UNIVERSITA' DEGLI STUDI DI TORINO

DIPARTIMENTO DI SCIENZE BIOMEDICHE

DOTTORATO DI RICERCA IN: "Complex systems applied to post-genomic biology"

CICLO: XXI

Genomic signatures of cancer progression: from in vitro experimental models to prediction of human breast cancer outcome.

TESI PRESENTATA DA:
Isella Claudio

TUTOR:
Prof. Medico Enzo

COORDINATORE DEL CICLO
Prof. Bussolino Federico

ANNI ACCADEMICI: 2005-2008

SETTORE SCIENTIFICO-DISCIPLINARE DI AFFERENZA*: BIO/11

Contents

Abstract

Breast cancer accounts for approximately 30% of all cancers diagnosed in women and about 16% of cancer deaths, due to metastatic progression. The treatment consists of tumour excision, possibly associated to radiation therapy and/or adjuvant chemotherapy. The latter strongly impairs the quality of life, and should be given only to patients likely to develop a relapse after surgery. However, current clinical and pathological parameters fail to correctly predict relapse in a significant fraction of the cases and chemotherapy is given to many patients who would not take advantage from it. Over the last few years, DNA microarrays have been employed to measure expression of thousands of genes in cancer samples from patients with different outcome, to construct *genomic classifiers* predicting disease outcome. Such empyrical approach suffered from overfitting, which led to suboptimal results in validation experiments. To overcome this problem, we limited the initial number of analyzed genes to those found regulated in experimental models of oncogenic transformation in-vitro: (i) induction of invasive-growth in mouse embryo liver cells by Hepatocyte Growth Factor; (ii) a microRNA circuit discovered to affect invasive growth; (iii) induction of anchorage-independent growth in human mammary cells by over-expression of the Gab2 gene. In these models, we identified sets of regulated genes and then constructed *in-silico* models of their Expression in breast cancer by extensive meta-analysis of published microarray datasets. All these *in-silico* models were challenged on several independent breast cancer microarray datasets, to classify patients for metastatic recurrence. These data show that genomic analysis of in-vitro models of cancer progression leads to the definition of signatures recapitulating and predicting metastatic progression of breast cancer.

Chapter 1

Introduction

Cancer (medical term: malignant neoplasm) is a class of diseases in which a group of cells display uncontrolled growth (division beyond the normal limits), invasion (intrusion on and destruction of adjacent tissues), and sometimes metastasis (spread to other locations in the body via lymph or blood). These three malignant properties of cancers differentiate them from benign tumors, which are self-limited and do not invade or metastasize.

Cancer may affect people at all ages, even fetuses, but the risk for most varieties increases with age [22] and causes about 13% of all deaths [144]. According to the American Cancer Society, 7.6 million people died from cancer in the world during 2007 [2].

Conventional therapy decision making in cancer therapy is based on histological and few molecular parameters according to which treatment is decided. These guidelines highlight a rudimentary understanding of the molecular origins of cancer and the different intracellular signalling pathways that are perturbed in the various types of cancer. In the past decades several genetic alterations beyond carcinogenesis have been identified, and now it is evident that cancer develops as a result of multiple genetic alterations which may differ from one patient to another. Therefore individuals with the same type of cancer often have dissimilar genetic defects in their tumours. This finding explains why patients who seem to have similar cancers respond

differently to anticancer agents. This heterogeneity is the main obstacle to obstacle to provide treatments effective on all cancers[131].

1.1 Cancer onset and progression as outcomes of multiple dysregulation events

Cancer onset is the process by which normal cells are transformed into cancer cells. Control of cell homeostasis and proliferation is physiologically maintained by cells according to signals derived from the environment and from the cell itself. Cancer progression is an evolutionary process in which cells loose constrain in cellular physiological processes. Hanahan and Weinberg in their epochal turning-point review in 2000 formalized of archetypal cellular traits altered during cancer onset and progression.

- **Self-Sufficiency in Growth Signals** - Normal cells require mitogenic growth signals (GS) before they can move from a quiescent state into an active proliferative state. These signals are transmitted into the cell by transmembrane receptors that bind distinctive classes of signaling molecules: diffusible growth factors, extracellular matrix components, and cell-to-cell adhesion/interaction molecules. While normal cell can not proliferate in the absence of such signals, many of the oncogenes in the cancer catalog act by recovering normal growth signaling.

- **Insensitivity to Antigrowth Signals** - Within a normal tissue, multiple antiproliferative signals operate to maintain cellular quiescence and tissue homeostasis; these signals include both soluble growth inhibitors and immobilized inhibitors embedded in the extracellular matrix and on the surfaces of nearby cells. These growth-inhibitory signals, like their positively acting counterparts, are received by transmembrane cell surface receptors coupled to intracellular signaling circuits.

- **Evading Apoptosis** - The ability of tumor cell populations to expand in number is determined not only by the rate of cell proliferation but also by the rate of cell attrition. Programmed cell death-apoptosis-represents a major source of this attrition. The evidence is mounting, principally from studies in mouse models and cultured cells, as well as from descriptive analyses of biopsied stages in human carcinogenesis, that acquired resistance toward apoptosis is a hallmark of most and perhaps all types of cancer.

- **Limitless Replicative Potential** - Three acquired capabilities-growth signal autonomy, insensitivity to antigrowth signals, and resistance to apoptosis-all lead to an uncoupling of a cell's growth program from signals in its environment. In principle, the resulting deregulated proliferation program should suffice to enable the generation of the vast cell populations that constitute macroscopic tumors. However, research performed over the past 30 years indicates that this acquired disruption of cell-to-cell signaling, on its own, does not ensure expansive tumor growth. Many and perhaps all types of mammalian cells carry an intrinsic, cell-autonomous program that limits their multiplication. This program appears to operate independently of the cell-to-cell signaling pathways described above. It too must be disrupted in order for a clone of cells to expand to a size that constitutes a macroscopic, life-threatening tumor.

- **Sustained Angiogenesis** - The oxygen and nutrients supplied by the vasculature are crucial for cell function and survival, obligating virtually all cells in a tissue to reside within 100 μm of a capillary blood vessel. During organogenesis, this closeness is ensured by coordinated growth of vessels and parenchyma. Once a tissue is formed, the growth of new blood vessels-the process of angiogenesis-is transitory and carefully regulated. Because of this dependence on nearby capillaries, it would seem plausible that proliferating cells within issue would have

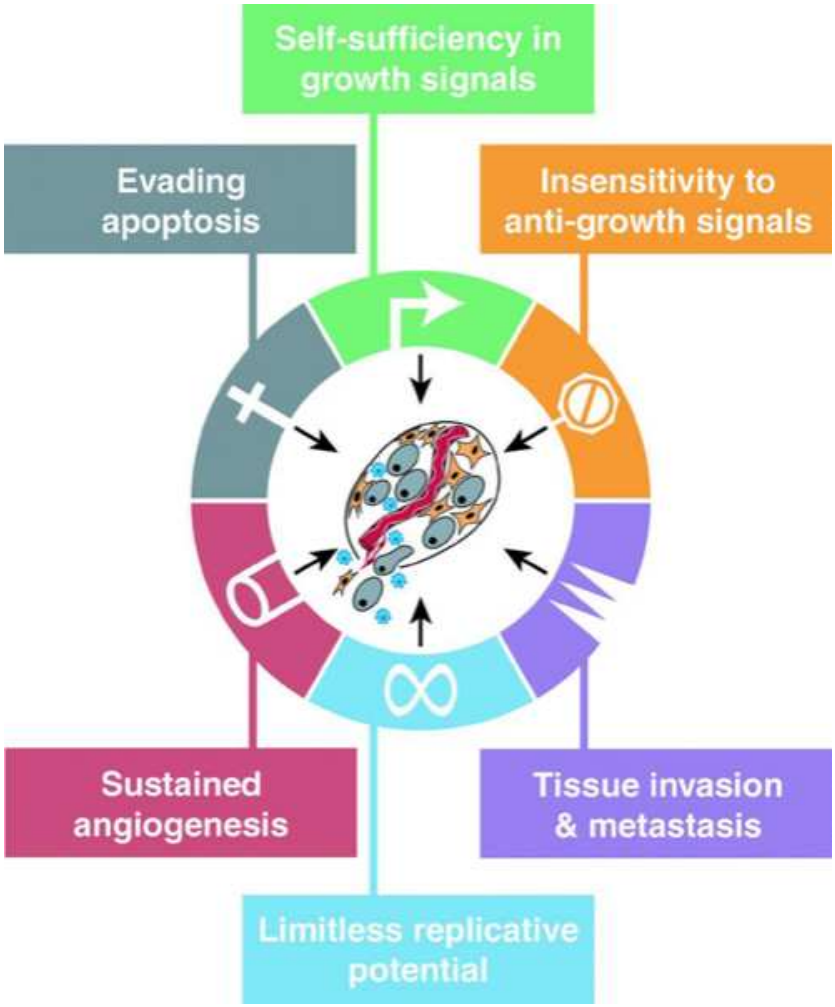
an intrinsic ability to encourage blood vessel growth. But the evidence is otherwise. The cells within aberrant proliferative lesions initially lack angiogenic ability, curtailing their capability for expansion. In order to progress to a larger size, incipient neoplasias must develop angiogenic ability[48].

- **Tissue Invasion and Metastasis** - Sooner or later during the development of most types of human cancer, primary tumor masses spawn pioneer cells that move out, invade adjacent tissues, and thence travel to distant sites where they may succeed in founding new colonies. These distant settlements of tumor cells-metastases-are the cause of 90% of human cancer deaths [115]. The capability for invasion and metastasis enables cancer cells to escape the primary tumor mass and colonize new terrain in the body where, at least initially, nutrients and space are not limiting.

Cells may have different option to acquire the above mentioned abilities; for instance, within a given cancer type, mutation of particular target genes such as RAS or p53 may be found in only a subset of otherwise histologically identical tumors. Furthermore, a specific genetic event may, on its own, contribute only partially to the acquisition of a single capability, while in others, this event may aid in the simultaneous acquisition of several distinct capabilities. Nonetheless, that independently of how the steps in these genetic pathways are arranged, the biological endpoints that are ultimately reach the hallmark capabilities of cancer will prove to be shared in common by all types of tumors.

1.2 Breast Cancer

Breast cancer is a prototypic example of clinical and biological heterogeneity of cancer derived from the same tissue of origin. Indeed, the term refers to a wide spectrum of diseases, with a large range of clinical outcomes. Furthermore, it is the most frequently diagnosed cancer in women in Western



1

Figure 1.1: *Model for cancer progression from Hanahan&Weinberg. Albeit with different mechanism most cancer acquire the same functionalities*

countries and accounting for approximately 30% of all cancers diagnosed and about 16% of all cancer deaths [45]. For this reason it is a major matter of interest for clinical investigations aimed at the dissection of biology underneath the disease to provide better treatments.

Breast cancer has been traditionally classified by histological type[18, 106, 109].

Whilst most of the concept regarding the morphologically defined clinical-

pathological parameters remains valid, immuno-histochemestry and molecular genetics have changed breast cancer classification perspective[18, 71, 106, 109]; nowadays breast cancers are described along four different classification schemes, or groups, each based on different criteria and serving a different purpose:

- **Pathology** - A pathologist will categorize each tumor based on its histological (microscopic anatomy) appearance and other criteria. The most common pathologic types of breast cancer are invasive ductal carcinoma, malignant cancer in the breast's ducts, and invasive lobular carcinoma, malignant cancer in the breast's lobules.

- **Grade of tumor** - The histological grade of a tumor is determined by a pathologist under a microscope. A well-differentiated (low grade) tumor resembles normal tissue. A poorly differentiated (high grade) tumor is composed of disorganized cells and, therefore, does not look like normal tissue. Moderately differentiated (intermediate grade) tumors are somewhere in between.

- **Protein & gene expression status** - Currently, all breast cancers should be tested for expression, or detectable effect, of the estrogen receptor (ER), progesterone receptor (PR) and HER2/neu proteins. These tests are usually done by immuno-histochemistry and are presented in a pathologist's report. The profile of expression of a given tumor helps predict its prognosis, or outlook, and helps an oncologist choose the most appropriate treatment.

- **Stage of a tumor** - The currently accepted staging scheme for breast cancer is the TNM classification.

Prognostic Indicators

Large meta-analyses have shown that recurrence is likely in only 20-30% of young women with early-stage (lymph-node-negative) breast cancer who

should undergo only surgery and localized radiation treatment; nevertheless, in the United States, 85-95% of women with this type of cancer receive adjuvant chemotherapy. Thus, 55-75% of women with early-stage breast cancer in the United States undergo a toxic therapy from which they will not benefit but will experience the side effects.

To predict disease outcome, various clinical and pathological parameters are considered, such as age, menopausal status, tumour size, lymphovascular invasion, histological grade, Estrogen receptor (ER) and ERBB2 receptor. There are five tumor classification values (Tis, T1, T2, T3 or T4) which depend on the presence or absence of invasive cancer, the dimensions of the invasive cancer, and the presence or absence of invasion outside of the breast (e.g. to the skin of the breast, to the muscle or to the rib cage underneath). Combination of these parameters are employed in different prognostic indicator such as the Nottingham Prognostic Index (NPI) and Adjuvant!Online, or included in algorithms used for the development of guidelines for treatment decision-making, such as those proposed by the St. Gallen consensus expert panel [124, 125, 86, 103].

Unfortunatelly, these prognostic indicators fail to predict individual patient outcomes as patients with the same clinical-pathological parameters may have different outcome.

As a consequence, in current clinical practice, the majority of patients with early breast cancer receive some form of systemic adjuvant therapy (chemotherapy and/or endocrine therapy), which may have important side effects and which puts considerable burden on health care costs. Thereby it is evident that current prognostic indicators inferred from clinical-pathological parameters fail to recapitulate the complexity of breast cancer disease;

Consequently it is top priority goal to gain new insight in the biology of breast cancer to develop tools usefull in medical decision making. Thanks to the full sequencing of human genome and the arise of OMICs era, new holistic technologies were developed in order to monitor large-scale molecular-level measurements to have an high resolution picture of the molecular machinery

in cancer progression.

1.3 OMICS

The explosion of high-throughput technologies available for generating large-scale molecular-level measurements in human populations has led to an increased interest in the discovery and validation of molecular new biomarkers in medical research. Uses of biomarkers in medical decision making is quite varied and includes such key features as surrogate endpoints, proxies for exposure [114], early detection of disease [43], and identification of predictive and prognostic factors in disease management [59].

A biomarker is formally defined as "a biological characteristic that is objectively measured and evaluated as an indicator of normal biologic processes, pathogenic processes, or pharmacologic responses to a therapeutic intervention" [14]. In translating this definition into the context of "omics" data (e.g., transcriptomic, proteomic, genomic) it is difficult to identify what is meant by a "biological characteristic" Often when omics data are evaluated for features associated with the medical condition of interest multiple molecular features emerge. Combined, these features may have biomarker potential and thus the biological characteristic of interest is in fact a set of features. This subtle change from considering a single molecular biomarker to considering a biomarker profile has motivated discussion on their proper reporting [88] and governmental regulation of clinical use [44]. Ultimately, a biomarker profile should undergo the same scrutiny required of single molecular markers.

In the following we explore current trends in biomarker research in the context of omics data. Examples will focus primarily on cancer, given our expertise, though we acknowledge that omics-based biomarker research is utilized in other areas[79].

1.3.1 DNA microarray and tumor gene expression profiles.

A DNA microarray is a high-throughput technology consisting of an arrayed series of thousands of microscopic spots of DNA oligonucleotides, called fea-

tures, each containing picomoles of a specific DNA sequence. This can be a short section of a gene or other DNA element that are used as probes to hybridize a cDNA or cRNA sample (called target) under high-stringency conditions. Probe-target hybridization is usually detected and quantified by fluorescence-based detection of fluorophore-labeled targets to determine relative abundance of nucleic acid sequences in the target[17].

DNA microarray are typically used in molecular biology and in medicine to monitor levels of thousands of genes simultaneously and study the effects of certain treatments, diseases, and developmental stages on gene expression. For example, microarray-based gene expression profiling can be used to identify genes whose expression is changed in response to pathogens or other organisms by comparing gene expression in infected to that in uninfected cells or tissues[122].

The initial enthusiasm for the application of microarray technology was tempered by the publication of several studies reporting contradictory results on the analysis of the same RNA samples hybridized on different microarray platforms[87]. Scepticism arose regarding the reliability and the reproducibility of this technique. In reality, many of these apparently divergent results reflect the complex nature of the data generated by high-throughput technologies and the analytical methods used, without necessarily meaning that these techniques are unreliable or inferior. Indeed, most of the discrepancies were attributed to inconsistent sequence fidelity and annotation, low specificity of the spotted cDNA microarrays, lack of probe specificity for different isoforms, or differences in the hybridization conditions, fluorescence measurement, normalization strategies and analytical algorithms applied[7, 12, 82, 57, 127]

1.3.2 Genomic signatures for breast cancer

DNA microarray-based technology has provided researchers an ideal opportunity to begin taking steps towards performing comprehensive molecular and genetic profiling of breast cancer. The simultaneous study of thousands of genes, rather than focusing on just a few with traditional methods, made

the microarray technology into a powerful, holistic analytical tool. Basic research and clinical investigation rapidly took advantage of the great potential of this technology, either to gain new insights into cell biology and for developing a biologically sounding and more clinically useful classification. Indeed, form genomic analysis it emerged a new category of biomarker the so called "Genomic signature". "Genomic signature" are typically composed of specific combination of genes which acquire significance when analyzed together. A first example of genomic classifier was proposed by van't Veer et al. in 2002[132], laster followed by several other classifiers

70 gene signature

In their first work van't Veer et colleague collected 78 lynophonode negative (N −) breast cancer patients with no adjuvant treatment and younger than 55 years of age, at the Netherlands Cancer Institute in Amsterdam and identified a 70-gene prognostic signature using the Agilent Rosetta platform[132]. This signature was then validated on a larger set of 295 young N − and N+) breast cancer patients from the same institution[133], and it is now FDA approved for clinical practices as a clinical kit "MammaPrint".

76 gene signature

Later on in 2005, a similar experiment design was applied by Wang et colleague [141], in collaboration with Veridex LLC (San Diego, USA) exploiting Affymetrix technology. They reported a 76-gene expression signature with discriminative power comparable to that of the Amsterdam signature in predicting the development of distant metastases in untreated patients of all age groups with node-negative breast cancer. The two main differences between the MammaPrint and the Rotterdam studies were the microarray platform and the study design applied for the development of the classifiers. Specifically, the Rotterdam group used Affymetrix technology and took into consideration two separate subgroups based on ER-status in order to build their predictor, with ER-status being determined by immuno-histochemistry.

Just like for the MammaPrint, they were able to validate this signature in 180 node-negative untreated patients derived from multiple institutions.

Genomic Grade Index

Another example of genomic classifier is the Genomic Grade Index, designed following hypothesis-driven approach was the study reported by Sotirou et col [121]. They focused on histological grade, a well-established pathological parameter rooted in the cell biology of breast cancer. Indeed, clinicians face a huge problem with respect to patients who have intermediate-grade tumors (grade 2), as these tumors, which represent 30% to 60% of cases, are the major source of inter-observer discrepancy and can display intermediate phenotype and survival, making treatment decisions for these patients a great challenge, with subsequent under- or over- treatment. To solve this ambiguity it was developed a Gene expression Grade Index (GGI) score based on 97 genes. Interestingly, the GGI was able to reclassify patients with histological grade 2 tumors into two groups with distinct clinical outcomes, similar to those with histological grade 1 and 3 tumors, respectively.

Basal-Like Signature

Sorlie et. al using an unsupervised approach, identified five clinically relevant subtypes of breast tumors[123], which have been further validated in independent data sets. Of these, the two main subtypes are associated with the most significant difference in clinical outcome: Patients with luminal A type tumors are facing a relatively good prognosis, whereas patients with basal-like tumors experience a much shorter overall-and disease-free survival period. Sorlie et al defined a prognositic classifier of 459 cDNA that distinguishes specifically these two groups.

1.4 In-vitro model for cancer progression

Most of the classifiers above mentioned, with the notable exception of GGI, are constructed according to a "top-down" experimental design which generates gene expression patterns associated with clinical outcome, without any a priori biological assumptions. This strategy is prone to over-fitting of the results, and the performance of the classifiers are sub-optimally validated on independent datasets. Conversely we exploited in house developed *in-vitro* model for cancer progression to construct classifiers, following a "bottom-up" strategy, in which genes previously associated with a specific phenotype, are then associated with cancer clinical outcome.

We developed 3 models.

- **HGF-driven invasive growth**
- **microRNA-24 circuit regulating Invasive growth**
- **Gab2-driven ancorage independence**

1.4.1 Invasive Growth: a complex biological program orchestrated by HGF

Invasive growth is a tightly orchestrated genetic program instructing the cells to dissociate from their neighbors, migrate through the extracellular matrix, colonize new sites, proliferate and differentiate[51], typically in in progenitor cells Developmental processes during which this program takes place include epithelial morphogenesis and angiogenesis. After embryonic development invasive growth is activated during acute injury repair, where it regulates inflammatory responses and wound healing[25].

When aberrantly activated invasive growth is involved in cellular mechanisms that cause local invasion and metastasis, having a central role in tumour progression[25].

This process has been well documented on epithelial cells and for this reason the first stages of invasive growth program are also defined epitelial-

mesenchimal transition (EMT). EMT includes release of junctions defining the tissue monolayer, change of polarity, amoeboid motility in extracellular matrix causing epithelial cells to acquire invasive mesenchymal phenotype and gene expression patterns. EMT and consequently invasive growth program are induced both physiologically and pathologically by several extracellular activators as cytokines and soluble growth factors[26]; among them, hepatocyte growth factor (HGF) and its receptor MET represents the prototype ligand-receptor couple and are the major players of a family that also includes MSP, RON, Semaphorins and Plexins[25]. Hepatocite growth factor (HGF), also known as Scatter Factor (SF), is a pleyotropic cytokine belonging to the plasminogen family. Secreted as inactive precursor it is activated locally upon proteolytic cleavage by the urokinase-type Plasminogen Activator[147]. SF was identified for its ability to induce scatter in epithelial cells[126]; while, independently, it was identified as a potent in vitro growth stimulator for primary hepatocites[94]. HGF elicits the invasive growth response binding its specific tyrosine kinase transmembrane receptor Met. Met receptor was firstly identified as the protein product of TPR-MET, a transforming gene derived from a chromosomal rearrangement in an osteosarcoma cell line after chemical carcinogen treatment[27]. After this first link with cancer the aberrant alteration of catalytic activity of Met has been demonstrated in several human tumours tissues [36, 37, 38, 39] and in mouse models[73, 84]. Finally Met gene was classified as a proto-oncogene [110] that when aberrantly activated contributes not only with the onset but also with the maintenance of malignancy leading to metastasis formation[72]. The invasive growth response to HGF requires days for fulfilment and involves transcriptional regulation of key effectors. Indeed, the Met tyrosine kinase receptor concomitantly regulates multiple signal transduction pathways controlling the cell transcriptional status[89]. Met activity impinges on the activation of several transducers such as the cytosolic tyrosine kinase SRC, the lipid kinase phosphatidynolinositol 3-kinase (PI3K) the GRB2 adaptor protein [100], the STAT3 transcription factor[15], SHC[97]

and GAB1[142]. These adaptors are responsible of Met's specific biological

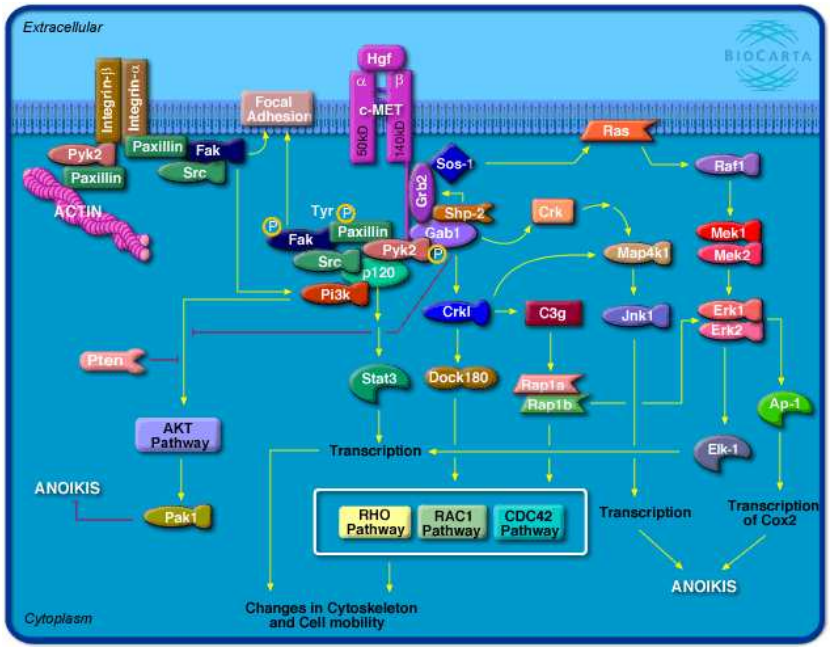


Figure 1.2: *HGF/Met Pathway*

activities resulting as a whole in the invasive growth behaviour that in normal conditions tightly depends on HGF availability. Many of these transducers play also important roles in the transduction of signals mediated by other couple growth factors-tyrosine kinase receptor (RTK) couples that trigger or not toward invasive growth program activation depending on tissue specific backgrounds. For this reason, despite of the growing knowledge about Met's transducers we continue to lack the comprehensive view of invasive growth program background needed to sustain the receptor induced response determining the final effects specific of each cell type. This insight can be achieved through widespread -omic studies and high-throughput functional genomic approaches.

Identification by expression profiling of HGF-regulated genes involved in Invasive Growth

The complexity of invasive growth program needs to be accuratelly described in by monitoring simultaneously multiple genes with high-throughput technology such as microarray platforms. As described before, HGF can orchestrate through Met activation in *in vitro* the whole spectrum of invasive growth-related processes including cell motility and scattering, proliferation, branching morphogenesis in collagen gels, matrix invasion and survival in other tissues [89].

