary gland structure; **b**. enlargment of terminal mammary duct.

Most breast cancers originate in the cells that line the ducts (ductal cancers), while same derive from the cells that line the lobules (lobular cancers).
It is important to considerate the lymph system is important to understand because it is one way breast cancers can spread. This system has several parts.
Lymph nodes are small, bean-shaped collections of immune system cells that are connected by lymphatic vessels. Lymph contains tissue fluid and waste products, as well as immune system cells. Breast cancer cells can enter lymphatic vessels and begin to grow in lymph nodes.
Most lymphatic vessels in the breast connect to lymph nodes under the arm (axillary nodes). Some lymphatic vessels connect to lymph nodes inside the chest (internal mammary nodes) and those either above or below the collarbone (supraclavicular or infraclavicular nodes) (Fig.2).

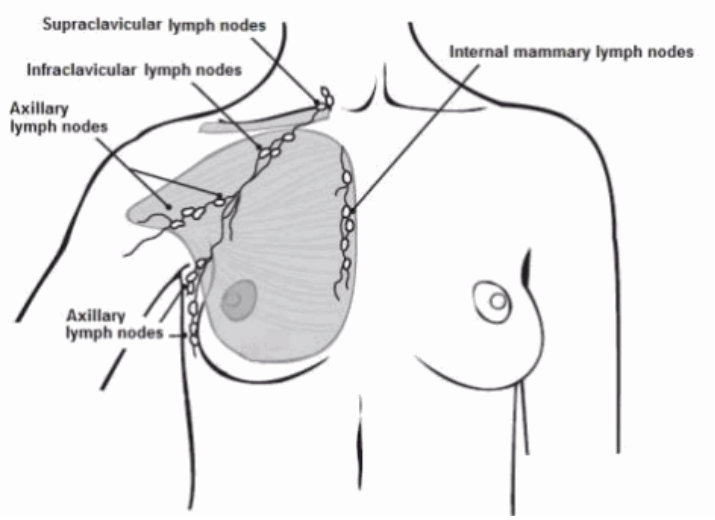


Figure 2: tree mammary lymph nodes

If the cancer cells have spread to lymph nodes, there is a higher chance that the cells could have also gotten into the bloodstream and spread (metastasized) to other sites in the body. The more lymph nodes contain breast cancer cells, the more likely it is that the cancer may be found in other organs as well. Because of this, the identification cancer cells in one or more lymph nodes impact on the treatment plan. Still, not all women with cancer cells in their lymph nodes develop metastases, and some women can have no cancer cells in their lymph nodes and later develop metastases.

1.1.1 FACTORS INVOLVED IN MAMMARY GLAND DEVELOPMENT

The mammary gland is a very dynamic tissue which undergoes dramatic changes during different developmental phases (1). In fact during puberty the mammary gland develops from a rudimentary tree to a branched epithelial network of ducts which can support alveolar development and subsequent milk production during pregnancy and lactation. All phases are under hormonal control: estrogen, progesterone and prolactin are the principal hormones that act on stromal cells to stimulate the branching process (2); estrogen and EGF, during puberty, control ductal elongation and branching; insulin growth factor (IGF), amphiregulin, neuregolin, (NRG) fibroblast growth factors (FGF) and receptors ERBB2 /3/4 are implicated in embryonic mammary gland development and branching morphogenesis; progesterone, prolactin and placental lactogens stimulate alveolar proliferation and differentiation during pregnancy. In addition to hormones, also the protease, as for

example the metalloproteases play a fundamental role in the branching process, these are both upstream and downstream of EGF receptor signaling by which they regulate growth factor function and branching.

1.1.2 MAMMARY GLAND MORPHOLOGY AND PHYSIOLOGY

1.1.2.1 Luminal and myoepithelial cells

Quiescent mammary ducts are bilayered, one layer of luminal epithelial cells, surrounded by a layer of myoepithelial cells (also referred to as basal cells). Luminal and myoepithelial cells originate from the same progenitor cell. This common progenitor differentiates into a luminal progenitor and a myoepithelial progenitor cell that finally differentiates to a myoepithelial cell (3). Luminal progenitors can differentiate to mature ductal, alveolar or secretory cells. Luminal epithelial cells differentiate to milk-producing cells upon stimulation with prolactin. Breast cancers mainly arise in the luminal epithelial compartment (4). This is most likely related to the dynamics of differentiation and dedifferentiation of these cells. Myoepithelial cells function as a guardian of tissue integrity of the mammary gland by maintaining tissue polarity (5) (6). To distinguish luminal and myoepithelial cells several markers can be used. Luminal epithelial cells are MUC-1 positive and express cytokeratin 8 (CK8). Myoepithelial cells on the other hand are smooth muscle actin-1 (SMA-1) positive, express cytokeratin 5 (CK5) and p63.

1.1.2.2 Basement membrane

Epithelial cells are dependent on adhesion to other cells and to the extracellular matrix in order to survive. Also in the mammary gland the basement membrane has an important role as a survival factor for mammary epithelial cells. The mammary basement membrane contains collagen IV, and laminin-1 and -11, which are cross-linked by nidogen-1 and 2 to form a gel-like structure to which mammary epithelial cells adhere (7). The adhesion of mammary epithelial cells to extracellular matrix via integrins, family of heterodimeric transmembrane glycoprotein receptors, has been shown to suppress apoptosis (8;9). Integrins regulate cell shape and facilitate migration by providing a structural link with the actin cytoskeleton. Apart from integrins, several other receptors

such as dystroglycans, syndecans and galactosyl transferases function as adhesion receptors for basement membrane proteins (10). Binding of integrins to the ECM promotes the formation of adhesion complexes at the plasma membrane. These adhesion complexes are important for cytoskeleton assembly. The expression and deposition of extracellular matrix proteins is frequently altered in breast cancer. Reduced expression of integrins and lack of basement membrane is often observed in metastatic disease.

1.1.2.3 Stroma e microenvironment

Stromal cells secrete a separate set of extracellular components such as fibronectin (fn). The stroma of mammary gland consists of extracellular matrix (ECM) and a variety of mesenchymal cells, including fibroblast, myofibroblast, endothelial cells, pericytes and leukocytes (11). Stromal-epithelial interactions play important roles during mammary gland development. Macrophages and eosinophils are recruited to the mammary gland simultaneous with the outgrowth of terminal end buds (TEBs). Macrophages are recruited mostly to the neck of the TEB while eosinophils are mainly located around the head of the TEB.

Leukocyte depletion from the mammary gland was shown to result in dramatic inhibition of ductal development. Survival and proliferation of macrophages as well as macrophage behavior, morphology and motility is regulated by colony-stimulating factor-1 (CSF-1). Regulation of the CSF-1/CSF-1R system seems to be crucial for ductal outgrowth. Since the CSF-1 receptor (CSF-1R) is exclusively expressed on macrophages in the mammary gland tissue, macrophages are important stromal factors in mammary gland development.

The extracellular matrix (ECM) is a dominant regulator of tissue development and homeostasis, it regulates growth, differentiation and apoptosis in human mammary epithelial cells (MEC) through a hierarchy of transcriptional events involving the intricate interplay between soluble and physical signaling pathways. The study of Weaver et al. shown that this pathway are influenced by the tissue structure that is directed by the cooperative interactions of the cell-cell and cell-ECM and can by modified by stromal factors (12).

1.2 BREAST CANCER

Breast cancer is the formation of a malignant tumor that has developed from cells in the breast. A malignant tumor is a group of cancer cells that may invade surrounding tissues or spread to distant areas of the body, thereby giving rise to metastases. Breast cancer affects predominantly women but may develop, in some cases, also in the men.

1.2.1 EPIDEMIOLOGY

Breast cancer is the most common cancer diagnosed in women and is the second most frequent cause of cancer mortality in western women. The past 20 year have seen a significant reduction in breast cancer-related mortality, largely due to improved early detection and the development of more effective adjuvant therapies. According to the current statistics of the AIRC (Associazione Italiana per la Ricerca sul Cancro), breast cancer accounts for 25% of all female cancers is responsible for 17% of cancer death in women in Italy and it is estimates that about 1 in 10 women will develop breast cancer during her lifetime. In Italy 37000 new cases are diagnosed each year, 152 per 100 000 women.

1.2.2 RISK FACTORS

Risk factors associated with the development of the tumor can be classified as lifestyle dependent or independent.

Every woman is at risk for developing breast cancer. Several relatively strong risk factors for breast cancer that affect large proportions of the general population have been known for some time. However, the vast majority of breast cancer cases occur in women who have no identifiable risk factors other than their gender (13). The “established” risk factors for breast cancer are female gender, age, previous breast cancer, benign breast disease, hereditary factors (family history of breast cancer), early age at menarche, late age at menopause, late age at first full-term pregnancy, postmenopausal obesity, low physical activity, race/ethnicity and high-dose exposure to ionizing radiation early in life.

Other putative risk factors for breast cancer include never having been pregnant, having only one pregnancy rather than many, not breast feeding after pregnancy, use of postmenopausal estrogen replacement therapy or postmenopausal hormone (estrogen/progestin) replacement therapy, use of oral contraceptives, certain specific dietary styles (high intake of fat and low intakes of fiber, fruits, and vegetables, low intake of phytoestrogens), alcohol consumption, tobacco smoking, and abortion. Although men develop breast cancer, the disease is 100 times more likely to occur in a woman than in a man (14).Women are at a higher risk of breast cancer because they have much more breast tissue than men do, and are more exposed during their life to estrogen stimulation.

Breast cancer affects women of all races and nationalities, and the incidence globally shows an increase in industrialized countries or in those with a recent industrialization (15).

However, most of the women who get breast cancer (approximately 80%) do not have a strong family history of disease (16). The effect of family history on breast cancer risk is believed to be due primarily to genetic factors. As much as 5-10% of all breast cancer cases are attributable to specific inherited single-gene mutation, and many other cases have some genetic component.

Epidemiological studies have shown that in women with a family history of breast cancer, the risk is increased two- to three fold.

Hereditary mutations of the BRCA1 and BRCA2 genes are responsible for 5% of all breast cancers. These genes are located on chromosome 17q21 and 13q13 respectively, and are involved in processes for cell survival, as the repair of breaks in DNA double helix, and through the formation of multiprotein complexes, remodeling and transcription of the DNA, thanks to the link with the RNA polymerase II RNA, helicase A, the histone-deacetylases (HDACs) and other proteins. BRCA1 that BRCA2 are also involved in maintaining the structural integrity of chromosomes and in the apoptotic process (17).

The diagnosis of breast cancer, especially in young women, is associated with the risk development of tumor in controlateral gland: in women who develop invasive cancer in the left breast within 30 years of life, the probability of a similar development in the right breast within in 40 years of age is 5% (0.5% per year). In addition, the likelihood of developing a primary tumor in the breast, increases proportionally with age, but the risk of controlateral breast cancer is reduced during the course of life. Several hypotheses have been put to explain these observations, although no one seems to give a full explanation (18). Additional risk factors are: early menarche (12 old or younger), late menopausal (55 or older>) and age-related factors in the first pregnancy. These are believed to me mediated through estrogen production (5).

An increased risk of breast cancer can be due to long-term exposure to estrogen and progesterone in women with early menarche and late menopause, or the use of long-term hormones therapy or contraceptives. Age at first pregnancy is another aspect of reproductive history that is associated with breast cancer risk. Women who have their first full-term pregnancy at a relatively early age have a lower risk of breast cancer than those who never have children or those who have their first child relatively late in life (19). To wait a long time between multiple pregnancies, miscarriages or abortion, twin births at first pregnancy with hypertension during this period (20) are other risk factors.

In addition there are several risk factors depending to lifestyle but is important to emphasize that the risk can be drastically reduced by the behavior of each person. A rich animal fats diet, but low fruits and vegetables is associated with an increase in the frequency of breast cancer, so different eating habits in different countries are therefore closely linked to the frequency of occurrence of the disease.

The metabolic syndrome is associated with an increased risk of breast cancer. Obese women in fact, with a biomass index > 40 kg/m2, show a greater susceptibility to cancer; the obesity cause 20% of deaths (21). These subjects has been shown not only a greater risk development of breast cancer, but also an increased risk of distant metastases (22;23).

Such observations are supported by the evidence of difference amount of circulating hormones, like estrogen and insulin, in obese person with respect to normal-weight person. Recently it has been shown how the adipocytes in the mammary gland through the secretion of cytokines can influence the proliferation and tumor invasion. (24)(25).

An association between risk of breast cancer and excessive body mass was found in postmenopausal women, indeed. Before menopause the ovaries produce most of the estrogens and the adipose tissue produces only a small amount, after menopause, most of estrogen are produced from periferic aromatization in fat.

However, the connection between obesity and breast cancer seems to be more complicated than described, the risk is increased for women who became obese in adulthood, while it may not increase in obese women after birth. In addition, an excess of fat in the abdominal circumference has a greater effect on risk than the same amount of fat in the area of the hips and thighs (26). Alcohol consumption was defined as a risk factor for the development of breast cancer: the risk increases is dose-dependent with the drink. A drink consumed daily (12once-5once of beer or wine) increases this parameter by 10%, two drinks consumed daily increases the risk to 20% (27). The

consumption of alcohol not only increases the risk of developing the disease but also the risk of relapses and deaths for breast cancer. The hypothesis proposed to explain how alcohol may induce this process are related to the increase in reactive oxygen species and acetaldehyde, and to hormonal (estrogen and insulin-like growth factor 1 (IGF-1)) pathways variation. These hypotheses however have not been confirmed (28-30).

Breast cancer is increasingly being diagnosed in Caucasian post-menopausal women or women born in the countries of northern Europe or America. A common denominator for the risk of developing breast cancer is related to the reproductive history. Numerous studies have shown that early pregnancies decrease the risk of developing breast cancer after 50 years of age, while a higher frequency of breast cancer is instead observed in nulliparous women between 30 and 34 years of age. The nulliparity can then be seen as a risk factor for the development of breast cancer while pregnancy would play a protective action in mammary epithelial level (15).

1.2.3 HISTOTYPES

Breast cancers are derived from the epithelial cells that line the terminal duct lobular unit. Cancer cells that remain within the basement membrane of the elements of the terminal duct lobular unit and the draining duct are classified as in situ or non invasive. In invasive breast cancer there is instead a dissemination of cancer cells outside the basement membrane of the ducts and lobules into the surrounding adjacent normal tissue. Both in situ and invasive cancers have characteristic patterns by which they can be classified.

Invasive breast cancers are divide into ductal carcinoma (DC) and lobular (LC).

Most carcinomas occurs in the duct lobular terminal units (UTDL) and subsequently, for mechanisms still not well known, gives rise to different tumors, not only for their morphology, but also for the biological behavior. Both types of tumors (DC and LC) can be in situ (remain confined in tissue in which it develops) or infiltrating. The ductal carcinoma accounts for about 75% of infiltrating tumors, while the lobular about 5%. Less frequent histotypes infiltrating are: medullary carcinoma (15%), colloid or mucinous (2%), tubular (1-2%), and other rare forms. The transition from normal structures to carcinoma in situ takes place through the formation of intermediate lesions, different for the two main types of tumor, called pre-neoplastic lesions.

1.2.4 STAGING AND CLASSIFICATION OF BREAST CANCER

Staging describes the severity of a person's cancer based on the extent of the original (primary) tumor and whether or not cancer has spread in the body. Staging is important for several reasons:

- Staging helps the physician to plan the appropriate treatment.
- The stage can be used to estimate the person's prognosis.
- Knowing the stage is important in identifying clinical trials that may be suitable for a particular patient.
- Staging helps health care providers and researchers exchange information about patients; it also gives them a common terminology for evaluating the results of clinical trials and comparing the results of different trials.

Staging is based on knowledge of the way cancer progresses.

The TNM system is one of the most widely used staging systems. This has been accepted by the International Union Against Cancer (UICC) and the American Joint Committee on Cancer (AJCC).

The TNM system is based on the extent of the tumor **(T)**, the extent of spread to the lymph nodes **(N)**, and the presence of distant metastasis **(M)**. A number is added to each letter to indicate the size or extent of the primary tumor and the extent of cancer spread (Tab.1).

- 'T' can be 1, 2, 3 or 4, with 1 being small and 4 large.
- 'N' can be between 0 (no positive nodes) and 3 (many of positive nodes).
- 'M' can either be 0 (the cancer hasn't spread) or 1 (the cancer has spread).

Primary Tumor (T)	
TX	Primary tumor cannot be evaluated
T0	No evidence of primary tumor
Tis	Carcinoma in situ (CIS; abnormal cells are present but have not spread to neighboring tissue; although not cancer, CIS may become cancer and is sometimes called preinvasive cancer)
T1, T2, T3, T4	Size and/or extent of the primary tumor
Regional Lymph Nodes (N)	
NX	Regional lymph nodes cannot be evaluated
N0	No regional lymph node involvement
N1, N2, N3	Involvement of regional lymph nodes (number of lymph nodes and/or extent of spread)
Distant Metastasis (M)	
MX	Distant metastasis cannot be evaluated
M0	No distant metastasis
M1	Distant metastasis is present

Table 1: TNM classification system.

The breast cancer stage is based on the results of testing that is done on the tumor and lymph nodes removed during surgery and other tests. For many cancers, TNM combinations correspond to one of five stages.

Stage 0 (carcinoma in situ)

There are 3 types of breast carcinoma in situ:

- Ductal carcinoma in situ (DCIS) is a noninvasive condition in which abnormal cells are found in the lining of a breast duct. The abnormal cells have not spread outside the duct to other tissues in the breast. In some cases, DCIS may become invasive cancer and spread to other tissues. At this time, there is no way to know which lesions could become invasive (Fig.3).

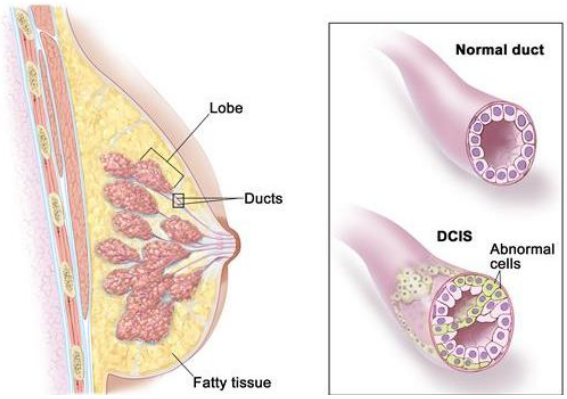


Figure 3: Ductal Carcinoma In Situ (DCIS)

- Lobular carcinoma in situ (LCIS) is a condition in which abnormal cells are found in the lobules of the breast. This condition seldom becomes invasive cancer. However, having LCIS in one breast increases the risk of developing breast cancer in either breast.
- Paget disease of the nipple is a condition in which abnormal cells are found in the nipple only.

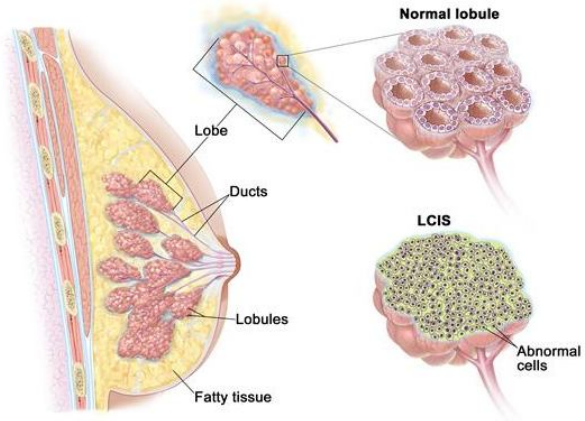
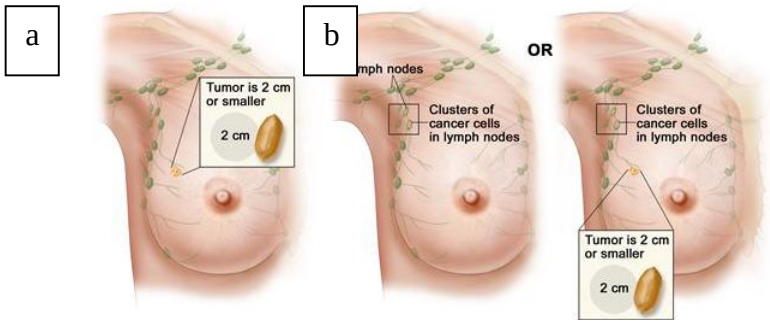


Figure 4: Lobular Carcinoma In Situ (LCIS)

Stage I

- In stage I, cancer has formed. Stage I is divided into stages IA and IB.
- In stage IA (Fig. 1a), the tumor is 2 centimeters or smaller. Cancer has not spread outside the breast.

- In stage IB (Fig. 1b), small clusters of breast cancer cells (larger than 0.2 millimeter but not larger than 2 millimeters) are found in the lymph nodes and either:
- no tumor is found in the breast; or
- the tumor is 2 centimeters or smaller.



Figures 5: a. Stage IA breast cancer; b. Stage IB breast cancer.

Stage II

Stage II is divided into stages IIA and IIB.

In stage IIA:

- no tumor is found in the breast or the tumor is 2 centimeters or smaller. Cancer (larger than 2 millimeters) is found in 1 to 3 axillary lymph nodes or in the lymph nodes near the breastbone (found during a sentinel lymph node biopsy); or
- the tumor is larger than 2 centimeters but not larger than 5 centimeters. Cancer has not spread to the lymph nodes (Fig.6).

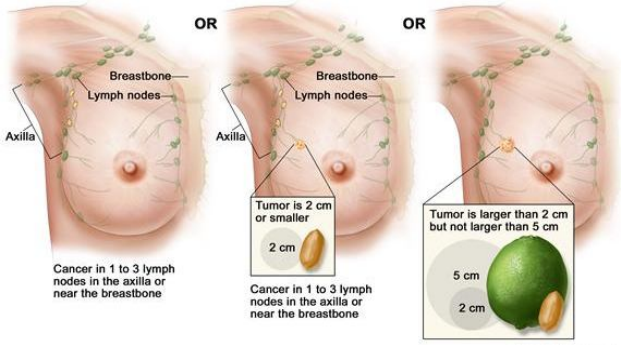


Figure 6: stage IIA breast cancer

In stage IIB (Fig.7), the tumor is:

- larger than 2 centimeters but not larger than 5 centimeters. Small clusters of breast cancer cells (larger than 0.2 millimeter but not larger than 2 millimeters) are found in the lymph nodes; or
- larger than 2 centimeters but not larger than 5 centimeters. Cancer has spread to 1 to 3 axillary lymph nodes or to the lymph nodes near the breastbone (found during a sentinel lymph node biopsy); or
- larger than 5 centimeters. Cancer has not spread to the lymph nodes.

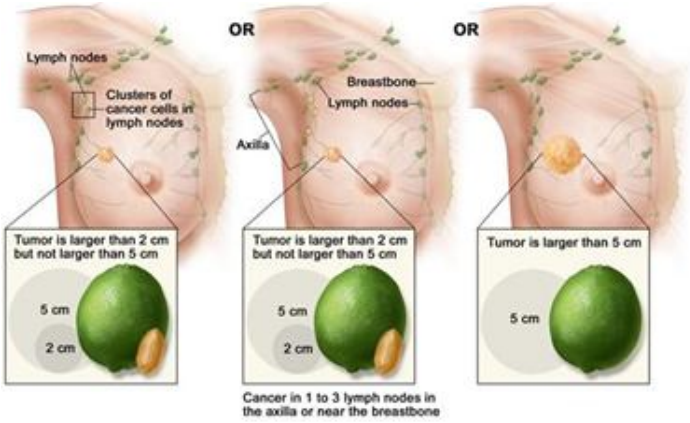


Figure 7: stage IIB breast cancer

Stage IIIA

In stage IIIA:

- no tumor is found in the breast or the tumor may be any size. Cancer is found in 4 to 9 axillary lymph nodes or in the lymph nodes near the breastbone (found during imaging tests or a physical exam); or
- the tumor is larger than 5 centimeters. Small clusters of breast cancer cells (larger than 0.2 millimeter but not larger than 2 millimeters) are found in the lymph nodes; or
- the tumor is larger than 5 centimeters. Cancer has spread to 1 to 3 axillary lymph nodes or to the lymph nodes near the breastbone (found during a sentinel lymph node biopsy) (Fig.8).

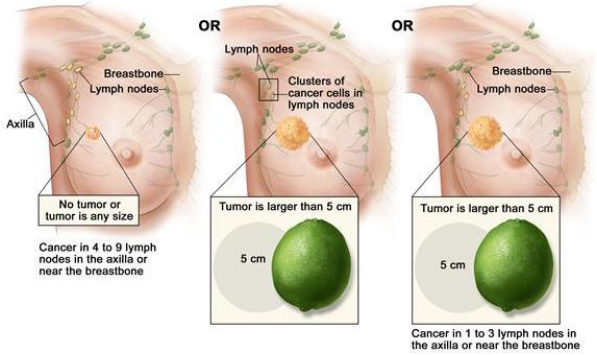


Figure 8: stage IIIA breast cancer

Stage IIIB

In stage IIIB (Fig.9), the tumor may be any size and cancer has spread to the chest wall and/or to the skin of the breast and caused swelling or an ulcer. Also, cancer may have spread to:

- up to 9 axillary lymph nodes; or
- the lymph nodes near the breastbone.

Cancer that has spread to the skin of the breast may also be inflammatory breast cancer.

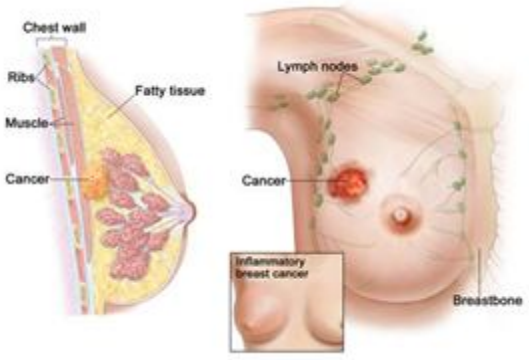


Figure 9: stage IIIB breast cancer

Stage IIIC

In stage IIIC, no tumor is found in the breast or the tumor may be any size. Cancer may have spread to the skin of the breast and caused swelling or an ulcer and/or has spread to the chest wall. Also, cancer has spread to:

- 10 or more axillary lymph nodes; or
- lymph nodes above or below the collarbone; or
- axillary lymph nodes and lymph nodes near the breastbone.

Cancer that has spread to the skin of the breast may also be inflammatory breast cancer. For treatment, stage IIIC breast cancer is divided into operable and inoperable stage IIIC (Fig.10).

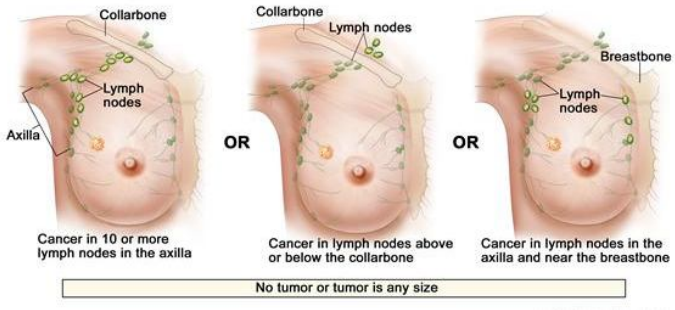


Figure 10: stage IIIC breast cancer

Stage IV

Stage IV breast cancer. The cancer has spread to other parts of the body, most often the bones, lungs, liver, or brain.

In stage IV, cancer has spread to other organs of the body, most often the bones, lungs, liver, or brain (Fig.11).

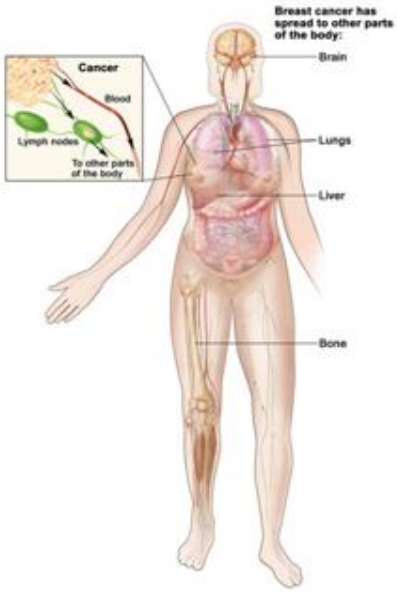


Figure 11: stage IV breast cancer

1.3 BREAST CANCER TREATMENT

Different types of treatment are available for patients with breast cancer. Some are standard (the currently used treatment), and some are being tested in clinical trials.

Most patients with breast cancer have surgery to remove the cancer from the breast. Some of the lymph nodes under the arm are usually taken out and looked at under a microscope to see if they contain cancer cells.

Breast-conserving surgery is an operation to remove the cancer but not the breast itself, includes the following:

- Lumpectomy: that is surgery to remove a tumor (lump) and a small amount of normal tissue around it.
- Partial mastectomy: to remove the part of the breast that has cancer and some normal tissue around it. The lining over the chest muscles below the cancer may also be removed. This procedure is also called a segmental mastectomy.

Patients who are treated with breast-conserving surgery may also have some of the lymph nodes under the arm removed for biopsy. This procedure is called lymph node dissection. Other types of surgery include the following:

- Total mastectomy: surgery to remove the whole breast that has cancer. This procedure is also called a simple mastectomy. Some of the lymph nodes under the arm may be removed for biopsy at the same time as the breast surgery or after. This is done through a separate incision.
- Modified radical mastectomy: surgery to remove the whole breast that has cancer, many of the lymph nodes under the arm, the lining over the chest muscles, and sometimes, part of the chest wall muscles.

Other treatment used for breast cancer is chemotherapy that may be given also before surgery to remove the tumor. In this case, chemotherapy will shrink the tumor and reduce the amount of tissue

that needs to be removed during surgery. Treatment given before surgery is called neoadjuvant therapy.

Even if the physician removes all the cancer that can be seen at the time of the surgery, some patients may be given radiation therapy, chemotherapy, or hormone therapy after surgery to kill any cancer cells that are left. Treatment given after the surgery, to lower the risk that the cancer will come back, is called adjuvant therapy.

Sentinel lymph node biopsy is the removal of the sentinel lymph node during surgery. The sentinel lymph node is the first lymph node to receive lymphatic drainage from a tumor. It is the first lymph node where cancer cells are likely to spread to from the tumor. A radioactive substance and/or blue dye is injected near the tumor. The substance or dye flows through the lymph ducts to the lymph nodes. The first lymph node to receive the substance or dye is removed. A pathologist views the tissue under a microscope to look for cancer cells. If cancer cells are not found, it may not be necessary to remove more lymph nodes. After the sentinel lymph node biopsy, the surgeon removes the tumor (breast-conserving surgery or mastectomy).

Radiation therapy is a cancer treatment that uses high-energy x-rays or other types of radiation to kill cancer cells or keep them from growing. There are two types of radiation therapy: external radiation therapy uses a machine outside the body to send radiation toward the cancer, and internal radiation therapy uses a radioactive substance sealed in needles, seeds, wires, or catheters that are placed directly into or near the cancer. The way the radiation therapy is given depends on the type and stage of the cancer being treated.

Chemotherapy is a cancer treatment that uses drugs to stop the growth of cancer cells, either by killing the cells or by stopping them from dividing. When chemotherapy is taken by mouth or injected into a vein or muscle, the drugs enter the bloodstream and can reach cancer cells throughout the body (systemic chemotherapy). When chemotherapy is placed directly into the cerebrospinal fluid, an organ, or a body cavity such as the abdomen, the drugs mainly affect cancer cells in those areas (regional chemotherapy). The way the chemotherapy is given depends on the type and stage of the cancer being treated.

Hormone therapy is a cancer treatment that removes hormones or blocks their action and stops cancer cells from growing. It is the first described target therapy for breast cancer and is effective

only for estrogen receptor positive (ER positive) cancers. A better treatment response rate is observed in patients with ER+ tumors which also express the progesterone receptor (PgR).

Hormone therapy can be used after surgery or other treatments to decrease the chance of recurrence. If the cancer has already spread, hormone therapy may shrink and control it.

Treatments that can be used in hormone therapy include:

- **Medications that block hormones from stimulating cancer cells.** Tamoxifen is the most commonly used selective estrogen receptor modulator (SERM). SERMs act by blocking estrogen from attaching to the estrogen receptor on the cancer cells, slowing the growth of tumors and killing tumor cells. Tamoxifen can be used in both pre- and postmenopausal women. Possible side effects include fatigue, hot flashes, night sweats and vaginal dryness. More significant risks include cataracts, blood clots, stroke and uterine cancer.
- **Medications that stop estrogen synthesis in peripheral tissues after menopause.** One group of drugs called aromatase inhibitors blocks the action of an enzyme that converts androgens in the body into estrogen. These drugs are effective only in postmenopausal women. Aromatase inhibitors include anastrozole (Arimidex), letrozole (Femara) and exemestane (Aromasin). Side effects of aromatase inhibitors include joint and muscle pain, as well as an increased risk of bone thinning (osteoporosis). Another drug, fulvestrant (Faslodex), directly blocks estrogen, which keeps tumors from getting the estrogen they need to survive. Fulvestrant is generally used in postmenopausal women for whom other hormone-blocking therapy is not effective or who cannot take tamoxifen. Side effects that may occur include fatigue, nausea and hot flashes. Fulvestrant is given by injection once a month.
- **Surgery or medications to stop hormone production in the ovaries.** In premenopausal women, surgery to remove the ovaries or medications to stop the ovaries from making estrogen can be an effective hormonal treatment. This type of surgery is known as prophylactic phorectomy and may be called surgical menopause.

For the treatment of early stage breast cancer, certain aromatase inhibitors may be used as adjuvant therapy instead of tamoxifen or after 2 or more years of tamoxifen. For the treatment of metastatic breast cancer, aromatase inhibitors are being tested in clinical trials to compare them to hormone therapy with tamoxifen.

Targeted therapy is a type of treatment that uses drugs or other substances to identify and attack specific cancer cells without harming normal cells. Monoclonal antibodies and tyrosine kinase inhibitors are two types of targeted therapies used in the treatment of breast cancer. PARP inhibitors are a type of targeted therapy being studied for the treatment of triple-negative breast cancer. Tyrosine kinase inhibitors are targeted therapy drugs that block signals needed for tumors to grow. Tyrosine kinase inhibitors may be used with other anticancer drugs as adjuvant therapy.

Monoclonal antibody therapy is a cancer treatment that uses antibodies made in the laboratory, from a single type of immune system cell. These antibodies can identify substances on cancer cells or normal substances that may help cancer cells grow. The antibodies attach to the substances and kill the cancer cells, block their growth, or keep them from spreading. Monoclonal antibodies are given by infusion. They may be used alone or to carry drugs, toxins, or radioactive material directly to cancer cells. Monoclonal antibodies may be used in combination with chemotherapy as adjuvant therapy.

- **Trastuzumab (Herceptin).** Some breast cancers contain excessive amounts of a protein called human growth factor receptor 2 (HER2). Trastuzumab targets this protein that is directly involved in breast cancer cells growth and survival. Side effects may include heart damage, headaches and skin rashes. About one-fourth of patients with breast cancer have tumors that may be treated with trastuzumab combined with chemotherapy.
- **Pertuzumab** is a monoclonal antibody that may be combined with trastuzumab and chemotherapy to treat breast cancer. It may be used to treat certain patients with HER2-positive breast cancer that has metastasized .
- **Lapatinib (Tykerb).** Lapatinib targets the HER2 protein and is approved for use in advanced metastatic breast cancer. Lapatinib is generally used for women who have already tried trastuzumab and their cancer has progressed. Potential side effects include nausea, vomiting, diarrhea, fatigue, mouth sores, skin rashes, and painful hands and feet.
- **Bevacizumab (Avastin).** Bevacizumab is a drug designed to interfere with VEGF-promoted neoangiogenesis. Without new blood vessels to bring oxygen and nutrients to the tumor, the cancer cells die. Possible side effects include fatigue, high blood pressure, mouth sores, headaches, slow wound healing, blood clots, heart damage, kidney damage, high blood

pressure and congestive heart failure. Recent data suggests that although this medication may help slow the growth of breast cancer, it does not appear to increase survival times probably due to induction of hypoxia-mediated tumor-promoting effects. Use of bevacizumab in breast cancer therefore still controversial.

Clinical trials

Clinical trials are used to test new and promising agents in the treatment of cancer. Clinical trials represent the cutting edge of cancer treatment, but they are by definition the administration of unproven treatments that may or may not be superior to currently available therapies.

Examples of treatments being studied in breast cancer clinical trials include:

- **New combinations of existing drugs.** Researchers are studying new ways of combining existing chemotherapy, hormone therapy and targeted-therapy drugs. Testing new combinations may help determine if certain breast cancers are more susceptible to specific combinations.
- **Bone-building drugs to prevent breast cancer recurrence.** Previous research found that adding a bone-building drug to hormone therapy treatment after surgery for premenopausal women reduced the risk of breast cancer recurrence. The drug used in the study, zoledronic acid (Reclast, Zometa), is a type of drug called a bisphosphonate that is used to treat osteoporosis and other bone diseases.
- **Using higher doses of radiation over a shorter period of time on a smaller portion of the breast.** Researchers are studying partial breast irradiation in women who have undergone lumpectomy. Partial breast irradiation involves higher doses of radiation aimed at only a portion of the breast, rather than the entire breast

1.4 ROLE OF BIOMARKERS

The TNM system, 50 years after its introduction, is still the main staging system to define the initial clinical aggressiveness of the tumor and consequently to modulate the treatment.

However, over the years, the identification of specific molecular alterations in tumor cells directly responsible and/or associated with transformation and tumor progression, has suggested the possibility of integrating the TNM anatomic staging with a bio- functional tumor characteristics.

A significant number of potential biomarkers of clinical utility were identified thanks to the increasing knowledge of tumor biology and to the availability of high-throughput technologies. Nevertheless, the real clinical utility of most of the biomarkers is still very controversial.

It is in fact important to know if therapeutic decisions guided by prediction based on biomarkers will add a direct benefit to the patient compared decisions taken based on conventional clinical pathological factors. Direct benefit means a better control of the tumor, a better therapeutic ratio linked to a reduction of the side effects or even the non-prescription of an ineffective treatment.

The National Cancer Institute (NCI) definition of biomarker is: “A biological molecule found in blood, other body fluids, or tissues that are a sign of a normal or abnormal process, or of a condition or disease. A biomarker may be used to see how well the body responds to a treatment for a disease or condition. Also called molecular marker and signature molecule."

It is necessary to distinguish between *disease-related* and *drug-related biomarkers*. Disease-related biomarkers give an indication of the probable effect of treatment on patient (risk indicator or predictive biomarkers), if a disease already exists (diagnostic biomarker), or how such a disease may develop in an individual case regardless of the type of treatment (prognostic biomarker). Predictive biomarkers help to assess the most likely response to a particular treatment type, while prognostic markers show the progression of disease with or without treatment (31). In contrast, drug-related biomarkers indicate whether a drug will be effective in a specific patient and how the patient’s body will process it.

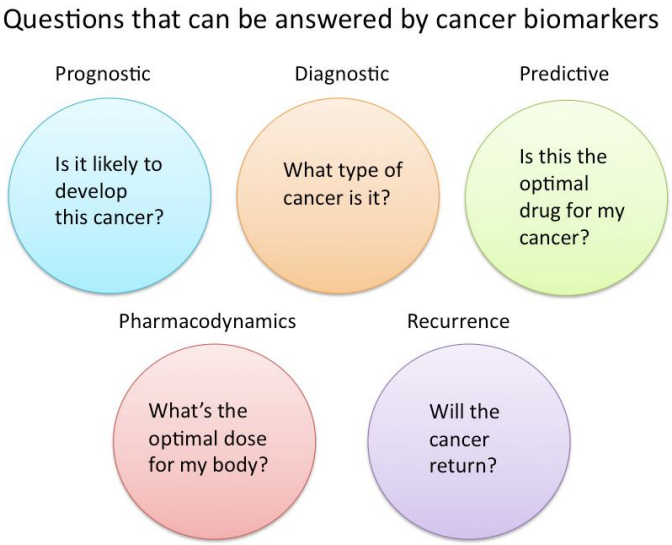


Figure 12: Biomarkers types

Prognostic

These cancer biomarkers help to assess the risk of developing a particular cancer and determine prognosis. They provide information to identify patients with different probabilities of survival or recovery of the disease, or tumors with highly invasive phenotype. In breast cancer patients with axillary lymph node metastases are considered at risk of relapse and therefore subjected to adjuvant treatments after surgical removal of the primary tumor. For patients that at the time of surgery do not have evidence of lymph node metastases, the decision is more difficult because a significant proportion will not experience a recurrence, while in about 30% of cases distant site disease will develop which will not be controlled by simple surgical eradication of the primary tumor. In such a situation the use of a biomarker is very useful.

Diagnostic

Diagnostic markers guide the diagnosis in particular type of cancers when pathologists are unable to identify the specific type of cancer based on morphology only.

Predictive

Predictive biomarkers predict the response to cancer drug(s) or treatment(s) in the patient (32). An example is given by the human epidermal growth factor receptor 2 (HER-2) overexpression due

to amplification of the HER-2 gene. Breast cancer patients with this abnormal increase of HER-2 respond to trastuzumab (Herceptin®). However, breast cancer patients only over-expressing HER-2 protein without gene amplification are not recommended to be treated with this drug. In breast cancer, the presence of receptors for estrogen (ER) and progesterone (PgR) is the strongest predictor of response to a hormone treatment with anti-estrogen (tamoxifen) or with aromatase inhibitor.

Pharmacodynamics and Pharmacokinetics

Cancer biomarkers under this category help determine the most effective dosage of drug or therapy is needed for that specific person (33). These biomarkers are just another tool aiding the field of personalized medicine.

Recurrence

Recurrence biomarkers are used to predict if cancer is likely to recur after treatment. An example is the Oncotype DX® breast cancer assay. This assay measures the expression of several genes within a breast tumor sample and allows to calculate a score (recurrence score, RR) associated to the likelihood of the patient to experience a relapse that the patient's cancer will return.

The sensitivity (ability to detect all positive cases) and specificity (ability to avoid false positives) of cancer biomarker determine its reliability. In addition to being reliable, provide a proved clinical benefit, a new biomarker must also provide independent and non-redundant information with respect to that provided by other markers/classifiers. The biomarker assay must be easily accessible to allow the widest possible clinical application and its determination must be standardized and reproducible. Finally, the optimal cancer biomarker is one that can be assessed on easily accessed biosamples (i.e. blood, urine, tissue from biopsy).

For breast cancer many markers based on gene expression signatures have been described, but only a few have been approved for clinical use in specific settings by the Food and Drug Administration (FDA).

1.5 TUMOR MICROENVIRONMENT

The tumor microenvironment composed of non cancer cells (immune cells and stromal cells) and extracellular matrix has been increasingly recognized as a major factor influencing the growth of cancer. It is not a simple bystander, but plays an important role in determining tumor progression and response to treatment (34).

While the normal cellular microenvironment con inhibit malignant cell growth, the modification that occur in the tumor microenvironment synergistically support cell proliferation. A variety of stromal cells in the surrounding environment are recruited by tumours, and these not only enhance growth of the primary cancer but also facilitate its metastatic dissemination to distant organs.

Tumor epithelial cells are surrounded by the tumor microenvironment, which is composed of the extracellular matrix (ECM) and various other cell types, such as endothelial cells, fibroblast, myofibroblast, leukocytes and myoepithelial and endothelial cells, that play a role in the initiation and progression of neoplasms (35;36). Cross-talk between epithelial and stromal cells is known to be essential for differentiation and development of normal organs and tissue as well as for the growth and progression of tumours.

Several studies have provided insight on the molecular characteristics differentiating tumor-associated stroma from normal stroma. Allinen and colleagues suggested that during cancer progression, striking changes in gene expression occurred in almost every cell type, with the most dramatic and consistent change (other than in the malignant epithelial cells) detected in cells expressing the surface marker CD10, which encompasse myoepithelial cells and myofibroblast (37).

Often in invasive carcinoma there is an expansion of tumor stroma, which is morphologically identified as desmoplasia, an area where tumor and stroma actively interact (38).

Fibroblast, among stromal cells, are the major cellular component, and have a profound influence on the phenotype of various carcinomas, inducing the production of various growth factors and chemokines and their release into the in the microenvironmental (39;40).