To disect the genetic program involved in the invassive growth, MLP-29 were treated for 1, 6 and 24 hours with HGF and EGF, a ligand for EGFR, another strong RTK involved in cancer progression; the RNA was extracted and analyzed on two independent gene expression technologies. This experiment led to the identification the invasive growth signature shown in figure.

The signature results enriched in genes specifically regulating cell growth and cell adhesion which are the main biological functions lying behind invasive growth. To better characterize the transcriptional modules induced by the two growth factors it was also performed a K-means clustering using cosine correlation. This analysis highlighted the high overlap of transcriptional responses of HGF and EGF which, although driving different phenotype in this cell-line, impinge on the same transcriptional targets. Clustering evidences also minor temporal or quantitative shifts between the induced responses in almost every sub-groups, but these differences are insufficient to explain how MET activation leads to the full invasive growth program stimulation. Despite this overlap, the study of genes having higher differential expression between HGF and untreated control has permitted the discovery of genes essential to produce the complete invasive growth program as Ostepontin and Arhgap12[51, 89].

Furthermore the translation of of expression of the 1132-gene signature in published microarray datasets of hepatocellular several human cancer

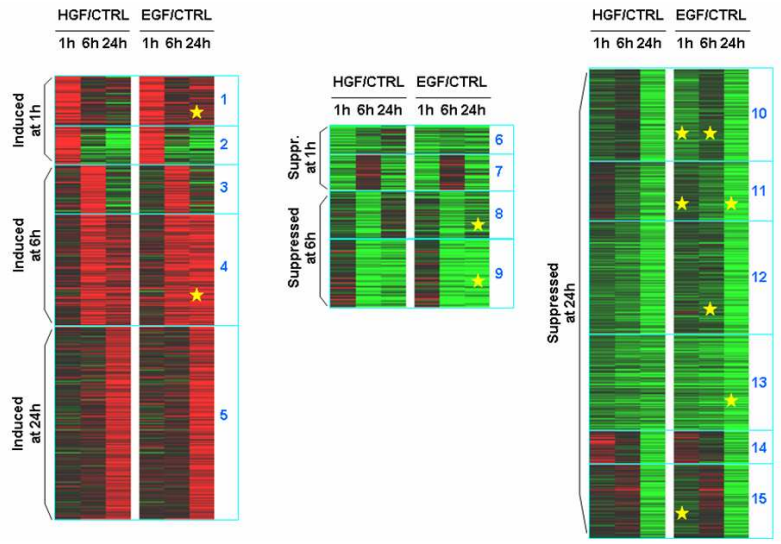


Figure 1.3: *K-means clustering and functional analysis of HGF- and EGF-regulated genes. Genes are grouped into 15 clusters, ordered by the timing of the response (early, delayed, late) and its sign (induction, suppression), as indicated. Positive log2ratio between stimulation and control is shown in red, negative ratio in green, as indicated by the calibration bar.*

[13, 132] found the invasive growth signature being correlated to cancer aggressiveness. The invasive growth signature was then explored in breast cancer and after mapping on NKI dataset obtained by merging two works from the Netherlands Cancer Institute (NKI)[132, 133].

To assess the signature ability in discriminate patients with poor and good prognosis, it was applied the NMC[143]. The optimal number of 27 discriminating genes was then defined by assessing the NMC performance using increasing number of genes taken from the ranked list.

Functional analysis of the 27-gene Invasive Growth signature (IG-27) was carried out by extensive database and literature screening, and highlighted a striking enrichment for two functional clusters (Fig1.4): (i) 7 genes upregulated in poor prognosis samples are involved in positive regulation of cell pro-

liferation; (ii) 6 genes downregulated in poor prognosis samples are involved in the control of ”inbound cell communication”, a super-cluster including functions like ”processing of extracellular ligands”, ”signal transduction”, and ”regulation of transcription”. The picture emerging is that poor prognosis is associated to a higher rate of cell proliferation and a ”desensitization” of the cancer cells to exogenous stimuli.

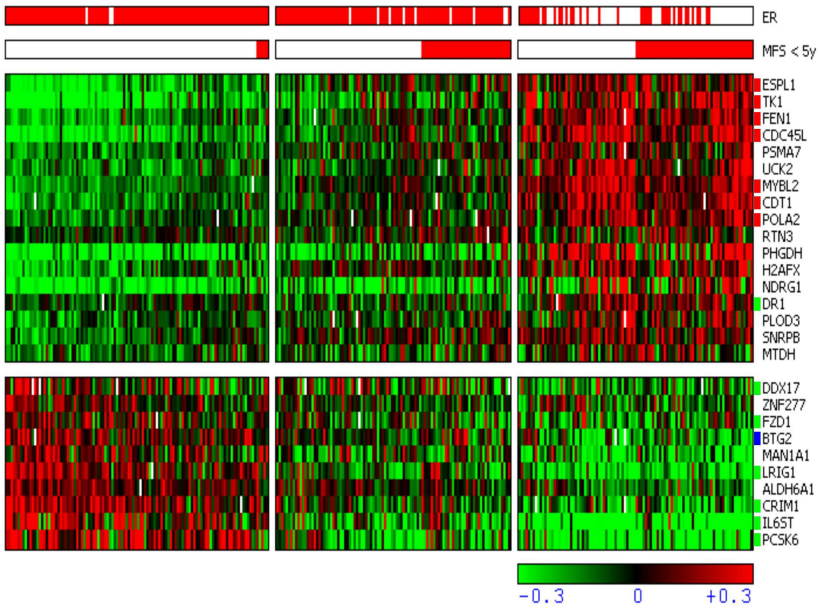


Figure 1.4: *The invasive growth signature predicts breast cancer metastasis. Specific functional groups in the IG-27 signature are diversely correlated with breast cancer prognosis. Genes belonging to the various functional clusters are counted as positive if their expression is positively correlated with metastatic relapse, or negative if they are downregulated in poor prognosis samples. Functional modules gene clustering. Samples (columns) are ordered from left to right according to increasing "Metastasis risk Index", calculated from the Invasive Growth signature by subtracting the Pearson correlation with the average good prognosis sample from the correlation with the average poor prognosis samples. The first horizontal bar, "ER", reflects the estrogen receptor status: light blue = positive, purple = negative. The second bar, "M<5Y", indicates the presence (red) or absence (green) of metastatic relapse within 5 years. Below the three annotation bars, the genes (rows) are clustered according to their overexpression (red) or downregulation (green) in the various samples.*

1.4.2 microRNA: a further level of gene expression regulation

microRNA's translation and their mechanism of action

miRNAs are small ncRNA transcripts initially produced from RNA polymerase II .They form a stem-loop structure and undergo processing by a protein complex containing the RNase III enzyme, Drosha and the double-stranded RNA binding protein Pasha in the nucleus. After initial cleavage, pre-microRNAs are exported into the cytoplasm by exportin-5, and subsequently cleaved by RNase III endonuclease Dicer. This mature microRNAs of approximately 22 nt of length are incorporated into microRNAISC (microRNA-induced silencing complex).Here, they act as negative regulators of gene expression through:

- **mRNA degradation** - microRNA find perfect complementary sequences in target mRNAs
- **translational inhibition** - microRNA find perfect match for nucleotides from 2 to 7 of the mature microRNA region, also known as seed sequence, and imperfect complementary for the remaining part of target mRNAs 3' UTR (untranslated region) [99]

As represented in 1.4.2, up to now mechanism of inhibition of translation, defective elongation, inhibition of initiation dependent on eIF4E, accelerated deadenilation and decay through normal degradation pathways were proposed and proved to explain microRNAs mechanisms of action [99]. The discovery of P-bodies, cell areas not defined by a membrane in which microRNA-mRNA complexes are seized and/or degradated had further complicated the puzzle of knowledge about microRNA's mechanisms of silencing.

MicroRNAs and the control of gene networks

Role of microRNAs in the regulation of biological programs miRNAs are small non coding RNAs of 20-22 nucleotides, typically excised from 60-110

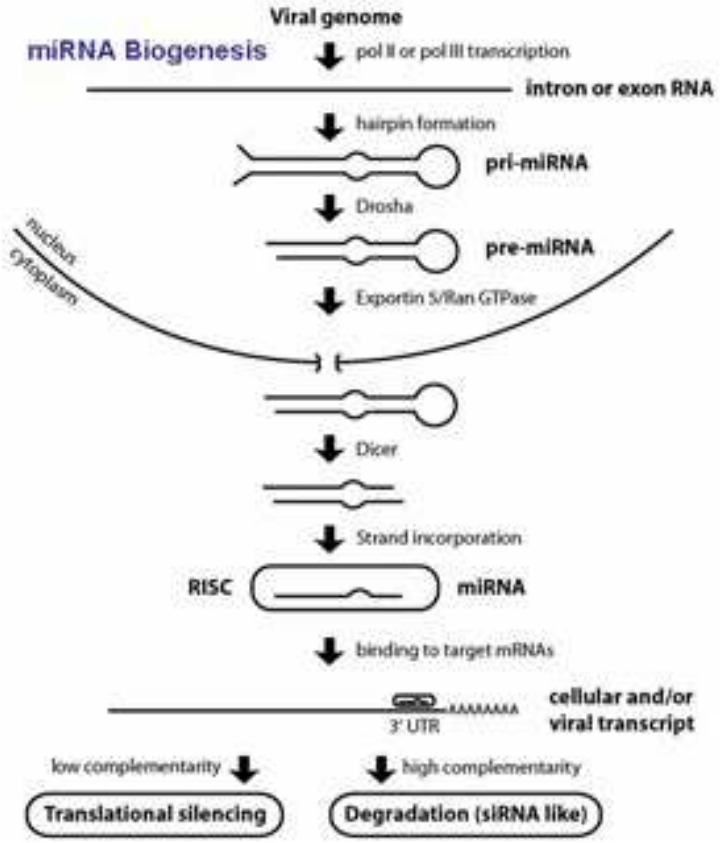


Figure 1.5: *Main mechanisms of microRNAs activity*

nucleotide fold-back RNA precursor structures. Bioinformatics data indicate that at least one third of known genes is a putative target of microRNAs [83] this means that each microRNA can control hundreds of gene targets, underscoring the potential influence of microRNAs on almost every genetic pathway. Many microRNAs are conserved in sequence between distantly related organisms suggesting for them the participation in essential biological functions [21]. Several groups actually have uncovered for them roles in the coordination of crucial biological processes including development, differentiation, apoptosis and proliferation [8, 58]. Recent evidences showed that microRNA mutations or mis-expression correlate with specific human

cancers raising the hypothesis that microRNAs can function as tumour suppressors and oncogenes [42]. Moreover it was demonstrated that microRNAs expression is a useful tool, having efficiencies higher than mRNAs, to classify cancers in groups with different characteristics such as cell type and aetiology and to define subclasses within a given tumor type [75, 65, 136]. Explaining their physiological role Cohen et al [24] proposed that microRNAs expression is essential to establish the identity of a new forming tissue. He proposed for them a role of genetic buffers acting on newly formed tissues by and/or translational inhibition of mRNAs that are foreign to that tissue although having residual leaky expression. Supporting these statements experiments on Drosophila scutellar bristles show that the action of microRNAs relies on large set of mRNAs that are related by their expression signature in time and space rather than on limited set of genes drastically down regulated.

1.4.3 Ancorage Independence Growth

Cell adhesion plays an important role in the regulation of cellular growth by giving consensus signals leading to cell proliferation [3]. Proliferation of mammalian cells is tightly regulated by multiple environmental influences, primarily adhesion to the extracellular matrix, cell-cell adhesion and soluble factors (i.e. polypeptide growth factors or inhibitors, mitogenic lipids, inflammatory cytokines and hormones). Among these environmental cues, soluble growth factors and integrin mediated adhesion are crucial [62] , as loss of adhesion generally results in complete G1 phase cell cycle arrest [5]. For susceptible cell types, loss of adhesion leads to anoikis [47], which might be regarded as the extreme case in which cell numbers decrease rather than remain stable. Both integrins and growth factor receptors use multiple cytoplasmic signaling pathways to regulate G1 phase cyclins and associated kinases that determine cell cycle progression [5, 92], strongly emphasizing the point that growth factors and the ECM are partners in cell cycle control, with each providing essential signals that allow for proper induction of the G1 cdks (cyclin-dependent kinases) and phosphorylation of their substrates.

Cyclin D1 is the primary D type cyclin for several anchorage- dependent cell types, and recent studies indicate that the ECM and mitogens are jointly required to induce cyclin D1 expression [102]: both stimuli are required for its mRNA expression and for its translation. In addition, the induction of cyclin A is strongly dependent on signals from the ECM, and several lines of evidence indicate that a large part of this effect is a consequence of adhesion-dependent pRb/p107 phosphorylation.

On the contrary, in suspended cells, mitogens are unable to induce the expression of cyclin D1 and subsequent formation of cyclins-cdk complexes which correlates with increased expression of cell cycle inhibitors (p21 and p27) contributing to cell cycle arrest. Indeed, it seems that the coordinated control of the G1 cyclin-cdks by growth factors and the ECM underlies the well-established anchorage requirement for the proliferation of untransformed cells, since formation and spread of tumors, is closely associated with decreased dependence on ECM for growth and survival [113]

We therefore carried out a functional screening to identify and characterize new genes conferring anchorage independence to human normal and neoplastic breast cells. MCF10A human mammary cells were transduced with a retroviral cDNA expression library and selected by growth in suspension. Microarray analysis targeted on library-derived transcripts, according to the recently developed "XenoArray" procedure [85], revealed strong and reproducible enrichment, after selection, of cDNAs encoding the scaffolding adaptor Gab2. The ability of Gab2 to promote anchorage independence was further characterized by gain-/loss-of-function and expression profiling experiments.

Chapter 2

Aim of the work

The aim of this work is to exploit gene expression profiling of cell-based models to define genomic signatures reliably predicting metastatic progression of breast cancer. In particular, transcriptome changes in *in-vitro* experimental models of regulated cell motility/invasiveness and anchorage independence are explored to generate lists of genes ("signatures") whose expression has prognostic value when analyzed in human breast cancer microarray datasets. In turn, successful translation of a model-derived signature into a clinically significant cancer classifier confirms the appropriateness of the chosen model. The experimental approaches chosen to find lists of genes whose expression is related to in vitro-modelled key steps of cancer onset and progression are here briefly outlined.(1) Genes transcriptionally regulated during induction of epithelial invasive growth by Hepatocyte Growth Factor (HGF), via the MET tyrosine kinase receptor, in MLP-29 mouse embryo liver cells. These genes are mapped onto breast cancer datasets and prioritized to generate a classifier predicting metastatic relapse. To construct such classifier, the Nearest Mean Classifier algorithm is implemented to calculate a Metastasis Score, proportional to the probability of metastatic relapse. The classifier is subsequently validated by calculating the Metastasis Score in independent datasets obtained in different microarray platforms.
(2) Identification of groups of genes that are both transcriptionally regulated

by HGF and targets of specific microRNAs, to highlight regulatory circuits involving miRNAs in invasive growth and cancer. This required the design of an algorithm (SeaMT, signature enrichment analysis for microRNA target) for the computational identification of microRNAs potentially regulating sets of genes transcriptionally co-regulated by HGF. After collecting the list of putative miRNA-binding sequences for each gene in the genome, a statistical analysis is performed to assess whether some subgroups of HGF-regulated genes are significantly enriched in binding sites for specific miRNAs. Such miRNAs are then functionally characterized, and the set of their putative target genes is further analyzed to build a classifier for breast cancer progression as in point (1).

(3) Functional screening for genes promoting anchorage-independent growth of MCF-10A human normal breast cells. The functional screening was based on the "XenoArray" technology, and highlighted Gab2, a scaffolding protein involved in signal transduction. Functional characterization of Gab2 provided evidence of its role in anchorage-independent growth. A Gab-2 signature, obtained by comparing Gab2-overexpressing, anchorage-independend MCF10A cells from control cells is obtained and is further analyzed to build a classifier for breast cancer progression as in point (1).

Finally, given the positive results obtained with all the three models, a standardized in-silico pipeline is delivered, allowing conversion of any list of genes of interest into a genomic classifier for human breast cancer, and assessment of the statistical significance of its classifying performance.

Chapter 3

Results

3.1 The HGF-driven invasive growth signature predicts metastatic relapse of breast cancer.

As previously described, genomic analysis of the invasive growth program in MLP-29 cells highlighted a signature composed of genes involved in several functions related to cancer progression. Transposition of the invasive growth signature on the NKI microarray dataset of 311 human breast cancers led to the definition of a minimum set of 27 genes that maximally differentiates patients displaying metastatic progression within five years from long-term disease-free patients.

To to construct a classifier for the 27 invasive growth genes, we assessed their performance on an independent dataset of 198 samples obtained on a different microarray platform (Affymetrix) and previously used to validate a 76-gene prognostic signature of breast cancer (the "Veridex Index") by the TRANSBIG consortium (TRANSBIG-198)[35].

We performed a leave-one-out analysis of NMC[143] in which each sample was assigned a MS based on its differential correlation with the good- and poor-prognosis centroids calculated in the other samples. We then defined

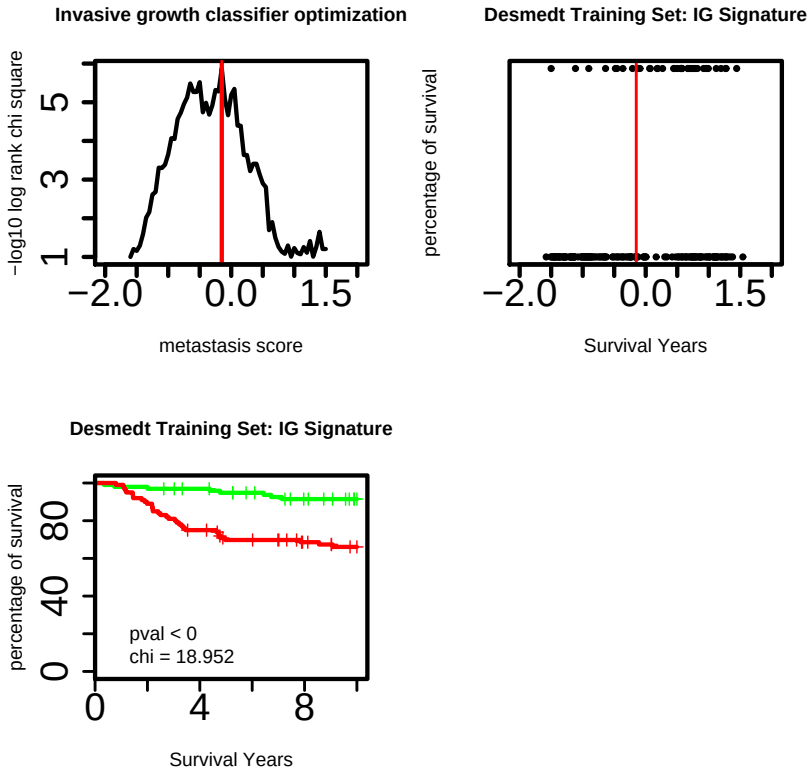


Figure 3.1: *Invasive growth classifier tuning in Transbig dataset: We constructed the metasamples for NMC in a set of 198 patients. In the Transbig dataset we calculated the median expression in good and poor prognosis. Metastasis score is calculated with Nearest mean classifier as a prognostic index, correlated with metastatic risk. We performed survival analysis with different threshold, and identified best solution at -0.15. Finally it is reported the kaplan- meier analysis on the Transbig breast cancer dataset according to IG-signature classification, patients classified as good prognosis, green line, poor prognosis ,red line, by the invasive growth signature.*

the optimal MS threshold for good- and poor prognosis class assignment by running a log rank chi square analysis for increasing threshold values (Fig3.1) Maximum significance was observed at the MS value of -0.15, thereby chosen as the class threshold (Fig3.1) Kaplan-Meier analysis showed that the IG-27 signature classifies patients with high accuracy using this threshold, despite

the platform change (p $<10^{-5}$, log-rank chi square = 18.95244, a specificity of 0.1382114 and a sensibility of 0.5856354). When samples were subdivided

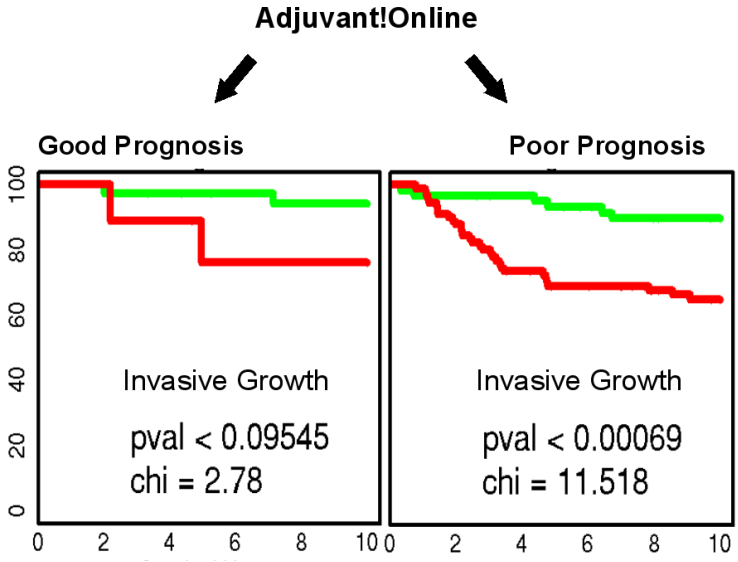


Figure 3.2: *IG stratification on clinical parameters*

by their prognostic class according to standard clinical/pathological parameters, as defined by the Adjuvant!Online (AOL) score[116], they could still reliably further subdivided in good- and poor-prognosis subgroup by IG-27. (Fig3.2). Multivariate Cox regression analysis showed that the IG-27 sig-

Table 3.1: TRANSBIG Multivariate analysis Invasive Growth vs Veridex

	p	coef	lower .95	upper .95
Invasive Growth	0.00	1.16	1.42	7.17
AOL	0.24	0.64	0.65	5.48
Veridex Index	0.03	1.33	1.15	12.51

nature largely outperforms the AOL score, and remains strongly significant and independent from the Veridex Index in predicting metastatic progres-

sion Table**??**. These results prompted us to further validation of the invasive growth-based classifier across additional Affymetrix dataset, using the NMC constructed on the TRANSBIG-198 dataset.3.1 The IG27 classifier achieved good performance in all the human cancer datasets tested. (Fig3.3)

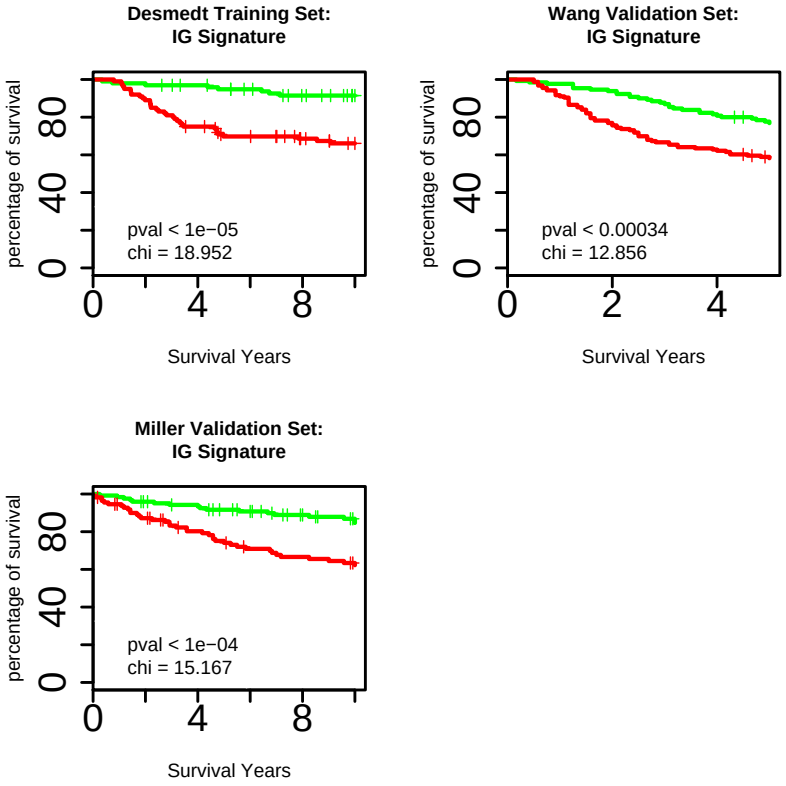


Figure 3.3: *Invasive growth classifier validation in several dataset*

To validate the signaute on a third microarray platform we carried out gene expression profiling using Illumina Human-8v1 expression Beadchips on 76 breast cancer frozen samples with long-term followup (ILM-76 dataset). Notably, all these samples are classified as poor prognosis by the AOL score 3.1.

Characteristics	All patients (n=76)	ER positive (n=50)	ER negative (n=26)
Age (years)			
Mean (SD)	52.7 (12.7)	53.1 (13.3)	51.8 (11.7)
≤40	11 (14%)	7 (14%)	4 (15%)
41-55	32 (42%)	20 (40%)	12 (46%)
56-70	24 (32%)	14 (28%)	10 (38%)
>70	5 (7%)	5 (10%)	
NA	4 (5%)	4 (8%)	
Menopausal Status			
Pre	35 (46%)	21 (42%)	14 (54%)
Post	35 (46%)	26 (52%)	9 (35%)
NA	6 (8%)	3 (6%)	3 (11%)
Grade			
1	1 (1%)		1 (4%)
2	34 (45%)	25 (50%)	9 (35%)
3	33 (43%)	20 (40%)	13 (50%)
NA	8 (11%)	5 (10%)	3 (11%)
Lymphonode status			
positive	52 (68%)	34 (68%)	18 (69%)
negative	24 (32%)	16 (32%)	8 (31%)
ER status			
positive	50 (66%)	50 (100%)	
negative	26 (34%)		26 (100%)
PR status			
positive	41 (54%)	36 (72%)	5 19%)
negative	35 (46%)	14 (28%)	21 (81%)
Overall survival (months)			
Mean (SD)	84.7 (44.1)	92.3 (38.4)	70 (51)
Disease free survival within 5 years			
yes	36 (47%)	19 (38%)	17 (65%)
no	40 (53%)	31 (62%)	9 (35%)
Adjuvant! Online			
Mean Adjuvant % alive at 10 years (SD)	49.2 (20)	51.1 (19.2)	45.8 (21.3)
Mean Adjuvant % alive and without cancer at 10 years SD)	34.5 (18.4)	33.2 (18.4)	36.8 (18.8)

Figure 3.4: table with clinical parameter of ILM-76 dataset

The IG-27 classifier was mapped via MAQC on the ILM-76 dataset together with four published genomic prognostic classifiers: the 76-gene "Veridex index" [141], the 70-gene signature developed by van't Veer and colleagues [132], the "Genomic Grade Index" [121] and the "Basal-like" signature [123].

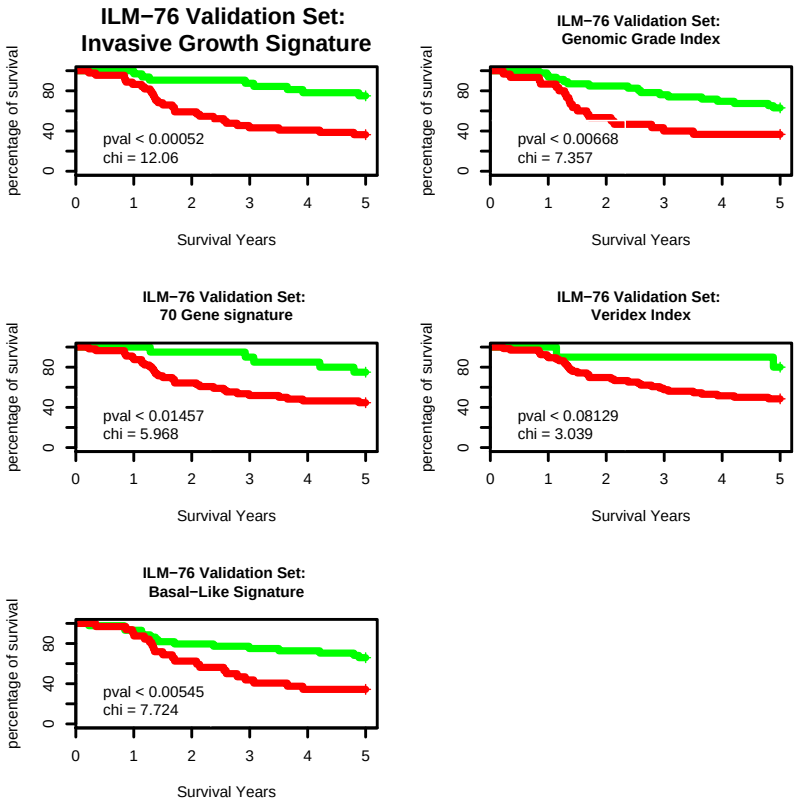


Figure 3.5: *Validation*

With the exception of IG-27, which maintained the -0.15 MS threshold set in the TRANSBIG-198 dataset, all the other classifiers had the threshold tuned to obtain maximal performances. Kaplan-Meier analysis (Fig3.5) showed that all the signatures maintained some ability to discriminate good- and poor prognosis samples, with the strongest significance obtained by the IG-27 classifier and the lowest by the Veridex Index. Interestingly, in multivariate cox regression, IG-27 was the only classifier remaining significant (Table 3.2). These data show that, being based on a biological model, the prognostic classifying performances of the IG-27 signature are less dependent on the microarray platform adopted, compared to empirically obtained genomic classifiers.

Table 3.2: ILM-76 Multivariate analysis of all genomic classifiers

	p	coef	lower .95	upper .95
GGI	0.81	-0.10	0.39	2.06
InvasiveGrowth	0.04	1.05	1.06	7.65
VeridexIndex	0.67	0.34	0.30	6.51
MamPrint	0.53	0.35	0.47	4.34
BasalLike	0.17	0.52	0.80	3.54

3.2 Identification of microRNAs involved in the invasive growth program

microRNAs (miRNAs) are cell regulators that act modulating half-life and/or translation of multiple target transcripts, adding a further level of complexity in the flow that leads from genes to proteins. In the context of invasive growth, one or more miRNAs could be key regulators capable of modulating the HGF-driven transcriptome, thereby sustaining or counteracting full activation of the program.

To search for such miRNAs, we conceived a computational approach based on the identification of miRNAs whose target genes are transcriptionally modulated during activation of the invasive growth program. As a source of genes involved in invasive growth, we chose gene expression data obtained from MLP-29 cells stimulated in vitro with HGF in a time-course experiment. miRNAs that bind a large fraction of genes regulated in this model are the best candidates for subsequent characterization.

To maximally benefit from the progress in genome annotation, the MLP-29 invasive growth microarray dataset was updated to a recent release of unigene. Reanalysis of the data led to the definition of a 2124-gene signature after cross-validation between platforms. Transcriptional waves of gene expression were resolved by gene clustering using the FLAME fuzzy clustering algorithm, developed in our lab[49]. Also this re-analysis highlighted

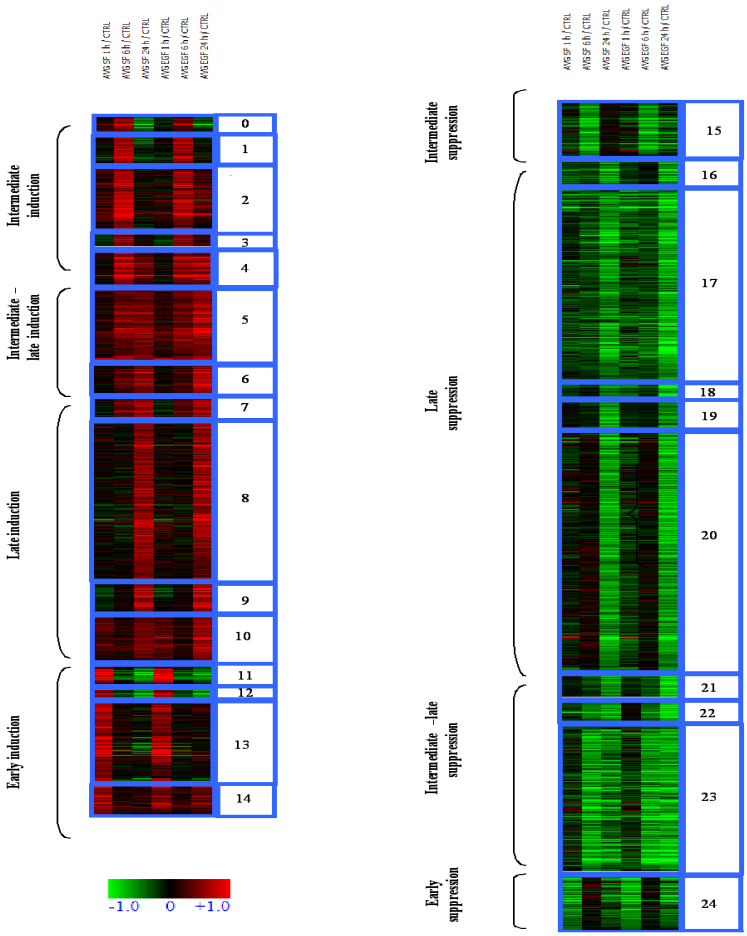


Figure 3.6: *HGF EGF time-course profile remapping*

small but consistent kinetic/quantitative differences in the transcriptional responses to HGF and EGF (Fig3.6).

3.2.1 SeaMT: Signature Enrichement Analysis for microRNA targets

Despite the important role recognized for miRNA's there wasn't a bioinformatic tool indicating miRNAs whose targets sites are enriched in gene lists emerging from expression analysis. We developed SeaMT (signature enrichment analysis for miRNA targets) using the hypergeometric function as core element to elaborate enrichment analysis. According to the work flow represented in (Fig3.7) the software proceeds in four steps:

- **Loading** Loading step initially creates the environment necessary for the software to work, this is achieved recalling Heatplus and biomaRt libraries form R [104, 52] package and the database Miranda, which contains human and mouse miRNAs targets genes lists, from SeaMT. Subsequently the following the external inputs are imported: (1)the matrixes containing expression values (2)hte list of the selected gene subset(signature) and the background gene list from which the signature has been uncovered. Finally, connecting to biomaRt's web site orthologous genes lists are generated and loaded for external input data.

- **Statistical analysis** After collecting the list of miRNAs targeting the background, for each miRNA hypergeometric parameters are defined as follow (Fig3.7)
  a)number of miRNA targets in the signature
  b)number of genes in the signature
  c)number of miRNA targets in the background
  d)number of genes in the background
  Having as input the parameters of the two side matrix represented in (Fig3.7) the hypergeometric function gives as output a p-value which mathematically estimates the probability of non existent association between differential expression and miRNA's target membership, which is the hypothesis of our test. If p-value is under Bonferroni cut off the hypothesis is rejected meaning it is very difficult to find randomly in

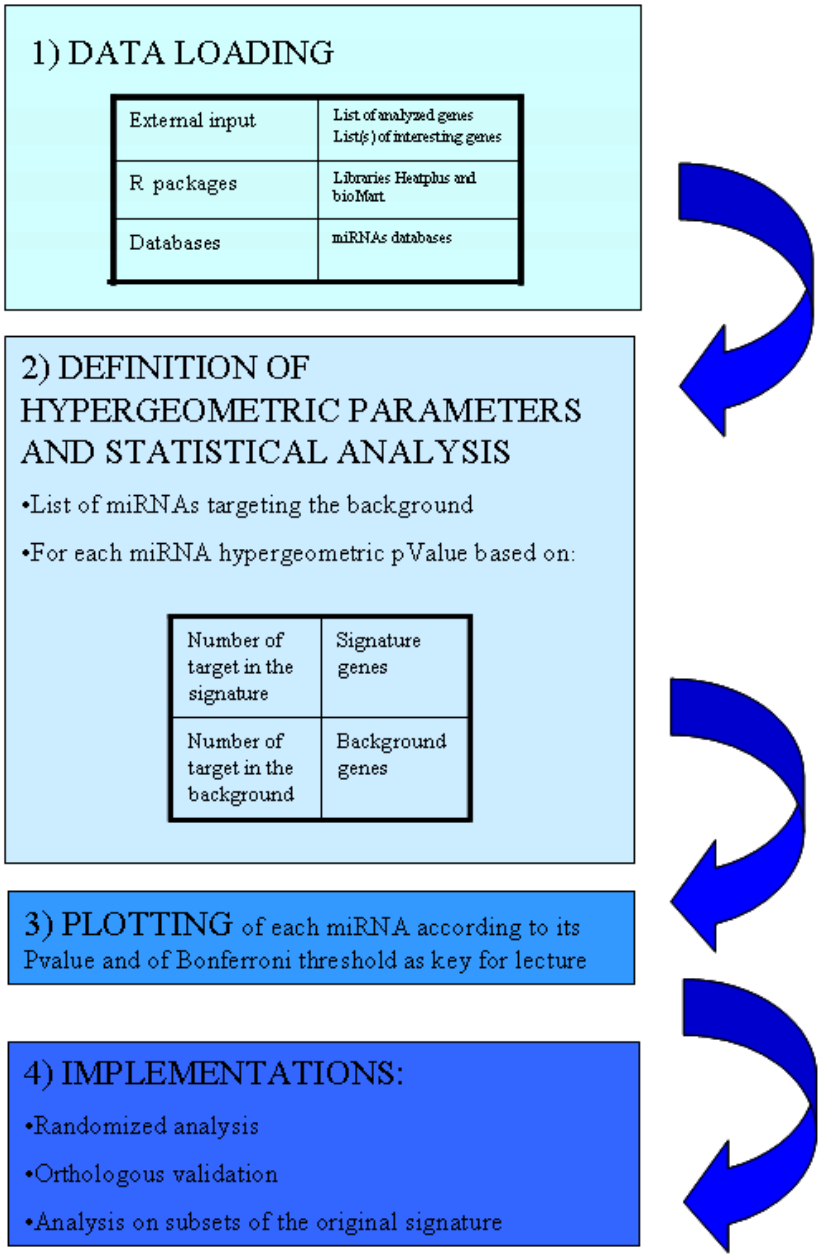


Figure 3.7: *Main steps of the SeaMT algorithm work-flow*

the universe the enrichment of target genes individuated among the

selected ones.

- **Plotting** For the results plotting representation we use a bidimensional space having on the x-axis the miRNAs having targets in the background and on the y-axis their p-values of target gene enriched in the signature. Thresholds of significance are displayed as horizontal lines to facilitate the individuation of miRNAs whose targets are enriched in the list of interesting genes.

- **Implementation**

SeaMT is based on a database of putative miRNA targets which is biased by high levels of false positives. To overcome this limit we implemented four functions:
1) Orthologues validation function: maps on an orthologous genome signature and background looking for conserved miRNA's enrichments. This function constitutes a first "biological validation" as we suppose that miRNAs involved in basic cell function regulation are conserved between species and the same is expected for their targets.
2) Randomization function: calculates target's occurrences on randomly chosen sub-groups permitting to eliminate miRNA's emerging by chance from the analysis.
3) Cluster analysis function: if the signature expression set contains cluster information SeaMT automatically performs the analysis on single clusters. As suggested by Thomson[128] miRNAs regulate groups of genes with similar tendencies more than genes drastically down-regulated, and clustering provides this information.
4) Threshold function: subdivides the signature according to threshold of differential expression defined by the users permitting to analyze the data in a more supervised manner than cluster.

Finally according to these implementations also plotting functions were modified to distinguish with colour codes miRNAs having or not an orthol-

ogous and to be able to plot analyzed genes expression values.

Bioinformatic analysis on invasive growth signature

Subsequently, the HGF-regulated transcripts had to be scanned for the presence of miRNA target sequences, using a miRNA-target database (Miranda algorithm, from mirGen portal [90]). After collecting the list of putative miRNA-binding sequences for each gene in the genome, a statistical analysis had to be performed to assess whether some subgroups of HGF-regulated genes were significantly enriched in binding sites for specific miRNAs. To this aim, an algorithm based on hypergeometric analysis was developed, and implemented into an R-based software.

Definition of hypergeometric groups

To exclude biases due to tissue-specific gene expression, the background gene list was restricted to the pool of genes expressed in MLP-29. For hypergeometric analysis, we considered the entire 2124-gene invasive growth signature, and also subsets of co-regulated genes defined by FLAME clustering[49]. We also analysed subsets of genes differentially expressed at individual time points above a positive or negative log2ratio threshold of 0.25. These lists were submitted to SeaMT analysis, the results of which are outlined below.

Candidates choice

To further consolidate these results, SeaMT was re-run with miRNA target lists generated by an indipendent database, PITA [**?**]. Only miR2-24 and miR-296 remained significant in both the human and murine genomic background.

It can be concluded that a significantly high fraction of the genes transcriptionally up-regulated at 24 hours by HGF is potentially targeted by miR-24 and miR-296, highlighting these two miRNAs as potential players in the invasive growth program.

No miRNA were found to have more targets then expected in the 2124 gene list. Conversely, analysis of expression clusters highlighted a strong enrichment in specific miRNA targets for the cluster containing genes up-regulated at 24 hours. These results are depicted in (Fig3.8), and show that miR-24, miR-296 and miR-484, target genes are enriched above a Bonferroni-corrected significance threshold, in both the human and murine genomic background. These results were confirmed when lists of genes regulated at individual time points where submitted to the analysis: only the list of genes upregulated at 24 hours displayed strong enrichment for target genes of various miRNAs, including again miR-24, miR-296, and miR-484. (Fig3.9).

For downregulated genes 3.2.1 we did not find significant microRNAs, apart from miR-377 which just reaches Bonferroni cut-off and was not confirmed by orthologous validation.

Candidates choice

To further consolidate these results, SeaMT was re-run with miRNA target lists generated by an indipendent database, PITA [**?**]. Only miR2-24 and miR-296 remained significant in both the human and murine genomic background.

It can be concluded that a significantly high fraction of the genes transcriptionally up-regulated at 24 hours by HGF is potentially targeted by miR-24 and miR-296, highlighting these two miRNAs as potential players in the invasive growth program.

3.2.2 Functional validation of candidate microRNAs

Real-time PCR analysis of miRNAs expression

Having chosen the candidates, before starting biological analysis, we tested their expression in MLP-29 cells trough real time PCR on a time-course for 1-6-24 hours with HGF (Fig3.10A) miR-296 was not detected in all time-points tested. After having validated the primer functionality in RNAs extracted

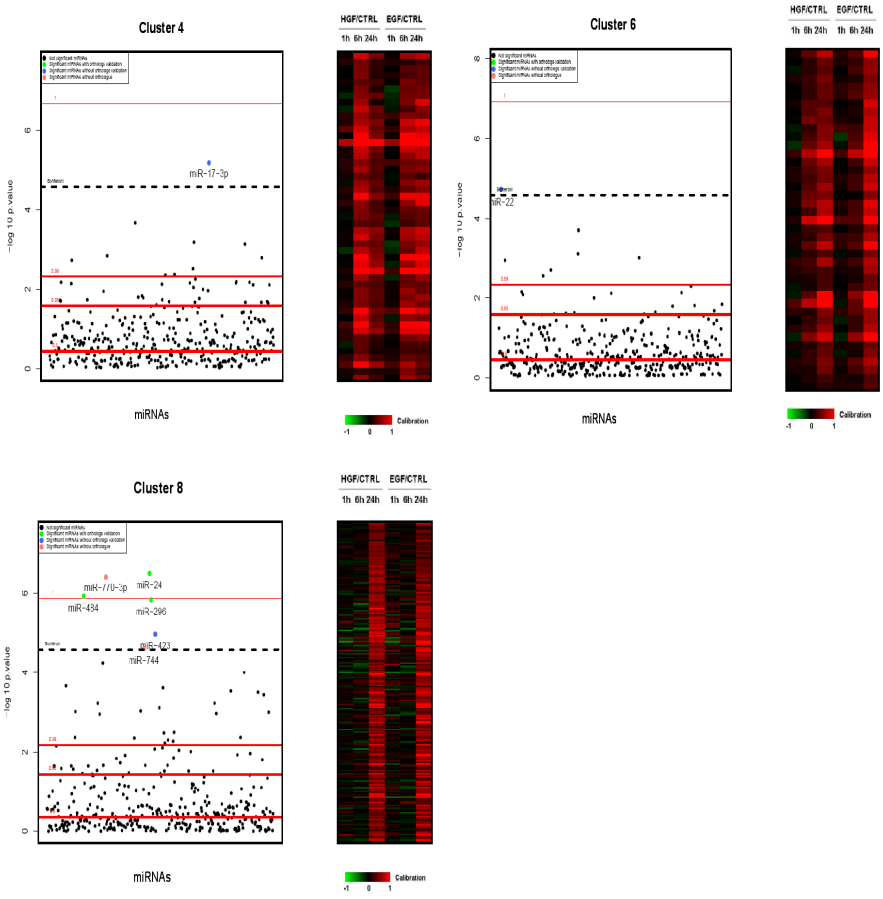


Figure 3.8: SeaMT on several gene expression clusters in MLP-29 cells upon HGF stimulation.

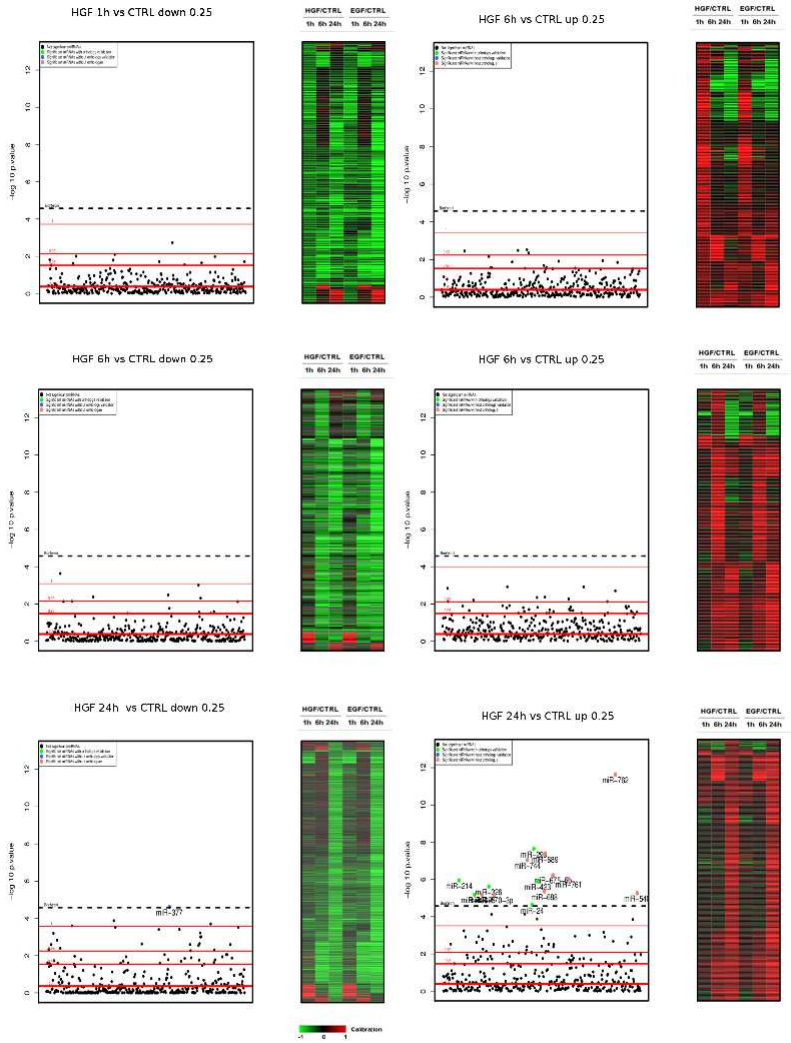


Figure 3.9: *SeaMT analysis performed on genes regulated by HGF at different time-points, only gene upregulated at 24h are enriched in miRNA targets*

from cells known to express miR-296 (data not shown) we concluded that it is not expressed expressed by HGF, as reported in (Fig3.10A). in the cell line nor regulated by HGF as reported in the next image. miR-24 indeed seemed to be induced by HGF stimulation at 1 hour. To confirm miR-24 regulation

by HGF we decided to perform a biological triplicate of HGF stimulation time-course. The result shown (Fig3.10B) confirms the induction of miR-24 expression after HGF stimulation. Basal expression and expression at the first time-point are highly consistent and reproducible while 6 hours and 24 hours time-points behaved quite differently exhibiting signals increase or maintenance at 6h hours and different levels of decrease at 24 hours.

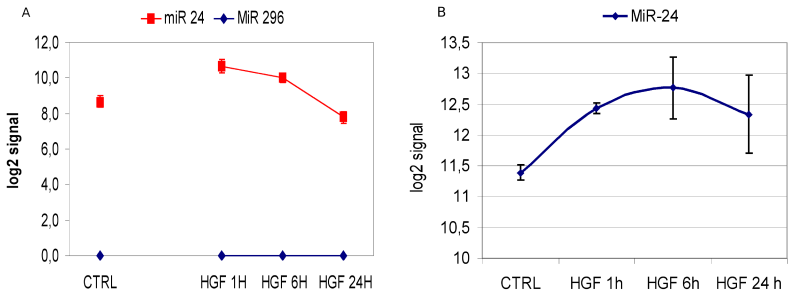


Figure 3.10: *miRs Differentialy dynamis*

miR-24 and miR-296 gain of function

After verifying that miR-296 was not expressed and miR-24 was induced by HGF we tried to deregulate the MLP-29 response to the growth factor by performing gain-of-function experiments with the two miRNAs. To this aim we produced two viral vectors expressing the respective pre-miRNAs. MLP-29 cells were infected with serial dilutions of viral production obtained from 293T cells. After selection we observed an MOI respectively of 2,67 for empty vector, 3,95 for miR-24 and for 2,25 miR-296. According to this result we used a viral dilution of 1:100 for the experimental infections. After 7 days of blasticidin selection, colonies were expanded, RNAs were extracted and tested respectively for miR-24 and miR-296 expression with quantitative real-rime PCR, obtaining the following Log2-signals: (miR-24 CTRL = 11.4, Transduced = 14.05, miR-296 CTRL = 0, Transduced = 12.13)

Biological assays on miRNAs-overexpressing cells

Scatter assay

To evaluate the effects of miRNAs overexpression, scatter assay was performed fixing the cells at 24 hour from stimulation with serial dilutions of HGF ranging from 40 U/mL to 0,5 U/mL. The scatter assay (Fig3.11) showed that both miR-24- and miR-296 sensitize MLP-29 cells to low doses of HGF, without having any effect on unstimulated cells. In particular, while control cells do not scatter when stimulated with HGF at concentrations under 5U/mL, both miRNA-expressing cells are responsive to concentrations up to ten-fold lower. The scatter assay involves a complex set of biological processes, including changes in cell shape, cell-cell and cell-matrix adhesion, migration, survival and proliferation. To have a better insight in miRNA's deregulation effects we investigated basic cellular processes behind it, like proliferation and migration, with specific assays.

Proliferation assay

To evaluate basal proliferation of miRNA transfected cells a MTT assay was performed every 24h for 4 days. MTT is a vital staining turning from yellow to violet when tetrazolium salt is catabolized from the cell. This enables the estimation of cell numbers with a 490 lambda absorbance reading. (Fig3.12) shows the relative growth data obtained for control cells and miRNAs expressing ones. miR-24 over-expressing cells displayed growth trends similar to the control while miR-296 ones exhibited a marked tendency to proliferate more than controls. *Wound healing To test the effects* of miR-24 and miR-296 on motility we performed a wound on confluent cell monolayers stimulated with different concentrations of HGF and cultured in low serum to minimize the proliferation effect on wound closure.