Stromal and other non-malignant cells create a supportive microenvironment for tumor growth, angiogenesis, and metastasis (40). The tumor recruits endothelial cells, fibroblasts, and inflammatory cells, pericytes, and these cells and the components of the extracellular matrix (ECM) secreted by them contribute to the microenvironment composition. Stromal cells generate both tumor enhancing and tumor suppressing signals. Fibroblast isolated from carcinoma, named

carcinoma-associated fibroblast (CAF), show biological function distinct from those of normal fibroblast in tumor progression and angiogenesis (41)**.** Cancer-associated fibroblasts (CAF) and myofibroblasts are abnormal stromal cells, but not malignant, and promote angiogenesis and proliferation. These fibroblasts secrete growth factors and cytokines that produce oncogenic signals. In a comparative study which isolated both CAF and normal fibroblasts from six patients with invasive breast cancer, Orimo, et al., demonstrated that CAF were able to promote tumor growth more than normal fibroblasts (41). These activated fibroblasts were able to promote angiogenesis via expression of stromal cell-derived factor-1 (SDF-1 or CXCL12). SDF-1 produces an endocrine effect by recruiting circulating endothelial progenitor cells (EPC) to the tumor. Tumor cells also contain CXCR4, the receptor for SDF-1, so stromal SDF-1 is able to directly stimulate tumor growth through a paracrine effect. The EPC recruited to the tumor by the cancer-associated fibroblasts are capable of being differentiated into tumor-associated vascular endothelial cells and used to construct new capillaries (41). Circulating EPC migrate to other tissues and form a "premetastatic niche" for the colonization of circulating tumor cells (42). SDF-1 is not the only factor that participates in recruiting cells to the tumor microenvironment. Growth factors secreted by the tumor also control the composition of the microenvironment. Transforming growth factor-β (TGF-β) recruits EPC to the microenvironment and is involved in activation of fibroblasts to CAF, while platelet derived growth factor (PDGF) is involved in recruiting fibroblasts and inducing their proliferation. VEGF does not directly recruit fibroblasts, but indirectly supports microenvironmental changes via creation of dysfunctional vascularization that allows plasma leakage, which attracts fibroblasts and other cells (42) .The microenvironment may convert recruited stromal cells to cancer-associated cells through epigenetic changes, including DNA methylation and chromatin remodeling, as changes in histone modification and DNA methylation have been found in tumor cells.

Proteins secreted by the tumor modify the microenvironment by contributing growth factors and proteases that degrade the extracellular matrix, and affect cell motility and adhesion. Stromal cells secrete ECM proteins, cytokines, growth factors, proteases, protease inhibitors, and endoglycosidases such as heparanase. These proteins modify the ECM in what is thought to be a systematic manner. Osteopontin, galectin-3, transforming growth factor-β and matrix metalloproteinases (MMP) are important secreted proteins closely associated with cancer development (43). MMP are expressed at higher levels by tumor-associated epithelial cells than by normal epithelial cells. These MMP revise the composition of the EMC by degrading the basement

membrane and other ECM proteins. This increase in proteolytic activity may also act to support tumor malignancy. Extreme shifts in the inhibition of protease activity by plasminogen activation inhibitor-1 (PAI-1), either by its absence or by excessive levels, demonstrate an anti-angiogenic effect in mouse cells, while physiological levels of PAI-1 support angiogenesism (44). Tissue inhibitors of metalloproteinases (TIMP) are endogenous inhibitors of MMP and may function to balance the protease activity of MMP to shift the balance from a pro-angiogenic to an inhibitory environment.

Targeting the tumor microenvironment may be a feasible therapeutic approach for cancer treatment and prevention. In contrast to genetically unstable tumor cells, the genetically stable cells composing the microenvironment are less plastic and less likely to acquire drug resistance, and therefore potentially make better targets for cancer therapy (45). In line with this, anti-angiogenic drugs have been developed to target tumor endothelial cells as a means of cancer treatment (46).

1.6 CANCER AND INFLAMMATION

Central to the development of cancer are genetic changes that endow these “cancer cells” with many of the hallmarks of cancer, such as self-sufficient growth and resistance to antigrowth and pro-death signals. However, while the genetic changes that occur within cancer cells themselves, such as activated oncogenes or dysfunctional tumor suppressors, are responsible for many aspects of cancer development, they are not sufficient. Tumor promotion and progression are dependent on ancillary processes provided by cells of the tumor environment but that are not necessarily cancerous themselves. Inflammation has long been associated with the development of cancer (47) (Fig.13).

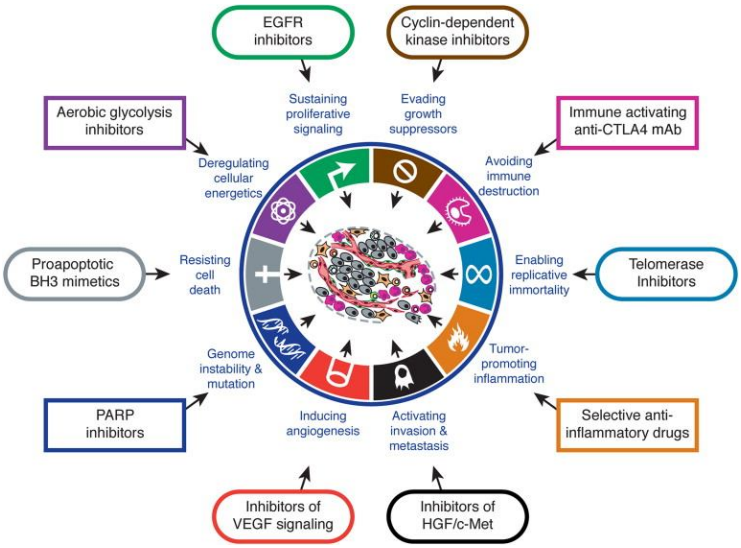


Figure 13: Hallmarks of Cancer.

Epidemiological studies have shown that chronic inflammation predisposes individuals to various types of cancer. It is estimated that underlying infections and inflammatory responses are linked to 15–20% of all deaths from cancer worldwide (48). There are many triggers of chronic inflammation that increase the risk of developing cancer. Such triggers include microbial infections (for example, infection with *Helicobacter pylori* is associated with gastric cancer and gastric mucosal lymphoma), autoimmune diseases (for example, inflammatory bowel disease is associated with colon cancer) and inflammatory conditions of unknown origin.

The main function of the mammalian immune system is to monitor tissue homeostasis, to protect against invading or infectious pathogens and to eliminate damaged cells. Therefore, it is surprising that cancer occurs with such a high frequency in humans. Clinical and experimental studies indicate that innate and adaptive immune cells are significant, albeit sometime paradoxical, determinants of epithelial tumorigenesis. Links between cancer and inflammation were first made in the nineteenth century, on the basis of observations that tumors often arose at sites of chronic inflammation and that inflammatory cells were present in biopsy samples from tumors.

Recent data have expanded the concept that inflammation is a critical component of tumour progression. It is now becoming clear that the tumour microenvironment, which is largely orchestrated by inflammatory cells, is an indispensable participant in the neoplastic process, fostering proliferation, survival and migration. The hallmarks of cancer-related inflammation include the presence of inflammatory cells and inflammatory mediators (for example, chemokines, cytokines and prostaglandins) in tumour tissues, tissue remodeling and angiogenesis similar to that

seen in chronic inflammatory responses, and tissue repair. These signs of ‘smouldering’ inflammation (49) are also present in tumours for which a firm causal relationship to inflammation has not been established (for example, breast tumours). Indeed, inflammatory cells and mediators are present in the microenvironment of most, if not all, tumours, irrespective of the trigger for development. The connection between inflammation and cancer can be viewed as consisting of two pathways: an extrinsic pathway, driven by inflammatory conditions that increase cancer risk (such as inflammatory bowel disease); and an intrinsic pathway, driven by genetic alterations that cause inflammation and neoplasia (such as oncogenes) (Fig.14).

The intrinsic pathway is activated by genetic events that cause neoplasia. These events include the activation of various types of oncogene by mutation, chromosomal rearrangement or amplification, and the inactivation of tumour-suppressor genes. Cells that are transformed in this manner produce inflammatory mediators, thereby generating an inflammatory microenvironment in tumours for which there is no underlying inflammatory condition (for example, breast tumours). By contrast, in the extrinsic pathway, inflammatory or infectious conditions augment the risk of developing cancer at certain anatomical sites (for example, the colon, prostate and pancreas). The two pathways converge, resulting in the activation of transcription factors, mainly nuclear factor-κB (NF-κB), signal transducer and activator of transcription 3 (STAT3) and hypoxia-inducible factor 1α (HIF1α), in tumour cells. These transcription factors coordinate the production of inflammatory mediators, including cytokines and chemokines, as well as the production of cyclooxygenase 2 (COX2) (which, in turn, results in the production of prostaglandins). These factors recruit and activate various leukocytes, most notably cells of the myelomonocytic lineage. The cytokines activate the same key transcription factors in inflammatory cells, stromal cells and tumour cells, resulting in more inflammatory mediators being produced and a cancer-related inflammatory microenvironment being generated. Smouldering cancer-related inflammation has many tumour-promoting effects (50).

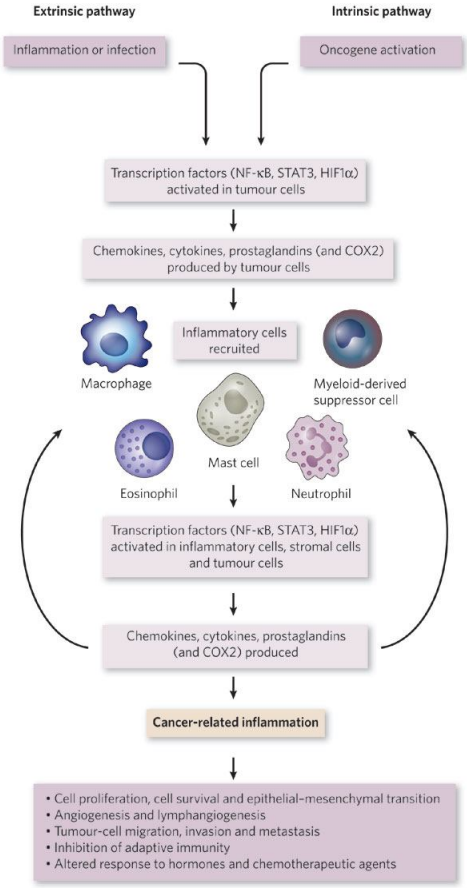


Figure 14: Pathway that connect inflammation and cancer.

To understand the role of inflammation in the evolution of cancer, it is important to understand what inflammation is and how it contributes to physiological and pathological processes such as wound healing and infection (Fig.15).

Normal tissues have a highly organized and segregated architecture. Epithelial cells sit atop a basement membrane separated from the vascularized stromal (dermis) compartment. Upon wounding or tissue assault, platelets are activated and form a haemostatic plug where they release vasoactive mediators that regulate vascular permeability, influx of serum fibrinogen, and formation of the fibrin clot. Chemotactic factors such as transforming growth factor-β and platelet-derived growth factor, derived from activated platelets, initiate granulation tissue formation, activation of fibroblasts, and induction and activation of proteolytic enzymes necessary for remodelling of the extracellular matrix (for example, matrix metalloproteinases and urokinase-type plasminogen activator). In combination, granulocytes, monocytes and fibroblasts are recruited, the venous

network restored, and re-epithelialization across the wound occurs. Epithelial and stromal cell types engage in a reciprocal signalling dialogue to facilitate healing. Once the wound is healed, the reciprocal signalling subsides (Fig. 15a). Invasive carcinomas, instead are less organized, in fact neoplasia-associated angiogenesis and lymphangiogenesis produces a chaotic vascular organization of blood vessels and lymphatics where neoplastic cells interact with other cell types (mesenchymal, haematopoietic and lymphoid) and a remodelled extracellular matrix. Although the vascular network is not disrupted in the same way during neoplastic progression as it is during wounding, many reciprocal interactions occur in parallel. Neoplastic cells produce an array of cytokines and chemokines that are mitogenic and/or chemoattractants for granulocytes, mast cells, monocytes/macrophages, fibroblasts and endothelial cells. In addition, activated fibroblasts and infiltrating inflammatory cells secrete proteolytic enzymes, cytokines and chemokines, which are mitogenic for neoplastic cells, as well as endothelial cells involved in neoangiogenesis and lymphangiogenesis. These factors potentiate tumour growth, stimulate angiogenesis, induce fibroblast migration and maturation, and enable metastatic spread via engagement with either the venous or lymphatic networks (Fig 15b) (51).

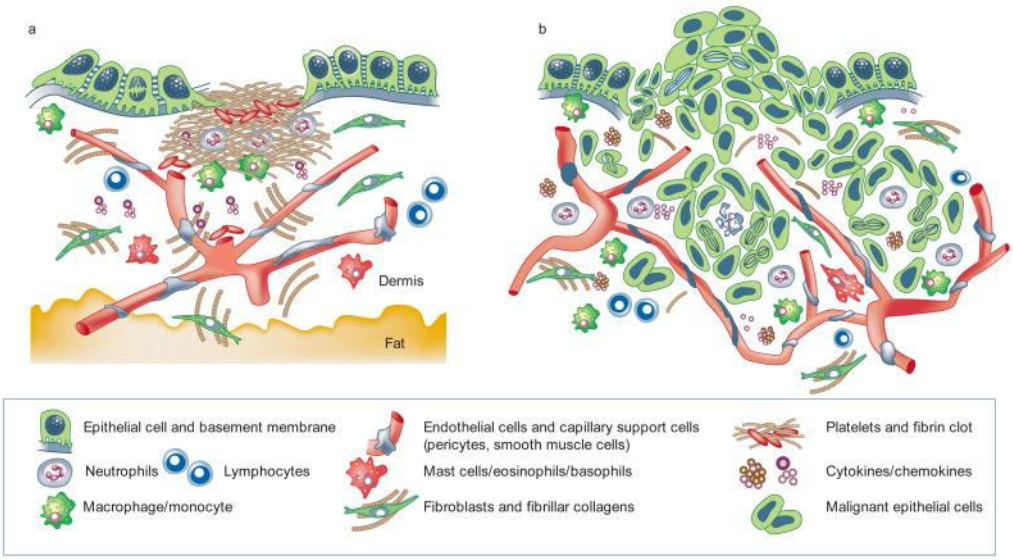


Figure 15: Wound healing versus invasive tumour growth. **a**, Normal tissues have a highly organized and segregated architecture; **b**, Invasive carcinomas are less organized.

In response to tissue injury, a multifactorial network of chemical signals initiate and maintain a host response designed to 'heal' the afflicted tissue. This involves activation and directed migration of leukocytes (neutrophils, monocytes and eosinophils) from the venous system to sites of damage

and tissue mast cells also have a significant role. For neutrophils, a four-step mechanism is believed to coordinate recruitment of these inflammatory cells to sites of tissue injury and to the provisional extracellular matrix (ECM) that forms a sCAFfolding upon which fibroblast and endothelial cells proliferate and migrate, thus providing a nidus for reconstitution of the normal microenvironment. These steps involve: activation of members of the selectin family of adhesion molecules (L- P-, and E-selectin) that facilitate rolling along the vascular endothelium; triggering of signals that activate and upregulate leukocyte integrins mediated by cytokines and leukocyte-activating molecules; immobilization of neutrophils on the surface of the vascular endothelium by means of tight adhesion through α4β1 and α4β7 integrins binding to endothelial vascular cell-adhesion molecule-1 (VCAM-1) and MadCAM-1, respectively; and transmigration through the endothelium to sites of injury, presumably facilitated by extracellular proteases, such as matrix metalloproteinases (MMPs).

The profile of cytokine/chemokines persisting at an inflammatory site is important in the development of chronic disease. The pro-inflammatory cytokine TNF-α (tumour necrosis factor-α) controls inflammatory cell populations as well as mediating many of the other aspects of the inflammatory process. TGF-β1 is also important, both positively and negatively influencing the processes of inflammation and repair (52). The key concept is that normal inflammation is usually self-limiting; however, dysregulation of any of the converging factors can lead to abnormalities and ultimately, pathogenesis. Neutrophils (and sometimes eosinophils) are the first recruited effectors of the acute inflammatory response. Monocytes, which differentiate into macrophages in tissues, are next to migrate to the site of tissue injury, guided by chemotactic factors. Once activated, macrophages are the main source of growth factors and cytokines, which profoundly affect endothelial, epithelial and mesenchymal cells in the local microenvironment. Mast cells are also important in acute inflammation owing to their release of stored and newly synthesized inflammatory mediators, such as histamine, cytokines and proteases complexed to highly sulphated proteoglycans, as well as lipid mediators.

Cytokine and chemokine balances regulate neoplastic outcome. The balance of cytokines in any given tumour is critical for regulating the type and extent of inflammatory infiltrate that forms. Tumours that produce little or no cytokines or an overabundance of anti-inflammatory cytokines induce limited inflammatory and vascular responses, resulting in constrained tumour growth. In contrast, production of an abundance of pro-inflammatory cytokines can lead to a level of inflammation that potentiates angiogenesis, thus favouring neoplastic growth. Alternatively, high

levels of monocytes and/or neutrophil infiltration, in response to an altered balance of pro-versus anti-inflammatory cytokines, can be associated with cytotoxicity, angiostasis and tumour regression. In tumours, interleukin-10 is generally a product of tumour cells and tumour-associated macrophages. (Fig.16) (51).

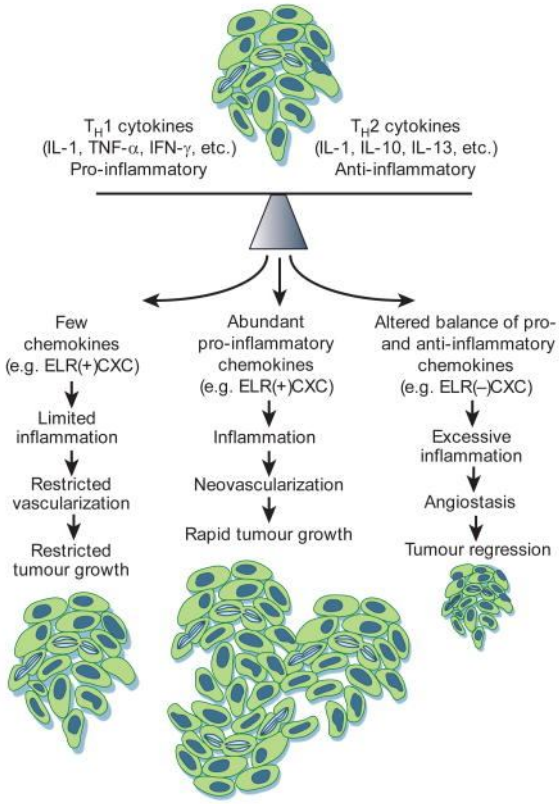


Figure 16: Cytokine and chemokine balances regulate neoplastic outcome.

Yet, the recruitment of inflammatory cells may also be counterproductive for tumour development, and also may represent an attempt by the host to suppress tumour growth (53). The pro-tumour actions of inflammatory cells include releasing growth and survival factors, promoting angiogenesis and lymphangiogenesis, stimulating DNA damage, remodelling the ECM to facilitate invasion, coating tumour cells to make available receptors for disseminating cells via lymphatics and capillaries, and evading host defense mechanisms. Although inflammatory responses should also be anti-tumour, cancer patients are often defective in their inflammatory responses. This may arise by two distinct tumour-mediated mechanisms: a failure to upregulate the anti-inflammatory cytokines,

or subversion of the host response resulting from desensitization of receptors owing to high chemokine and cytokine concentrations that then blunt systemic responses.
The challenge for the future is to normalize the inflammatory network to regain a normal host response overall: decreasing the high levels of tumour-promoting properties of the infiltrating cells, such as pro-inflammatory cytokines, while increasing their tumour-suppressing properties, such as anti-inflammatory cytokines. In this way, later in tumour progression, we can harness the activities that are antitumour while suppressing those that are pro-tumour.

1.6.1 INTERFERONS

Interferons (IFNs) are a diverse family of pleiotropic cytokines in humans typically divided among three IFN classes: Type I IFN, Type II IFN, and Type III IFN. All classes are very important for fighting viral infections. Interferons (IFNs) belong to the large class of glycoproteins known as cytokines. Interferons are named after their ability to "interfere" with viral replication within host cells but they have other functions: activate immune cells, such as natural killer cells and macrophages; increase recognition of infection or tumor cells by up-regulating antigen presentation to T lymphocytes; and increase the ability of uninfected host cells to resist new infection by virus. They allow for communication between cells to trigger the protective defenses of the immune system that eradicate pathogens or tumors.
IFNs play an essential role in innate immunity by inhibiting the replication and spread of viral, bacterial, and parasitic pathogens. They also modulate immune responses and exert antiproliferative effects in some cell types. As a result of these functions, IFNs are used in the clinic to treat certain viral infections, some cancer types, and multiple sclerosis. IFNs mediate their effects by binding to cell surface receptors activating members of the JAK kinase family of proteins. Activated JAK kinases phosphorylate the signal transducers and activators of transcription (STAT) family of transcription factors. The STAT proteins homo- or heterodimerize and form complexes with other transcription factors to activate transcription of IFN-stimulated genes (ISGs) (54). The gene products regulated by IFNs are the primary effectors of the IFN response. Although the functions of most ISGs remain to be elucidated, some of the best studied ISGs play pivotal roles in host defense. Experiments using mice indicate the IFNs are essential for innate immunity against viral infections. Mice with a targeted disruption of the type I or II IFN receptor genes are extremely susceptible to

viral infections. IFNs induce the production of several known antiviral proteins including the double-stranded RNA dependent kinase “protein kinase RNA-regulated” (PRKR), a family of 29,59-oligoadenylate synthetases that lead to the activation of RNase L and the Mx proteins, all of which have been shown to restrict the growth of certain viruses (55). However, the inhibition of viral replication induced by IFNs is only partially dependent on these particular ISGs, because mice triply deficient for PRKR, RNase L, and Mx1 genes retain partial responsiveness to the antiviral effects of IFNs (56). These results imply that other as-yet-unidentified ISGs are also potent antiviral effectors.

1.7 MICROARRAY TECHNOLOGY

Microarray technology evolved from Southern blotting, where fragmented DNA is attached to a substrate and then probed with a known DNA sequence. A microarray (also commonly known as DNA chip or biochip) is a collection of DNA spots attached to a solid surface. This technology is used to measure the expression levels of large numbers of genes simultaneously or to genotype multiple regions of a genome. Each DNA spot contains picomoles (10^{-12} moles) of a specific DNA sequence, known as *probes* (or *reporters or oligos*). These can be a short section of a gene or other DNA element that are used to hybridize a cDNA or cRNA (also called anti-sense RNA) sample (called *target*) under high-stringency conditions. Probe-target hybridization is usually detected and quantified by detection of fluorophore-, silver-, or chemiluminescence-labeled targets to determine relative abundance of nucleic acid sequences in the target.

Since an array can contain tens of thousands of probes, a microarray experiment can accomplish many genetic tests in parallel. Therefore arrays have dramatically accelerated many types of investigation. In standard microarrays, the probes are synthesized and then attached via surface engineering to a solid surface by a covalent bond to a chemical matrix (via epoxy-silane, amino-silane, lysine, polyacrylamide or others). The solid surface can be glass or a silicon chip, in which case they are colloquially known as an *Affy chip* when an Affymetrix chip is used. Other microarray platforms, such as Illumina, use microscopic beads, instead of the large solid support. Alternatively, microarrays can be constructed by the direct synthesis of oligonucleotide probes on solid surfaces. DNA arrays are different from other types of microarray only in that they either measure DNA or use DNA as part of its detection system. Microarrays also differ in fabrication, workings, accuracy,

efficiency, and cost. Many types of arrays exist and the broadest distinction is whether they are spatially arranged on a surface or on coded beads:

- The traditional solid-phase array is a collection of orderly microscopic "spots", called features, each with a thousands of identical and specific probes attached to a solid surface, such as glass, plastic or silicon biochip (commonly known as a *genome chip*, *DNA chip* or *gene array*). Thousands of these features can be placed in known locations on a single DNA microarray.
- The alternative bead array is a collection of microscopic polystyrene beads, each with a specific probe and a ratio of two or more dyes, which do not interfere with the fluorescent dyes used on the target sequence.