In (Fig3.12) shown the percentages of wound closure obtained after 12 hours from wounding. Both miRNAs seem to have effects on motility in concomitance with with HGF stimulation. The trend induced by miRNAs over-expression is similar, although the maximum difference from the controls seems to be at the lower ligand concentration for miR-24, and high for miR-

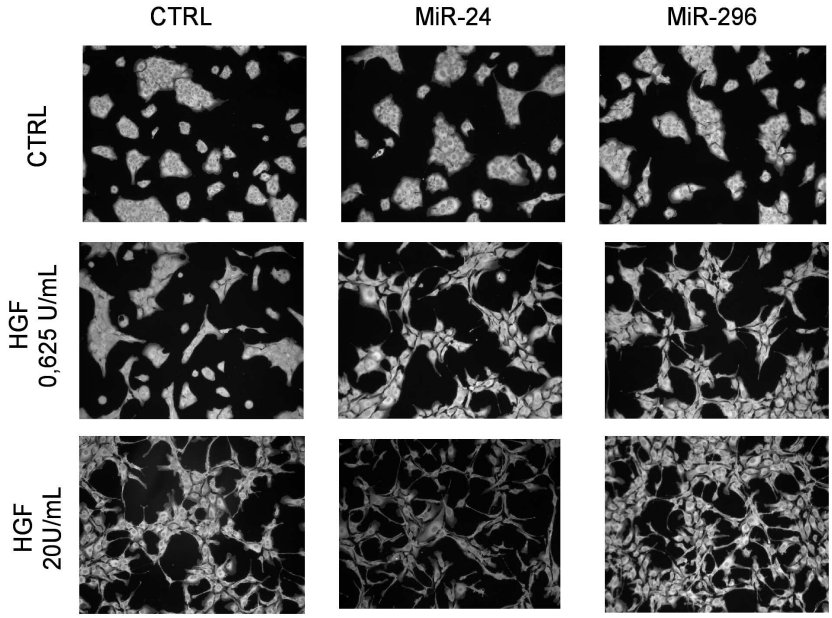


Figure 3.11: *miRs overexpression sensitize MLP-29 to HGF scattering*

296. T-test statistic are however significant for both miR-24 and miR-296 upon HGF stimulations of 40 U/mL, the respective p.values were 0,0241 and 0,01. Both miRNAs expressing cells do not exhibit a well characterized trend if compared with the control when HGF stimulation is performed at 16 U/mL.

Functional annotation of common HGF/miR-24 target genes

The experimental results obtained led to an apparent paradox, as they show that to promote invasive growth, HGF upregulates both miR-24 and its target genes, which in turn are expected to be negatively modulated by the miRNA. To find a possible explanation for this paradox, a careful analysis of

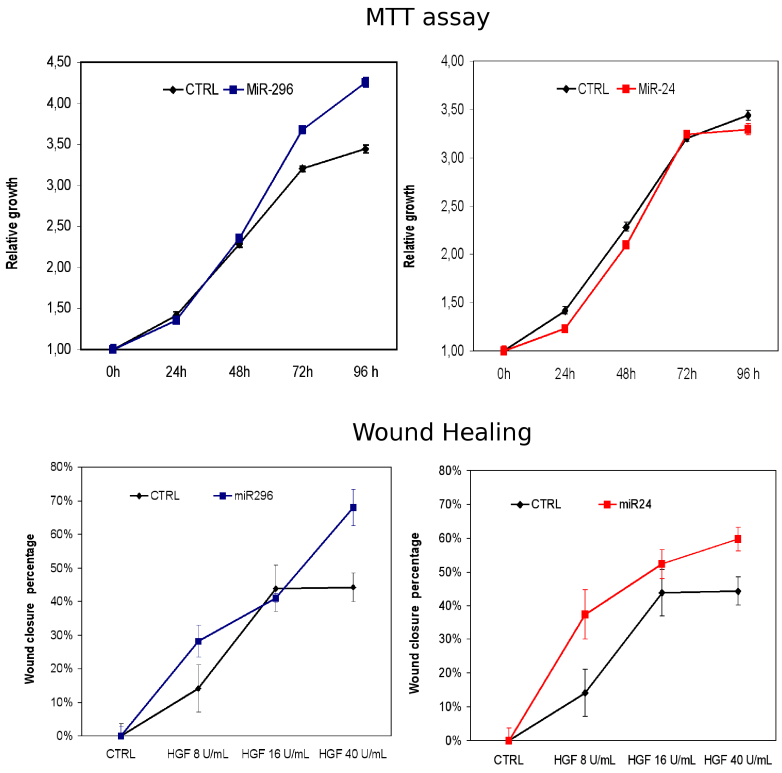


Figure 3.12: *miRs overexpression sensitize MLP-29 to HGF scattering*

genes targeted by miR-24 was performed, and among them we found CSK (C-terminal SRC kinase) and Talin, whose downmodulation by the miRNA could explain the experimental results described above. (Fig3.13)

Csk is an endogenous inhibitor of tyrosine kinases of the Src-family, whose main components are c-Src, Fyn, Yes, Lck, Lyn, Hck, Fgr, Blk[117]. Interestingly, it is known that, upon HGF stimulation, c-Src is activated by the Met receptor and contributes to promote motility [100, 118]. Therefore, Csk transcriptional upregulation by HGF may be part of a negative feedback loop, whereby its downmodulation by miR-24 may enhance the response to low doses of HGF and ”buffer” the negative loop. To have a better insight of the

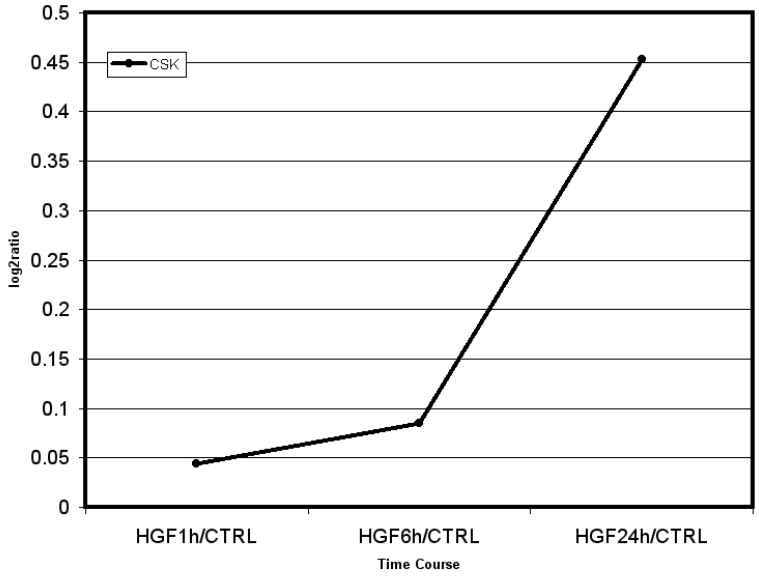


Figure 3.13: *miRs overexpression sensitize MLP-29 to HGF scattering*

possible miRNA "buffering" activity we investigated the expression levels of mRNAs encoding c-Src and its activator Shp2[119]. While c-Src was upregulated at 1h from HGF stimulation, Shp2 was found to be down regulated at 24h, consistently with the negative feedback loop involving Csk upregulation.

Talin is a protein involved in focal adhesion formation; it acts as a bridge between integrins and actin having also a central role at the leading edge of migration [119]. The requirement of Talin for leading edge formation could explain its transcriptional up-regulation by HGF at 24 hours, when filipodia and lamellipodia are clearly visible on scattered cells. Conversely, miR-24 induction by HGF at 1-6 hours may lead to transient Talin down-modulation, which could permit the initial disassembly of focal adhesion and epithelial shape disruption necessary for the beginning of the migratory response.

3.2.3 Construction of the Genomic classifier in breast cancer.

From functional validation it emerged that miR-24 could be involved in the regulation of the invasive growth program in MLP-29, while the results obtained on mir-296 are biased by the fact that miR-296 it is not expressed in our in vitro model. Albeit interesting, the strong phenotype observed for miR-296 could be due to the introduction of a non-physiological perturbation in the system. Thus we focused on the characterization of the miR-24 program in human breast cancer. In frame with the procedure previously defined

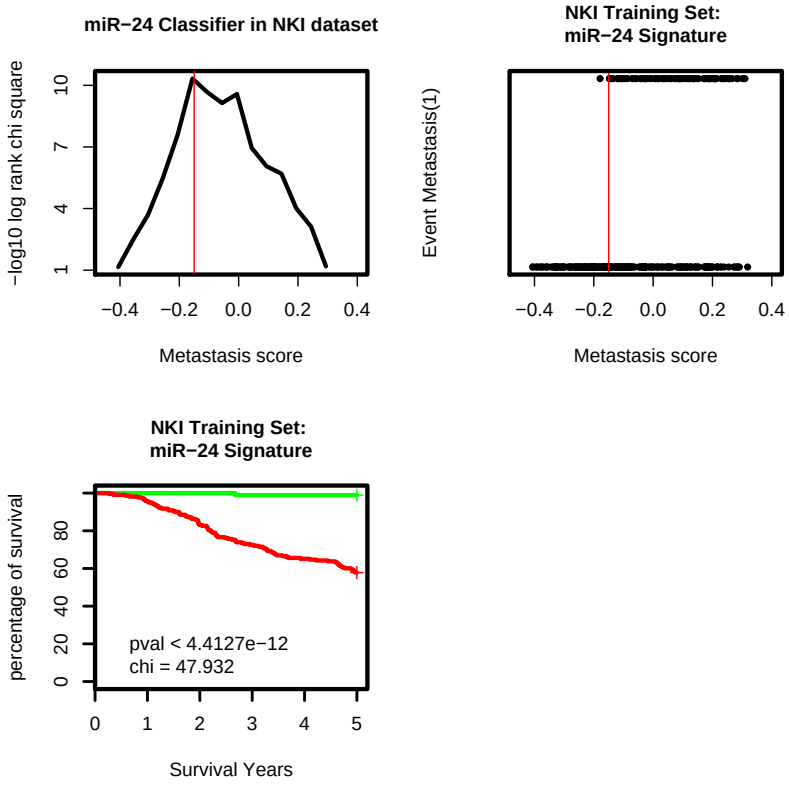


Figure 3.14: *NKI Training set: miR-24 training in the full NKI cohort of patients.*

for the invasive growth signature we mapped all the genes predicted for being target of miR-24 by Miranda, according to mirGen database to NKI human

breast cancer dataset and constructed the NMC in the NKI dataset (Fig3.13) we report the construction of the genomic classifier and its performance in the NKI datasets. In the NKI breast cancer dataset dataset it was possible to define a threshold to discriminates patients with good performance in the prediction of good prognosis group: the classifier that finally leads to log-rank p-value chi square $< 4.412692^{-12}$, with a precision of 0.01075269, and a sensitivity of 0.4220183.

Validation of the miR-24 genomic classifier.

To validate the significance of the miR24 classfier, we verified if it maintains significance also in other breast cancer gene expression data.

The miR-24 classifier was mapped according to MAQC cross-mapping table to breast cancer datasets [141, 35, 93], and confirmed its discriminating power. Indeed, in all the datasets miR-24 classifier was able to discriminate good and poor prognosis. Furthermore in the ILM-76 dataset the classifier after a retuning of the MS threshold yelded a subgroup of aggressive cancer, however the classifier could not match the performance of other genomic classifier in multivariate analysis (data not shown). The reason for this result could be attributed to the fact that, differently from what we did for IG-27, we did not prioritized genes maximally discriminating patients outcome, and to the fact that microRNA target predictors are affected by a relatively high rate of false positive[78].

3.3 Gab2-driven AIG

After successfully constructing genomic classifiers for breast cancer starting from an *in vitro* model for cancer progression based on mouse live cells, we aimed to create a model more directly fitting the development of breast cancer. In this view the ability to grow in the absence of anchorage to the extracellular matrix represents a key oncogenic property of cancer cells. To screen for genes conferring anchorage independence, we exploited XenoArray

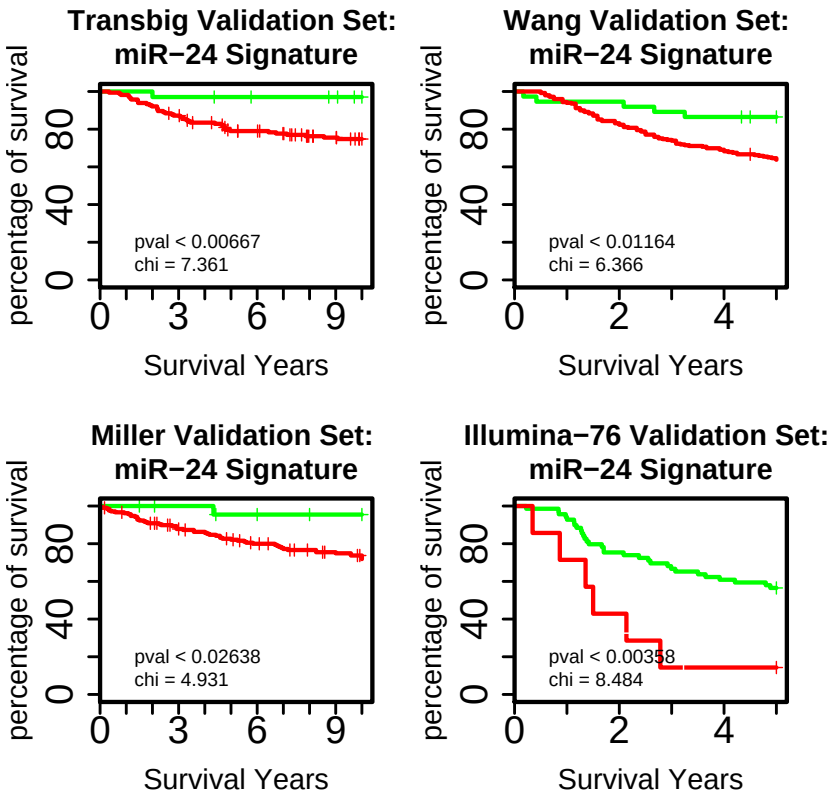


Figure 3.15: *miR-24 classifier validation*

analysis, a recently developed strategy combining expression library transduction with DNA microarray analysis in MCF10A normal human breast cells.

3.3.1 Setup of a gain-of-function screening for anchorage independence in MCF10A cells.

For the functional screening, MCF10A cells were transduced with a commercial mouse testis retroviral expression library (Stratagene) or with GFP as a control. To increase the screening robustness, infections were performed in duplicate (A and B), using an estimated multiplicity of infection of 1, to avoid multiple integrations in the same cell. This led to the generation of

four populations of cells: (i) GFP-transduced A, (ii) GFP-transduced B, (iii) library-transduced A, and (iii) library-transduced B. To detect and quantify library-derived transcripts we performed XenoArray analysis [85], by extracting total RNA from the four cell populations and hybridizing the resulting cRNAs on murine expression arrays, to allow specific detection of library-derived transcripts of murine origin. The scatter plot in (Fig3.16) shows the expression measurements obtained in GFP-transduced A cells (x-axis) versus the measurements obtained in library-transduced A cells. Both library-transduced populations clearly showed a consistent number of detectable murine transcripts (945 and 1125 probes in infection A and B, respectively, with a detection p-value <0.01).

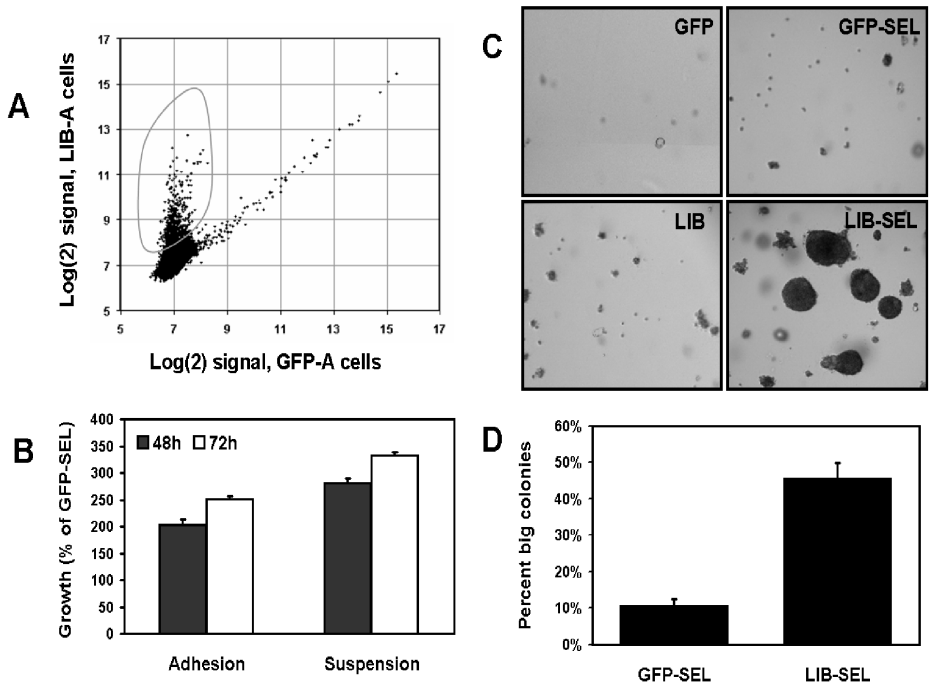


Figure 3.16: *XenoArray analysis on MCF10A cells and acquisition of anchorage independence by library-transduced selected cells. (A) Dot plot comparing the signal intensity for GFP-transduced and library-transduced cells corresponding to infection A. (B) MTT growth assay on polyhema-selected populations after 48h and 72h in adhesion or suspension, as indicated. Cell growth is expressed as a ratio between library-transduced and GFP-transduced cells, after normalization to the amount of viable plated cells at day 0. The data represent the mean and standard error of triplicate values. (C) Soft agar assay on GFP- and library-transduced cells, unselected or selected on polyhema, as indicated. Phase-contrast images were captured by a BD Pathway microscopic station (BD biosciences) after 3 weeks in agar. (D) Quantification of colony formation as the fraction of colonies bigger than the 95th percentile of colony size of the control cells (GFP-transduced non-selected), as described in Materials and Methods. The data represent the mean and standard error of triplicate values.*

Conversely, very few probes (around 100) cross-hybridized to endogenous transcripts and were detected also in GFP-transduced cells. Based on our

previous observations on XenoArray analysis sensitivity [85], we estimated that a selection-driven 20-fold enrichment of even a rare transcript, bringing it from 8 to 160 parts per million, should be enough to render it clearly detectable by XenoArray analysis. We exploited the anchorage-dependence of MCF10A cells for a selective screening based on culturing GFP- or library-transduced cells on polyhema-coated plates. The four transduced populations were each split in two sub-lines: one was grown in adherence, the other underwent six cycles of selection, each cycle consisting of 48h of culture on polyhema followed by 24h of recovery on regular plates. Cells recovered from GFP- and library-transduced cells after selection were named, respectively, "GFP-SEL" and "LIB-SEL", and assayed for their ability to grow in the presence or absence of anchorage. LIB-SEL, but not GFP-SEL cells displayed significantly higher growth rate than unselected cells, in both adherence and suspension (Fig3.16B). Moreover, as shown in (Fig3.16C-D), only LIB-SEL cells could form large colonies in soft agar, an in vitro hallmark of cell transformation. These findings confirmed a "library effect" not explainable with insertional mutagenesis but likely deriving from the expression of advantageous exogenous transcripts.

Identification by "XenoArray" analysis of cDNAs conferring anchorage independence to MCF10A cells.

To identify library-derived transcripts likely to promote anchorage-independent growth, we conducted XenoArray analysis on library-transduced cells, before and after selection (Fig3.17A-B). We observed a significant number of probes giving a higher signal in selected cells, indicating that cells expressing the respective transcripts were enriched by the selection. To identify the genes that were reproducibly enriched in both selections we calculated, for each transcript, the log2ratio between its signal before and after selection. Interestingly, the Gab2 transcript showed a strong enrichment in both selections (average enrichment = 14-fold). The enrichment was observed with 3 different probes, each designed in a different region of the Gab2 transcript. Other

genes, including Ntrk3 and Cyp11a1, displayed a stronger enrichment in selection A (respectively, 41- and 49-fold in selection A, and 2- and 1.3-fold in selection B). Quantitative Real-Time PCR analysis with mouse-specific primers confirmed that Gab2 was the most enriched transcript, followed by Ntrk3 and Cyp11a1 (Fig3.17C). Therefore, we focused on this gene and validated its enrichment also at the protein level (Fig3.17D), thereby showing that the exogenous cDNA enriched after the selection actually encodes the full-length Gab2 protein. To verify whether exogenous Gab2 is essential for anchorage-independent growth of the LIB-SEL population, we downregulated it by RNA interference. Murine Gab2-shRNA transduction strongly reduced the growth advantage of LIB-SEL cells, both in adhesion and in suspension. Most importantly, these cells lost the ability to grow in soft agar, highlighting a potentially remarkable role for Gab2 in anchorage-independent growth of human mammary epithelial cells (data not shown).

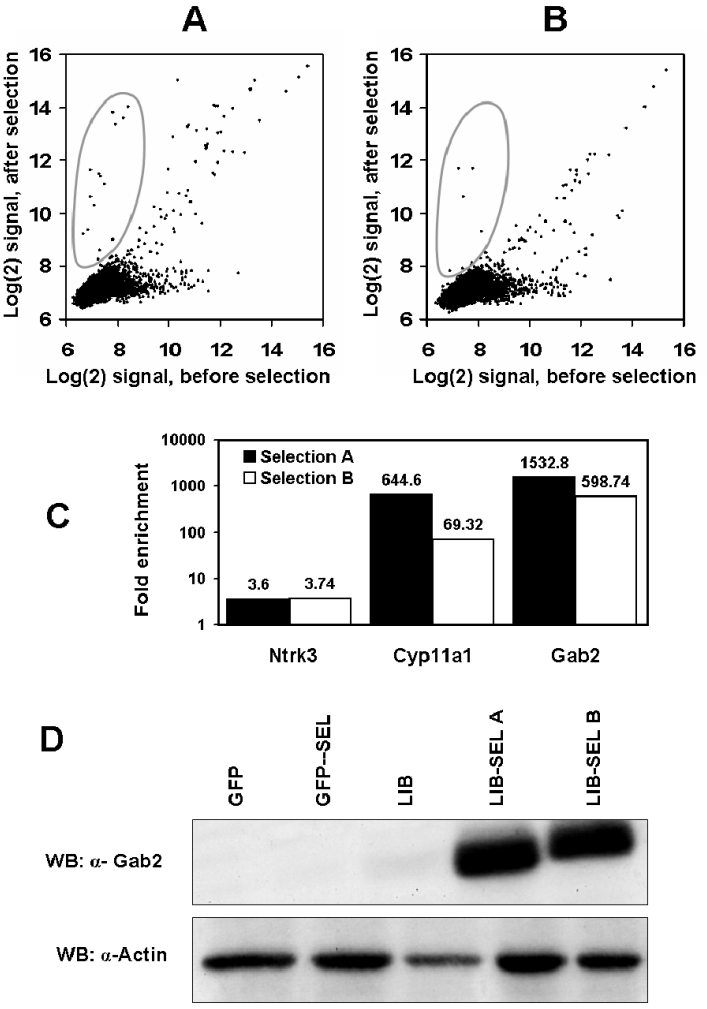


Figure 3.17: *Identification of enriched cDNAs in anchorage-independent, library-transduced MCF10A cells. (A-B) XenoArray analysis on library-transduced MCF10A cells before (x-axis) and after (y-axis) selection, performed on first infection (A) and second infection (B). The gray circles highlight enriched transcripts. (C) Real-time PCR validation of enriched transcripts in both selections. The y-axis represents the relative increase in abundance of the transcripts in selected cells compared to unselected cells. (D) Western blot analysis on GFP- and library-transduced cells before and after selection to detect Gab2 protein enrichment.*

3.3.2 Validation and characterization of Gab2-driven anchorage-independence.

Gab proteins, including mammalian Gab1, Gab2 and Gab3, comprise a growing family of scaffolding/docking molecules involved in multiple signaling pathways mediated by receptor tyrosine kinases (RTKs) and non-RTK receptors. As such, they function as assembly platforms that typically associate to the plasma membrane and bring together the various signaling proteins necessary to elicit a defined response to receptor activation.

To directly assess whether Gab2 may promote anchorage-independent growth, we transduced MCF10A cells with the human Gab2 coding sequence, cloned in a retroviral vector (gift of R. Daly;[138]). The levels of exogenous Gab2 were comparable to those observed in the LIB-SEL population.

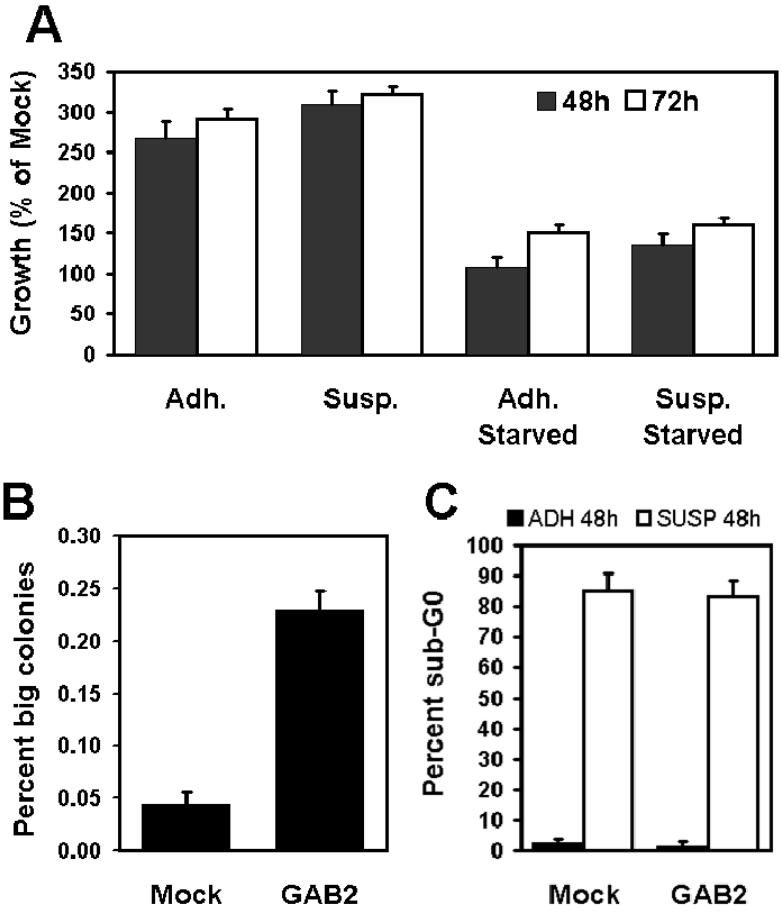


Figure 3.18: *Gab2 overexpression promotes anchorage-independent growth of MCF10A cells. (A) MTT growth assay on control and Gab2-expressing MCF10A cells in adhesion or suspension, in the presence of complete medium or of starving medium (no EGF and 2% serum) for 48h and 72h, as indicated. Cell vitality was normalized to the amount of viable plated cells at time 0, and then expressed as a ratio between mock and Gab2-expressing cells. The data represent the mean and standard error of triplicates. (B) Quantification of colony formation according to the 95th percentile of the colony size of control cells (Mock), as described in Materials and Methods. (C) Flow cytometry analysis of apoptosis induction for Mock and Gab2-expressing MCF10A. Cell death was measured after 48h either in adhesion or suspension, by assessing the number of hypodiploid nuclei with the DNAcon3 kit. The percent of apoptotic cells is reported on the y-axis.*

As shown in (Fig3.18A), adherent GAB2-overexpressing cells showed a significant increase in proliferation (2.6-fold and 2.9-fold, respectively at 48h and 72h), which was further enhanced in the absence of anchorage (3,1-fold and 3,2-fold, respectively at 48h and 72h). Notably, Gab2-driven growth advantage was almost totally lost when cells were kept in starving medium (no EGF, and serum lowered to 2%), indicating that Gab2 promotes proliferation independently from cell anchorage to the ECM, but dependently from the presence of EGF and/or serum. Accordingly, GAB2-overexpressing cells formed significantly larger and more abundant colonies in soft agar, compared to wild-type cells (Fig3.18B). This result is of particular significance because these cells were transduced with GAB2 but not selected in any way for growing in suspension. To evaluate whether Gab2 promotes anchorage independence by promoting survival of detached cells, we estimated the fraction of dead cells after 48h of suspension culture, by cytofluorimetric measurement of the propidium iodide (PI) signal, which detects plasma membrane integrity (Fig3.18C). Surprisingly, after 48h of polyhema plating, we detected a comparable extent of cell death between wild-type (85,21%) and Gab2-expressing cells (83,23%). These data indicate that Gab2 is not involved in the protection of MCF10A cells from anoikis, but rather allows their proliferation even in the absence of the consensus coming from adhesion to the ECM. As a further control, Gab2 downregulation by RNA interference abrogated growth in suspension of MCF10A cells transduced with exogenous Gab2. (not shown)

3.3.3 Gab2-driven anchorage independence requires Src and involves Stat3.

Scaffolding proteins have been shown to play a pivotal role in transducing signals from activated RTKs. Biochemical analyses and yeast two-hybrid screens have identified several signaling molecules that can bind to Gab2 upon receptor activation, including the tyrosine phosphatase Shp2, leading to activation of Erk and Jnk [16], the p85 subunit of PI3K, leading to Akt

activation [81], and Src family kinases [98].

Therefore, to dissect the signaling pathways downstream Gab2 that could mediate anchorage-independent growth, we examined the effects on cell vitality of a panel of small molecule inhibitors targeting the above mentioned signaling kinases (Fig3.19a).

The PIK3 inhibitor was the most effective, but with no differential between anchorage-dependent and independent growth, or between control and Gab2-expressing cells, showing a general requirement of this pathway for survival of MCF10A cells. The Mek inhibitor was slightly more effective on Gab2-expressing cells in suspension, but still inhibited growth quite well in WT cells. On the contrary, the Jnk inhibitor displayed modest effects in all conditions. Finally , the Src inhibitor displayed the highest specificity towards Gab2-expressing cells in suspension, while the growth inhibition observed on adherent cells indicated that Src is required for anchorage-dependent growth, regardless of the GAB2 status. A more detailed analysis of the effects of Src inhibition is shown in (Fig3.19B). According to these data, Gab2-driven anchorage independence requires Src, which typically is activated by integrins when cells are adherent and becomes inactivated upon detachment (17).

Consistently, western blot analysis on cell lysates from control and Gab2 expressing cells kept in adhesion or suspension confirmed Gab2-driven activation of Src and of one of its downstream targets, Stat3 (Fig3.19C). In adhesion, Gab2-expressing cells displayed a stronger basal phosphorylation of Src. Active Src levels were reduced in cells kept in suspension, but while in control cells Src activation was completely abolished at 48h, Gab2-expressing cells maintained some phosphorylation. Previous studies have demonstrated a biochemical and functional link between Src and Gab2 in driving EGF-dependent and independent proliferation.

However, this is the first evidence reporting a key role of Gab2 in promoting anchorage-independent growth via Src. Analysis of Stat3 activation highlighted a similar but more pronounced effect of Gab2 expression, indicat-

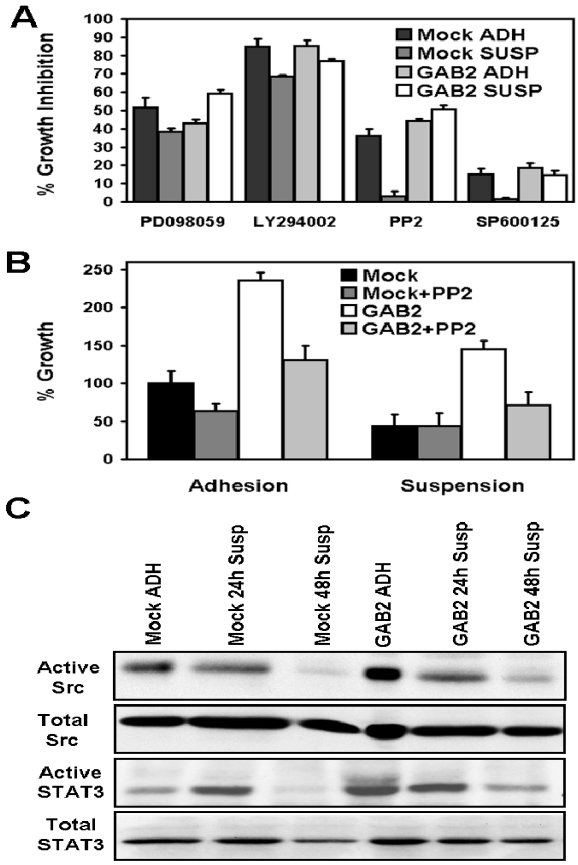


Figure 3.19: *Evaluation of the contribution of different signaling pathways to Gab2-mediated enhancement of cell growth. (A) Mock and GAB2-overexpressing (GAB2) MCF10A cells were incubated in adhesion (ADH) or suspension (SUSP) in the presence or absence of MEK inhibitor (PD98059, 40ÎijM), PI3K inhibitor (LY294002, 50ÎijM), Src inhibitor (PP2, 10ÎijM), or JNK inhibitor (SP600125, 10ÎijM). Cell vitality was assessed with the MTT assay after 24h from the treatment and the drug effect was expressed as percent growth inhibition (with respect to untreated cells). The data represent the mean and standard error of triplicate values from two independent experiments. (B) Detailed MTT analysis of the effects of Src inhibition by PP2 on cell growth in various conditions. Cell growth is expressed as percent of Mock adherent cells. (C) Western blot analysis on Mock and Gab2-expressing cells in adhesion or after 24h and 48h in suspension. Antibodies directed against the activated form of Src (phosphorylated at tyrosine 416) and Stat3 (phosphorylated at tyrosine 705), or total Src or Stat3 were used.*

ing the capacity of Gab2 to sustain the activation of Stat3 also in the absence of a substratum consensus. Since many studies provided evidence for Stat3 involvement in Src-mediated oncogenesis [138] and anchorage-independent growth [120], this data suggest that Gab2 could signal through Src and Stat3 to accomplish anchorage-independent growth.

3.3.4 Endogenous Gab2 is essential for anchorage-independent growth of normal and neoplastic cells.

The data shown so far demonstrate that constitutive, exogenous expression of Gab2 promotes anchorage-independent growth. To verify if this effect is mirrored by physiologically controlled Gab2 expression, we silenced by RNAi the endogenous Gab2 in MCF10A and in human cancer cells.

In MCF10A cells, Gab2 silencing markedly reduced their growth both in adhesion and in suspension (Fig3.20A) MDA-MB-231 breast cancer cells and MDA-MB-435 melanoma cells responded to Gab2 silencing with a modest reduction of proliferation (Fig3.20B), indicating a minor role of Gab2 in the presence of attachment. Strikingly however, the ability of these cells to form colonies in soft agar was almost completely abrogated by Gab2 silencing (Fig3.20C).

This result reflects a prominent role of Gab2 in anchorage-independent growth and possibly in the transformed phenotype of human cancer cells. In accordance with our previous western blot data, Gab2 loss determined a significant and concomitant decrease in Src and Stat3 activation (Fig3.20D). Interestingly, the most evident reduction of Src-Stat3 phosphorylation was observed for MDA-MB 231 cells, which endogenously express the highest levels of Gab2 and most strongly reduce their soft agar growth upon Gab2 silencing.

Altogether, these data confirmed the Gab2-Src-Stat3 axis as a key promoter of anchorage-independent growth of neoplastic cells.

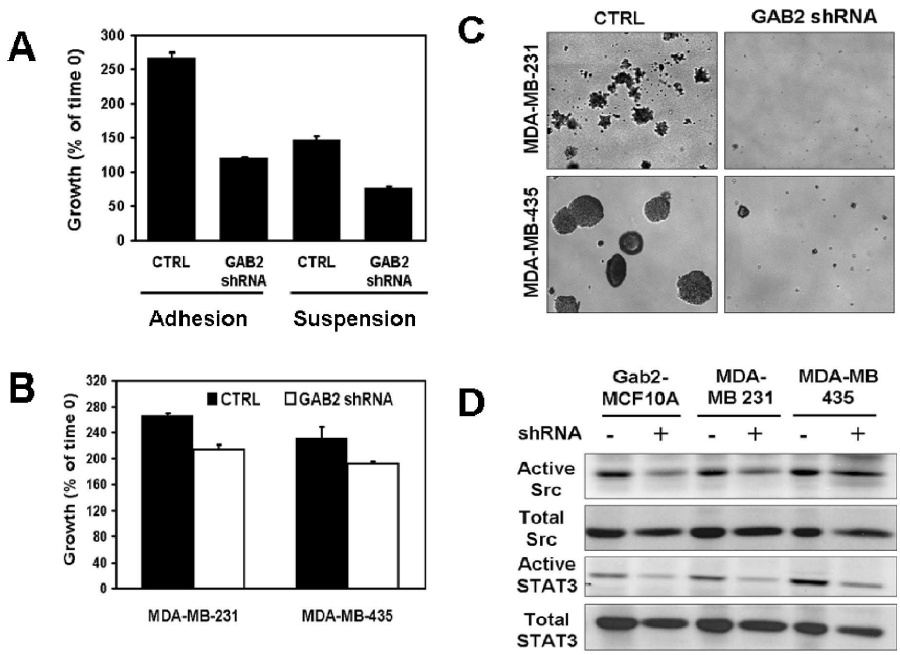


Figure 3.20: *Knock-down of endogenous Gab2 impairs MCF10A growth and anchorage-independent growth of human neoplastic cells. (A) MTT growth assay on wild-type MCF10A cells transduced with a scramble vector (CTRL) or a Gab2-shRNA in adhesion or suspension for 48h. Cell vitality was normalized to the amount of viable plated cells at time 0 and visualized independently for both cells. The data represent the mean and standard error of triplicate values. (B) MTT growth assay on MDA-MB-231 and MDA-MB-435 cells transduced with scramble vector (CTRL) or Gab2-shRNA in adhesion or suspension for 48h. Cell vitality was normalized to the amount of viable plated cells at time 0 and visualized independently for both cells. The data represent the mean and standard error of triplicate values. (C) Soft agar growth of cells expressing Gab2 shRNA or scramble vector (CTRL). Phase-contrast images were captured by a BD Pathway microscopic station (BD biosciences) after 3 weeks in agar. (D) Western blot analysis of Src and Stat3 activation in control and GAB2 shRNA-transduced cells, as indicated.*

3.3.5 A Gab2 transcriptional signature correlates with outside-inside signaling

Functional characterization Gab2 confirmed its role in promoting MCF10A growth in suspension when overexpressed. Interestingly, endogenous Gab2 downregulation in neoplastic cells abrogated their growth in soft agar, while only minimally affecting their adherent growth. These results support Gab2 as a pivotal player in the signal transduction involved in anchorage independent growth in mammary epithelial cells.

Therefore, to gain further insights on Gab2-driven anchorage-independence, we performed gene expression on MCF10A cells transduced with Gab2 and selected by growth in the absence of anchorage in respect to GFP transduced MCF10A after selection. This analysis highlighted 221 probes, corresponding to 205 independent genes: the "GAB2-signature". The signature revealed a significant enrichment ($p < 2.5 \times 10^{-5}$) in genes related to cell proliferation.

Among these genes we found several cyclins and cell cycle regulators essential for G1/S and G2/M phase transitions of cell cycle. Also gene involved in intracellular signaling were found to be enriched, and leucocyte trans-endothelial migration

Interestingly, the GAB2-signature was also significantly enriched ($p < 0.0023$) in genes whose expression distinguishes Dasatinib-sensitive and resistant breast cancer cell lines [63]. Dasatinib is a small molecule inhibitor of Src-family kinases employed for the treatment of leukemias, currently being assessed in multiple clinical trials on various solid tumors, including breast cancer and melanoma.

3.3.6 Construction of a Gab2 signature based classifier for breast cancer prognosis

To extend the value of the in vitro model of Gab2-driven anchorage independence, we verified if the "Gab2-signature" could be associated to human breast cancer aggressiveness. To this aim, the signature was mapped on a 311-sample breast cancer dataset generated at the Netherlands Cancer Institute on 2-color oligonucleotide microarrays (NKI dataset) and published in two works [132, 133].

After proper filtering, 150 array's probes were mapped to the Gab2-signature. Interestingly, the signature resulted to be strongly enriched in genes discriminating breast cancer patients with or without metastatic recurrence within five years from the initial diagnosis ($p < 10^{-5}$; see Supplementary Methods).

On the basis of these results, using the nearest-mean classifier approach [143], we built a classifier in the NKI dataset, which provides a "Metastasis Score" discriminating patients with good and poor prognosis. Functional data mining revealed two specific gene functional modules differentially associated with breast cancer prognosis in this dataset: (i) a sizeable proliferation module positively correlated to metastatic progression, and (ii) an interesting module composed of negative regulators of cell-matrix interaction, migration and invasion, expressed at higher levels in good prognosis samples.

To minimize the rate of false negatives (patients classified as "good prognosis" that instead developed metastasis within 5 years), we set the threshold for the "good prognosis" class at a Metastasis Score equal or lower than -0.15, (Fig3.21), with very significant p-value, ($p < 3.959715^{-06}$, log-rank chi square = 21.28425) and good specificity and sensbility for the good prognosis group 0.06349206 and 0.5412844.

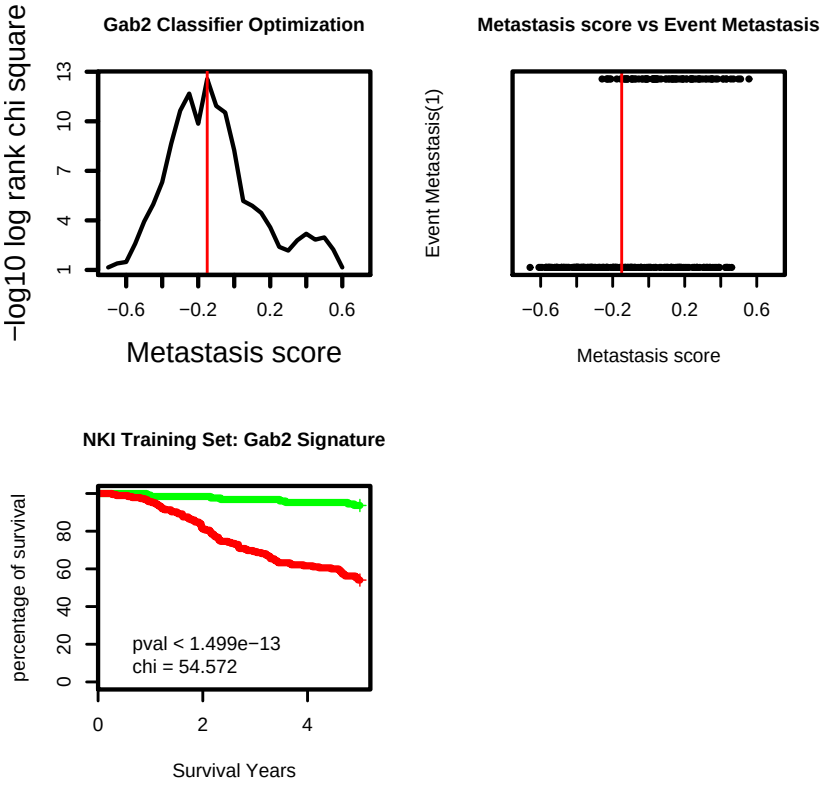


Figure 3.21: *Gab2 classifier tuning in Neatherland Cancer Institute dataset: We constructed the Gab2 classifier in a set of 311 patients. Here we calculated the median expression in good and poor prognosis. Metastasis score is calculated with Nearest mean classifier as a prognostic index, correlated with metastatic risk. We tried different threshold, and identified best thresholding at -0.15 MS. Finally it is reported the kaplan-meier analysis on the NKI breast cancer dataset according to gab2-signature classification [patients classified as good prognosis (green line) o poor prognosis (red line) by the GAB2-signature.]*

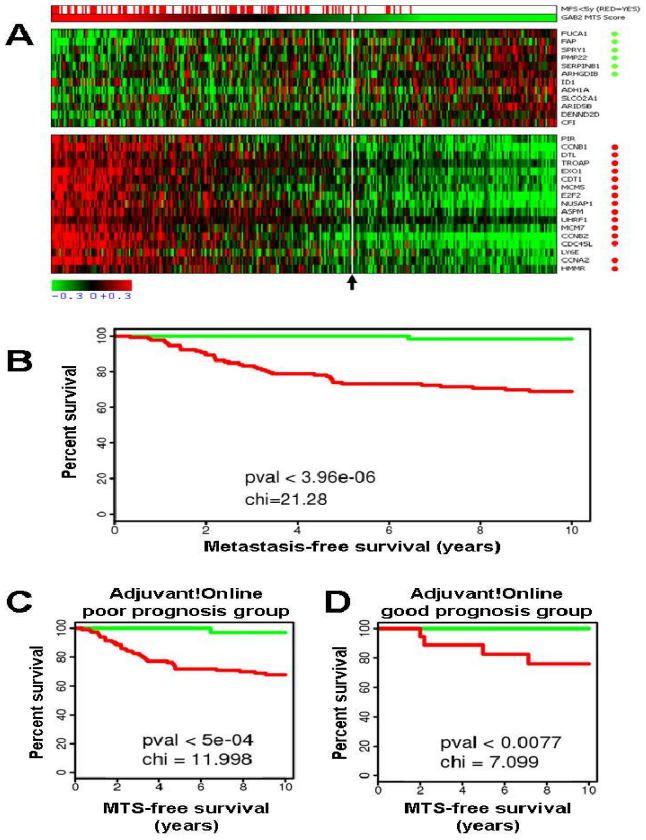


Figure 3.22: *The GAB2-signature predicts breast cancer metastatic relapse. (A) Heatmap showing the expression of the two main gene functional modules in the NKI-311 breast cancer dataset. The samples (columns) are ordered by decreasing GAB2-signature metastasis score (GAB2 MTS Score), which is graphically reported in the second row. The first row shows the occurrence of metastatic relapse within five year. The arrow at the bottom indicates the -0.15 threshold of metastasis score discriminating good and poor prognosis samples, also highlighted by a white vertical line crossing the heatmap. Green and red dots on the right highlight the genes annotated to the two functional modules, respectively downregulated and upregulated in poor prognosis samples. (B) Kaplan-Meier analysis on a dataset of 198 breast cancer samples classified as good prognosis (green line) or poor prognosis (red line) by the GAB2-signature. (C, D) Kaplan-Meier analysis on the same 198-samples dataset subdivided in two prognostic subgroups (C = poor prognosis, D = good prognosis) by the Adjuvant!Online clinical score. Each subgroup is then further subdivided by the GAB2-signature in good prognosis (green line) or poor prognosis (red line) samples.*

Validation of the Gab2 genomic classifier in independent breast cancer datasets.

It is possible that the good performance obtained in the training set would would reflect of systematic overfitting of the models, in the datasets. It is good practice in class prediction analysis to validate the results in independent datasets, as proposed by many statistical works [105]. Thus we acquired several Affymetrix breast cancer datasets from literature [35, 141, 93] and an in house built dataset, where we applied our genomic classifier to check the robustness of the performance. Finally we ensured the clinical relevance of our genomic classifier in respect to other published genomic classifiers.

The Gab2 classifier was mapped according to MAQC cross-mapping table to breast cancer dataset previously mentioned and proved of being capable to classify patients with high accuracy using the same threshold of Metastasis Score, despite the platform change ($p < 10^{-5}$). When samples were subdivided by their prognostic class according to standard clinical/pathological parameters, as defined by the Adjuvant!Online score [116], they could still be reliably further subdivided in good- and poor-prognosis subgroups by the GAB2-signature. Accordingly, when compared in univariate and pairwise multivariate analysis, the GAB2-signature largely outperformed the Adjuvant!Online score in predicting metastatic progression. Univariate and multivariate statistics were also calculated, on the same 198-sample dataset, for the Gab2 signature and three other published genomic prognostic classifiers: the 76-gene "Veridex index" [141], the 70-gene "MammaPrint"[132], for which the 198-sample classification is provided by the Transbig Consortium [35] and the "Genomic Grade Index", calculated according to the published methodology [121]. Also these analyses showed that the GAB2-signature provides strong and independent prognostic information. Clinical significance of the Gab2 signature was further confirmed on an additionals breast cancer datasets assessing its predictive value[93, 141]. (Fig3.23)

Table 3.3: TRANSBIG Paired multivariate analysis Gab2 vs Genomic Grade Index

	p	coef	lower .95	upper .95
Gab2	0.02	2.61	1.62	115.46
Ggi	0.15	0.69	0.79	5.06

Table 3.4: TRANSBIG Paired multivariate analysis Gab2 vs Veridex

	p	coef	lower .95	upper .95
Gab2	0.00	2.87	2.39	130.47
veridex	0.08	1.06	0.89	9.48

Table 3.5: TRANSBIG Paired multivariate analysis Gab2 vs Mammaprint

	p	coef	lower .95	upper .95
Gab2	0.05	2.07	0.97	64.44
MamPrint	0.05	2.07	0.97	64.65

Table 3.6: TRANSBIG Global multivariate analysis for all signature

	p	coef	lower .95	upper .95
Gab2	0.14	1.66	0.57	49.00
MamPrint	0.09	1.90	0.77	57.89
veridex	0.09	1.01	0.84	9.00
Ggi	0.61	0.25	0.49	3.33

To obtain an optimal performance in the ILM-76 samples, we needed to change the MS threshold to classify patients. There are may explanation for this results, it could be due to the change of platform standard, thereby a strategy similar to the one adopted for the invasive growth signature could lead to a better performance. Otherwise it could be a specific feature of the Gab2 classifier index that is able to specifically recognize a specific group

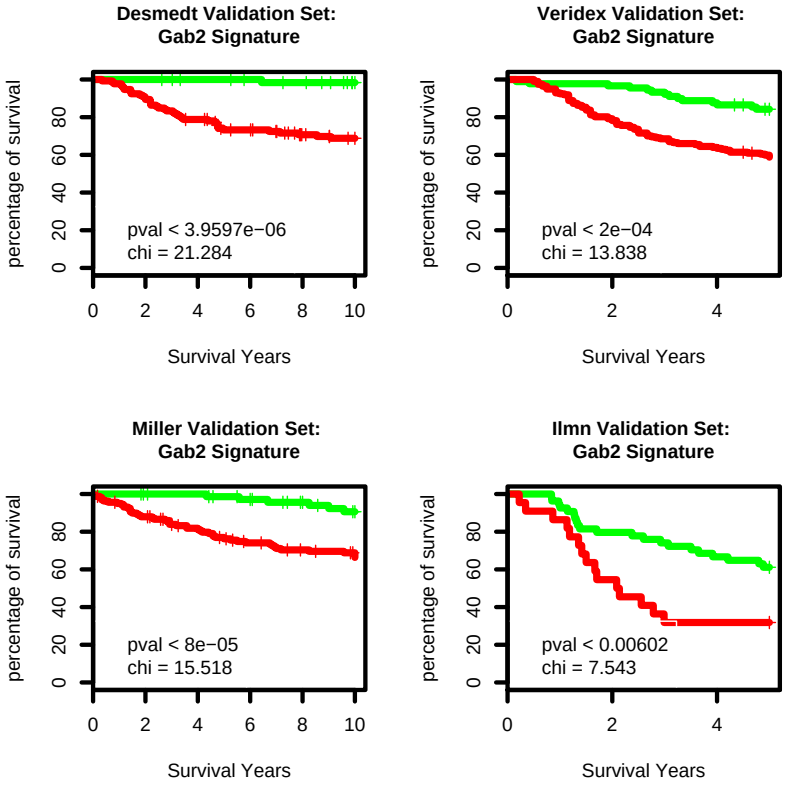


Figure 3.23: *Breast cancer datasets: Gab2 classifier performance on a large cohort of patients*

good prognosis patients, while this dataset is characterized by high frequency of aggressive disease.

Nevertheless in all the published dataset the Gab2-classifier proved to be a strong predictor with the high precision in the definition of good prognosis group, thereby userfull ”biomarker” in therapy decision making application.