DNA microarrays can be used to detect DNA (as in comparative genomic hybridization), or detect RNA (most commonly as cDNA after reverse transcription) that may or may not be translated into proteins. The process of measuring gene expression via cDNA is called expression analysis or gene expression profiling.

Depending upon the kind of immobilized sample used to construct the arrays and the information fetched, the Microarray experiments can be categorized in three ways:

1. **Microarray Expression Analysis:** in this experimental setup, the cDNA derived from the mRNA of known genes is hybridized. Gene expression microarray can be single-or double channel. In the former case all samples are labeled with the same dye and independently hybridized. Single gene expression levels are derived by comparing intensities between the different samples after appropriate normalization procedures. In the two channel arrays, each sample is compared to a reference (e.g. an RNA sample obtained by an appropriate mixture of tumor samples, cell lines, etc) or to a different type of sample (e.g. normal tissue *vs* tumor tissue). Samples and reference are labeled with different dyes, and the obtained colored spots reflects the relative abundance for each gene among the two samples. The sample has genes from both the normal as well as the diseased tissues. Spots with more intensity are obtained for diseased tissue gene if the gene is over expressed in the diseased condition. This expression pattern is then compared to the expression pattern of a gene responsible for a disease.

2. **Microarray for Mutation Analysis:** for this analysis, the researchers use gDNA. The genes might differ from each other by as less as a single nucleotide base.

A single base difference between two sequences is known as Single Nucleotide Polymorphism (SNP) and detecting them is known as SNP detection.

3. **Comparative Genomic Hybridization:** it is used for the identification in the increase or decrease of the important chromosomal fragments harboring genes involved in a disease.

Microarrays can be used for different purposes:

Gene Discovery: DNA Microarray technology helps in the identification of new genes, gain knowledge about their function and expression levels under different conditions.

Disease Diagnosis: DNA Microarray technology helps researchers learn more about different diseases such as heart diseases, mental illness, infectious disease and especially the study of cancer. Until recently, different types of cancer have been classified on the basis of the organs in which the tumors develop. Now, with the evolution of microarray technology, it will be possible for the researchers to further classify the types of cancer on the basis of the patterns of gene activity in the tumor cells. This will tremendously help the pharmaceutical community to develop more effective drugs as the treatment strategies will be targeted directly to the specific type of cancer.

Drug Discovery: Microarray technology has extensive application in *Pharmacogenomics.* Pharmacogenomics is the study of correlations between therapeutic responses to drugs and the genetic profiles of the patients. Comparative analysis of the genes from a diseased and a normal cell will help the identification of the biochemical constitution of the proteins synthesized by the diseased genes. Specific genes coding for specific proteins may emerge this way as druggable targets.

Toxicological Research: Microarray technology provides a robust platform for the research of the impact of toxins on the cells and their passing on to the progeny. Toxicogenomics establishes correlation between responses to toxicants and the changes in the genetic profiles of the cells exposed to such toxicants.

In the recent past, microarray technology has been extensively used and over the years, an enormous amount of data related to gene expression has been generated . This data is scattered and

is not easily available for public use. To facilitate accessibility to this data, the **National Center for Biotechnology Information** (**NCBI**) has created the **Gene Expression Omnibus** or **GEO**. It is a data repository facility which includes data on gene expression from varied sources .Many of the gene expression data are also linked to clinical information.

Nevertheless the heterogeneity of platforms used has raised some skepticism about the reliability, reproducibility and comparability of microarray data. A study conducted by the Food and Drug Administration as a project called Microarray Quality Control (MAQC) has dispelled many doubts thanks to the analysis of data from 137 participants from 51 academic and industrial partners obtained from the use of different platforms (Illumina, Agilent and Affymetrix). The reproducibility of microarray data was found to be comparable to that of commonly used diagnostic tests, such as immunohistochemistry and evaluation of the histological grade (57).

1.7.1 GENE EXPRESSION PROFILES OF BREAST CANCER: STATE OF THE ART

In cancerous disease many genes are altered. The definition of the gene expression profile of the tumor allowed a molecular classification of cancer to better understand the prognostic classification, a more accurate prediction of response to treatment and the aggressiveness of the disease (58).

Gene expression analysis offers a systematic approach for the prediction and identification of new gene functions, and molecular characteristics (59) useful for the development of new drugs and new potential therapeutic strategies (60).

The successes achieved by the administration of trastuzumab and tamoxifen in patients respectively ERBB2-positive and ER-has demonstrated the importance to identify *a priori* the patients who would benefit of a particular treatment (59). Finally, of great interest from the point of view pharmacogenomics, is the study of gene expression in the course of a specific treatment.

- Development of multigene expression signatures in breast cancer

One of the first studies undertaken with microarray in breast cancer has led to the definition of a molecular classification of this highly heterogeneous neoplasm (61;62). Based on the gene expression profile 5 molecular classes were identified:

1) luminal A,
2) luminal B,
3) basal-like tumors,

4) ERBB2 positive tumors and

5) normal breast-like tumors (63).

The recognition of a distinct expression profile between different tumor subtypes has allowed us to better understand the biology and has contributed to a new vision of breast cancer, not as a single disease, but as a set of several biologically distinct diseases.

This observation is changing the design of trials of new drugs because it is now clear that the molecular type stratification is a necessary strategy to improve the response to each treatment.

Clinical pathological variables are not able to describe the heterogeneity of breast cancer, as patients with the same clinical stage and pathological grade show a wide spectrum of variability in treatment response and prognosis.

It is therefore necessary to develop classifiers based on multigene expression signature, allowing a better patient stratification. The classes considered for stratification of patients for treatment purposes are generally good and bad prognosis, but such classes may be any other feature for which it is possible to define a training set of data.

- Prediction of prognosis

With microarray technique it investigated the heterogeneity of disease in order to develop tools (gene signature) to improve outcome prediction compared traditional prognostic factors.

In literature there are numerous examples, mainly based on two different approaches:

• TOP-DOWN approach: comparison of gene expression profiles of tumors from groups of patients with known disease outcome, in order to identify differentially expressed genes (class comparison) between the groups without any assumption biological *a priori*.

• BOTTOM-UP approach or hypothesis-driven approach: selected gene associated to specific phenotypes or pathways are *a priori* identified based on previous biological knowledge pathway and their correlation with disease outcome is tested. The so-called CANDIDATE GENES approach is a specific type of bottom up approach where a multivariate predictor is built combining gene expression data of a limited series of selected genes on the basis of biological knowledge.

-Prediction of treatment response

The improvement of the prognostic tools will play a fundamental role in the future both for the identification of patients at high risk of recurrence disease to be subjected to aggressive treatments and to avoid unnecessary treatments and toxicity in low-risk patients.

However, is also important to identify which is the optimal therapy for each patient. In oncology the ultimate goal of functional genomics is the definition of personalized therapy.

The molecular markers that are routinely used in breast cancer study are the estrogen receptor (ER), progesterone receptor (PgR) and ERBB2 also known as HER-2/neu. The ER presence predicts the tamoxifen and other anti-estrogens response. Due to its low toxicity tamoxifen is used in the long term (5 years) in the treatment of ER-positive tumors. Approximately the 70% of tumors ER/PgR-positive, the 45% of tumors ER-negative and PgR-positive and only the 34% of tumors ER-positive/PgR-negative are responsible to treatment with tamoxifen. ErbB2 is a tyrosine kinase receptor overexpressed in 20-30% of cases of breast cancer and is associated with a high risk of recurrence and death. The presence of ERBB2 is predictive of the response to trastuzumab (Herceptin) (64).

Therefore the prediction of treatment response based on ER status is not satisfactory. This is mainly due to the development of resistance mechanisms that may be of two types: intrinsic (already present at the beginning of the treatment) or acquired (induced by the treatment itself). The gene profile study has so far been mainly used to understand the intrinsic resistance, and only a few studies have faced the problem of acquired resistance through the study of the gene profile after treatment in patients treated with neoadjuvant therapy (65).

These studies have shown the activation of cytoprotective / antiapoptotic mechanisms in response to treatment that may have a general meaning and not just limited to endocrine therapy (66).

Several randomized studies have also demonstrated the role of neoadjuvant endocrine therapy in postmenopausal women with ER+ breast cancer treated with tamoxifen, aromatase inhibitors or their combination. These studies offer the possibility to obtain a signature of predictive genes of response to treatment (67;68) using the pathological response as a surrogate of the long-term benefit.

With a hypothesis-driven approach (bottom-up approach) going to assess in cell lines modulated genes in oncogenic pathways altered in tumors, it was possible to predict the response to treatments that had specific targets (69).

For the hormonal response prediction, in addition to the approach for the development of *a priori* signature, there is the possibility of trying to improve the ER predictability, through the definition of a set of genes that better predict the ER status, compared with the immunohistochemistry determination or the qPCR that are subject to errors and interlaboratory variability (70).

At the moment, few studies have identified predictors of response to chemotherapy. The problem is particularly complex because the treatment response depends both of prognosis and of individual sensitivity and it is therefore important to stratify patients according to the risk. This approach is

possible only by development of risk molecular predictors that allow a uniform stratification of patients and especially layering on a continuous scale.

The data are still very preliminary, but they are encouraging. It is also important emphasize that significant contributions can also arise from the gene profile retrospective study of patients involved in the first chemotherapy historical trials. The limit to study partially overcome treatments is largely offset by the possibility of having available a control arm. Moreover these studies have still technical difficulties related to the poor quality of RNA extracted from archival material of decades ago, but rapid progress in the design of new technologies for the study of the transcriptome will certainly overcome this obstacle.

1.7.2 SUB-TYPES BREAST CANCER IDENTIFIED BY GENE EXPRESSION PROFILING

Perou and Sorlie group, through the analysis of gene expression profiles have identified five molecular subtypes of breast cancer (62;71;72).

These studies represented a milestone in the biology of breast cancer study. Sorlie *et al.* have characterized the variation in gene expression patterns in a set of 65 surgical specimens of breast tumours from 42 different individuals, using complementary DNA microarrays representing 8,102 human genes. These patterns provided a distinctive molecular portrait of each tumour. Twenty of the tumours were sampled twice, before and after a 16-week course of doxorubicin chemotherapy, and two tumours were paired with a lymph node metastasis from the same patient.

Gene expression patterns in two tumour samples from the same individual were almost always more similar to each other than either was to any other sample. By grouping the samples into clusters using those genes which were more variable among samples from different patients compared to replicate samples from the same patient, it was observed that the majority of tumors belonging to the same pair clusterized together on the terminal branches of the dendrogram, pointing out that, despite the potential sources of variability due to different sample preparation and chemotherapy treatment, each tumor was more similar to its paired tumor, compared to all others. This implies that each cancer is unique and has a distinctive pattern of gene expression. The researchers then exploited the advantage linked to the availability of the pairs of tumors to explore the possibility of identifying signature of genes able to discriminate between the different subtypes tumor. A pattern of gene expression, useful for the classification of tumors, it must include genes

expressed in similar samples taken from the same tumor, but expressed in a different way compared to other cancers. The pairs of tumors of the same patients represented a unique opportunity for the systematic search of genes whose expression levels reflected the intrinsic characteristics of these tumors. When variation in expression of this set of genes was used to order the tissue samples, 17 of the 20 'before and after' doxorubicin pairs were grouped together as were both of the tumour/lymph node metastasis pairs.

Sets of co-expressed genes were identified for which variation in messenger RNA levels could be related to specific features of physiological variation. From the genes whose expression was well measured in the 65 tissue samples was selected a subset of 496 genes (termed the 'intrinsic' gene subset) that consisted of genes with significantly greater variation in expression between different tumours than between paired samples from the same tumour.

The tumours could be classified into subtypes distinguished by pervasive differences in their gene expression patterns. The dendrogram appear in two main branches, which are subdivided into smaller clusters, for which they can be identified specific biological characteristics. Branches are colored and each colors represent one of the four subtypes: basal-like, orange; *Erb-B2* +, pink; normal-breast-like, light green; and luminal epithelial/ER+, dark blue and the small black bars beneath the dendrogram identify the 17 pairs that were matched by this hierarchical clustering; larger green bars identify the positions of the three pairs that were not matched by the clustering (71) (Fig.17).

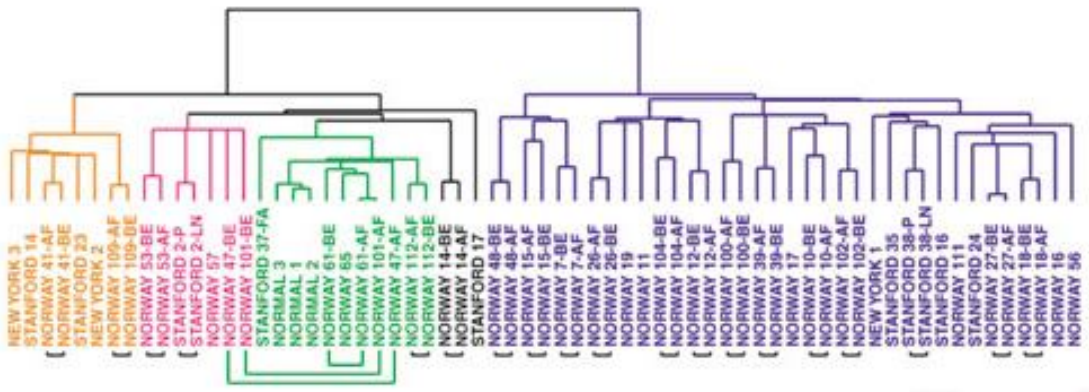
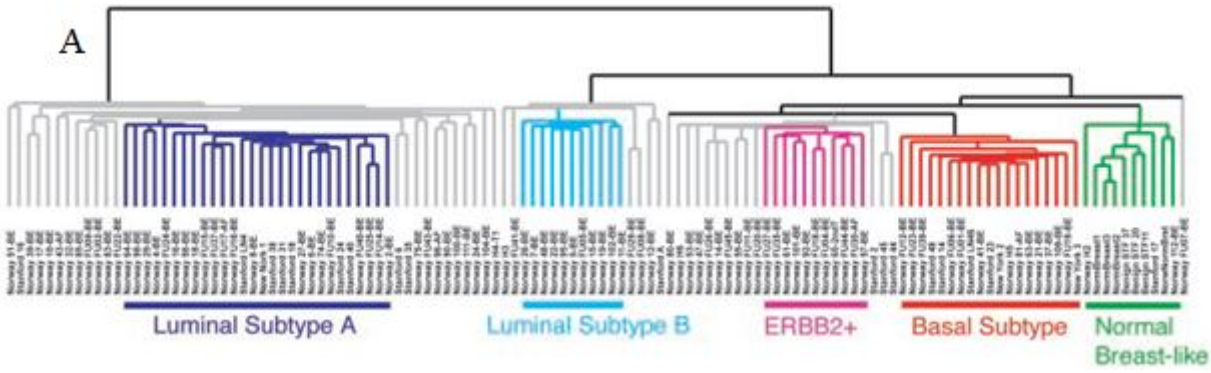


Figure 17: Experimental sample associated cluster dendrogram.

In their subsequent studies Perou and Sorlie, have further refined the division into subtypes and redefined the list of genes intrinsic (62;72).

They produced a new list of 540 genes intrinsic, many of which are in common with the first one. The overall pattern of expression of these genes showed the existence of two major groups of tumors that express or not the genes characteristic of luminal cells, including of course ESR1.

The largest number of analyzed samples also made it possible to divide the group into two subgroups of tumors luminal (Fig.18-A). Among these, the larger group, called Luminal-A, is represented by tumors with high expression of the gene of the estrogen receptor (ESR1), LIV-1 protein, regulated by ESR1, and GATA-binding protein 3 (GATA3) (Fig. 18-F). The second group of cancers, called Luminal-B, show a medium or low expression of specific luminal tumors genes, including the cluster of genes associated with the ER, but is further distinguished from the luminal-A for the high expression of a new set of genes as GGH, LAPTMB4, NSEP1 and CCNE1, but whose function is unknown (Fig.18-C). The subtype to the basal epithelial cells linked is characterized by the high expression of KRT5 and KRT17 and is completely negative for the cluster of luminal/ER genes (Fig. 18-D), while the ERBB2 subtype is characterized by a high expression of many genes contained ERBB2 oncogene, including ERBB2 same, GRB7 and TRAP100 (Fig18-B). Finally has been highlighted a group of tumors normal-like, similar to normal tissue, showing high expression of many genes known to be expressed by adipose tissue and other types of non-epithelial cells (Fig. 18-E). These tumors also showed strong expression of the genes of the basal cells and low expression of luminal genes. However it is not yet clear whether these tumors represent a real molecular subtype or simply be a reflection of sample contamination by normal cells.



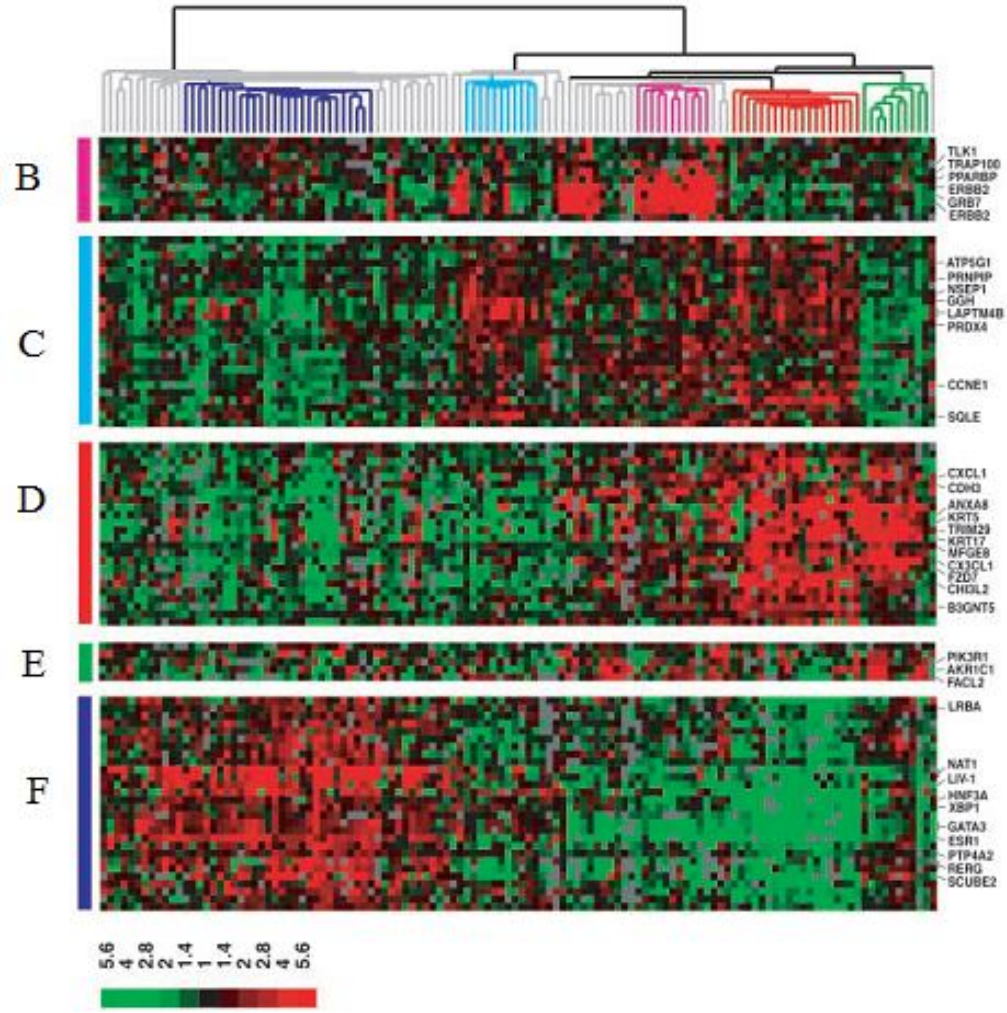


Figure 18: Hierarchical Clustering with the subset of the 540 genes intrinsic. A-Dendrogram of samples. The branches that correspond to samples with a low correlation to each subtype are colored in gray. B-cluster of genes associated with ERBB2. C-cluster associated with the luminal subtype-B. D-cluster associated with the basal epithelial cells. E-cluster of genes important for normal-like tumors. F-cluster of genes associated with the luminal subtype-A, containing the gene for the ER.

To understand whether the five different tumor subtypes identified by hierarchical clustering could represent clinically distinct groups of patients in the same study was performed univariate survival analysis, comparing the subtypes with respect to overall survival and disease-free survival. The Kaplan-Meier analysis, constructed by excluding the group of normal-like, showed a significant difference in overall survival between patients belonging to different subtypes (Fig.19-A). Specifically, the basal-like and ERBB2 subtypes are associated with a lower survival. Over-expression of onco-ERBB2 protein is a prognostic factor recognized, associated with poor survival

in breast cancer and this is also verified for subtype ERBB2 defined in the study. The most interesting result is the significant difference in survival observed between tumors luminal-A and luminal-B. The luminal-B tumors may represent a clinically distinct group with a worse course of the disease, particularly with respect to relapse. Perhaps this subtype may reflect a group of patients who are less likely to benefit from adjuvant therapy with tamoxifen, despite positivity for the estrogen receptor. Further the clinical relevance of the subtypes is highlighted by similar expression of some genes between luminal-B and ER-negative tumors, and basal-like subtypes and ERBB2, suggesting that high levels of expression of these genes may be associated with poor prognosis.

After Perou and Sorlie identification of the five molecular subtypes, it was necessary a validation in independent datasets. The robustness of the tumor subtypes has been tested conducting similar analyzes on published dataset by van't Veer and colleagues (73). The hierarchical clustering was used exactly as described in previous work, to highlight the expression patterns of intrinsic gene in 97 tumor samples. Of the 540 intrinsic genes previously identified, 461 were evaluable in van't Veer dataset also. As expected, the clearer distinction was between tumors with high levels of ER-related genes cluster and tumors that were negative for these genes, which exhibited a gene expression profile characteristic of cancer basal-like and ERBB2. Even the division into luminal-A and B was observed, but was less clear, while the basal-like tumors were reconfirmed as the more homogeneous group, with a high correlation degree between them and a well separated cluster from other subtypes. To test whether the subtypes were associated with significant differences in prognosis, even in this cohort of patients, the investigators performed a Kaplan-Meier univariate analysis with respect to time of development of distant metastases. As shown in Fig.19-B, the probability of remaining disease-free is significantly different between the subtypes: patients with a tumor-luminal A have a longer disease-free interval, while the groups basal-like and ERBB2 show an interval time shorter disease-free.

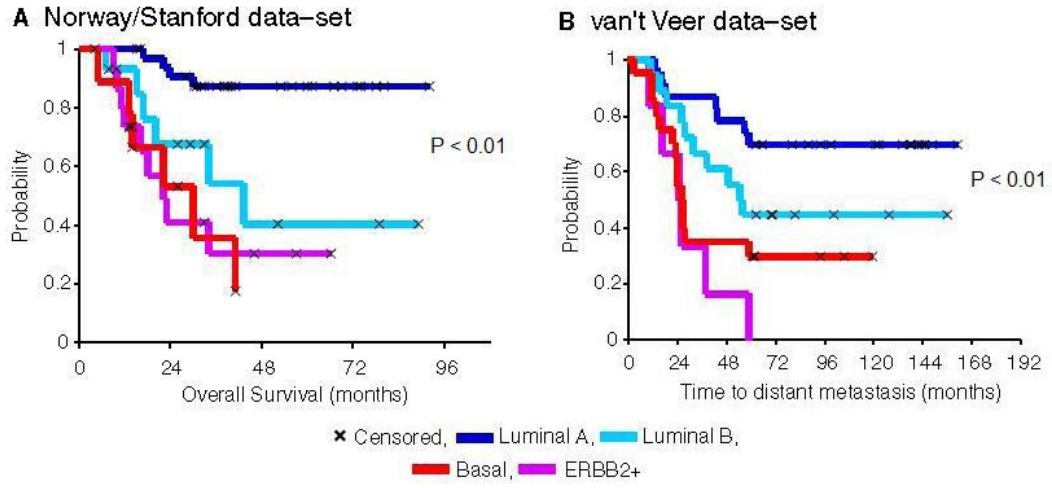


Figure 19: Kaplan–Meier analysis of disease outcome in two patient cohorts. (*A*) Time to development of distant metastasis in the 97 sporadic cases. Patients were stratified according to the subtypes as shown in Fig. 2*B*. (*B*) Overall survival for 72 patients with locally advanced breast cancer in the Norway cohort. The normal-like tumor subgroups were omitted from both data sets in this analysis.

The reproducibility of the subtypes was further confirmed in another study, which derived a third list of 1300 intrinsic genes. Although this list has been shown to be associated with prognosis in a combined dataset of 311 samples from three independent microarray studies. In this study also, it was identified a new subtype can be associated with a high expression of genes regulated by interferon, but it has not been confirmed in other studies molecular (74). To highlight the differences between the different subtypes of breast cancers were used genetic approaches that show how different subtypes are associated with specific chromosomal aberrations (75-77). For example, on chromosome 16q the loss of genetic material seems to be an event associated with low-grade tumors and a good prognosis. Roylance and colleagues used the comparative genomic hybridization (CGH) for comparing the DNA content of infiltrating ductal carcinomas, grade I, predominantly ER-positive, with the DNA content of carcinomas of grade III ER-negative. The authors found that 65% of grade I tumors had lost the long arm of chromosome 16, compared to 16% of grade III tumors, concluding that this chromosomal region contains genes important for tumor progression (75). A second study showed that changes in the number of copies of DNA are associated with tumor grade, ER status and mutations in the TP53 gene and identified several loci associated with poor prognosis.

p53 is a protein that plays an important role in the control of cellular responses to genotoxic damage and regulates the activation of other genes involved in the processes of apoptosis, arrest of cell cycle and DNA repair (78). Previous studies have shown that mutations in the TP53 gene, predict a

poor prognosis and are associated with a poor response to systemic therapies (79;80).To investigate the role of p53 in determining the expression pattern of different subtypes, Sorlie et al performed a screening for mutations in the coding region of TP53 in a series of patients. The frequency distribution of mutation between the different subtypes was significantly different ($p < 0.001$). The group of Luminal-A contained only 16% of mutated tumors, while the luminal-B, ERBB2, and basal-like had 71%, 86% and 75% of tumors mutated TP53, respectively (62). The identification of TP53 mutations in tumors with high levels of ERBB2 gene, supports the idea of a interdependent role between TP53 and ERBB2 (81). To further investigate the TP53 mutations effect on the gene expression profile of tumor subtypes, the researchers looked gene differentially expressed between TP53 mutated and wild-type tumors. 158 genes correlated with the status of TP53 (FDR <1%) was identified (63). As expected, many of the genes highly expressed in cancer basal-like and luminal-B genes are regulated by the cell cycle, such as PLK1, CCNA2, STK6, and MAPK13. GRB7 and STARD3,two genes co-expressed with the oncogene ERBB2, are included in this list but not the gene ERBB2 same. Many genes whose expression is able to distinguish between luminal-A and luminal-B, as GGH, and LAPTMB4 MYBL2, are expressed at high levels in mutated tumors. Among the genes over-expressed in the wild-type tumors are most of the genes belonging to the ER cluster and this summarizes the low rate of mutation present in luminal-A tumors. Further studies are needed to determine which of these genes are targets of TP53 and which are only associated with a particular phenotype expression. Mutations in the BRCA1 gene predispose to the development of early breast cancer. When Sorlie and colleagues added 18 tumors mutated BRCA1 to the samples used in the clustering analysis, have noticed few differences in expression patterns, but all mutated tumors have included in the basal-like cluster (72). This indicates that BRCA1 gene mutation predisposes to basal-like tumors development, which is associated with lack of expression of estrogen receptor and poor prognosis. As also previously reported, the mutated BRCA1 tumors are generally highly proliferative, have mutated TP53 and do not express ER and ERBB2 (82) (83). Recent studies have shown that BRCA1 regulates the estrogen and androgen receptor activity, providing a link between BRCA1 and hormone-dependent tumors (84).

Recently Caldas group presents an integrated genomic/transcriptomic analysis of copy number and gene expression in breast cancers with long-term clinical outcomes composed of a discovery set of 997 primary tumours and a validation set of 995 tumours. Inherited variants (copy number variants and single nucleotide polymorphisms) and acquired somatic copy number aberrations (CNAs) were associated with expression in 40% of genes, with the landscape dominated by *cis* and *trans*-acting

CNAs. By delineating expression outlier genes driven in *cis* by CNAs, they identified putative cancer genes, including deletions in PPP2R2A, MTAPand MAP2K4. Unsupervised analysis of paired DNA–RNA profiles revealed novel subgroups with distinct clinical outcomes, which reproduced in the validation cohort. These include a high-risk, estrogen-receptor-positive 11q13/14 *cis*-acting subgroup and a favorable prognosis subgroup devoid of CNAs. Their results provide a 10 novel molecular stratification of the breast cancer population, derived from the impact of somatic CNAs on the transcriptome (85).

1.7.3 IMMUNOHISTOCHEMISTRY CLASSIFICATION OF LUMINAL AND BASAL-LIKE TUMORS

Gene expression profiling has led a major subdivision of mammary tumors respect to expression of hormone receptors (luminal tumors), with a gene expression pattern similar to that of the luminal cells of the mammary ducts, and tumors that do not express these receptors. This latter category include the HER2/neu subtype, the basal-like (with gene expression pattern similar to basal epithelial cells of stratified epithelia and my-epithelial cells of the mammary ducts) subtype and the normal breast-like subtype, with characteristics similar to normal breast tissue. However, there is a complete overlap between the immunohistochemical and gene expression profiling tumors classification.

The basal-like and the normal breast-like are triple negative tumors (ER-, PgR-, HER2/neu-). The basal-like subtype is well characterized and they represent about 80% of triple negative tumors and 10-15% of all breast cancers. The term "basal" is used to indicate a cell population present in normal breast tissue expressing high molecular weight cytokeratins (CK) as CK5/6 and CK17, which are found at the level of basal cells of various stratified epithelia. The cytokeratins, in particular, are the components of cytoskeletal filaments type of that comprise microfilaments, intermediate filaments and microtubules, and ensure proper organization and structuring of the three-dimensional cytoskeleton inside the epithelial cells (86). Mammary tumors expressing high molecular weight cytokeratins were identified by immunohistochemical analysis, already before microarray techniques, such as poor prognosis tumors compared to those expressing simple cytokeratins.

The genes that encode for two markers of cell proliferation, generally used in immunohistochemistry, Ki-67 and PCNA, can be also used for the tumor subtypes characterization. Basal-like and luminal B subtypes express high levels of these genes associated with proliferation, while luminal-A, normal-like and in part also the ERBB2 tumors, are predominantly negative for the expression of this gene cluster. This may indicate proliferative differences for each tumor subtype, indicating the basal-like subtype as particularly aggressive. The classification at the level of this histological subtype, and is then linked to the lack of hormone receptors, EGFR (epidermal growth factor c-kit, p63, P cadherin, Ki-6, PCNA, CK5/6, CK14, 17 expression (62). As already seen, the luminal tumors, less aggressive compared to basal tumors are characterized by the expression of ER, PgR, BCL2, CK8 and CK18 and low expression of CK5, CK5 / 6, CK17, EGFR and vimentin (87-89). Cells belonging to the class luminal B, show a low expression of ER and PgR and high expression of genes associated with HER2 as ERBB2 and GRB7 that are not found in the luminal A tumors, that are low-grade tumors and more favorable prognosis.

1.8 RNA EXPRESSION ANALYSIS FROM FORMALIN FIXED-PARAFFIN EMBEDDED TISSUES

Reports from different clinical studies have shown that DNA microarray technology allows the identification of different subtypes within apparently homogeneous diseases (71;90;91) and that the identified subtypes often have important clinical implications (62;92). However, the direct application this powerful method to the clinical setting has been limited due to its reliance on fresh tissue. The majority of studies carried out up to date have used high quality RNA obtained from frozen samples and therefore, such studies have been limited by the small number of frozen sample collections linked to clinical data. On the other hand, there is instead a huge resource of FFPE tissues specimens held in histopathology departments around the world. Formalin-fixed, paraffin-embedded (FFPE) tissue is the most common specimen available for molecular assays on tissue after diagnostic histopathological examination. It represents an invaluable source of information on diseases where the patient outcomes are already known. Older archives contain many unique FFPE tissue specimens that would be impossible to replicate today due to changes in medical practice and technology. These samples provide an invaluable resource for studying the molecular basis of disease, making it possible to perform large retrospective studies correlating molecular features with therapeutic response and clinical outcome. However extraction of RNA from these tissues has

proved to be problematic due to the detrimental effects of formalin-fixation. Over recent years, researchers attempting to understand these adverse effects on RNA and identify solutions to remove or diminish it have published many reports on the subject. So far gene expression profiling from FFPE tissues has been problematic, as the retrieval of RNA from FFPE material is challenging (93). Although tissue architecture and proteins are preserved with formalin fixation and paraffin embedding, this method does not preserve nucleic acids very well resulting in RNA that is often significantly degraded (94). Moreover, formalin fixation causes cross-linkage between nucleic acids and proteins (95;96) and covalently modifies RNA by the addition of monomethylol groups to the bases, making subsequent RNA extraction, reverse transcription and quantitation problematic (97;98). Nearly 40% of adenines as opposed to 4% of uracils acquire monomethylol additions following incubation in buffered formalin (99). As a result it has been speculated that the poly A tail of fixed mRNA is heavily modified (100) inhibiting oligo (dT) primer annealing to the polyA tail and consequently the reverse transcription reaction. In addition, the degraded mRNA may not contain a poly A tail for substrate binding by oligo (dT) (101).

Consequently, significant efforts to improve extraction of RNA from formalin-fixed tissue have been made by introducing various modifications to the extraction steps. Masuda et al. showed that solubilization of FFPE tissue was not possible by chaotropic agents such as guanidinium thiocyanate, which is used in the classic method for RNA extraction from fresh tissue. Instead solubilization of FFPE tissue using proteinase K enabled the release of RNA from the cross-linked matrix and resulted in almost the same RNA recovery as from a fresh tissue. However following extraction such RNAs were still a poor substrates for reverse transcription and subsequent PCR was limited to the amplification of small targets. This poor result was attributed to the formalin induced monomethylol additions to the nucleic acid bases preventing successful priming of oligo (dT) for cDNA synthesis. It was found that approximately half of the monomethylol groups could be removed simply by incubating the RNA in formalin-free buffer (Tris, pH 8.5) at 70°C resulting in more ‘free’ RNA that could act as template in reverse transcription or PCR. There has been a great deal of research into modifying methods for successfully extracting usable RNA from FFPE samples (102) and these modifications have been implemented in almost all commercially available FFPE extraction methods today. Furthermore, RNA can degrade in the tissue prior to fixation if conserved at room temperature and many hospitals do not fix the surgically removed tissue immediately. It is therefore important to introduce a change in clinico-pathological routine practice for the development of a standardized, consistent and short time interval between removal of

specimen and fixation. Additionally, degradation of RNA over time continues whilst the RNA is stored in the paraffin block through processes such as oxidation (103;104). New guidelines were developed to protect the RNA by impacting on several parameters. Firstly, the pre-fixation time (time between surgical incision to fixation), which should be kept as short as possible to reduce degradation of RNA (101;105). Secondly, it has been recommended that the fixation time should be kept within 12–48 h as longer fixation times can cause further degradation of the samples (106). Finally it was suggested that formalin should be buffered to a neutral pH as prolonged tissue hypoxia reduces pH in tissues locally, decreasing the yield of nucleic acids. It is clear therefore that the adoption of standard operating procedures, at least within the confines of defined clinical trials and preferably in all cases, may markedly improve the quality of RNA extracted from FFPE tissue.

The major advantage of qRT-PCR is the ability to use very short RNA fragments for amplification, which suggests that this method is well suited to the amplification of highly degraded RNA in FFPE tissue. RNA fragments extracted from FFPE tissue can be as short as 200 bp and it is therefore suggested that amplification of fragments longer than this should not be attempted. Several studies have shown that amplicons of less than »130 bp are optimal and show very high success rates (106;107). Reverse transcription is the most important part of the qRT-PCR process and success at this stage is essential for the generation of cDNA of sufficient yield and quality. Unfortunately, monomethylol groups are never removed completely following RNA extraction and purification, and these groups, which preferentially bind to adenines, can inhibit substrate annealing of oligo (dT) to the poly A tail. It is recommended, that the reverse transcription step is primed either with a gene specific primer or random hexamers to ensure detection of a transcript of interest.

However, despite these optimized procedures RT-PCR from FFPE tissue consistently obtains lower Ct values (»5 Cts lower) than for matched fresh frozen tissues with the same input RNA (105;107;108). Indicating that a large portion of the RNA is still not accessible for cDNA amplification despite improvements to RNA extraction methods from FFPE tissue.

The preferred source of mRNA for microarray profiling has always been snap frozen tissue material (109). And the poor quality associated with RNA from FFPE has often been cited as a primary reason for the continued use of fresh frozen tissues, in the clinical application of microarray technology (110). However in spite of these difficulties, there have been a few recent reports of microarray analysis of RNA extracted from FFPE tissues using T7 in-vitro transcription (IVT) oligo (dT) methodologies which have generated valid, reproducible biological data. FFPE extracted mRNA is more susceptible to cross hybridization than high quality due to its high degradation. In

recent years, there has been a significant drive to develop tools and protocols that enable mRNA profiling from more readily available FFPE tissue. It has been demonstrated that the addition of random primers to the cDNA synthesis reaction will give higher detection rates from FFPE than oligo (dT) priming alone (111). As a result, other commercial amplification kits are becoming available specifically for use with degraded RNA or RNA extracted from FFPE tissues. These kits utilize not only the oligo (dT) primer in the reverse transcription reaction but also random primers to ensure the amplification of RNAs where it has been previously impossible due to the lack of a poly A tail or chemical modifications of this sequence. One of these protocols is the recently released NuGEN WT-Ovation FFPE system, which has been especially developed by NuGEN Technologies Inc for use with RNA extracted from FFPE tissue. This method uses WT-RiboSPIA, which is a whole transcriptome amplification system using an isothermal linear nucleic acid amplification method. This whole transcriptome amplification strategy was previously used to obtain good correlation in gene expression profiles between paraffin-embedded and fresh frozen cell pellets (112). Another technology used in measuring gene expression from FFPE samples is the DASL (cDNA-mediated annealing, selection, extension and ligation) assay developed by Illumina (Illumina Inc., San Diego, CA, USA). The mechanism behind the DASL assay also uses priming with random hexamers in the cDNA synthesis stage but with no oligo-d(T) priming. The assay works from as little as 50 ng of total RNA to analyze 300–400 transcripts. This is a significant increase over what is readily achievable by multiplex RT-PCR, but is still 10–20 fold less than can be achieved by using microarray analysis. An initial study by Bibikova et al. found that 90% of the genes detected in Fresh frozen samples were detected in FFPE tissue; however, the gene expression profiles from FFPE did not correlate exactly with the profiles from fresh samples ($R2 = 0.69$) (113). With the development of high-density microarray platforms that work with FFPE tissue the DASL assay may not be an ideal discovery tool. However, while not directly applicable in the biomarker discovery process, the DASL assay may have a role in the validation of transcripts previously identified by high-density microarray platforms. Almac Diagnostics have adopted a different approach to extract robust data from FFPE derived RNA. Almac have developed a range of high density Disease Specific MicroArrays (DSAs™) manufactured on the gold standard Affymetrix platform, which capture as completely as possible all transcripts transcribed in a specific disease setting such as breast, colorectal or non small cell lung cancer. Technically this is extremely important as most linear amplification technologies include a reverse transcription step using an oligo (dT) primer. This has the effect of enriching for 3’ information, particularly from degraded

RNA typically derived from FFPE. It has previously been suggested that a more 3'biased design may improve the identification of deregulated genes from FFPE samples (112). The combination of extra disease specific content and actual 3'design results in substantially greater detection of robust gene expression data from FFPE samples.

As the reproducibility and reliability of gene expression profiling from FFPE tissue grows, we are likely to see a marked increase in the number mRNA based companion diagnostics in routine clinical use.

It is however important to recall that as most clinical studies with gene expression data have been performed using the widely-used Affymetrix chips, profiling of FFPE samples with such technology has the advantage to allow a direct comparability with previous studies.

2. PURPOSE OF THIS THESIS

The clinical problem

The goal of medicine has been, since its inception, to provide treatment recommendations tailored to the illness of an individual. Some of the earliest medical writings clearly document that different empirical therapies were recommended for different constellations of symptoms. Thus **personalized medicine** is not *per se* a new concept. The history of medicine is intricately intertwined with technologic developments in diagnostic methods that aim to define disease ever more narrowly and to predict with ever increasing precision the clinical outcome with or without particular therapies.

Current medical decision making in oncology takes place in a four-dimensional decision space as physicians and patients need to consider the clinical outcome:

- in the absence of treatment (i.e., prognosis),
- the probability of benefit from therapy;
- the risk of adverse events from an intervention.

An important fourth dimension is patient preference. A person's willingness to accept therapy is influenced by her or his risk tolerance for adverse events from the disease and from the treatment It is assumed that the **more accurate the prognostic and response predictions and toxicity estimates are the more personalized treatment recommendation can be made for an individual** (Fig.20).

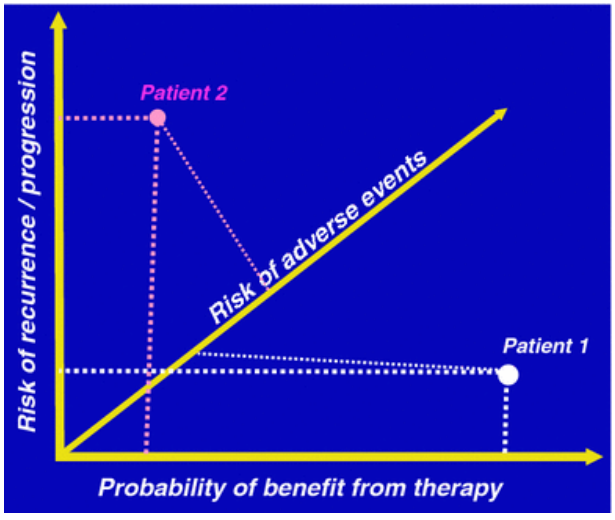


Fig.20: Three-dimensional decision space of predicted prognosis, predicted response to therapy, risk of toxicity, and patient preference (114).

Several diagnostic tools are commonly used in the context of breast cancer to gage patient preference, estimate the risk of recurrence, and estimate the probability of benefit from endocrine or trastuzumab therapy (Table 2). Patient preferences are elicited through the medical interview. The risk of adverse events is estimated, rather subjectively, on age, co-morbidities, and results from simple organ function tests. The risk of recurrence is primarily determined from lymph node status, tumor size, and histological grade. Estrogen receptor (ER) and progesterone receptor (PR) immunohistochemical results are used to define the subset of individuals who may benefit from endocrine therapy, and HER-2 fluorescent in situ hybridization or immunohistochemical analysis results are used to select patients for trastuzumab treatment. Several of these clinical and pathologic features can be combined into the practically useful and validated multivariable outcome prediction model Adjuvant Online (www.adjuvantonline.com). This freely available web-based tool estimates the risk of recurrence (or death) with loco regional therapy alone and with various systemic adjuvant treatments, including endocrine therapy and chemotherapy for stage I–III breast cancer.

Historical tools to aid medical decision making in breast cancer:

Variable	**Method to elicit outcome estimates**
Patient preferences	Medical interview
Risk of adverse events	Medical history Simple organ function tests (BUN, creatinine, liver enzymes, and complete blood count) Electrocardiogram Echocardiogram
Prognosis	TNM stage Histologic grade Adjuvant online software
Probability of benefit from therapy	ER and PR immunohistochemical analysis (for endocrine therapy) HER-2 fluorescent in situ hybridization or immunohistochemical analysis (for trastuzumab therapy)

Table 2: diagnostic breast cancer tools

However, current prediction models are suboptimal, individual predictive variables have limited accuracy, and **the actual clinical outcomes remain heterogeneous in any given prognostic group**. ER status and HER-2 status are helpful in identifying patients who are not eligible for endocrine or trastuzumab therapies by virtue of their high negative predictive values (NPVs) and high sensitivities. However, only a minority of ER-positive or HER-2-positive patients respond to receptor-targeted therapy. The positive predictive values (PPVs) of these tests are <50%.

Currently, there are no accepted molecular predictors of response to various chemotherapeutic drugs. We also have limited ability to predict adverse events despite the relatively high toxicity and modest activity of cytotoxic drugs. These limitations have driven biomarker research to develop more accurate molecular predictors of clinical outcome.

The clinical purpose of our study was therefore to obtain a better stratification of patients based on risk and treatment benefit to avoid unnecessary toxicity to low risk patients and to plan for alternative treatments for high risk patients.