3.4 NeoAdjuvant Reponse Prediction

In the previous part of the workwe could construct genomic prognostic classifiers for breast cancer metastatization, useful for therapy decision making, defining a group of good prognosis patients, in most of the cases with

good precision, and a group of poor prognosis patients, with good sensibility. These genomic "biomarker" along with published genomic classifiers, provide a rationale to spare adjuvant therapy to breast cancer patients and reduce overtreatment. Nevertheless, we still do not know which patients will respond

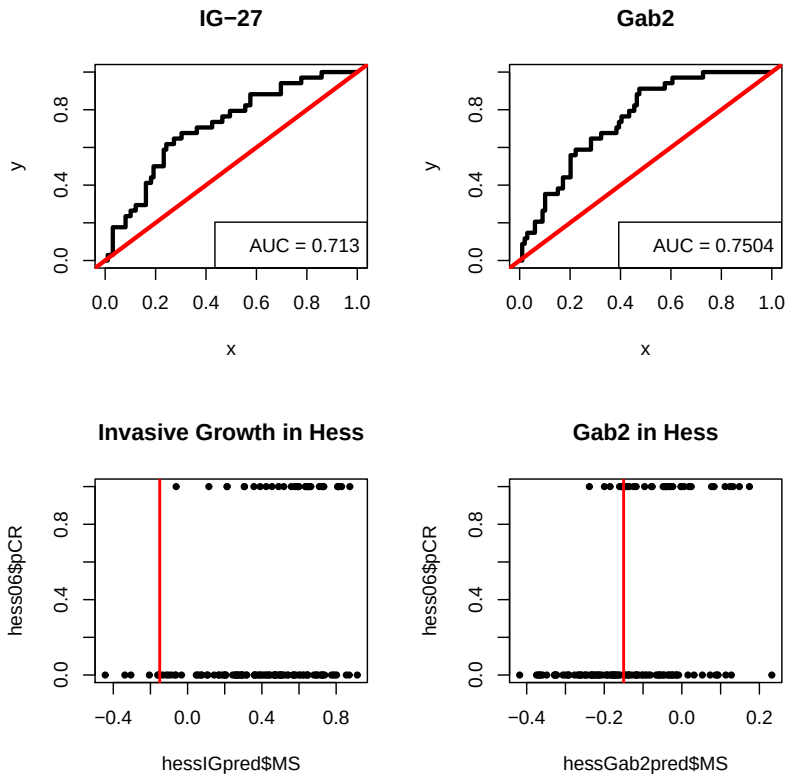


Figure 3.24: *neo-adjuvant prediction of the in vitro model classifier. ROC curves from both the invasive growth and the Gab2 signature show that the classifiers are able to discriminate RC patients from pCR. Application of the genomic prognositic classifiers permits the identification of a group of patients with low metastatic risk and low probability to respond to neo-adjuvant therapy*

to the standard treatment.

Recently Hess et colleagues [60] released a study in which they carried out gene expression profiling of 133 fine-needle aspiration specimens taken before neo-adjuvant treatment, in particular paclitaxel and fluorouracil-doxorubicin-

cyclophosphamide (T/FAC), After completion of the treatment, at surgery, they checked whether the patient responded or not to the therapy, and subdivided the response in two categories: pCR(34 patients), pathological complete response, for the patients that do not show any *in-situ* tumor at the end of the neo-adjuvant treatment and RC(99 patients), for patients with residual cancer. Thus we checked our prognostic classifiers are also associated with response to neo-adjuvant therapy (Fig3.25). Notably the patients with low MS, thereby the patients predicted to be good prognosis, with little metastatic recurrence probability curiously are also the patients that do not respond to neo-adjuvant therapy. Conversely pCR patients are mostly classified as poor prognosis. Indeed using the Gab2 classifier, we if we apply the Gab2 classifier we found that only 6 of the 60 patients classified as good prognosis responded to the therapy, while we could classify only 6 patients as good prognosis with the invasive growth signature and none responded the neo-adjuvant treatment. This result is controversial: it is quite obvious that response to neo-adjuvant therapy is an hallmark of good prognosis, however few patients with good prognosis will respond to neo-adjuvant therapy, it shound be noticed that as previously mentioned, our genomic classfiers are enriched for genes associated with celle proliferation, a key function for cancer onset and necessary feature for chemotherapy response.

Therefore we can suppose that the MS derived from our models are guided by a strong module of cell proliferation, which finally lead to a useful prognostic index to spare chemotherapy to patients, because it identifies two groups of patients: one characterized by low proliferation with low metastatic risk and therefore low opportunity to respond chemotherapy treatment, the other with high metastatic risk and higher probability to respond to neo-adjuvant therapy.

Chapter 4

Discussion

In the present study we characterized by gene expression profiling various *in vitro* model and assess involvement related to human cancer progression, leading to the construction genomic classifiers potentially useful in clinical therapy decision making for breast cancer.

The invasive growth signature, was defined in vitro on mouse liver stem / progenitor cells, was composed of genes whose expression is correlated to aggressiveness and metastatic relapse of various types of human cancer, thereby crossing species (mouse vs human) and tissue (liver vs breast, lung and prostate). Moreover, the signature yielded a robust classifier for metastatic relapse of human breast cancer and correctly classifying breast cancer samples also in a microarray platform different from the one used for training the classifier. The minimal overlap observed between the invasive growth signature and the other published signatures analyzed is not surprising, as poor overlap is typical also when the signatures are obtained with strategies much more similar to each other[141, 132]. In our case, such a divergence it is likely to reflect not only the differences in the strategies, but also the underlying biological significance. In this view, the functional readout of the IG-27 signature is of particular interest, as it highlights metastasis-prone tumours having a higher proliferation rate combined with a sort of "desensitization" to exogenous stimuli, be they extracellular ligands, steroid hormones or cell-cell

and cell-matrix contacts. This peculiar functional status is likely to derive from strong constitutive oncogenic stimulation, promoting both proliferation and negative feedbacks on inbound signals as recently highlighted by Amit and colleagues[**?**], and possibly suggests a therapeutic rationale impinging on both cell proliferation and cell signaling.

The second task of this work has to identify and functionally characterize miRNAs potentially modulating the HGF-driven invasive growth program in epithelial cells. To this aim, we conceived an innovative approach to prioritize candidate miRNAs, based on the analysis of expression profiles performed on MLP-29 cells activating invasive growth in response to HGF stimulation. It is commonly accepted that many miRNA target genes display a common regulation at the transcript level and in some cases they also exhibit common transcript decrease in response to miRNA expression. We reasoned that at least part of the transcript level changes in response to HGF could be further modulated by HGF-driven miRNA regulation. If this was the case, specific groups of HGF-regulated genes should contain abundant target sequences for specific miRNAs. To detect miRNAs potentially modulating HGF-target genes and thus the invasive growth response, we developed an algorithm calculating the abundance of miRNA-target sequences in the whole set of explored genes and estimating statistical enrichment of such sequences in small gene subgroups, namely, clusters of HGF-responsive genes. This enrichment analysis revealed that target sequences for miR-24 and miR-296 are particularly abundant in genes induced at 24 hours from HGF stimulation. The obtained enrichment leads to an interesting point as it is demonstrated that miRNA regulation typically drives transcriptional inhibition, not induction, we could partially explain this result if we consider that miRNAs are not the only regulators in the pathway driving from genes to proteins. Maybe in this case their transcriptional regulatory effect is covered by those triggered by HGF that result to be globally stronger. Real-time PCR analysis showed that only miR-24 was expressed in the cell line and more interestingly that its expression was regulated by met activation. Regarding possible robe for

miR-296 role in invasive growth, although such anyway is suggested by the functional experiments and this miRNA could be physiologically expressed in other cell lines. To test the role of our candidates on HGF-induced invasive growth activation we overexpressed them in MLP-29 cell line and we performed a scatter assay. Both miR-24 and miR-296 produced a higher sensitivity to HGF in scattering induction. To further dissect this information we tested separately proliferation and invasion capabilities produced by miRNAs deregulation. miR-24 revealed to be an inducer of cell motility in an HGF-dependent manner having only minimal proliferative effects, while miR-296 exhibited on both remarkable inducing effects. This is probably due to the fact that miR-296 is not expressed in MLP-29, therefore its exogenous expression gives stronger effects. Experimental results led to the necessity of further analysis of miRNA targets as they highlighted a possible circuit of coordinated up-regulation of one miRNA and of its targets that is somehow puzzling if we thought to miRNAs as simple cooperator in mRNA down-regulation. Experimental validations on miR-17 upon c-myc induction although demonstrated the existence of these kind of circuit[34, 101]. Loops like the one supposed in this study found a biological role in the prevention of noise-driven cell program activation and can be classified as feed-back miRNAs circuitry according to the criteria individuated by Tsang in a study about miRNA driven networks [129]. The careful analysis of miR-24 targets identified Csk and Talin as interesting candidates to explain this paradoxical concomitant induction of a gene and a miRNA repressing it, leading to hypothesize the circuit displayed in (Fig4.1). At 1-6 hours (Fig4.1A), both miR-24 and c-Src are induced by HGF. This led to a strong Src activation via a direct biochemical pathway, downstream Met, and an indirect pathway involving down-modulation of its inhibitor Csk by miR-24. This activation is tightly time-dependent, since at later times (Fig4.1B) miR-24 returns to basal level and its target Csk is upregulated at the transcript level. A further enhancement of the negative feedback loop impinging on Src is produced through the transcriptional down-modulation by HGF of one of its main ac-

tivators, Shp2.

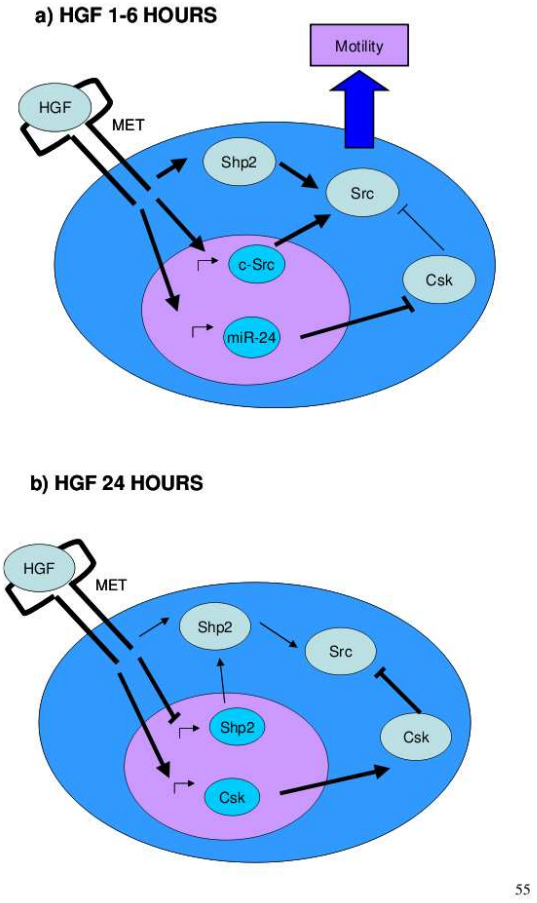


Figure 4.1: *Invasive growth feed back loop involving miR-24 and its target Csk*

As represented in the next picture, c-Src induction at 1 hour altogether with the concomitant Talin down-modulation driven by miR-24 could lead to a transient increase of cell fibres formation, taking advantage also on focal adhesion relative decrease, to establishing the dynamic background needed to produce invasive phenotype at 24 hours.

At this time miR-24 decrease altogether with transcriptional induction of Talin and c-Src inhibitor produces the inversion of the equilibrium between

the pathways leading to focal adhesion and stress fibres formation. The new equilibrium settled by HGF at later time-points could help in the establishment of invasive phenotype which becomes clearly visible on MLP-29 after 24 hours of HGF treatment. Therefore the thigthy time regulated miRNA expression could be essential to modulate transcriptional level in order to permit the initial higher motility propension needed to produce epithelial disruption and then to establish the mesenchimal transition driven by HGF.

This study validated SeaMT software results and demonstrated miR-24 and miR-296 involvement in invasive growth regulation through the increase of HGF sensitivity and the induction of higher proliferation and motility abilities. Finally the effects uncovered in experimental assays had to the definition of a possible miRNA-driven circuitry needing to be tested in further experiments, altogether demonstrating the role of miRNA as cell specific modulators of RTK response.

Future developments will certainly involve the targets validation through c-Src inhibition and western blots of Talin, p-c-Src and Csk during the HGF stimulation time-course.

We would also like to test the role of our miRNAs in other cell models expressing met receptor and possibly displaying basal expression levels of miR-296 and not of miR-24, to extend the validation of our results and compare the effects of these miRNAs starting from similar physiological expression levels. From the bioinformatics point of view we already start to look for the enrichment of miR-24 and miR-296 targets on other gene expression profiling and to look for their expression level on tumoral expression profilings.

In supporting to these results miR-24, targets were also found to be enriched in subsets related with poor prognosis. Together with the enrichment of invasive growth this results suggest that the miR-24 could be involved in the regulation of physio-patological program involved in cancer onset and progression. The classifier constructed on these genes was found to partition breast cancer, albeit not better than available classifiers. In agreement with

the construction of the invasive growth classifier it would be possible that prioritization of the genes most discriminating human cancer pathologies would improve the performance of the genomic classifier. This could be achieved either with the prioritization of the genes inside the cancer dataset, or by the interpolation of predicted target for miR-24 from different prediction algorithms.

After successfully constructing genomic classifiers for breast cancer starting from *in vitro* models quite far from the human disease, we focused on a closer *in vitro* cell model, MCF10A, to construct a model for anchorage independece; to ensure successful metastatic dissemination, malignant cells must acquire the ability to grow in the absence of their environment of origin. In fact, the capacity of cells to survive and proliferate in vitro in the absence of integrin-mediated adhesion strongly correlates with tumorigenesis in vivo and may enable tumor cells to metastasize and grow at inappropriate sites in the body[32].

We hereby describe a key role *in vitro* anchorage-independent growth of Gab2, a multiadaptor protein devoid of enzymatic activity. Gab2 is a member of the Grb2-associated binding protein (GAB) gene family [55]. They are so called "scaffolding" or "docking" proteins because of the presence of multiple functional motifs mediating interactions with many other signaling molecules [96]. GAB proteins are involved in signaling events triggered by a variety of stimuli, including GFs, cytokines, G-coupled receptors and T- and B-lymphocyte antigens, ultimately regulating cell growth and differentiation [16, 74]. In particular, Gab2 was found to mediate transformation driven by oncogenes impinging on the PI3k/Akt and Ras/Erk pathways [111]. Among the Gab2 direct interactors are proteins with key roles in human cancer when mutated, such as PI3K and the tyrosine phosphatase Ptpn11 (also known as Shp2). Recent work suggested that the oncogenic properties of Shp2 mutant proteins require signal enhancement by Gab2 [148]. Interestingly, Gab2 maps to a chromosomal region (11q13) amplified in 10-15% of breast cancers and its overexpression was confirmed in several breast cancer

cell lines [31]. The role of Gab2 in mammary tumor metastasis was also explored and confirmed in mouse models [77]. We now show that in vitro Gab2 strongly promotes anchorage-independent growth of MCF10A normal breast cells. This information extends the previously described growth-promoting activity of Gab2 in the same cells when adherent [**?**]. We also found that Gab2-driven anchorage independence is not due to a protection from cell death upon detachment. This finding was unexpected, given the fact that Gab2 potentiates the PI3k/Akt and Ras/Erk pathways, but it is in line with previous reports indicating that Gab2 does not prevent apoptosis of luminal cells during morphogenesis of MCF10A cells [10]. Moreover, our experiments using small molecule inhibitors showed that the PI3k/Akt and Ras/Erk pathways are required for MCF10A survival and proliferation independently of the adhesion status and of Gab2 expression. A much more specific role was found to be played by Src, whose inhibition had no effect on wild-type cells in suspension, but strongly impaired their adherent growth, confirming that Src conveys the proliferative consensus provided by integrin engagement [98]. However, Gab2-expressing cells required Src activity also in suspension, providing a strong rationale for Src involvement in Gab2-driven anchorage-independence. The biochemical link between Gab2 and Src can be provided by Shp2, previously described to directly bind Gab2 [81] and to activate Src[149]. It is of particular interest that Gab2-sustained growth in suspension was impaired when cells were cultured in the absence of EGF, indicating that Gab2 can only overcome the lack of adhesive consensus in the presence of an upstream signal from GFs. Recently, Gab2 was found to promote GF independence, but in cooperation with oncogenic Src [9]. Therefore, Gab2 can rescue cells from the need of two concomitant proliferative stimuli - activated GF receptors and activated Src when at least one of the two is present at sufficiently high levels. It should be noted that the dual consensus system involving GF receptors and integrins/Src is not the result of two parallel and totally independent pathways. Rather, several points of reciprocal influence and cross-talk have been described [46], and Gab2 can be

a key rheostat and integrator, allowing for a "spillover" of the signal across the two pathways.

In this view, proliferation of non-adherent MCF10A cells could be promoted by a GF-driven direct activation of the Erk and Akt pathways and indirect, Gab2-mediated activation of Src and of its downstream signaling molecules, in particular Stat3. Indeed, Stat3 has been already involved in Src and Jak1-driven proliferation of human breast carcinoma cells[50] and in anchorage-independent growth of cancer cells [150]. Moreover, Gab2 was found to contain a functional Stat3 binding motif promoting its recruitment and activation [95]. Our biochemical data confirmed the contribution of Src and Stat3 to Gab2-mediated anchorage-independent growth. Gab2 expression increased Src and Stat3 phosphorylation both basally and after prolonged suspension culture. Moreover, Gab2 downregulation by RNAi led to reduction of Src and Stat3 activation not only in MCF10A cells constitutively expressing exogenous Gab2, but also in neoplastic cells loosing anchorage independence as a consequence of endogenous Gab2 silencing. In line with this, our gene expression analysis showed that a significant fraction of the genes whose expression is regulated by Gab2 are also differentially expressed in breast cancer cells resistant or sensitive to Dasatinib, a Src-family kinase inhibitor. Overall, these findings configure a central role of Gab2 in anchorage independence, and could provide a rationale for developing a predictor of response to Dasatinib or other Src-family inhibitors in breast cancer based on the levels of Gab2 and/or of its target genes.

It is noteworthy that the role of Gab2 in anchorage-independent growth emerged within the context of a high-throughput selective functional screening, in which this gene competed with several thousand others. Apart from GAB2, the analysis revealed a reproducible enrichment also for a well-known transforming gene, NTRK3, previously found to play a key role in anoikis resistance [53], and for a cytochrome P450 superfamily member, CYP11A1, whose polymorphisms are associated with breast cancer risk [139]. Therefore, these two genes could represent internal controls of the screening and further

confirm its effectiveness.

Finally, it is of particular significance that the in vitro-defined Gab2 transcriptional signature yielded a robust classifier for metastatic relapse of human breast cancer. Of the two key functional modules found in the signature, the proliferation module, positively correlated with metastasis, adds further informative genes to the already described core of proliferation genes associated to breast cancer progression [146]. More novel is the module of genes negatively correlated with metastasis. For six genes of the module, existing information allows drafting a unifying view, comprising extracellular, transmembrane and intracellular components that negatively affect cell-matrix interaction, migration and invasion: (i) two extracellular proteins that, respectively, inhibit matrix-degrading proteases and remove fucose from ECM glycans, thereby impairing ECM binding and invasion by cancer cells [140]; (ii) the gene, encoding an integral membrane protease whose expression by stromal cells is negatively correlated with metastatic progression [4]; (iii) that encodes a 4-spanning integral membrane protein widely expressed at apical junctions of epithelial cells, increasing transepithelial electrical resistance and decreasing migration[108]; (iv) negative regulators of key signaling systems promoting, respectively, cell migration and cancer progression: Rho-family GTPases and tyrosine kinase receptors like EGFR and FGFR. These two functional modules, within the context of the GAB2-signature, altogether generate a prognostic classifier predicting metastatic progression with high accuracy, independently from existing genomic signatures and outperforming the clinical-pathological prognostic parameters currently integrated into the Adjuvant!online web tool to provide indication for adjuvant chemotherapy [116]. These results pave the way to further validation of the GAB2-signature on large breast cancer cohorts to confirm its potential clinical usefulness.

4.1 Perspective and conclusions

we successfully constructed genomic signature from our *in vitro* models that yielded robust classifiers for metastatic relapse of human breast cancer and correctly classifying breast cancer samples also in a microarray platform different from the one used for training the classifier. These classifiers are of particular interest considering that they are not only able to discriminate breast cancer progression, but they could also predict the response to neo-adjuvant therapy. This aspect is extremely useful in the context of therapy decision making being able to identify a large subgroup of patients to which chemotherapy should be spared, having low metastatic risk and low probability to respond to neo-adjuvant treatment.

Actually these results validated also the *in-vitro* models, and our "bottom-up" workflow to construct genomic classifiers driven by a cellular model. In fact it is notable that starting from a restricted number of gene we could achieve classifying performance competing with and in some cases overcoming the published genomic classifiers, FDA approved, designed with "top-down" approaches.

Exploiting the experience developed in the process, we are standardizing our pipeline in which gene lists derived from in vitro biological models related to cancer progression enter an automated workflow (implemented in R-bioconductor) exploiting several published datasets to construct and validate classifiers predicting metastatic relapse. Significant relationship between the signature and breast cancer prognosis is established by confronting the performance of the signature-derived classifier with a distribution of random performances obtained in a Montecarlo simulation carried out on random gene lists of the same size of the signature. Here we propose a prototype in which we assess the performance of a Gab2 classifier, constructed on several dataset and finally validated on the Transbig validation dataset. On the basis of this simulation we can confirm the validity of the classifier, and set a reference for comparison of the classifier performance.

All these classifiers need extensive validation of their performance in a

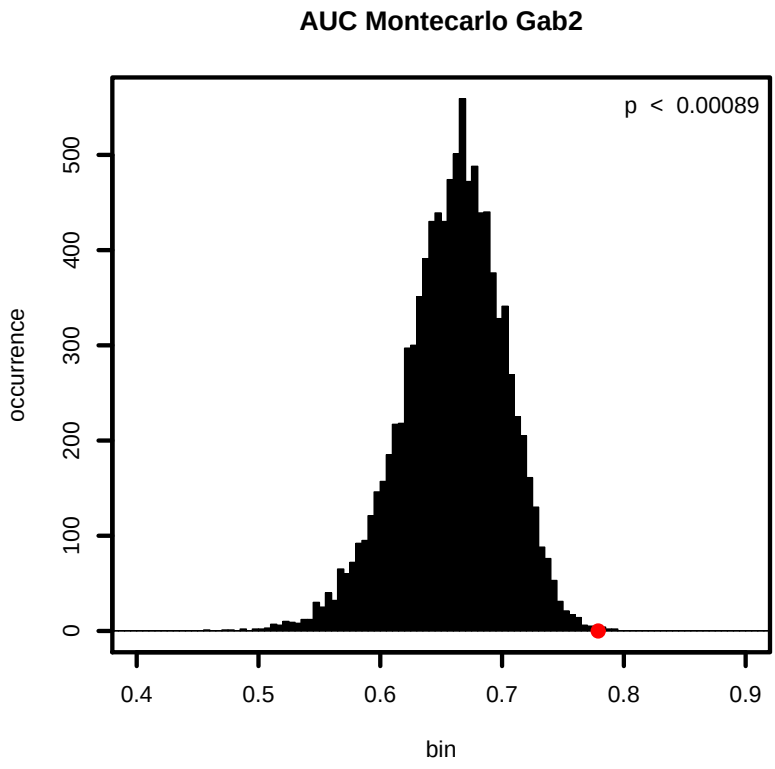


Figure 4.2: *pipeline Prototype: gene from Gab2 signature are exploited to construct a NMC calculated on several dataset. The classifiers AUC obtained with this strategy outperforms the results obtained from the construction on one single dataset (data not shown), and it is not a common event out of a randomization of 10.000 permutations*

larger cohorts of patients, for their prognostic significance and for their ability to predict response to chemotherapy, in particular we want to verify if we have a specific predictor for paclitaxel and fluorouracil-doxorubicin-cyclophosphamide (T/FAC), or if our signature yield a wider spectrum of sensibility. Essentially it would be interesting to characterize the patients subgroups and find correlation between genomic signature and clinical-pathological parameters that may address the activity of RTK activity driven by invasive growth or by Gab2 activation. In this frame it is a major goal to construct genomic classifier for specific subgroups of clinical relevance, such as ER- patients. Finally we may be interested in investigating for the rebuilt of classifiers for specific breast cancer subsets. The final aim of the work should be providing a tool for clinical routine, therefore we are planing to validate the performance of the classifiers in microfluidic cards, a technology which is much more portable in clinic in respect to microarray technology.

A broader perspective would be to extend our investigation to other level of gene regulation. Albeit we designed a bioinformatics tool to identify microRNAs involved in the regulation of gene expression, in this study we mainly focused on construction of genomic classifier on the basis of gene expression. Actually only a small fraction of -OMIC technologies have been explored and there are more than few evidences, [135, 131, 21] that integration of several feature in cell biology, such as protein translation, microRNA expression, could provide best insight in complexity of pathological transformation of normal cell during cancer onset and progression.

Chapter 5

Material and Methods

5.1 ILM-76 dataset

RNA extraction and processing Total RNA was extracted using the Trizol Plus purification Kit (Invitrogen, cat.no.12183555), according to the manufacturer's protocol. RNA quantification and quality assessment was performed on a Bioanalyzer 2100 (Agilent). Synthesis of cDNA and biotinylated cRNA was performed using the Illumina TotalPrep RNA Amplification Kit (Ambion Cat. n. IL1791), according to the manufacturer's protocol. Quality assessment and quantification of cRNAs were performed on Bioanalyzer 2100. Hybridization of cRNAs (1500 nanograms) was carried out using Illumina Beadarrays (Human_Ref8_V1). Array washing was performed using Illumina High-stringency wash buffer for 10 min at 55, and followed by staining and scanning according to standard Illumina protocols. Probe intensity data were obtained using the Illumina BeadStudio software, and further processed with R-Bioconductor, and lumi package[104, 52] and Excel software.