The above stated purpose was pursued by building gene expression-based predictors. Besides the problem of distinguishing pure prognostic from treatment response signatures, we were faced with two completely different situations regarding the type of sample used for building the predictors: fresh frozen samples on one side and formalin-or Bouin-fixed and paraffin embedded samples on the other side (BFPE/FFPE). The type of sample dramatically affects the technical approach and led us to work in parallel in two completely different technical contexts.

The technical challenge

The technical challenge is completely different depending upon the type of sample. In the case of frozen sample it is important to obtain high quality RNA samples and to maintain as much as possible an uniformity within the technical quality of samples. Degraded samples are therefore usually discarded to avoid biases.

On the contrary fixed sample do not allow to obtain good quality RNA and in such situation specific technical protocols to deal with such problem must be developed.

As we were mainly trying to define a context specific signature, we had to use samples derived from clinical trials in a prospective-retrospective design. The majority of samples collected in clinical trials are formalin-fixed and paraffin embedded (FFPE) and represents therefore a challenge for obtaining good quality gene expression data. The technical purpose of this thesis was therefore

to develop a robust pipeline for ‘unlocking’ the biological information contained in FFPE samples. This was addressed through:

Development of specific technical protocols.

Careful sample pre.-analytical assessment.

Development of an *ad hoc* data analysis protocol.

3. MATERIALS AND METHODS

3.1 CLINICAL CASES

The FROZEN clinical case study consisted of 127 women with operable breast cancer who are underwent conservative or radical surgical resection of the primary tumor at the National Cancer Institute of Milan (INTM) from 1983 to 1998. For each patient axillary lymph nodes were analyzed, and only women with a negative nodal status were included in the study. 123 samples were available for follow-up information. 59 samples were obtained from primary tumors of patients who developed metastases within 5 years after surgery and 64 samples from patients that have no metastases for a period of more than 5 years. The two groups were selected in order to have similar distribution with respect to classical prognostic factors such as ER status, PgR, tumor size and age.

3.2 CASISTICA ECTO1

Publicly available Datasets

We used a set of publicly available datasets downloaded from the Gene Expression Omnibus website (http://www.ncbi.nlm.nih.gov/geo/). All datasets contained data from the Affymetrix platform. The GSE identification number and the main features of all datasets are listed in the table below (Table 3).

GSE_ID	NAME	PLATFORM	TISSUE	TOT SAMPLES	SUITABLE SAMPLES
GSE2109	expO	Affymetrix PLUS2	breast cancer	353	301
GSE3744	Richardson	Affymetrix PLUS2	breast cancer	47	40
GSE5460	Lu	Affymetrix PLUS2	breast cancer	127	127
GSE10780	Chen	Affymetrix PLUS2	breast cancer	185	42
GSE12276	Massague	Affymetrix PLUS2	breast cancer	204	204
GSE12763	Hoeflich	Affymetrix PLUS2	breast cancer	30	30
GSE13787	Marty	Affymetrix PLUS2	breast cancer	23	23
GSE16391	Desmedt	Affymetrix PLUS2	breast cancer	55	55
GSE18864	Silver	Affymetrix PLUS2	breast cancer	84	75
GSE19615	Li	Affymetrix PLUS2	breast cancer	115	115
GSE20711	Dedeurwaerder	Affymetrix PLUS2	breast cancer	90	88
GSE21653	Sabatier	Affymetrix PLUS2	breast cancer	266	266
GSE2034	Wang	Affymetrix U133A	breast cancer	286	286
GSE2990	Sotiriou	Affymetrix U133A	breast cancer	189	84
GSE5327	Minn	Affymetrix U133A	breast cancer	58	58
GSE7390	Transbig	Affymetrix U133A	breast cancer	198	198
GSE11121	Mainz	Affymetrix U133A	breast cancer	200	200
GSE9195	Loi	Affymetrix PLUS2	breast cancer	77	77
GSE6532	Loi2	Both	breast cancer	277	277
GSE12093	Zhang	Affymetrix U133A	breast cancer	136	136
GSE17705	Symmans	Affymetrix U133A	breast cancer	298	195
GSE26971	Muller	Affymetrix U133A	breast cancer	277	277
GSE19246	William	Affymetrix PLUS2	DLBCL	2x59	2x56

Table 3: List of Affymetrix dataset

3.3 RNA EXTRACTION FROM FROZEN TISSUE

Fifty to 100 mg of tissue sample were cut from the surgical frozen specimen stored at -80°C, homogenized in Teflon previously cooled in liquid nitrogen. Tissue was pulverized using a Mikrodismembrator (Braun Biotech International, Germany) and then 1 mL Trizol was added.

- Phase separation

The homogenized samples were incubated for 5 minutes at room temperature (RT) to permit the complete dissociation of nucleoprotein complexes and then centrifuged 10 min at 4°C at 12,000 g and the surnatant was transferred in a fresh tube. 0.2 mL of chloroform per 1 mL of TRIZOL® Reagent were added. Tubes were shaked vigorously by hand for 15 seconds and incubated at RT for 2 to 3 minutes. The samples were centrifuged at no more than 12,000 × g for 15 minutes at 4°C.

Following centrifugation, the mixture was separated into a lower red, phenol-chloroform phase, an interphase, and a colorless upper aqueous phase. RNA remained exclusively in the aqueous phase. The volume of the aqueous phase was about 60% of the volume of TRIZOL® Reagent used for homogenization (~600 µL).

- RNA precipitation

The aqueous phase was transfered to a fresh tube. The RNA was precipitated from the aqueous phase by mixing with isopropyl alcohol. 0.5 mL of isopropyl alcohol, per 1 mL of TRIZOL® Reagent used for the initial homogenization, was added. Samples were incubated at RT for 10 minutes and centrifuged at no more than 12,000 × *g* for 10 minutes at 4°C. The RNA precipitate forms a gel-like pellet on the side and bottom of the tube.

- RNA wash

Supernatant was removed and the RNA pellet was washed with 1,5 mL of 75% ethanol. The sample was mixed by vortexing and centrifuged at no more than 7,500 × *g* for 5 minutes at 4°C.

- Redissolving the RNA

At the end of the procedure, the RNA pellet was dried (air-dry or vacuum-dry for 5-10 minutes). RNA was dissolved in RNase-free water by passing the solution a few times through a pipette. RNA was stored at -80°C.

3.4 RNA DNase TREATMENT AND CLEAN UP

To eliminate any DNA contamination in RNA samples the RNeasy kit by Qiagen was used. With RNase-free water the samples were brought to final volume of 87.5 µl. To each sample we added 10 µl Buffer RDD and 2.5 µl DNase I stock solution and incubated a RT for 10 min. Thereafter we proceeded according to the kit protocol with series of washes to clean up the RNA:

1. 350 µl Buffer RLT were added and mixed well.
2. 250 µl ethanol (96–100%) were added to the diluted RNA and mixed well by pipetting
3. The sample (700 µl) was transfered to an RNeasy Mini spin column placed in a 2 mL collection tube and centrifuged for 15 s at 8000 x g. The flow-through was discarded.

4. To wash the spin column membrane, two consecutive washes with 500 µl of buffer RPE were made and centrifuged for 15 s at 8000 x g the first time, and 2 min at 8000 x g for the second one. The flow-through was discard.
30 µl RNase-free water was added directly to the spin column membrane to elute the RNA.
and centrifuged for 1 min at 8000 x g.

3.5 NUCLEIC ACID QUANTIFICATION

To quantify the RNA extract and cDNA single strand product with WT-Ovation ™ FFPE RNA Amplification System V2 (NuGEN) we used Nanodrop ND-2000C (Thermo Scientific). To evaluate the concentration of sRNA we used the reading mode 'ssDNA'. 1µL of the sample was placed directly onto the lower measurement pedestal of the NanoDrop. The upper optical pedestal was lowered and the measurement was taken. After each reading, the sample was wiped away from both the upper and lower pedestals using a clean paper. Each RNA sample was quantified in this way.
In its simplest form the absorbance is measured at 260 and 280 nm. The concentration was expressed as ng/ µL. RNA has its absorption maximum at 260 nm and the ratio of the absorbance at 260 and 280 nm is used to assess the RNA purity of a RNA preparation. Pure RNA has an A260/A280 of 2.1. the ratio between absorbance at 260 and 230 nm is used to evaluate a Phenol contaminations. The ratio should be 1.7-1.9. It is important that the value was higher than 1.2.

3.6 RNA AND cDNA QUALITY

The integrity of RNA samples is essential in the context of gene expression analysis via microarray technology or real-time PCR. Total RNA electrophoretic profile was analyzed by the Agilent RNA 6000 NanoLabChip kit on the Agilent 2100 Bioanalyzer. This analysis quantifies the 28S and 18S rRNA species using capillary electrophoresis and applies the RIN (RNA integrity number) algorithm to assess RNA quality. RIN ranges from 1 (poor) to 10 (best) (Fig.21). We decided to use samples with RIN higher than 6.5.

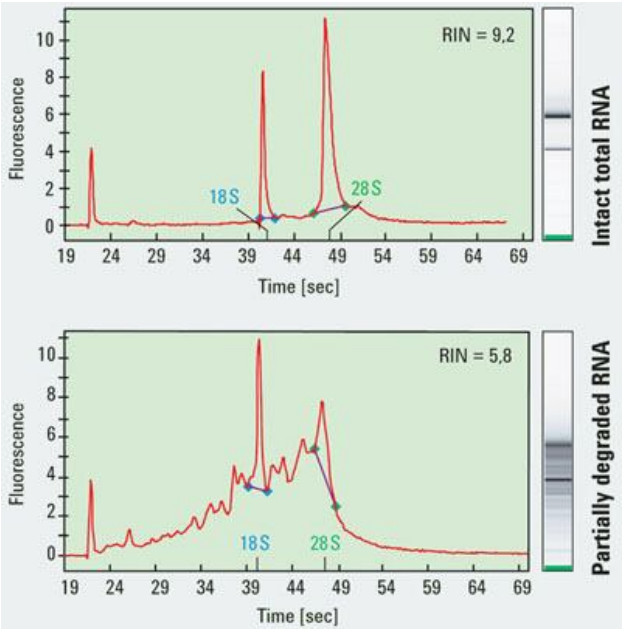


Figure 21: RNA electropherograms: upper figure is a RNA of good quality with high RIN value; bottom figure correspond to degraded RNA with low RIN value.

RNA chip was also used to evaluate cDNA single strand product with WT-Ovation ™ FFPE RNA Amplification System V2 (NuGEN). For this analysis we did not consider the RIN value, but the length of produced amplifications, shown in the electropherogram. If the size of aDNA is higher than 200 nt the amplification product can be used for the hybridization on Affymetrix chip (Fig.22a), if lower the aDNA was considered inadequate for further microarray analysis (Fig.22b).

a b

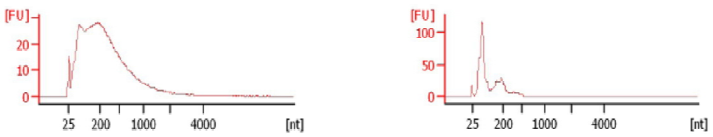


Figure 22: a. elettropherogram of good quality aDNA; b. elettropherogram of degraded aDNA.

The chip was run according to manufacture instructions. In summary

All reagents were taken out of the fridge to warm at room temperature for at least 30 minutes.

- Preparation of Gel matrix

550 µL of gel matrix was added to a spin filter and centrifuged for 10 min at 4000 rpm and 65 µL filtered gel was aliquoted into 0.5mL RNase free microfuge tubes.

- Preparation of Gel-dye matrix

Dye was vortex for 10 seconds and spun down; 1 µL of dye was added to a 65 µL tube of gel matrix and vortexed thoroughly for 10 seconds and then centrifuged at maximum speed (13000g or 14000rpm) for 10 minutes.

- Loading chip with the gel matrix and RNA 6000 Nano Marker:

A new chip was placed on priming station and 9 µL of gel-dye matrix were loaded into the well marked G. The chip priming station was closed and after 30 second the plunger with clip release mechanism was release. 9µL of the gel dye matrix were pipetted in the two additional wells marked G and later 5 µL of marker were pipette into the well marked ladder and each of the 12 RNA wells.

- Loading the RNA sample and the RNA 6000 ladder:

RNA sample and RNA ladder were incubated the in the 70° C heating block for 2 minutes and put on ice for 5 minutes. For the aDNA this step was not necessary because the samples were a single strand.

1 µL of ladder was load into the well marked “ladder” and 1 µL of RNA was load into in the chip sample wells.

The chip was place in the adapted vortexer and mixed for 1 minute at 2400rpm. The loaded Chip was Place in the Bioanalyzer for the run.

3.7 LYMPHOCYTE INFILTRATION SCORE (LI)

Tissue sections stained with haematoxylin and eosin were histologically evaluated by an expert pathologist with respect to the extent of inflammatory infiltrate with a 0 to 3 score (none, mild, moderate or marked).

3.8 IMMUNOHISTOCHEMICAL DETERMINATIONS AND EVALUATION OF IHC

IHC was used to evaluate the expression levels of ISG (MX-1, ISG15, OAS2 and IFIT3) and to study the different sub/population cells infiltrating tumors.

Formalin-fixed, paraffin-embedded whole tumor sections (4µm thick) were dehydrated in xylene, rehydratated and processed for autoclave antigen retrieval at 95 ºC in 10 mM Na-citrate buffer for 6 min. The table reports the characteristics of antibody used. Negative controls were run on each slide by omitting the primary antibody. The Ultravision detection system HPR Polymer (Thermo Scientific, Astmoor, UK) was used for development. Slides were counterstained with haemotoxylin and examined under a light microscope by two independent investigators.

Antibody	Clone	Dilution	Incubation time	Company
CD4	1F6	1:40	1h TA	*Novocastra* Laboratories, Newcastle upon. Tyne, U.K.
CD8	C8/144b	1:20	1h TA	(Dako Corp, Carpinteria, CA
CD68	KP1	1:1000	1h TA	(Dako Corp, Carpinteria, CA
CD56	123C3.D5	1:400	1h TA	Thermo-Scientifics Waltham,USA
CD57	VC1C	1:50	1h TA	Sigma Life Science, St Louis, Mo
CD3	SP7	1:400	1h TA	Thermo-Scientifics, Waltham,USA
FOXP3	259D/C7	1:200	1h TA	BD Pharmingen, NJ, USA
CD45	PD7/26	1:200	1h TA	(Dako Corp, Carpinteria, CA
CD20	L26	1:400	1h TA	(Dako Corp, Carpinteria, CA
GRANZYME	11F1	1:80	1h TA	*Novocastra* Laboratories, Newcastle upon. Tyne, U.K.
HLADR	LN3	1:800	1h TA	Thermo-Scientifics, Waltham,USA
MX1	MPA030917	1:200	1h TA	Sigma Life Science, St Louis, Mo
ISG15	RBO2791	1:100	1h TA	Abgent, San Diego, CA
IFIT3	E-22	1:75	1h TA	Santa Cruz Biotechnology, Santa Cruz, CA
OAS92	H2556/3	1:75	1h TA	R&D SYSTEM, Minneapolis, MN USA

Table 4: Antibodies used for immunohistochemical determinations

3.9 REVERSE TRANSCRIPTION

To synthesize single-stranded cDNA from total RNA obtained from frozen tissue samples, the High-Capacity cDNA Reverse Transcriptions kit (Applied Biosystems) was used. A standard RT

protocol was used; 900 ng of RNA in 7,5 µL RNase –free water were reverse transcribed with oligo (dT) primer in 15 µL of total reaction volume containing:

Component	**Volume/reaction (µL)**
10X RT Buffer	1.5
25X dNTP Mix (100nM)	0.6
10X RT Random Primers	1.5
Multiscribe™ Reverse Transcriptase	0.75
Nuclease-free H2O	3.15
Total per reaction	7.5

Table 5: reaction mix

Reverse transcription was performed using the following thermal cycler program:

	STEP1	**STEP2**	**STEP3**	**STEP4**
Temperature (°C)	25	42	85	4
Time (min)	10	60	5	∞

Table 6: thermal cycler RT profile

3.10 REAL TIME PCR

- In order to investigate the ability of the conditioned medium obtained from each type of co-culture to trigger the interferon response in monoculture of cancer epithelial cells, we assess the expression of some interferon-stimulated genes (ISG), using relative Taqman Gene Expression Assays (Applied Biosystems): *OAS 2* (2'-5'oligoadenylate synthetase 2), *IFIT 3*

(Interferone induced protein with tetratricopeptide repeats 1), *MX1* (gene encoding the mixovirus resistance proteins 1) and also *STAT 1*, an activator of transcription; *ACTB* and *GAPDH* were used as housekeepimg genes (Table 7). All data were referred to co-culture control sample: epithelial cells/epithelial cells. To evaluate the expression of each gene 10 ng of cDNA for sample was used and the analysis was performed in triplicate.

The data were analyzed by SDS 2.4 (Applied Biosystems) using a threshold of 0.2 and referred as relative quantity respect to a calibrator sample using the $2^{-\Delta\Delta Ct}$ method.

GENE	**TaqMan Gene expression Assay (20X)**
OAS 2	Hs00942650_m1
IFIT 3	Hs00155468_m1
MX 1	Hs00182073_m1
STAT 1	Hs01014002_m1
ACTB	Hs03023943_g1
GAPDH	Hs00266705_g1

Table 7: TaqMan Gene expression Assays

- In the pre-analytical sample pre-assessment of the pilot study the expression of the housekeeping gene *RPL13a* was assessed to evaluate the quality of the obtained aDNA, using the relative Taqman assay, which allows the amplification of a short sequence of 105 pb (Hs01926559_g1). 3,3 ng of aDNA were used for each sample and the analysis was performed in triplicate; the experiments were run using 'Absolute Quantification' set-up. The data were analyzed by SDS 2.4 using a threshold of 0.2.

PCR reaction mix component	Volume (µL)
TaqMan Fast Universal PCR master mix (2X), No Amperase UNG	7.50
TaqMan® Gene expression Assay (20X)	0.75
Template cDNA	1.00
H2O RNAsi free	5.75
Total volume for reaction	15

Table 8: Mix of PCR reaction

		Number of cycles: 40	
Temperature (°C)	95	95	60
Time (min)	0:20	0:01	0:20

Table 9: Thermal cycler PCR profile

Real-time PCR was performed using 96-well plates on 7900HT Real-Time PCR System (Applied Biosystems).

3.11 RNA ISOLATION FROM FFPE TISSUES

RNA from FFPE material was extracted using the miRNeasy FFPE kit (Qiagen). A 4 µm pre-cut section was stained with haematoxylin and eosin and reviewed by a pathologist. Only blocks with > 70%tumor cells were used.

From FFPE blocks, 2 sections of 20 µm were cut. For extraction we used manufacture's instruction:

- Deparaffination: 1mL of xylene was added to the section, vortexed for 10 sec and centrifuged at 10,000 rpm for 2 min. 1mL of Ethanol 100% was added to the pellet,

vortexed for 10 sec and centrifuged at 10,000 rpm for 2 min. 150 µL of PKD buffer was added to the pellet.

- Proteinase K digestion: 10 µL of proteinase K were added and incubated at 56°C for 2 h and then 80°Cfor 15 min, then the samples was keep on ice for 3 min. and centrifuged for 20 min at 13,500 rpm.
- DNAse treatment: 16 µL of DNase Booster Buffer and 10 µL of DNase I were added to the pellet and incubate a RT for 15 min.
- RNA isolation: added 320 µL of RBC buffer to adjust binding conditions and than 1120 µL of Ethanol 100%.
- Eluition: total volume was transferred on RNeasy Min Elute column and centrifuged 15 s at 10,000 rpm. Two washes were done: first with 500 µL of RPE and centrifuged 15 s at 10,000 rpm, second with 500 µL of RPE and centrifuged 2min at 10,000 rpm.
  RNA was eluted with 15 µL of RNase-free water with centrifugation of 5 min at 14,000 rpm.

3.12 RNA AMPLIFICATION

The NuGEN Ovation FFPE WTA System is a whole transcriptome amplification method that enables gene expression studies using precious archive samples of small and degraded RNA in FFPE samples

With the NuGEN protocol, RNAs are reverse transcribed to cDNAs, and amplified during the so-called SPIA amplification, a linear isothermal DNA amplification process and the finally amplified products, consisting of single-strand DNA (ssDNA), are biotin labeled and fragmented according to the manufacturer's guidelines.

Description of Ribo-SPIA® Technology

Ribo-SPIA technology is a three-step process that generates amplified cDNA from as little as 50 nanograms of FFPE derived total RNA (Fig.23).

1. Generation of First Strand cDNA

First strand cDNA was prepared from total RNA using a unique first strand DNA/RNA chimeric primer mix and reverse transcriptase (RT). The primers have a DNA portion that hybridizes either

to the 5´ portion of the poly(A) sequence or randomly across the transcript. RT extends the 3´ DNA end of each primer generating first strand cDNA. The resulting cDNA/mRNA hybrid molecule contains a unique RNA sequence at the 5´ end of the cDNA strand.

2. Generation of a DNA/RNA Heteroduplex Double-stranded cDNA

Fragmentation of the mRNA within the cDNA/mRNA complex creates priming sites for DNA polymerase to synthesize a second strand, which includes DNA complementary to the 5´ unique sequence from the first strand chimeric primers. The result is a double-stranded cDNA with a unique DNA/RNA heteroduplex at one end.

3. SPIA® Amplification

SPIA is a robust isothermal strand displacement amplification process developed by NuGEN. It uses a DNA/RNA chimeric SPIA primer, DNA polymerase and RNase H in a homogeneous isothermal assay that provides highly efficient amplification of DNA sequences. RNase H is used to degrade RNA in the DNA/RNA heteroduplex at the 5´ end of the first cDNA strand. This results in the exposure of a DNA sequence that is available for binding the first SPIA primer. DNA polymerase then initiates replication at the 3´ end of the primer, displacing the existing forward strand. The RNA portion at the 5´ end of the newly synthesized strand is again removed by RNase H, exposing part of the unique priming site for initiation of the next round of cDNA synthesis. The process of SPIA DNA/RNA primer binding, DNA replication, strand displacement and RNA cleavage is repeated, resulting in rapid accumulation of SPIA cDNA.

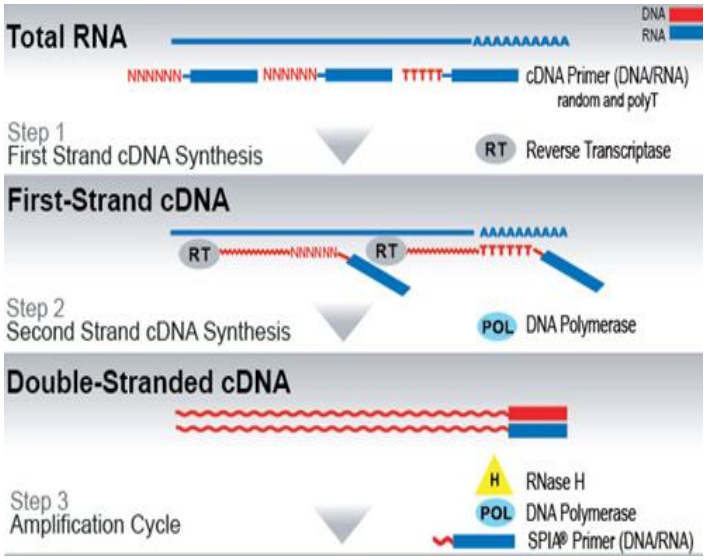


Figure 23: Ribo-SPIA technology

We amplified 100 ng of total RNA in 5 µL of RNA-free water. To the synthesized first strand cDNA 2 µL of the First Strand Enzyme Primer mix were added and incubated for primer annealing in pre-warmed thermal cycler at 65°C for 2 min and then the samples were cooled to 4 °C.
The reaction mix containing (for each sample) 2,5 µL of the First Strand Buffer Mix and 0,5 µL of First Strand Enzyme Mix was added to the samples. The reaction was performed using the following thermal cycler program: 4°C – 2 min, 25°C – 30 min, 42°C – 15 min, 70°C – 15 min, hold at 4°C.
To the synthesized second strand cDNA a mix containing 9,75 µL of Second Strand Buffer Mix and 0.25 µL of Second Strand Enzyme mix was added. The reaction was performed with the following thermal cycler program: 4°C – 1 min, 25°C – 10 min, 50°C – 30 min, 80°C – 20 min, hold at 4°C.

Next step was the purification of cDNA that consist of three steps after the addition of 32 µL of magnetic beads and 10 min of incubation.

1. Binding of cDNA to magnetic bead with 5 min of incubation on the magnet.
2. Separation of cDNA bound to magnetic beads from contaminants, removing and discarding the supernatant.
3. Three washings, each with 200 µL of ethanol 70%.

After air dried the beads on the magnet for a minimum of 20 min, the kit proceeds with Spia Amplification using:

- a master mix 1 with 50 µL of Buffer Mix +20 µL of Spia Primer Mix 1+ 0.7 µL Spia Enhancer +10 µL Spia Enzyme Mix. Reaction was ran in thermo-cycler: 4°C – 1 min, 47°C – 30 min and cool to 4°C .
- a master mix 2 with 30 µL of Buffer Mix +20 µL of Spia Primer Mix 2+ 2.3 µL Spia Enhancer +30 µL Spia Enzyme Mix. Reaction was ran in thermo-cycler: 4°C – 1 min, 47°C – 60 min, heat at 95°C and cool to 4°C.

The beads were separated and discarded from amplified cDNA using the magnet.

3.13 PURIFIATION OF AMPLIFIED CDNA

To purify the cDNA obtained with NuGEN amplification we used the QIAquick PCR purification Kit (Qiagen), with some modifications introduced by the NuGEN protocol.
To 160 µL of amplified cDNA, 800 µL of PB buffer were added. The total volume was loaded onto the column and centrifuged for 1 min at 13,000 rpm. Thereafter the columns were washed 2 times, each with 700 µL of ethanol 80% and centrifuged 1 min at 13,000 rpm.
Elution of purified cDNA was done with 30 µL of nuclease-free water with centrifugation 1 min at 13,000 rpm.

3.14 cDNA FRAGMENTATION AND BIOTIN LABELING

Target preparation for Affymetrix GeneChip arrays was performed using 5 µg in 25µL of amplified cDNA as input to the NuGEN Encore™ Biotin Module kit.
The first step was the cDNA fragmentation: a mix containing 5 µL of Buffer Mix +2 µL of Fragmentation Enzyme Mix was added to the samples.
The second step was the labeling, which was performed by adding a mix containing 15 µL of Labeling Buffer Mix +1,5 µL of labeling reagent + 1,5 µL of Labeling Enzyme Mix and incubation in thermal cycler at 37 °C for 60 min, 70 °C for 10 min and cooling at 4°C.

3.15 HYBRIDATION CHIP

Frozen samples

- RNA frozen samples was processed for microarray hybridisation using the Illumina chips. 800 ng of total RNA were reverse transcribed, labeled with biotin and amplified overnight using the Illumina RNA TotalPrep Amplification kit (Ambion, Austin, TX) according to manufacturer's protocol. One µg of the biotinylated cRNA sample was mixed with the Hyb E1 hybridization buffer containing 37.5% (w/w) formamide and then hybridized at 58 °C overnight to the HumanRef-6_v3 (Illumina, Inc, San Diego, CA) in the case of the *Discovery Dataset* and on the Human-HT 12 v3 in the case of the *Metagene Discovery dataset*. Array chips were washed with manufacturer's E1BC solution, stained with 1 ug/mL Cy3-streptavidine (Amersham Biosciences, Buckinghamshire, UK) and eventually scanned with Illumina BeadArray Reader.
- After purification of the amplified single stranded cDNA obtained from NuGEN Ovation FFPE WTA System, 5 µL of UNG Buffer [10 mM Phosphate buffer ($K_2HPO_4 \cdot 3H_20$, Sigma-ALDRICH, Inc) and 4 mM $MgCl_2$ (USB)] and 5 µL of UNG enzyme (Epicentre Biotechnologies) were added to 25 µL of total volume containing 5 µg of cDNA. The solution was incubated at 50°C for 30 min in a thermal cycler. Following the incubation 5 µL of labeling buffer [0.952 M Acetic Acid (Sigma-ALDRICH, Inc) and 28 mM $MgCl_2$ (USB)] 5 µL of ARP solution [11.2 mg/mL ARP (N-aminooxyacetyl)-N'-(D-biotinoyl) hydrazine, trifluoroacetic acid salt (Molecular Probes) in 22.4 mM Phosphate buffer] were added a the samples and incubated at 50°C for 60 min in a thermal cycler. After purification, 750 ng biotin-labeled sample were used per Illumina Chip array. The hybridization cocktail was prepared according to the Bead Chip manufacturer's instructions with the exception for the hybridation temperature that was reduced to 48°C to accommodate the altered hybridization kinetics of cDNA /DNA pairs relative to cRNA/ DNA pairs.

Fixed samples

FFPE and BFPE samples were profiled comparing different technical approaches in order to define the optimal approach to be applied to the clinical samples.

- DASL

Gene expression profiles of RNA derived from FFPE samples were obtained using Illumina DASL Cancer Panel (502 genes) or HumanRef-8 WG-DASL v3.0 (24000 genes) (Illumina, San Diego, CA, USA) developed specifically to analyze partially degraded RNA. In brief, 200 ng total RNA were converted to cDNA using biotinylated oligo-dT18 and random nonamer primers, followed by immobilization to a streptavidin-coated solid support. The biotinylated cDNAs were then simultaneously annealed to a set of assay-specific oligonucleotides. Extension and ligation of the annealed oligonucleotides generated PCR templates that were amplified using fluorescence-labeled and biotinylated universal primers. The labeled PCR products were then captured on streptavidin paramagnetic beads, washed, and denatured to yield single-stranded fluorescent molecules which were hybridized to the Sentrix Array Matrix (Cancer Panel, Illumina) or the Whole Genome gene expression BeadChips (HumanRef-8, Illumina) for 16 h at 58 °C. Illumina BeadArray Reader was used for scanning the arrays and Illumina BeadScan was used for image acquisition and recovery of primary data.

- AFFYMETRIX

Gene expression profiles of RNA from FFPE samples were obtained using Affymetrix chip HG-U133 2.0 Plus.

A total of 5 µg in 50 µL of ssDNA labeled and fragmented with the NuGEN kit were used for each sample in the hybridization reaction. The hybridization cocktail was prepared according to the manufacturer instructions, using 5 µL of control oligonucleotide B2 (3 nM), 15 µL of 20x eukaryotic hybridization controls (bioB, bioC, bioD, cre), 3 µL herring sperm DNA (10 mg/mL), 3 µL of acetylated BSA (50 mg/mL), 150 µL of 2x hybridization buffer; 30 µL of 100% DMSO and 44 µL of water to a final volume of 250 µL. Hybridization cocktail was denaturated at 99°C for 5 min, then placed at 45°Cfor 5 and finally spun at 14,000 rpm for 5 min before loading them on arrays. A total of 250 µL of the hybridization cocktails were loaded on arrays after a short GeneChip pre-hybridization in 1x hybridization buffer for 10 min in a rotating oven at 45°C and 60 rpm. Hybridization lasted 18 h in a rotating oven. GeneChips were scanned using the GeneChip Scanner 3000 (Affymetrix) using manufacturer's recommendations.

3.16 CELL CULTURES

In our study different epithelial cell lines derived from human breast adenocarcinomas were used; they were all purchased from the American Type Culture Collection (ATCC):

- SkBr3 cells were cultured in McCoy's 5A Medium (Lonza) supplemented with 10% Fetal Bovine Serum (FBS).
- MCF-7 cells and MDA-MB-231 were maintained in Dulbecco's Modified Eagle's Medium , (DMEM-F12) supplemented with 5% FBS.

Human fibroblast cell lines of different origin were also used:

- HTB-125 cell line was purchased from ATCC and derives from human normal breast tissue peripheral to an infiltrating ductal carcinoma. It was maintained in ATCC Hybri-Care Medium supplemented with 10% FBS and EGF (30 ng/mL).
- CCL-171 cell line was purchased from ATCC; it derives from human normal lung tissue and was cultured in Eagle's Minimum Essential Medium (EMEM) supplemented with 10% FBS.
- HNDF cell line was purchased from Lonza and derives from human normal derma; it was cultured in Fibroblast Basal Medium (FBM) supplemented with Fibroblast Growth Medium-2 (FGM-2) Bullet kit (Lonza), which consists of 2% FBS, 0.1% Insulin, 0.1% GA 1000 and 0.1% FGF.
- B-CAF MS132 cell line was isolated in our laboratory from a clinical breast tumor and it is α-SMA positive. It was maintained in FBM supplemented like HNDF cell line.
- CAF- Red line dyed with PLVX-DS RED-EXPRESS2-N1 VECTOR (CLONOTECH, Mountain View, CA, USA). It was maintained in FBM supplemented like HNDF cell line.

All these cell lines were cultured at 37°C in a 5% carbon dioxide and 95% air humidified atmosphere.

For co-culture experiments MIX medium was used: *DMEM F-12+5%FBS* + FBM 2%FBS 1:1

3.17 TRANSWELL CO-CULTURE SYSTEM

For experiments of Transwell co-cultures were used: MCF-7, SkBr3 and MDA MB231 as epithelial cells and HNDF, CCL-171, HTB 125 and B-CAF MS132 as fibroblast cells.

Co-cultures were performed in the 24-well plates with 6.5mm Transwell® with 0.4 µm Pore Polycarbonate Membrane Insert (Costar, Corning Life Science, NY, USA).

On the bottom of wells 50 000 epithelial cells were seeded. On top of Transwell chambers only 10000 fibroblast cells were seeded or 10 000 epithelial cells or 5 000 fibroblast cells +5 000 epithelial cells (Fig.24). The cells were cultivated for 96 h in serum-free MIX medium. Then the conditioned media were collected for cytokines evaluation by Bio- Plex and migration and invasion assays while the epithelial cells in the bottom of the wells were collected for RNA extraction.

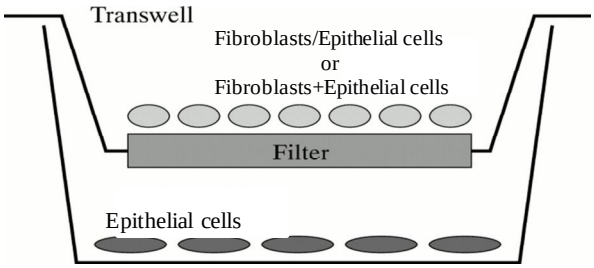


Figure 24: Transwell co-culture model.

3.18 MIGRATION/INVASION CO-CULTURE SYSTEMS

This type of assay was performed with MCF-7 and HNDF.

Only for invasion assay, sterile polycarbonate filters with 8 µm Pore (Costar, Corning Life Science, NY, USA) were coated with 60 µl of Matrigel (200 µg/mL) and the membrane were air-dried at 4°C. The filters were hydrated with 200 µL of serum free media at room temperature for 60 min before use. 100 µL o medium containing 38000 MCF-7 cells were added into top Trans chambers, and 57000 HNDF were cultured in the lower chambers with 0.5 mL serum-free MIX medium (Fig.25). The 24-well plates were incubated in a 5% CO_2 humidified incubator at 37°C for 72h. The conditioned media were collected to evaluate cytokines concentration by E.L.I.S.A. assays. The cells on the upper surface were gently removed with a cotton swab and the filters were fixed with ethanol 100% for 30 minutes at -20°C and then washed with distilled water.

The migrated or invaded cells were stained with sulforhodamine (SRB) for 30 minutes at room temperature. After the incubation time, SRB was removed and each well was washed with acetic acid 1%. The 24-well plates were dried over night.

The cells on the lower surface of the filters were quantified with microscope; each test group was assayed in triplicate.

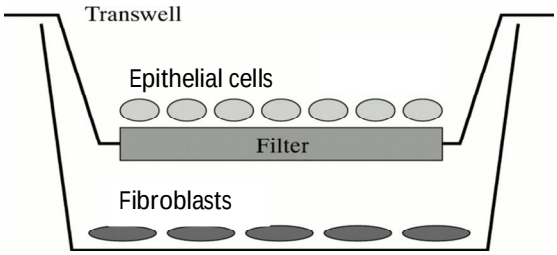


Figure 25: migration/invasion co-culture model.

3.19 3D CO-CULTURED SYSTEM

This assay was performed in 24-well plates, subdivided in ON TOP and EMBEDDED.
Matrigel (BD BIOSCENCES, San Jose, California) was put at 4° C overnight to thaw.
Matrigel was prepared on ice in order to have 7 mg/mL diluted in MIX medium.
The wells were coated with Matrigel: 110 µl for ON TOP and 80 µl for EMBEDDED experiments.
The plate was incubated for one hour at 37°C.
During the incubation, MCF-7 and HNDF were trypsinized from monolayer to a single-cell suspension and counted with coulter counter.
For ON TOP assay:
4 x 10^3 cells in 100 µl of MIX medium were put in the coated wells and incubated for 30 minutes at 37°C. During the incubation, MIX serum medium was cooled at 4°C and 1500 µl of MIX medium plus 10% of Matrigel were prepared. 100 µl of this medium were placed on top of the cells.
For EMBEDDED assay:
4 x 10^3 cells were mixed with 200 µl of Matrigel (7 mg/mL). This solution was put on coated cells and incubated for one hour at 37°C. Then 200 µl of MIX medium were added on top of cells.
Both assays were incubated for 10 days in a 5% CO2 humidified incubator at 37°C. Later the experiment was starved and after 48h the conditioned medium was collected. During the experiment some picture were taken under the microscope.

3.20 MIGRATION ASSAY

Migration assay was performed in the 24-well plates with 6.5mm Transwell® with 8.0µm Pore Polycarbonate Membrane Insert (Costar, Corning Life Science, NY, USA).

- Epithelial cell line SKBR3 and MCF-7 were starved for 96 hours to obtain conditioned Medium (C.M.) in different co-culture conditions (see Transwell co-cultures).
- Under the Transwell insert 1 mL of conditioned medium was added or for the positive control medium supplemented with 10% FBS.
- On the insert 240000 cells were seeded (cells were resuspended in medium without serum).
- Migration assay was incubated in a 5% CO_2 humidified incubator at 37°C for 48h.
- Then the serum-free medium into each well was removed and wells and filters were washed with saline solution.
- The cells on the upper surface were gently removed with a cotton swab.
- Then the filters were fixed with ethanol 100% for 30 minutes at -20°C and then washed with distilled water.
- The migrated cells were stained with sulforhodamine (SRB) for 30 minutes at room temperature.
- After the incubation time, SRB was removed and each well was washed with acetic acid 1%.
- The 24-well plates were dried over night.
- Finally the filter was quantified under a microscope; each test group was assayed in duplicate.

3.21 INVASION ASSAY

Invasion assay was performed in the 24-well plates with 6.5mm Transwell® with 8.0µm Pore Polycarbonate Membrane Insert (Corning Life Science) additioned with Matrigel (BD BIOSCENCES San Jose, California)

- Epithelial cell lines SkBr3 and MCF-7 were starved for 96 hours to obtain conditioned Medium' (C.M.) in different co-culture conditions (see Transwell co-cultures).
- On top of Transwell, the pores were coated with 60 µL of Matrigel (208 µL/mL) and the membrane were air dried a RT for 1h.

- Under the Transwell insert 1 mL of conditioned medium was added or for the positive control medium supplemented with 10% FBS.
- On the insert 480000 cells were seeded (cells were resuspended in medium without serum)
- Invasion assay was incubated in a 5% CO_2 humidified incubator at 37°C for 48h or for 72h.
- Then the serum-free medium into each well was removed and wells and filters were washed with saline solution.
- The cells on the upper surface were gently removed with a cotton swab.
- Then the filters were fixed with ethanol 100% for 30 minutes at -20°C and then washed with distilled water.
- The migrated cells were stained with sulforhodamine (SRB) for 30 minutes at room temperature.
- After the incubation time, SRB was removed and each well was washed with acetic acid 1%.
- The 24-well plates were dried over night.
- Finally the filter was quantified under a microscope; each test group was assayed in duplicate.

3.22 RNA EXTRACTION FROM CULTURED CELLS

- RNA from cultured cells was extracted with Agencourt RNAdvance Cell v2 kit (Beckman Coulter), which is based on Solid Phase Reversible Immobilization (SPRI) paramagnetic bead technology and is known for its ability to consistently deliver pure nucleic acids of the highest quality and yield. This system reliably delivers high RNA recovery and purity without the need for filtration or centrifugation.

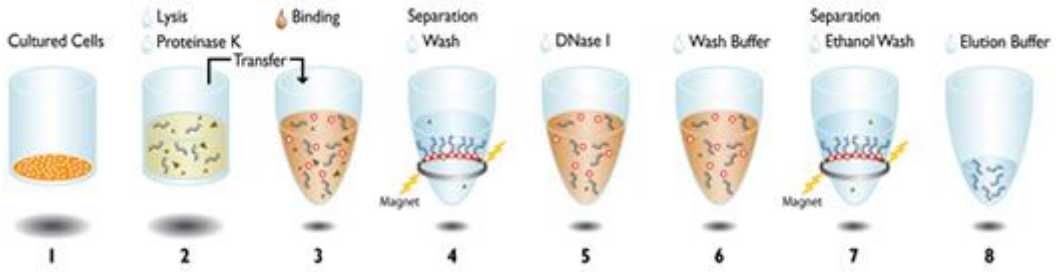


Figure 26: overview of extraction process

- **1.** The culture medium was removed from the cells as completely as possible.

- **2.** Cultured cells were lysed with Lysis Buffer, digested with Proteinase K and transferred into new plate.
- **3.** Total RNA was bound to paramagnetic beads.
- **4.** The beads were washed from contaminants with Wash Buffer and Ethanol.
- **5.** DNase was added to digest genomic DNA.
- **6.** RNA was re-bound to beads with Wash Buffer.
- **7.** The magnetic beads were washed with 70% Ethanol to remove residual contaminants.
- **8.** RNA was eluted in 20µL of RNase-free water from magnetic particles and stored at 80°C.

3.23 BIO-PLEX CYTOKINE ASSAY

Initially, for the first co-culture experiments, to investigate the cytokines in the conditioned medium, we used a multiplex assay system which permits the simultaneous detection of multiple cytokines in a single well of a 96-well microplate.

The Bio- Plex cytokine assay consists of 3 core technologies:

1. Family of fluorescently dye microspheres (beads) to which biomolecules are bound.
2. Flow cytometer with two lasers and associated optics to measure biochemical reactions that occur on the surface of beads.
3. High-speed digital signal processor that efficiently manages the fluorescent output.

Principle:

It is designed in a capture sandwich immunoassay format. An antibody specifically directed against the cytokine of interest is covalently coupled to color-coded 5.5µm polystyrene beads. The antibody coupled beads are allowed to react with a sample containing an unknown amount of cytokine, or with a standard solution containing a known amount of cytokine.

A biotinylated detection antibody specific for a different epitope on the cytokine is added to the beads. The result is the formation of a sandwich of antibodies around cytokine. The reaction mixture is detected by the addition of Strepavidin-phycoerithrin (Strepavidin-PE) which binds to the biotinylated detection antibodies.

The constituents of each well are drawn up into the flow-based Bio-Plex suspension array system, which identifies and quantifies each specific reaction based on the bead color and fluorescence. The magnitude of the reaction is measured using fluorescently labeled reporter molecules associated with each target protein. Unknown cytokine concentrations are automatically calculated by Bio-Plex Manager Software using a standard curve derived from a recombinant cytokine standard.

Procedure:

We evaluated 2 panels of cytokines for a total of 48:

1. Bio-Plex PRO™ Human Cytokine Standard Group I 27-PLEX.
2. Bio-Plex PRO™ Human Cytokine Standard Group II 21-PLEX.

1. To prepare cytokine standard (Multiplex cytokine standard) 1 tube of the lyophilized cytokine standard was reconstituted with 500 µL of MIX MEDIUM (*DMEM F-12+5%FBS* + FBM 2%FBS 1:1). After gently mixing by flicking the bottom of the tube 3-5 times, it was incubated on ice for 30 min. (50,000 pg/mL per cytokine).

2. Cytokine standard curve was prepared in following range 0.2- 3,200 pg/mL by serial dilution in MIX MEDIUM.

3. The vacuum filter was calibrated using 96 well plate and fix the Hg level in between 1 and 2.

4. Working bead solution was prepared mixing 5,760 µL of Bio-plex assay buffer and 240 µL of 25X stock beads in a tube covered with aluminum foil and place in dark.

5. Perform Multiplex Assay procedure:

I. 150 µL of Bio-Plex assay buffer were added to pre-wet wells and the buffer was removed by vacuum filtration;
II. the multiplex bead working solution was vortexed for 15-20 sec at medium speed and 50 µL was pipetted into each well than vaccum filtered;
III. 100 µL of Bio-Plex wash buffer were dispensed to each well and were removed by vacuum filtration;
IV. 50 µL standard or sample were added per well;
V. the filter plate was placed onto a microplate shaker. The shaker speed was regulated to 1,000 rpm and maintained for 30 sec and reduced to 3,00 rpm for 30 min.

6. Using Detection antibody (50X):

I. detection antibody diluent A (2940 µL) was pipetted and place in a tube covered with foil;
II. 60 µL of 50X stock detection antibody were added and placed in dark;
III. 25 µL of detection antibody were added to each well of filter plate (after 3 washes with wash buffer);
IV. filter plate was placed onto a shaker after sealing with tape and followed the same rpm (1000 rpm for 30 sec. and 3,00 rpm for 30 min.).