5.2 Dunnett's T-test

To identify genes differentially expressed in multiple timepoints we employed the Dunnett's T-test [40], an inferential parametric test designed to compare

the mean of each of several experimental groups with the mean of a control group. A simple description of the properties of the Dunnett's T-test can be found at http://davidmlane.com/hyperstat/B112114.html The test evaluates the hypothesis, in our case the change of log2expression values, by means of an estimation of the mean square error within groups, corrected by the harmonic mean of the sample numbers. The test was performed on log2scaled values, with m = 1. In the standard DunnettâĂŹs test m is absent and the t threshold for significance can be derived from the Dunnett's t tables, available for example at http://davidmlane.com/hyperstat/table_Dunnett.html. In our case, with m different from zero, we had to estimate the correct t value by running the test iteratively on groups from permutations of experiments, thereby estimating the False Discovery Rate (FDR;. Indeed, our FDR analysis showed that the Dunnett's test with m different from zero is more powerful, and much more reliable than classical T-test. To prioritize differentially regulated genes, we choose t = 2, with an estimated alpha (FDR) of <0.05 according to the median distribution of 5000 randomly permutated datasets.

5.3 RNA extraction and processing for microarray analysis of Gab2 unselected and selected cells

RNA from GFP transduced cells, GAB2 transduced cells before and after the anchorage independent-growth selection was extracted using the TRIzol reagent (Invitrogen), according to the manufacturer's protocol, and then further purified using the RNeasy Mini kit from Qiagen. The quantification and quality analysis of RNA was performed on a Bioanalyzer 2100 (Agilent). Synthesis of cDNA and biotinylated cRNA was performed using the Illumina TotalPrep RNA Amplification Kit (Ambion Cat. n. Probe.Illumina1791), according to the manufacturer's protocol. Quality assessment and quan-

tification of cRNAs were performed on Bioanalyzer 2100. Hybridization of cRNAs (20 micrograms) was carried out using Illumina Human 48k gene chips (Human_Ref8_<V2).Array washing was performed using Illumina High-stringency wash buffer for 30 min at 55, and followed by staining and scanning according to standard Illumina protocols. Probe intensity data were obtained using the Illumina BeadStudio software, and further processed with R-Bioconductor [52]. Probe intensity data were obtained using Illumina BeadStudio 1.5.1.3 software, and further processed with R-Bioconductor and Excel software. Data was normalized according to Rank-Invariant and a filter for detection was applied in order to discard genes with a Detection Score beyond 0.99 in at least one condition. Filtered data were scaled to remove negative expression values from the analysis.

5.4 SeaMT, signature enrichment analysis for microRNA targets

We developed an R-Bioconductor [104, 52] package for microRNA binding sites enrichment analysis, SeaMT, depending on two available libraries, Heatplus and biomaRt. This package takes advantage from the published database for microRNA binding sites such as Miranda, in the version provided by Diana [90]. The core function of the package compares the microRNA targets occurrence in a set of genes in respect to a background defined by the user. The analysis is subdivided in 4 steps:

1. Definition of the background according to genes expressed in the model
2. Definition of the signature of interesting genes
3. Mapping the background and the signature to microRNAs binding site database and count the number of occurrences of each microRNAs binding site in the two lists of genes.
4. Enrichment analysis of microRNA targets according to hypergeometric distribution.

SeaMT is based on database of putative targets biased by high level of false

positive. To overcome this limits we applied very stringent selection criteria. First of all we corrected for multiple testing with Bonferroni correction according to the number of microRNA in the database; second we considered only those microRNA emerging from enrichment analysis also in the human orthologous genes. Finally we exploited the analysis onto independent microRNA binding site database "PITA" [78] to verify the significance of the candidates.

5.5 Functional annotation and functional enrichment analysis

Probe identifiers contained in the annotation manifest provided by Illumina were loaded on the David Ease portal (5) to generate a background list (all probes) and the GAB2-signature list. Enrichment in biological functions for the GAB2-signature genes was evaluated using the "functional annotation chart" function on the portal.

5.6 Enrichment in genes discriminating the Dasatinib response

To indentify molecular markers predictive of response to Dasatinib, Huang and colleagues conducted gene expression profiling using Affymetrix HGU133 arrays on a panel of 23 breast cancer cell lines, either resistant (16 lines) or sensitive (7 lines) to Dasatinib (GSE6569). For each probe in the array, the differential expression between Dasatinib-sensitive and resistant cells was calculated using the Signal to Noise ratio [54]. To assess the enrichment of the GAB2-signature in genes with high SNR, Gene Symbols corresponding to the signature were mapped on the dataset, resulting in 356 Affymetrix probe sets. Subsequently, the number of signature genes with absolute SNR values falling in the top 5% (or 10% for miRna enrichment) of all the dataset

was counted.

5.7 Enrichment in genes discriminating good and poor prognosis breast cancer.

For metanalysis on breast cancer microarray data, two public available data sets from the Netherlands Cancer Institute [132, 133] (NKI; http://www.rii.com/publications/2002 were used and merged into a unique 311-sample dataset (NKI-311). The data were filtered to remove probes whose signal and standard deviation never reached the 50th percentile of their respective distributions. Further filtering was applied on probes for which more than 99% of the expression values were missing. The probes were annotated with gene symbols obtained via Unigene (release Hs 204), and for each of them the SNR between good- and poor-prognosis samples (absence or presence of metastatic relapse within 5 years) was calculated in the NKI-311 dataset as described above. After mapping the GAB2-signature on this dataset via Gene Symbols, its enrichment in genes with high SNR was calculated as described above (observed/expected = 2.8, hypergeometric p.val $< 10^{-5}$).

5.8 Classification of Breast cancer samples

To generate a classifier for breast cancer patients, we applied the nearest mean classifier approach [143]. Briefly, we calculated for each gene of the GAB2-Signature the median expression in the good and poor prognosis subgroups of the NKI-311 dataset. For a more accurate calculation of the median expression, the data were bootstrapped (1000 bootstraps each including a random selection of 80% of subgroup samples). The classifier is therefore composed of the lists of the signatures' genes mapped on the NKI dataset and, for each gene, the median expression values (means) for the good and poor prognosis groups. To classify samples, a "Metastasis Score" (MS) is then calculated, based on the GAB2-signature genes, by subtracting the Pearson correlation

with the good prognosis mean from the Pearson correlation with the poor prognosis mean. The MS is therefore directly proportional to the risk of metastatic relapse within five years. To further map the GAB2-signature on independent breast cancer datasets obtained on different microarray platforms, we used a univocal cross-mapping table generated by the Microarray Quality Control (MAQC) consortium [87] and applied it on an independent dataset of 198 samples from the TRANSBIG consortium [35]. To reach homogeneity in data structure, to properly apply the NMC obtained in the NKI-311 dataset, the Affymetrix log2 expression signals of the 198-sample dataset were converted to log2ratios, using as reference the median expression calculated in the same dataset. Univariate and multivariate analyses conducted on the validation datasets.

5.9 Cell Culture and Reagents.

MCF10A cells were obtained from ATCC and cultured as described [138]. MDA-MB-231 and MDA-MB-435 cells were obtained from ATCC and cultured in DMEM (Gibco) supplemented with 10% fetal bovine serum (Sigma). The antibodies used were: anti-Gab2 (Upstate Biotechnology), anti- Tyr416-phosphorylated Src (Cell Signaling), anti-total Src (Cell Signaling), anti-Tyr705-phosphorylated Stat3 (Cell Signaling), anti-total Stat3 (Cell Signaling), goat anti-actin (Santa Cruz). A mouse testis retroviral expression library together with pFB-hrGFP retroviral supernatant, packaged in the VSV envelope were purchased from Stratagene (ViraPort, Cat n. 972300) and used to infect 1.5x105 MCF10A cells in 60mm tissue culture plates using 10μg/ml DEAE-dextran (Amersham Bioscence). GFP expression analysis was performed after 48 hours using a FACS Calibur flow cytometer (Becton Dickinson). Retroviral expression vector for GAB2 in pMIG (also known as pMSCV-IRES-GFP) was a gift of R. Daly. Virus production and transduction were performed as described [138]. Lentiviral shRNA expression vectors against murine or human GAB2 were purchased from Sigma (MISSIONâĎć

TRC shRNA Target Set), together with the pLKO.1-puro Control Vector. Viral supernatants were obtained according to the manufacturer's protocol. Infected cells were selected by puromycin treatment (2μg/ml for one week). Specific shRNA sequences which efficiently downregulated Gab2 protein, as assessed in Western blot, were the following: (i) Murine: CCGGCCGACACAATACAGAATTCAACTCGAGTTGAATTCTGTATTGTGTCGGTTTTTG; (ii) Human: CCGGCAGCCAACTCTGTTCACGTTTCTCGAGAAACGTGAACAGAGTTGGCTGTTTTTTG.

5.10 XenoArray analysis

RNA was extracted using the TRIzol reagent (Invitrogen), according to the manufacturer's protocol, and then further purified using the RNeasy Mini kit from Qiagen. Quantification and quality analysis of RNA was performed on a Bioanalyzer 2100 (Agilent). Synthesis of cDNA and biotinylated cRNA was performed using the Illumina TotalPrep RNA Amplification Kit (Ambion Cat. n. IL1791), according to the manufacturer's protocol, with previously reported variations [85]. Hybridization was carried out on Illumina Mouse6_V1 arrays, using 1.5 μg of cRNA.

5.11 Anchorage-independent growth selection.

Polyhema-coated 10cm Petri dishes were prepared by applying 4ml of a 12mg/ml solution of poly-hydroxy-ethyl-methacrylate (polyhema; Sigma) in ethanol, drying under tissue culture hood, repeating the application once and incubating the plates overnight at 37. For the selection, 3 x 106 trypsinized cells were plated onto polyhema-coated plates in complete growth medium. Cells were cultured in suspension for 48h then were let to recover on regular plates for 24h before repeating the selection cycle.

5.12 Western Blot.

Cell lysates from 2-5ÃŮ106 cells were prepared in RIPA buffer (150 mM NaCl, 1% NP40, 0.5% DOC, 50 mM TrisHCL at pH 8, 0.1% SDS, 10% glycerol, 5 Mm EDTA, 20 mM NaF and 1 mM Na3VO4) supplemented with 1 μg/ml each of pepstatin, leupeptin, aprotinin, and 200 μg/ml phenyl methylsulphonyl fluoride (PMSF). Lysates were cleared by centrifugation at 12,000 rpm for 20 min at 4and normalized with the BCA Protein Assay Reagent Kit (Pierce). Extracts were run on SDS-polyacrylamide gels, transferred onto nitrocellulose membranes (Hybond; GE Healthcare) and incubated with different antibodies overnight at 4. Nitrocellulose-bound antibodies were detected by the ECL system (GE Healthcare).

5.13 Real-time PCR.

Two micrograms of total RNA extracted with TRIzol reagent (Invitrogen) were reverse transcribed with the High Capacity cDNA Reverse Transcription Kit (Applied Biosystems, Foster City, CA). Quantitative Real-time PCR with Sybr Green was performed on the ABI Prism 7900HT Sequence Detection System (Applied Biosystems, Oak Brook, IL). PCR primers were designed with Primer Express software (Applied Biosystems) against mouse-specific regions of the transcripts (except for the PGK housekeeping gene) in order to monitor the expression of only library-derived transcripts: hPGK: sense 5'-CTTATGAGCCACCTAGGCCG-3'; antisense 5'-CATCCTTGCCCAGCAGAGAT-3'; mNtrk3: sense 5'- TGGCAACTACACCCTCATTGC-3'; antisense 5'-GAAATCTGTGCTCTCTGGAAAGG-3'; mCyp11a1: sense 5'- GGACTTAAGGCAGAAGCGA
3': antisense 5'- AATGTTGGCCTGGATGTTCTTG-3'; mGab2: sense 5'-CTGCTGAACCTCCAGGAAAGA-3'; antisense 5'- GCCAGCAGGGTAGAAGAACCT-3'.

5.14 Cell-based assays

For MTT cell growth assays, 103 cells of each cell line were seeded in triplicate in regular or polyhema-coated 96-well plates. Cells were cultured in growth medium containing all supplements or in starving medium (serum reduced to 2%, no EGF). At the indicated times, a tetrazolium saltâĂŞbased reagent (CellTiter96 Aqueous One Solution, Promega) was added to each well according to the instructions provided by the manufacturer. After an incubation of 2 h, absorbance was read at 490 nm on a DTX 880 plate reader (Beckman Coulter). A control of cell plating, estimated after 4h from plating, was used as a reference to adjust subsequent acquisitions of each cell line. For soft agar growth, 3x104 cells were resuspended in 2ml of 0.5% top agar (SeaPlaque Agarose from Cambrex) in growth medium and seeded in 6-well plates previously filled with 3ml of 1% basal agar in growth medium. The assay was performed in duplicate. After 3 weeks, phase-contrast pictures were captured by a BD Pathway microscopic station (BD biosciences). Image analysis and quantification of single colonies number and size were performed by the Attovision 1.5 software (BD biosciences). Briefly, a threshold of colony size was defined either as the 95th or 5th percentile of the colony size in pictures taken from the control cells. Then the number of colonies above the threshold was calculated for all fields (8 fields/sample). The bar chart indicates the average and standard error of the number of colonies/field above the threshold. For detachment-induced cell death analysis, 3x105 cells were plated on regular or polyhema-coated 35mm plates for 48 hours. Cell death was then measured by assessing the number of hypodiploid nuclei with the DNAcon3 kit (ConsulTS, Rivalta, Italy), according to the manufacturer's protocol, and with cytofluorimetric analysis using a FACSCalibur (Becton Dickinson, San Diego, CA). Hypodiploid, subG0/G1 nuclei were defined as those displaying a PI staining value lower than that of cells in the G0/G1 cell cycle phase (diploid DNA peak).

Bibliography

[1] American Cancer Society *Report sees 7.6 million global 2007 cancer deaths* Reuters. Retrieved on 2008-08-07.

[2] Amit I, Citri A, Shay T, Lu Y, Katz M, Zhang F, Tarcic G, Siwak D, Lahad J, Jacob-Hirsch J, Amariglio N, Vaisman N, Segal E, Rechavi G, Alon U, Mills GB, Domany E, Yarden Y. *A module of negative feedback regulators defines growth factor signaling.* Nat Genet. 2007 Apr;39(4):503-12. Epub 2007 Feb 25. PMID: 17322878

[3] Aplin, A. E. and Juliano, R *Integrin and cytoskeletal regulation of growth factor signaling to the MAP kinase pathway* J. Cell Sci. 112,695 -706.

[4] Ariga,N., Sato,E., Ohuchi,N., Nagura,H. and Ohtani,H. *Stromal expression of fibroblast activation protein/seprase, a cell membrane serine proteinase and gelatinase, is associated with longer survival in patients with invasive ductal carcinoma of breast* Int.J Cancer, 95: 67-72, 2001.

[5] Assoian RK *Control of the G1 phase cyclin-dependent kinases by mitogenic growth factors and the extracellular matrix* Cytokine Growth Factor Rev. Sep;8(3):165-70.

[6] BÃűcker W. *WHO classification of breast tumors and tumors of the female genital organs:pathology and genetics* Verh Dtsch Ges Pathol. 2002;86:116-9. PMID: 12647359

[7] Barczak A, Rodriguez MW, Hanspers K, Koth LL, Tai YC, Bolstad BM, Speed TP, Erle DJ. *Spotted long oligonucleotide arrays for human gene expression analysis.* Genome Res. 2003 Jul;13(7):1775-85. Epub 2003 Jun 12. PMID: 12805270

[8] Bartel, D.P. *MicroRNAs: genomics, biogenesis, mechanism, and function.* Cell, 2004. 116(2): p. 281-97.

[9] Bennett,H.L., Brummer,T., Jeanes,A., Yap,A.S. and Daly,R.J. *Gab2 and Src co-operate in human mammary epithelial cells to promote growth factor independence and disruption of acinar morphogenesis* Oncogene, 2007

[10] Bentires-Alj M, Gil SG, Chan R, Wang ZC, Wang Y, Imanaka N, Harris LN, Richardson A, Neel BG, Gu H. *A role for the scaffolding adapter GAB2 in breast cancer.* Nat Med. 2006 Jan;12(1):114-21. Epub 2005 Dec 20. PMID: 16369543

[11] Bos JL *Epac proteins: multi-purpose cAMP targets.* Trends Biochem Sci.Dec;31(12):680-6.

[12] Validation of a novel, fully integrated and flexible microarray benchtop facility for gene expression profiling. *Validation of a novel, fully integrated and flexible microarray benchtop facility for gene expression profiling.* Nucleic Acids Res. 2003 Dec 1;31(23):e151. PMCID: PMC290286 PMID: 14627841

[13] Bhattacharjee,A., Richards,W.G., Staunton,J., Li,C., Monti,S., Vasa,P., Ladd,C., Beheshti,J., Bueno,R., Gillette,M., Loda,M., Weber,G., Mark,E.J., Lander,E.S., Wong,W., Johnson,B.E., Golub,T.R., Sugarbaker,D.J. and Meyerson,M. *Classification of human lung carcinomas by mRNA expression profiling reveals distinct adenocarcinoma subclasses* Proc.Natl.Acad.Sci.U.S.A, 98: 13790-13795, 2001.

[14] Biomarkers Working Group *Biomarkers and surrogate endpoints: preferred definitions and conceptual framework* Clin. Pharmacol. Ther. 2001 89âĂŞ95.

[15] Boccaccio C, AndÃš M, Tamagnone L, Bardelli A, Michieli P, Battistini C, Comoglio PM. *Induction of epithelial tubules by growth factor HGF depends on the STAT pathway.* Nature. 1998 Jan 15;391(6664):285-8. PMID: 9440692

[16] Bouscary,D., Lecoq-Lafon,C., Chretien,S., et al *Role of Gab proteins in phosphatidylinositol 3-kinase activation by thrombopoietin (Tpo)* Oncogene, 20: 2197-2204, 2001.

[17] Brown PO, Botstein D. *Exploring the new world of the genome with DNA microarrays.* Nat Genet. 1999 Jan;21(1 Suppl):33-7. Review. PMID: 9915498

[18] Buerger H, Mommers EC, Littmann R, Simon R, Diallo R, Poremba C,Dockhorn-Dworniczak B, van Diest PJ, Boecker W. *Ductal invasive G2 and G3 carcinomas of the breast are the end stages of at least two different lines of genetic evolution.* J Pathol. 2001 Jun;194(2):165-70. PMID: 11400144

[19] Buyse M, Molenberghs G, Burzykowski T, Renard D, Geys H. *The validation of surrogate endpoints in meta-analyses of randomized experiments.* Biostatistics. 2000 Mar;1(1):49-67. PMID: 12933525

[20] Buyse M, Loi S, van't Veer L, Viale G, Delorenzi M, Glas AM, d'Assignies MS, Bergh J, Lidereau R, Ellis P, Harris A, Bogaerts J, Therasse P, Floore A, Amakrane M, Piette F, Rutgers E, Sotiriou C, Cardoso F, Piccart MJ; TRANSBIG Consortium. *Validation and clinical utility of a 70-gene prognostic signature for women with node-negative breast cancer.* J Natl Cancer Inst. 2006 Sep 6;98(17):1183-92. PMID: 16954471

[21] Calin, G.A. and C.M. Croce *Chromosomal rearrangements and microRNAs: a new cancer link with clinical implications.* J Clin Invest, 2007. 117(8): p.2059-66.

[22] Cancer Research UK (January 2007) *UK cancer incidence statistics by age* Retrieved on 2007-06-25.

[23] Chong, Y.P., T.D. Mulhern, and H.C. Cheng *C-terminal Src kinase (CSK) and CSK-homologous kinase (CHK)–endogenous negative regulators of Src-family protein kinases.* Growth Factors, 2005. 23(3): p. 233-44.

[24] Cohen, S.M., J. Brennecke, and A. Stark *Denoising feedback loops by thresholding–a new role for microRNAs* Proc Natl Acad Sci U S A, 2006. 103(7): p.2257-61.

[25] Comoglio, P.M. and L. Trusolino *Invasive growth: from development to metastasis* J Clin Invest, 2002. 109(7): p. 857-62.

[26] Cross, M. and T.M. Dexter *Growth factors in development, transformation, and tumorigenesis* Cell, 1991. 64(2): p. 271-80.

[27] Cooper CS, Park M, Blair DG, Tainsky MA, Huebner K, Croce CM, Vande Woude GF. *Molecular cloning of a new transforming gene from a chemically transformed human cell line.* Nature. 1984 Sep 6-11;311(5981):29-33.PMID: 6590967

[28] Chong, Y.P., T.D. Mulhern, and H.C. Cheng *C-terminal Src kinase (CSK) and CSK-homologous kinase (CHK)–endogenous negative regulators of Src-family protein kinases.* Growth Factors, 2005. 23(3): p. 233-44.

[29] Cutrupi, S., et al. *Src-mediated activation of alpha-diacylglycerol kinase is required for hepatocyte growth factor-induced cell motility.* Embo J, 2000. 19(17): p. 4614-22.

[30] Critchley, D.R., *Focal adhesions - the cytoskeletal connection.* Cell Biol, 2000. 12(1): p. 133-9.

[31] Daly,R.J., Gu,H., Parmar,J., et al *The docking protein Gab2 is overexpressed and estrogen regulated in human breast cancer* Oncogene, 21: 5175-5181, 2002.

[32] Danen,E.H. and Yamada,K.M. *Fibronectin, integrins, and growth control* J Cell Physiol, 189: 1-13, 2001.

[33] Danen,E.H. and Yamada,K.M. *Fibronectin, integrins, and growth control* J Cell Physiol, 189: 1-13, 2001.

Dennis G Jr, Sherman BT, Hosack DA, Yang J, Gao W, Lane HC, Lempicki RA. DAVID: Database for Annotation, Visualization, and Integrated Discovery. Genome Biology 2003;4(5):P3.

[34] DeRisi, J.L., V.R. Iyer, and P.O. Brown *Exploring the metabolic and genetic control of gene expression on a genomic scale.* Science, 1997. 278(5338): p.680-6.

[35] Desmedt C, Piette F, Loi S, Wang Y, Lallemand F, Haibe-Kains B, Viale G, Delorenzi M, Zhang Y, d'Assignies MS, Bergh J, Lidereau R, Ellis P, Harris AL, Klijn JG, Foekens JA, Cardoso F, Piccart MJ, Buyse M, Sotiriou C; TRANSBIG Consortium. *Strong time dependence of the 76-gene prognostic signature for node-negative breast cancer patients in the TRANSBIG multicenter independent validation series.* Clin Cancer Res. 2007 Jun 1;13(11):3207-14. PMID: 17545524

[36] Di Renzo MF, Olivero M, Giacomini A, Porte H, Chastre E, Mirossay L, Nordlinger B, Bretti S, Bottardi S, Giordano S *Overexpression and amplification of the met/HGF receptor gene during the progression of colorectal cancer.* Clin Cancer Res. 1995 Feb;1(2):147-54.PMID: 9815967

[37] Di Renzo MF, Olivero M, Ferro S, Prat M, Bongarzone I, Pilotti S, Belfiore A, Costantino A, Vigneri R, Pierotti MA *Overexpression of the c-MET/HGF receptor gene in human thyroid carcinomas.* Oncogene. 1992 Dec;7(12):2549-53. PMID: 1334253

[38] Di Renzo MF, Olivero M, Giacomini A, Porte H, Chastre E, Mirossay L, Nordlinger B, Bretti S, Bottardi S, Giordano S *Overexpression and amplification of the met/HGF receptor gene during the progression of colorectal cancer.* Clin Cancer Res. 1995 Feb;1(2):147-54. PMID: 9815967

[39] Di Renzo MF, Olivero M, Martone T, Maffe A, Maggiora P, Stefani AD, Valente G, Giordano S, Cortesina G, Comoglio PM. *Somatic mutations of the MET oncogene are selected during metastatic spread of human HNSC carcinomas.* Oncogene. 2000 Mar 16;19(12):1547-55. PMID: 10734314

[40] Dunnett, C. *New tables for multiple comparisons with a control.* Biometrics 20:482-491.

[41] Du, P., Kibbe, W.A. and Lin, S.M. *'lumi: a pipeline for processing Illumina microarray* Bioinformatics 24(13):1547-1548

[42] Esquela-Kerscher, A. and F.J. Slack *Oncomirs - microRNAs with a role in cancer.* Nat Rev Cancer, 2006. 6(4): p. 259-69.

[43] R. Etzioni, N. Urban, S. Ramsey, M. McIntosh, S. Schwartz, B. Reid, J. Radich, G. Anderson and L. Hartwell *The case for early detection* Nat. Rev. Cancer. 2003 243-52

[44] Draft Guidance for Industry, Clinical Laboratories, and FDA Staff *In Vitro Diagnostic Multivariate Index Assays.* U.S. Food and Drug Administration. Web site: http://www.fda.gov/cdrh/ovid/guidance/1610.html. August 25, 2008.

[45] Feuer EJ, Wun LM, Boring CC, Flanders WD, Timmel MJ, Tong T. *The lifetime risk of developing breast cancer.* J Natl Cancer Inst. 1993 Jun 2;85(11):892-7.PMID: 8492317

[46] ffrench-Constant C, Colognato H. *Integrins: versatile integrators of extracellular signals.* Trends Cell Biol. 2004 Dec;14(12):678-86. PMID: 1556404

[47] Frisch, S. M. and Ruoslahti *Integrins and anoikis* Curr. Opin. Cell Biol. 9, 701-706.

[48] J. Folkman *Tumor angiogenesis.* In: J.F. Holland, R.C. Bast, D.L. Morton, E. Frei, D.W. Kufe and R.R. Weichselbaum, Editors, Cancer Medicine, Williams and Wilkins, Baltimore, MD (1997) 181âĂŞ204.pp .

[49] Fu L, Medico E. *FLAME, a novel fuzzy clustering method for the analysis of DNA microarray data.* BMC Bioinformatics. 2007 Jan 4;8:3.PMID: 17204155

[50] Garcia,R., Bowman,T.L., Niu,G., et al. *Constitutive activation of Stat3 by the Src and JAK tyrosine kinases participates in growth regulation of human breast carcinoma cells* Oncogene, 20: 2499-2513, 2001.