7. Using Strepavidin-PE preparation:

I. 5940 µL of Bio-Plex assay buffer were pipetted in a tube covered with foil;
II. 60 µL of 100X strepavidin-PE were added into the tube and placed in dark;
III. 50 µL of strepavidin-PE preparation were added to the filter plate after washed 3X with Bio-Plex wash buffer;
IV. filter plate was placed onto a shaker and followed the same rpm and incubated for 10 min at room temperature.
V. The buffer was removed by vacuum filtration and 3 washes with 100 µL of Bio-Plex wash buffer were done. The buffer was removed by vaccum filtration.

VI. The beads were resuspended in each well with 125 µL of Bio-Plex assay buffer. Filter plate was placed onto a shaker at 1000 rpm for 30 sec immediately before to read the plate on Bio-Plex system.

The plate was read with the Bio-Plex Reader setting the machine to ready 100 beads/region in a volume of 50 µL. After running protocol, Bio-Plex manager was gave standard curve and value of each unknown cytokine in samples.

3.24 E.L.I.S.A. TEST

Four cytokines were evaluated in the conditioned medium obtained from co-cultures between MCF-7 with HNDF and MCF-7 with B-CAF Red.

For the IL-6 and the IL-8 we used Quantikine® High Sensitivity ELISA (R&D Systems, Abingdon, UK), and for HGF (Hepatocyte Grow Factor) and SCF (Stem Cell Factor) we used Quantikine® ELISA (R&D Systems, Abingdon, UK).

These kits were performed in according to the manufacturer's instruction. The assay employs the quantitative sandwich enzyme immunoassay technique: a monoclonal antibody specific for each cytokine has been pre-coated onto a microplate.

Standards and samples were pipetted into the wells and cytokines were bounded by the immobilized antibody. After washing away any unbound substance, an enzyme-linked polyclonal antibody specific for those cytokines was added to the wells. Following a wash to remove any un bound antibody-enzyme reagent, a substrate solution was added to the wells. After an incubation period (an amplifier solution was added to the wells, for high sensitivity assay) color developed in proportion to the amount of cytokines in the initial step. The color development was stopped and the intensity of the color was measured with reader plate.

Reagent preparation:

Wash buffer: 100 mL of Wash Buffer Concentrate were diluted into 900 mL of de-ionized water.

Substrate Solution: the lyophilized Substrate was reconstituted with 6 mL of Substrate diluents.

Amplifier Solution: the lyophilized Amplifier was reconstituted with 6 mL of Amplifier Diluents.

Calibrator Diluent: 10 mL of Calibrator Diluent Concentrate were diluted into 10 mL of de-ionized or distilled water.

To quantify the cytokines it was necessary prepare a standard curve.

Concentration of Standard cytokines:

IL-6: 5 - 2.5 – 1.25 – 0.625 – 0.312 – 0.156 (pg/mL) in Calibrator Diluent.

IL-8: 64 – 32 – 16 – 8 – 4 – 2 – 1 (pg/mL) in Calibrator Diluent.

HGF: 8000 – 4000 – 2000 – 1000 – 500 – 250 – 250 (pg/mL) in Calibrator Diluent.

SCF: 1000 – 500 – 250 – 125 – 62.5 – 31.2 (pg/mL) in Calibrator Diluent.

Assay procedure:

all samples and standard were assayed in duplicated.

Specific cytokine diluent was added to the wells:

- assay Diluent RD1W for HGF ELISA.
- Assay Diluent RD1-1 for SCF ELISA.
- Assay Diluent RD1-115 for Il-8 ELISA.
- Assay Diluent RD1-75 for Il-6 ELISA.

100 µl of Standard and samples were added per well and incubated for 2 hour at room temperature.

After the incubation period, liquid from the wells was removed inverting the plate, each well was filled with 400 µl of Wash Buffer. This step was repeated 5 times.

200 µl of Cytokine Conjugates were added to all wells and incubated for 2 hour at room temperature.

Then the aspiration and wash steps were repeated.

50 µl of Substrate Solution were added for Il-6 and Il-8 ELISA and incubated for one hour while 200 µl of Substrate Solution were added for HGF and SCF ELISA and incubated for 30 minutes. It was important protect the plate from light.

Only for High Sensitivity assay (Il-6 and Il-8) 50 µl of Amplifier Solution were added and incubate for 30 minutes.

Finally 50 µl of Stop Solution were added to each well and determined the optical density of each well within 30 minutes, using a microplate reader set to 450 nm for High sensitivity assay and 490 nm for the other assay.

3.25 GREEN FLUORESCENT PROTEIN (GFP) TRANSFECTION PROTOCOL FOR VISUALIZING MCF-7 CELL LINE IN 3-D CO-CULTURE

1. One day before the transfection, MCF-7 were seeded in 3 T25 cell culture flasks in DMEM-F12 with 5% serum so that the cells will be at ~80% of confluence at the time of transfection: two to transfect with the plasmid pEGFP-N1 (Clontech, Palo Alto, CA, USA) and one to treat only with Lipofectamine 2000 reagent (Invitrogen, San Diego, CA, USA), to act as control.
2. 4.0 μg GFP-actin vector were diluted in 500 μL of Opti-MEM I reduced serum medium (Invitrogen) and then mixed gently.

3. 5 μL of Lipofectamine 2000 was diluted in 500 μL of Opti-MEM and mixed gently.

4. The diluted GFP-actin vector was combined with the diluted Lipofectamine 2000 (total volume is 1 mL), mixed and then incubated for 15 minutes at room temperature to allow the GFP-actin vector- Lipofectamine 2000 complexes to form.

5. 1 mL GFP-actin vector- Lipofectamine 2000 complexes was added to each T25 flask containing cells to treat and then incubated for 5 hours at 37°C; after that the medium was removed and replaced with fresh DMEM-F12 with 5% serum.

6. Cells were then cultured for an additional 48 hours before adding 500 μg/mL of Geneticin G-418 Sulphato (Gibco, BRL, Gaithersburg, MD), used for the selection and maintenance of the GFP-expressing clones.

3.26 DATA ANALYSIS AND STATISTICS

3.26.1 ILLUMINA DATA PROCESSING

Microarray raw data where generated using the Illumina BeadStudio 3.8 software and processed using the *lumi* package (115) of Bioconductor. After quality control, the Robust Spline

Normalization was applied and probes with a detection $P < 0.01$ in at least one sample were selected.

3.26.2 METAGENES IDENTIFICATION

After applying the normalization and filtering procedures described above, the most variable probes (IQR>0.5;11334 probes) were selected and subjected to hierarchical clustering using 1-Pearson correlation as distance metrics and an average linkage method. A correlation threshold of r>0.7 was used and only clusters containing at least 7 probes were considered.

Each probe cluster identified in the above described independent dataset was re-clustered in our discovery dataset. Probes correlating 0.7 or less were removed and only clusters with at least 3 probes left after re-clustering were considered. Probe signals were median scaled and for each probes cluster, a metagene was computed as the mean expression of the probes in the cluster.

To define clusters on Affymetrix data, probesets targeting the genes belonging to each cluster of interest were re-clustered on the validation dataset and only those with a correlation > 0.7 were retained and used to compute the metagenes values as described.

3.26.3 MOLECULAR SUBTYPES

Patients were subdivided into three molecular subtypes based on *ESR1* and *ERBB2* expression levels in a way similar to what described in Bianchini et al (116): basal (*ESR1-/ERBB2-*), ERBB2+(*ESR1+/-/ERBB2+*) and luminal (*ESR1+/ERBB2-*). The ILMN_15142 and ILMN_28003 probes for Illumina data and the “205225_at” and “216836_s_at” probesets for Affymetrix data were considered as reporters for respectively, *ESR1* and *ERBB2* gene expression. The threshold values to define gene expression positivity were selected according to the strong bimodal distribution observed.

3.26.4 CLASS COMPARISON

The class comparison analysis was performed using a linear modelling approach and empirical Bayes methods as implemented in the *limma* Bioconductor package (R version 2.14). A $P<0.05$ was considered significant if not otherwise specified.

3.26.5 SURVIVAL ANALYSIS

Univariate and multivariate Cox regression, as implemented in the *survival* Bioconductor package, was used to correlate metagenes or other scores with outcome. Metagenes or score values were also dichotomized to plot Kaplan-Maier survival curves. Distant Metastasis Free Survival (DMFS) was the main endpoint and survival differences were evaluated using the log-rank test.

3.27 ALTERNATIVE CDFS AND NORMALIZATION

Using the *altcdenvsf* Bioconductor package (version 2.16.0, R version 2.14.1) (Gautier et al, 2004; Gentleman et al, 2004), we created an alternative Chip Description File (CDF) where only probes unambiguously mapping RefSeq transcripts (hg19, GRCh37) were retained. For genes with multiple RefSeq transcripts, the longest isoform was considered (hereafter we refer to this alternative CDF as *RefSeq_all*).

The affy Bioconductor package (version 1.32.1) (117) was used for data import and MAS5 and RMA normalization, while for fRMA we used the package of the same name (version 1.6.0) (118). fRMA is based on pre-computed parameters estimated using a large set of CEL files derived from the Gene Expression Omnibus repository (119). As parameters need to be re-estimated when using an alternative CDF, we carried out our own estimation using the same set of 200 batches of 5 samples used by the authors, taking advantage of the functions implemented in the frmaTools Bioconductor package (version 1.6.0) (120).

4. RESULTS AND DISCUSSION

4.1 THE PROGNOSTIC ISSUE

A prognostic marker is defined as a measure able to predict the natural behavior of the disease in the absence of any treatment. Prognosis and prognostic markers are generally evaluated in terms of time to event as final outcome. In breast cancer pure prognostic studies have been run in node negative patients (N-) as they represent low-risk patients where loco-regional treatment alone can be considered curative. Prognostic studies on N- patients were available only prior to the demonstration of survival benefit of adjuvant treatment in high risk patients.

Current routinely used prognostic models for node-negative breast cancer that rely on tumor size and histologic grade are useful but imperfect. At least two distinct gene expression profiling-based tests were recently developed that may complement the prognostic prediction for these patients. One of these, Mammaprint (Agendia Inc., Amsterdam, Netherlands), was recently cleared by the U.S. Food and Drug Administration (FDA) to aid prognostic prediction in node-negative breast cancer and may become available commercially shortly in this country. This assay measures the expression of 70 genes and calculates a prognostic score that can be used to categorize patients into good or poor prognostic risk groups (121). Other investigators also identified genes that were associated with relapse in node-negative breast cancer; markers were selected separately from ER-negative and ER-positive tumors and were combined into a single 76-gene prognostic signature (VDX2; Veridex, LLC, Warren, NJ). This test was also evaluated with two cohorts of patients that received no systemic adjuvant therapy and were not included in the development of the test (122).

Both of these microarray-based assays provide prognostic prediction with moderately high precision and seem to have some complementary value to tumor size- and grade-based predictions. However, what constitutes a low enough risk to forgo systemic adjuvant chemotherapy is influenced not only by the absolute risk of relapse but also by the risk of adverse events, the probability of benefit from therapy, and personal preferences. Many patients seem to be willing to accept adjuvant chemotherapy for rather small gains in survival. Molecular prognostic markers may provide little clinical value for these individuals because no predictive test is accurate enough to completely rule out risk of relapse or some potential benefit from

adjuvant therapy. However, many other patients are reluctant to accept the toxicities, inconvenience, and costs of chemotherapy for a small and uncertain benefit. For these individuals, more precise prediction of risk of recurrence and sensitivity to adjuvant therapy with genomic tests can assist in making a more informed decision.

4.1.1 PREDICTING DISTANT METASTASES IN NODE NEGATIVE PATIENTS

Many hypotheses have been postulated to explain the nature of the metastatic process, but none of them completely accounts for all clinical and biological observations. One important and intriguing question is whether the events that modify genes and consequently cellular processes required for metastasis are already predetermined in the primary tumor or whether a subset of cells gains their metastatic ability in a micro-evolutionary process acquiring additional alterations or through selection of the most aggressive clones (123).

Most gene profiling signatures predicting prognosis in node negative breast cancer link the metastatic risk to ER- and/or high proliferating tumors and do not include, with few exceptions, the genes reported to be associated to metastases in experimental models.

This part of study was designed to identify genes linked to the risk of developing metastases in node negative patients selected to control the effect of classical prognostic factors on the clinical outcome, and all having tumors of the luminal subtype in order to avoid inter-subtype confounding effects. The study included a discovery dataset (profiled at Fondazione IRCCS Istituto Nazionale dei Tumori (INT) of Milano) and a validation dataset obtained from the literature.

The discovery dataset consisted in 123 node negative breast cancer patients developing distant metastases within 5 years from surgery (59 patients) and being free of distant metastases (64 patients) for at least 5 years. In order to control the effect of classical prognostic factors on the clinical outcome, patients were selected to obtain similar distributions across metastatic and metastasis-free group for age (*t*-test, P=0.21, N=123) and tumor size (*t*-test, P=0.64, N=123). In keeping with the criticism associated to identification of prognostic signatures on patients not subdivided by molecular subtype, we ran all subsequent analyses focusing on the subset of 92 patients with luminal (ER+/ERBB2-) tumours. To identify such subset in a possibly homogeneous way, ER and ERBB2 status were defined based on the strong bimodal expression of the corresponding probes (ILMN_15142, ILMN_28003) overcoming this way problems

associated to technical variations in ER and ERBB2 determinations over the years. Again, age (*P*=0.211, *t*-test, N=92) and size (*P*=0.64, *t*-test, N=92) were not significantly different between the two subgroups of patients developing (N=59) or not (N=64) distant metastases. Data are summarized in Table 10.

	All patients			Patients with ER+/ERBB2- tumor		
	Patients developing metastasis (n=59)	Disease free patients (n=64)	*P*	Patients developing metastasis (n=42)	Disease free patients (n=50)	*P*
Age (years)			0.82			0.21
Range	30 - 85	37 - 82		37-78	37-82	
Median	58	53		60	52	
≤ 50	18	27		9	22	
> 50	41	37		33	28	
Size (cm)			0.26			
Range	0.7 - 9	0.6 - 4.5		0.7-3.8	0.6-4.5	0.64
Median	2.2	2		2	2	
≤ 2	27	38		21	33	
> 2	32	26		21	17	
Histotype						
CDI	48	49		33	36	
CDI+CLI	2	5		0	5	
CLI	9	3		9	3	
Other	0	7		0	6	
Follow-up time (month)						
Range	8 - 58*	60 - 185		9-58	71-185	
Median	28	117		28	117	

Table 10. Clinico-pathological characteristics of patients.

All analyses were carried out considering expression of metagenes rather than single genes. The metagene approach is very effective in simplifying biological interpretation of data, reducing multiple testing-associated statistical issues and it allows an easier cross-platform analysis.

Eight-four clusters of highly correlated genes were first identified by hierarchical clustering and 79 were validated on the discovery dataset. For each cluster, a metagene was computed as the mean expression of all probes after median centring of the signals.

By class comparison analysis, only the IFN metagene (metagene #42) was found to be differentially expressed at statistically significant level (*P*=0.029, modified *t*-test, N=92) between primaries from patients with or without metastases and it showed a higher expression level in the

poor prognosis group. Such metagene (Table11) includes several genes, collectively known as interferon-stimulated genes (ISGs), whose products are involved in the host innate immune response to viral infections. They are classically upregulated in response to type I interferons, IFNα and IFNβ and to lesser extent to Type II, IFNγ in response to viral infection (124).

Illumina probe ID	Gene Symbol	Cytoband	Gene Name
ILMN_1797001	*DDX58*	9p21.1a	DEAD (Asp-Glu-Ala-Asp) box polypeptide 58
ILMN_1729749	*HERC5*	4q22.1b	hect domain and RLD 5
ILMN_1654639	*HERC6*	4q22.1b	hect domain and RLD 6, transcript variant 4
ILMN_1835092	*HS.125087*		
ILMN_2058782	*IFI27*	14q32.13a	interferon, alpha-inducible protein 27
ILMN_1760062	*IFI44*	1p31.1e	interferon-induced protein 44
ILMN_1723912	*IFI44L*	1p31.1e	interferon-induced protein 44-like
ILMN_2347798	*IFI6*	1p35.3b	interferon, alpha-inducible protein 6
ILMN_1781373	*IFIH1*	2q24.2d	interferon induced with helicase C domain 1
ILMN_1699331	*IFIT1*	10q23.31b	interferon-induced protein with tetratricopeptide repeats 1
ILMN_1739428	*IFIT2*	10q23.31b	interferon-induced protein with tetratricopeptide repeats 2
ILMN_2239754	*IFIT3*	10q23.31b	interferon-induced protein with tetratricopeptide repeats 3
ILMN_2054019	*ISG15*	1p36.33b	ISG15 ubiquitin-like modifier
ILMN_1662358	*MX1*	21q22.3a	myxovirus (influenza virus) resistance 1, interferon-inducible protein p78 (mouse)
ILMN_2410826	*OAS1*	12q24.13b	2',5'-oligoadenylate synthetase 1, 40/46kDa, transcript variant 3
ILMN_1672606	*OAS1*	12q24.13b	2',5'-oligoadenylate synthetase 1, 40/46kDa, transcript variant 1
ILMN_1736729	*OAS2*	12q24.13b	2'-5'-oligoadenylate synthetase 2, 69/71kDa , transcript variant 2
ILMN_2184262	*OAS3*	12q24.13b	2'-5'-oligoadenylate synthetase 3, 100kDa
ILMN_1681721	*OASL*	12q24.31a	2'-5'-oligoadenylate synthetase-like, transcript variant 1
ILMN_1657871	*RSAD2*	2p25.2a	radical S-adenosyl methionine domain containing 2
ILMN_1814305	*SAMD9*	7q21.2b	sterile alpha motif domain containing 9

Table 11: Annotation of the IFN cluster.

Among those are IFIT (IFN-induced protein with tetratricopeptide repeats) genes encoding *IFIT1*, *IFIT2* and *IFIT3*, recently shown to form complexes able to bind and inhibit specific viral

nucleic acids (125); OAS (2'-5'-oligoadenylate synthetase) genes encoding *OAS1*, *OAS2* and *OAS3*, proteins which cause viral RNA degradation through their ability to synthesize 2',5'-oligoadenylates; the *MX1* (myxovirus resistance 1) gene encoding a protein that physically inhibit the assembly of viral particles; genes encoding members of the IFN-regulated ubiquitin like protein response (ISGylation; *HERC5* and *ISG15*) which bind numerous protein substrates, modulate pleiotropic cellular response and negatively affect viral protein synthesis (126).

The relevance for interferon-regulated genes among primary human breast tumors had already been pointed out in gene expression data published since 1999 by Perou and colleagues (74;127;128) and a subsequent study, which considered the van't Veer dataset, reported overexpression of a group of ISGs by 40% of breast cancer samples (129). Furthermore higher expression of ISGs such as IFI27, IFIT1, IFI35, MX1, GP2 and ISG15 was reported to be predictive of adverse clinical outcome following radiation and chemotherapy (130).

To validate the prognostic role of the IFN metagene in luminal tumors and to investigate its role in the other breast cancer molecular subtypes we used a combined dataset obtained by pooling 3 publicly available untreated node negative breast cancer datasets. This allowed us to separately test 467 luminal, 104 ERBB2+ and 123 basal tumors, characterized by different clinical outcome overcoming the size limits of our discovery dataset concerning non luminal tumors.

Univariate Cox regression analysis for the IFN metagene (Tab.12) was performed separately for each molecular subtype (Fig.30).

IFN cluster	
Gene	***Probesets***
MX1	202411_at
IFI27	203153_at
OAS1	204415_at
IFIT1	205483_s_at
IFI6	213797_at
IFI44L	204439_at
IFIT3	214453_s_at
OAS2	202086_at
ISG15	204747_at
OAS1	202869_at
RSAD2	204972_at
IFI44	205552_s_at
OAS3	218400_at

Table 12: IFN metagene

A slight, but still significant increase in metastatic risk (HR=1.19, 95% CI 1.01-1.41, *P*=0.04) was observed for luminal patients with high IFN metagene expression confirming the results found in the discovery dataset. Interestingly, IFN metagene expression was instead associated to a 1.69-fold reduction of risk (*P*=0.0039) in patients whose tumors were classified as ERBB2+, whereas only a trend towards protection from distant metastases risk was observed in women with basal tumor (HR=0.87; 95% CI 0.68-1.10; *P*=0.2442).

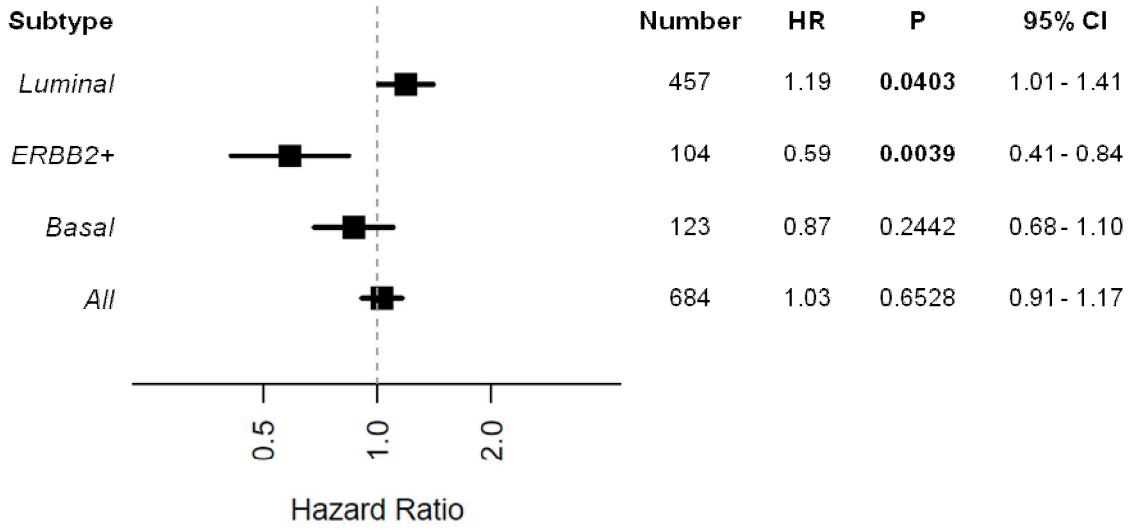


Figure 27: Forest plots showing the HRs of the DMFS univariate Cox regression analyses for the IFN metagene in the different molecular subtypes and in the overall population.

The same type of association between IFN metagene expression and clinical outcome was observed by Kaplan-Meier analyses defining as IFN metagene positive the 30% of all patients with the highest expression values of IFN metagene (Fig. 28). The time function analysis confirmed the irrelevant role of IFN metagene expression on outcome in basal tumors (*P*=0.32, log-rank test N=123), its protective role in patients with ERBB2+ tumors (*P*=0.014, log-rank test, N=104) and its association to a shorter DMFS in women with tumors classified as luminal (*P*=0.012, log-rank test, N=457).