[51] Gentile A, D'Alessandro L, Lazzari L, Martinoglio B, Bertotti A, Mira A, Lanzetti L, Comoglio PM, Medico E. *Met-driven invasive growth involves transcriptional regulation of Arhgap12* Oncogene. 2008 May 26. [Epub ahead of print]PMID: 18504429

[52] Robert C Gentleman and Vincent J. Carey and Douglas M. Bates and others *Bioconductor: Open software development for computational biology and bioinformatics* http://genomebiology.com/2004/5/10/R80

[53] Geiger,T.R. and Peeper,D.S. *Critical role for TrkB kinase function in anoikis suppression, tumorigenesis, and metastasis* Cancer Res., 67: 6221-6229, 2007

[54] Golub, T.R., D.K.Slonim, P.Tamayo, C.Huard, M.Gaasenbeek, J.P.Mesirov, H.Coller, M.L.Loh, J.R.Downing, M.A.Caligiuri, C.D.Bloomfield, and E.S.Lander *Molecular classification of cancer: class discovery and class prediction by gene expression monitoring.* Science 286:531-537.

[55] Gu,H., Pratt,J.C., Burakoff,S.J. and Neel,B.G. *Cloning of p97/Gab2, the major SHP2-binding protein in hematopoietic cells, reveals a novel pathway for cytokine-induced gene activation* Mol.Cell, 2: 729-740, 1998.

[56] Hanahan D, Weinberg RA. *The hallmarks of cancer.* Cell. 2000 Jan 7;100(1):57-70. Review.

[57] Hardiman G. *Microarrays Technologies 2006: an overview.* Pharmacogenomics. 2006 Dec;7(8):1153-8. PMID: 17184202

[58] Harfe, B.D. *MicroRNAs in vertebrate development.* Curr Opin Genet Dev,2005. 15(4): p. 410-5.

[59] D.F. Hayes *Prognostic and predictive factors revisited* Breast 2005 493-9

[60] Hess KR, Anderson K, Symmans WF, Valero V, Ibrahim N, Mejia JA, Booser D, Theriault RL, Buzdar AU, Dempsey PJ, Rouzier R, Sneige N, Ross JS, Vidaurre T, GÃşmez HL, Hortobagyi GN, Pusztai L. *Pharmacogenomic predictor of sensitivity to preoperative chemotherapy with paclitaxel and fluorouracil, doxorubicin, and cyclophosphamide in breast cancer.* J Clin Oncol. 2006 Sep 10;24(26):4236-44. Epub 2006 Aug 8.PMID: 16896004

[61] Hess,V. *[Adjuvant!Online–an Internet-based decision tool for adjuvant chemotherapy in early breast cancer]* Ther.Umsch., 65: 201-205, 2008.

[62] Howe, A. K. and Juliano, *Regulation of anchorage-dependent signal transduction by protein kinase A and p21-activated kinase.* Nat. Cell Biol. 2, 593-600.

[63] Huang,F., Reeves,K., Han,X., Fairchild,C., Platero,S., Wong,T.W., Lee,F., Shaw,P. and Clark,E. *Identification of candidate molecular markers predicting sensitivity in solid tumors to dasatinib: rationale for patient selection* Cancer Res., 67: 2226-2238, 2007.

[64] Jackson, R.J. and N. Standart *How do microRNAs regulate gene expression?* Sci STKE, 2007. 2007(367): p. re1.

[65] Iorio, M.V., et al. *MicroRNA gene expression deregulation in human breast cancer.* Cancer Res, 2005. 65(16): p. 7065-70.

[66] Hess,V. *[Adjuvant!Online–an Internet-based decision tool for adjuvant chemotherapy in early breast cancer]* Ther.Umsch., 65: 201-205, 2008.

[67] Chong, Y.P., T.D. Mulhern, and H.C. Cheng *C-terminal Src kinase (CSK) and CSK-homologous kinase (CHK)–endogenous negative regulators of Src-family protein kinases.* Growth Factors, 2005. 23(3): p. 233-44.

[68] Cutrupi, S., et al. *Src-mediated activation of alpha-diacylglycerol kinase is required for hepatocyte growth factor-induced cell motility.* Embo J, 2000. 19(17): p. 4614-22.

[69] Critchley, D.R., *Focal adhesions - the cytoskeletal connection.*

[70] Laird,A.D., Li,G., Moss,K.G., et al *Src family kinase activity is required for signal tranducer and activator of transcription 3 and focal adhesion kinase phosphorylation and vascular endothelial growth factor signaling in vivo and for anchorage-dependent and -independent growth of human tumor cells* Mol.Cancer Ther., 2: 461-469, 2003.

[71] Lakhani SR. *The transition from hyperplasia to invasive carcinoma of the breast.* J Pathol. 1999 Feb;187(3):272-8. PMID: 10398078

[72] Lesko, E. and M. Majka, *The biological role of HGF-MET axis in tumor growth and development of metastasis.* Front Biosci, 2008. 13: p. 1271-80.

[73] Liang TJ, Reid AE, Xavier R, Cardiff RD, Wang TC. *Transgenic expression of tpr-met oncogene leads to development of mammary hyperplasia and tumors.* J Clin Invest. 1996 Jun 15;97(12):2872-7. PMID: 8675700

[74] Liu,Y., Jenkins,B., Shin,J.L. and Rohrschneider,L.R. *Scaffolding protein Gab2 mediates differentiation signaling downstream of Fms receptor tyrosine kinase* Mol.Cell Biol, 21: 3047-3056, 2001.

[75] Lu, J., et al *MicroRNA expression profiles classify human cancers* Nature, 2005. 435(7043): p. 834-8.

[76] Kato, M. and F.J. Slack *microRNAs: small molecules with big roles - C.elegans to human cancer.* Biol Cell, 2008. 100(2): p. 71-81.

[77] Ke,Y., Wu,D., Princen,F., et al. *Role of Gab2 in mammary tumorigenesis and metastasis* Oncogene, 26: 4951-4960, 2007

[78] Kertesz, M., et al. *The role of site accessibility in microRNA target recognition.* Nat Genet, 2007. 39(10): p. 1278-84.

[79] Kriete A, Sokhansanj BA, Coppock DL, West GB. *Systems approaches to the networks of aging* Ageing Res Rev. 2006 Nov;5(4):434-48. Epub 2006 Aug 14. Review.PMID: 16904954

[80] Kooistra MR, Dube N, Bos JL *Rap1: a key regulator in cell-cell junction formation.* J Cell Sci. Jan 1;120(Pt 1):17-22.

[81] Kong,M., Mounier,C., Dumas,V. and Posner,B.I. *Epidermal growth factor-induced DNA synthesis. Key role for Src phosphorylation of the docking protein Gab2* J Biol Chem, 278: 5837-5844, 2003.

[82] Kothapalli R, Yoder SJ, Mane S, Loughran TP Jr. *Microarray results: how accurate are they?* BMC Bioinformatics. 2002 Aug 23;3:22. PMCID: PMC126254 PMID: 12194703

[83] Jackson, R.J. and N. Standart *How do microRNAs regulate gene expression?* Sci STKE, 2007. 2007(367): p. re1.

[84] Jeffers M, Fiscella M, Webb CP, Anver M, Koochekpour S, Vande Woude GF. *The mutationally activated Met receptor mediates motility and metastasis.* Proc Natl Acad Sci U S A. 1998 Nov 24;95(24):14417-22. PMID: 9826715

[85] Martelli,M.L., Isella,C., Mira,A., Fu,L., Cantarella,D. and Medico,E. *Exploiting orthologue diversity for systematic detection of gain-of-function phenotypes* BMC.Genomics, 9: 254, 2008.

[86] Masood S. *Prediction of recurrence for advanced breast cancer. Traditional and contemporary pathologic and molecular markers.* Surg Oncol Clin N Am. 1995 Oct;4(4):601-32. PMID: 8535901

[87] MAQC Consortium, Shi L, Reid LH, Jones WD, Shippy R, Warrington JA, Baker SC, Collins PJ, de Longueville F, Kawasaki ES, Lee KY, Luo Y, Sun YA, Willey JC, Setterquist RA, Fischer GM, Tong W, Dragan YP, Dix DJ, Frueh FW, Goodsaid FM, Herman D, Jensen RV, Johnson CD, Lobenhofer EK, Puri RK, Schrf U, Thierry-Mieg J, Wang C, Wilson M, Wolber PK, Zhang L, Amur S, Bao W, Barbacioru CC, Lucas AB, Bertholet V, Boysen C, Bromley B, Brown D, Brunner A, Canales R, Cao XM, Cebula TA, Chen JJ, Cheng J, Chu TM, Chudin E, Corson J, Corton JC, Croner LJ, Davies C, Davison TS, Delenstarr G, Deng X, Dorris D, Eklund AC, Fan XH, Fang H, Fulmer-Smentek S, Fuscoe JC, Gallagher K, Ge W, Guo L, Guo X, Hager J, Haje PK, Han J, Han T, Harbottle HC, Harris SC, Hatchwell E, Hauser CA, Hester S, Hong H, Hurban P, Jackson SA, Ji H, Knight CR, Kuo WP, LeClerc JE, Levy S, Li QZ, Liu C, Liu Y, Lombardi MJ, Ma Y, Magnuson SR, Maqsodi B, McDaniel T, Mei N, Myklebost O, Ning B, Novoradovskaya N, Orr MS, Osborn TW, Papallo A, Patterson TA, Perkins RG, Peters EH, Peterson R, Philips KL, Pine PS, Pusztai L, Qian F, Ren H, Rosen M, Rosenzweig BA, Samaha RR, Schena M, Schroth GP, Shchegrova S, Smith DD, Staedtler F, Su Z, Sun H, Szallasi Z, Tezak Z, Thierry-Mieg D, Thompson KL, Tikhonova I, Turpaz Y, Vallanat B, Van C, Walker SJ, Wang SJ, Wang Y, Wolfinger R, Wong A, Wu J, Xiao C, Xie Q, Xu J, Yang W, Zhang L, Zhong S, Zong Y, Slikker W Jr. *The MicroArray Quality Control (MAQC) project shows inter- and intraplatform reproducibility of gene expression measurements.* Nat Biotechnol. 2006 Sep;24(9):1151-61. PMID: 16964229

[88] McShane LM, Altman DG, Sauerbrei W, Taube SE, Gion M, Clark GM *Reporting recommendations for tumor marker prognostic studies.* J Clin Oncol. 2005 Dec 20;23(36):9067-72.PMID: 16172462

[89] Medico E, Gentile A, Lo Celso C, Williams TA, Gambarotta G, Trusolino L, Comoglio PM. *Osteopontin is an autocrine mediator of hepatocyte growth factor-induced invasive growth.* Cancer Res. 2001 Aug 1;61(15):5861-8. PMID: 11479227

[90] M. Megraw, P. Sethupathy, B. Corda, and A.G. Hatzigeorgiou *miRGen: A database for the study of animal microRNA genomic organization and function.* Nucleic Acids Research, 35: D149-D155.

[91] Menard RE, Jovanovski AP, Mattingly RR. *Active p21-activated kinase 1 rescues MCF10A breast epithelial cells from undergoing anoikis.* Neoplasia. 2005 Jul;7(7):638-45. PMID: 16026643

[92] Mettouchi A, Klein S, Guo W, Lopez-Lago M, Lemichez E, Westwick JK, Giancotti FG. *Integrin-specific activation of Rac controls progression through the G(1) phase of the cell cycle.* Mol Cell. 2001 Jul;8(1):115-27.

[93] Miller,L.D., Smeds,J., George,J., Vega,V.B., Vergara,L., Ploner,A., Pawitan,Y., Hall,P., Klaar,S., Liu,E.T. and Bergh,J. *An expression signature for p53 status in human breast cancer predicts mutation status, transcriptional effects, and patient survival* Proc Natl.Acad.Sci.U.S.A, 102: 13550-13555, 2005.

[94] Nakamura, T., H. Teramoto, and A. Ichihara. *Purification and characterization of a growth factor from rat platelets for mature parenchymal hepatocytes in primary cultures.* Proc Natl Acad Sci U S A,1986. 83(17): p. 6489-93.

[95] Ni,S., Zhao,C., Feng,G.S., Paulson,R.F. and Correll,P.H. *A novel Stat3 binding motif in Gab2 mediates transformation of primary hematopoietic cells by the Stk/Ron receptor tyrosine kinase in response to Friend virus infection* Mol.Cell Biol, 27: 3708-3715, 2007.

[96] Nishida,K., Yoshida,Y., Itoh,M., et al *Gab-family adapter proteins act downstream of cytokine and growth factor receptors and T- and B-cell antigen receptors* Blood, 93: 1809-1816, 1999.

[97] Pelicci G, Giordano S, Zhen Z, Salcini AE, Lanfrancone L, Bardelli A, Panayotou G, Waterfield MD, Ponzetto C, Pelicci PG, et al. *The motogenic and mitogenic responses to HGF are amplified by the Shc adaptor protein.* Oncogene. 1995 Apr 20;10(8):1631-8. PMID: 7731718

[98] Playford,M.P. and Schaller,M.D. *The interplay between Src and integrins in normal and tumor biology* Oncogene, 23: 7928-7946, 2004.

[99] Pillai, R.S., S.N. Bhattacharyya, and W. Filipowicz *Repression of protein synthesis by microRNAs: how many mechanisms?* Trends Cell Biol, 2007. 17(3):p. 118-26.

[100] Ponzetto C, Bardelli A, Zhen Z, Maina F, dalla Zonca P, Giordano S, Graziani A, Panayotou G, Comoglio PM. *A multifunctional docking site mediates signaling and transformation by the hepatocyte growth factor/scatter factor receptor family.* Cell. 1994 Apr 22;77(2):261-71. PMID: 7513258

[101] O'Donnell, K.A., et al. *c-Myc-regulated microRNAs modulate E2F1 expression* Nature, 2005. 435(7043): p. 839-43.

[102] Ohtani K, DeGregori J, Nevins JR. *Regulation of the cyclin E gene by transcription factor E2F1.* Proc Natl Acad Sci U S A. 1995 Dec 19;92(26):12146-50. PMID: 8618861

[103] Ravdin PM, Siminoff LA, Davis GJ, Mercer MB, Hewlett J, Gerson N, Parker HL. *Computer program to assist in making decisions about adjuvant therapy for women with early breast cancer.* J Clin Oncol. 2001 Feb 15;19(4):980-91. PMID: 11181660

[104] R Development Core Team *R: A Language and Environment for Statistical Computing* http://www.R-project.org

[105] Reid JF, Lusa L, De Cecco L, Coradini D, Veneroni S, Daidone MG, Gariboldi M, Pierotti MA. *Limits of predictive models using microarray data for breast cancer clinical treatment outcome.* J Natl Cancer Inst. 2005 Jun 15;97(12):927-30.

[106] Reis-Filho JS, Lakhani SR. *The diagnosis and management of pre-invasive breast disease: genetic alterationsin pre-invasive lesions.* Breast Cancer Res. 2003;5(6):313-9. Epub 2003 Oct 9. PMID: 14580249

[107] Retta SF, Balzac F, Avolio M *Rap1: a turnabout for the crosstalk between cadherins and integrins.* Eur J Cell Biol. Apr;85(3-4):283-93.

[108] Roux,K.J., Amici,S.A., Fletcher,B.S. and Notterpek,L. *Modulation of epithelial morphology, monolayer permeability, and cell migration by growth arrest specific 3/peripheral myelin protein 22* Mol.Biol Cell, 16: 1142-1151, 2005.

[109] Roylance R, Gorman P, Harris W, Liebmann R, Barnes D, Hanby A, Sheer D. *Comparative genomic hybridization of breast tumors stratified by histological grade reveals new insights into the biological progression of breast cancer.* Cancer Res. 1999 Apr 1;59(7):1433-6. PMID: 10197608

[110] Rubin, J.S., D.P. Bottaro, and S.A. Aaronson *Hepatocyte growthfactor/scatter factor and its receptor, the c-met proto-oncogene product.* Biochim Biophys Acta, 1993. 1155(3): p. 357-71.

[111] Sattler,M., Mohi,M.G., Pride,Y.B., et al *Critical role for Gab2 in transformation by BCR/ABL* Cancer Cell, 1: 479-492, 2002.

[112] Savagner, P., K.M. Yamada, and J.P. Thiery *The zinc-finger protein slug causes desmosome dissociation, an initial and necessary step for growth factor-induced epithelial-mesenchymal transition.* J Cell Biol, 1997. 137(6):p. 1403-19.

[113] Schwartz MA, Ginsberg MH *Networks and crosstalk: integrin signalling spreads.* Nat Cell Biol, 4:E65-E68.

[114] M.J. Slotnick and J.O. Nriagu *Validity of human nails as a biomarker of arsenic and selenium exposure: a review* Environ. Res. 2006 125-139

[115] M.B. Sporn *The war on cancer.* Lancet 347 (1996), pp.

[116] Hess,V. *[Adjuvant!Online–an Internet-based decision tool for adjuvant chemotherapy in early breast cancer]* Ther.Umsch., 65: 201-205, 2008.

[117] Chong, Y.P., T.D. Mulhern, and H.C. Cheng *C-terminal Src kinase (CSK) and CSK-homologous kinase (CHK)–endogenous negative regulators of Src-family protein kinases.* Growth Factors, 2005. 23(3): p. 233-44.

[118] Cutrupi, S., et al. *Src-mediated activation of alpha-diacylglycerol kinase is required for hepatocyte growth factor-induced cell motility.* Embo J, 2000. 19(17): p. 4614-22.

[119] Critchley, D.R., *Focal adhesions - the cytoskeletal connection.*

[120] Laird,A.D., Li,G., Moss,K.G., et al *Src family kinase activity is required for signal tranducer and activator of transcription 3 and focal adhesion kinase phosphorylation and vascular endothelial growth factor signaling in vivo and for anchorage-dependent and -independent growth of human tumor cells* Mol.Cancer Ther., 2: 461-469, 2003. Cell Biol, 2000. 12(1): p. 133-9.

[121] Sotiriou C, Wirapati P, Loi S, Harris A, Fox S, Smeds J, Nordgren H, Farmer P, Praz V, Haibe-Kains B, Desmedt C, Larsimont D, Cardoso F, Peterse H, Nuyten D, Buyse M, Van de Vijver MJ, Bergh J, Piccart M, Delorenzi M. *Gene expression profiling in breast cancer: understanding the molecular basis of histologic grade to improve prognosis.* J Natl Cancer Inst. 2006 Feb 15;98(4):262-72.PMID: 16478745

[122] Sotiriou C, Piccart MJ. *Taking gene-expression profiling to the clinic: when will molecular signatures become relevant to patient care?* Nat Rev Cancer. 2007 Jul;7(7):545-53. PMID: 17585334

[123] SÃÿrlie T, Perou CM, Tibshirani R, Aas T, Geisler S, Johnsen H, Hastie T, Eisen MB, van de Rijn M, Jeffrey SS, Thorsen T, Quist H, Matese JC, Brown PO, Botstein D, Eystein LÃÿnning P, BÃÿrresen-Dale AL. *Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications.* Proc Natl Acad Sci U S A. 2001 Sep 11;98(19):10869-74. PMID: 11553815

[124] Goldhirsch A, Glick JH, Gelber RD, Coates AS, Thürlimann B, Senn HJ; Panel members. *Meeting highlights: international expert consensus on the primary therapy of early breast cancer 2005.* Ann Oncol. 2005 Oct;16(10):1569-83. Epub 2005 Sep 7.

[125] Goldhirsch A, Coates AS, Gelber RD, Glick JH, Thürlimann B, Senn HJ; St Gallen Expert Panel Members. *First–select the target: better choice of adjuvant treatments for breast cancer patients.* Ann Oncol. 2006 Dec;17(12):1772-6. Epub 2006 Oct 27. PMID: 17071934

[126] Stoker M, Gherardi E, Perryman M, Gray J. *Scatter factor is a fibroblast-derived modulator of epithelial cell mobility.* Nature. 1987 May 21-27;327(6119):239-42.PMID: 2952888

[127] Tan PK, Downey TJ, Spitznagel EL Jr, Xu P, Fu D, Dimitrov DS, Lempicki RA, Raaka BM, Cam MC. *Evaluation of gene expression measurements from commercial microarray platforms.* Nucleic Acids Res. 2003 Oct 1;31(19):5676-84. PMID: 14500831

[128] Thomson, J.M., et al., *Extensive post-transcriptional regulation of microRNAs and its implications for cancer.* Genes Dev, 2006. 20(16): p.2202-7.

[129] Tsang, J., J. Zhu, and A. van Oudenaarden, *MicroRNA-mediated feedback and feedforward loops are recurrent network motifs in mammals.* Mol Cell,2007. 26(5): p. 753-67.

[130] Vadlamudi RK, Bagheri-Yarmand R, Yang Z, Balasenthil S, Nguyen D, Sahin AA, den Hollander P, Kumar R. *Dynein light chain 1, a p21 activated kinase 1-interacting substrate, promotes cancerous phenotypes.* Jun;5(6):575-85.

[131] van't Veer LJ, Bernards R. *Enabling personalized cancer medicine through analysis of gene-expression patterns.* Nature. 2008 Apr 3;452(7187):564-70. PMID: 18385730

[132] van 't Veer LJ, Dai H, van de Vijver MJ, He YD, Hart AA, Mao M, Peterse HL, van der Kooy K, Marton MJ, Witteveen AT, Schreiber GJ, Kerkhoven RM, Roberts C, Linsley PS, Bernards R, Friend SH. *Gene expression profiling predicts clinical outcome of breast cancer.* Nature. 2002 Jan 31;415(6871):530-6. PMID: 11823860

[133] van de Vijver MJ, He YD, van't Veer LJ, Dai H, Hart AA, Voskuil DW, Schreiber GJ, Peterse JL, Roberts C, Marton MJ, Parrish M, Atsma D, Witteveen A, Glas A, Delahaye L, van der Velde T, Bartelink H, Rodenhuis S, Rutgers ET, Friend SH, Bernards R. *A gene-expression signature as a predictor of survival in breast cancer.* N Engl J Med. 2002 Dec 19;347(25):1999-2009. PMID: 12490681

[134] Varner JA *The sticky truth about angiogenesis and thrombospondins.* J Clin Invest.Dec;116(12):3111-3.

[135] Vidal, M. *A biological atlas of functional maps.* Cell, 2001. 104(3): p. 333-9.

[136] Volinia, S., et al. *A microRNA expression signature of human solid tumors defines cancer gene targets* Proc Natl Acad Sci U S A, 2006. 103(7): p.2257-61.

[137] Vogelstein B, Fearon ER, Hamilton SR, Kern SE, Preisinger AC, Leppert M, Nakamura Y, White R, Smits AM, Bos JL. *Genetic alterations during colorectal-tumor development.* N Engl J Med. 1988 Sep 1;319(9):525-32. PMID: 2841597

[138] Yu,M., Luo,J., Yang,W., Wang,Y., Mizuki,M., Kanakura,Y., Besmer,P., Neel,B.G. and Gu,H. *The scaffolding adapter Gab2, via Shp-2, regulates kit-evoked mast cell proliferation by activating the Rac/JNK pathway* J Biol Chem, 281: 28615-28626, 2006.

[139] Yaspan,B.L., Breyer,J.P., Cai,Q., et al. *Haplotype analysis of CYP11A1 identifies promoter variants associated with breast cancer risk* Cancer Res., 67: 5673-5682, 2007.

[140] Yuan,K., Listinsky,C.M., Singh,R.K., Listinsky,J.J. and Siegal,G.P. *Cell surface associated alpha-L: -fucose moieties modulate human breast cancer neoplastic progression* Pathol.Oncol.Res., 14: 145-156, 2008.

[141] Wang Y, Klijn JG, Zhang Y, Sieuwerts AM, Look MP, Yang F, Talantov D, Timmermans M, Meijer-van Gelder ME, Yu J, Jatkoe T, Berns EM, Atkins D, Foekens JA. *Gene-expression profiles to predict distant metastasis of lymph-node-negative primary breast cancer.* Lancet. 2005 Feb 19-25;365(9460):671-9. PMID: 15721472

[142] Weidner KM, Di Cesare S, Sachs M, Brinkmann V, Behrens J, Birchmeier W. *Interaction between Gab1 and the c-Met receptor tyrosine kinase is responsible for epithelial morphogenesis.* Nature. 1996 Nov 14;384(6605):173-6.PMID: 8906793

[143] Wessels LF, Reinders MJ, Hart AA, Veenman CJ, Dai H, He YD, van't Veer LJ. *A protocol for building and evaluating predictors of disease state based on microarray data.* Bioinformatics. 2005 Oct 1;21(19):3755-62. Epub 2005 Apr 7.PMID: 15817694

[144] WHO *Cancer* World Health Organization. Retrieved on 2007-06-25.

[145] Wittchen ES, van Buul JD, Burridge K, Worthylake RA *Rap, Rac, and Rho as architects of transendothelial migration* Curr Opin Hematol. Jan;12(1):14-21.

[146] Wirapati,P., Sotiriou,C., Kunkel,S., et al. *Meta-analysis of gene expression profiles in breast cancer: toward a unified understanding of breast cancer subtyping and prognosis signatures* Breast Cancer Res., 10: R65, 2008.

[147] Zanetti A, Stoppacciaro A, Marzullo A, Ciabatta M, Fazioli F, Prat M, Comoglio PM, Baroni CD, Ruco LP. *Expression of Met protein and urokinase-type plasminogen activator receptor (uPA-R) in papillary carcinoma of the thyroid.* J Pathol. 1998 Nov;186(3):287-91.PMID: 10211118

[148] Zatkova,A., Schoch,C., Speleman,F., et al *GAB2 is a novel target of 11q amplification in AML/MDS* Genes Chromosomes.Cancer, 45: 798-807, 2006.

[149] Zhang,S.Q., Yang,W., Kontaridis,M.I., et al. *Shp2 regulates SRC family kinase activity and Ras/Erk activation by controlling Csk recruitment* Mol.Cell, 13: 341-355, 2004.

[150] Zhang,W., Zong,C.S., Hermanto,U., Lopez-Bergami,P., Ronai,Z. and Wang,L.H *RACK1 recruits STAT3 specifically to insulin and insulin-like growth factor 1 receptors for activation, which is important for regulating anchorage-independent growth* Mol.Cell Biol, 26: 413-424, 2006.