UNIVERSITÀ DEGLI STUDI DI TRIESTE

XXVII CICLO DEL DOTTORATO DI RICERCA IN

BIOMEDICINA MOLECOLARE

ROLE OF HMGA1 IN BREAST CANCER AGGRESSIVENESS

Settore scientifico-disciplinare: BIO/10

DOTTORANDA
GLORIA ROS

COORDINATORE
PROF. GUIDALBERTO MANFIOLETTI

SUPERVISORE DI TESI
PROF. GUIDALBERTO MANFIOLETTI

CORRELATORE DI TESI
SILVIA PEGORARO, PhD

ANNO ACCADEMICO 2014 / 2015

I am among those who think that science has great beauty.
A scientist in his laboratory is not only a technician:
he is also a child placed before natural phenomena
which impress him like a fairy tale.

Marie Curie
French (Polish-born) chemist & physicist (1867 – 1934)

To my hydrogens that make me a complete molecule of water.

Index

Summary

Breast cancer is the most common cancer in women and a leading cause of cancer mortality worldwide, mainly due to metastatic progression. Although in the last years cancer has become more curable, the path to eradicate it, is still long and convolute. Understanding molecular mechanisms that induce and support cancer progression and aggressiveness is crucial to find new targets and drugs to treat it. High Mobility Group A1 (HMGA1) protein is an oncofetal architectural chromatin factor that promotes neoplastic transformation and progression but its role in breast cancer aggressiveness is still unclear. The aim of this thesis was to unravel the involvement of HMGA1 in this disease highlighting which pathways are intersected. Thus, working on basal-like breast cancer subtype, we demonstrated that HMGA1 plays a crucial role in conferring metastatic traits. Indeed, HMGA1 silencing reduces migration and invasion properties *in vitro* and metastasis formation *in vivo* with a concomitant mesenchymal to epithelial transition and decreases stem cell properties and self-renewal activity. We performed microarray analysis in cells expressing or depleted for HMGA1 and we identified a specific 130-HMGA1 gene signature associated with poor prognosis. This analysis allowed us to identify pathways, controlled by HMGA1 and known to be involved in aggressiveness traits, such as Wnt/β-catenin and Notch. Moreover, among the genes present in the 130-HMGA1 gene signature we deepen the relationship between HMGA1 and CCNE2, one of the gene most correlated with clinical outcome. We showed that CCNE2 acts downstream HMGA1 to regulate the migration and invasion properties of basal-like breast cancer cells. Moreover, we demonstrated that CCNE2 action is mediated by the oncogene YAP, the downstream effector of the Hippo pathway. Indeed, knock down of both HMGA1 and CCNE2 impaired nuclear localization and activity of YAP, acting upstream of the Hippo pathway core kinases MST1/2 and LATS1/2. Moreover, in breast cancer patients, HMGA1 and CCNE2 expression was associated with YAP/TAZ signature further supporting this connection. Because CDKs are the main partners of CCNE2, we blocked their activity using CDK inhibitors in order to impair HMGA1-CCNE2-YAP axis and we demonstrated a decrease in cell migration and an induction of translocation of YAP from nucleus to cytoplasm. Therefore, this thesis highlights the involvement of HMGA1 in breast cancer metastasis through the interplay with different pathways. In particular, we identified for the first time a role of HMGA1 in regulating YAP through the modulation of the Hippo pathway.

Introduction

1 CANCER HALLMARKS

Cancer is a complex class of diseases in which the aberrations of key regulatory pathways, that govern normal cell proliferation and homeostasis, are fundamental for its development. There are more than 100 distinct types of human cancer, and subtypes of tumors can be found within specific organs (Hanahan and Weinberg, 2000). Cancer can be summarized simplistically as an abnormal growth of cells caused by multiple changes in gene expression evolving into a population of cells that can invade tissues and metastasize to distant sites, causing illness and, often death of the host (Cooper, 2000). In a more complex view cancer resemble many diseases with related features, and these features have been called hallmarks of cancer. The importance of cancer hallmarks, outlined for the first time in 2000 and then revised and expanded in 2011, by Hanahan and Weinberg, dwells in the finally identification of ten traits (Fig. 1), shared by all type of cancer, that characterize the biology of cancer development and aggressiveness beyond its extremely complexity (Hanahan and Weinberg, 2000, 2011).

Figure 1. The ten hallmarks shared by cancers described in 2011 by Hanahan and Weinberg.

1.1 SUSTAIN CELL PROLIFERATION AND SURVIVAL AS THE FIRST STEP IN CARCINOGENESIS

An essential trait of cancer cells involves their ability to sustain proliferation signals. Normal tissues control the production and release of growth-promoting signals that regulate cell growth and division cycle, ensuring the homeostasis of cell number and thus the maintenance of normal tissue architecture and function. Cancer cells loose this control and they acquire the ability to be self-sufficient in growth signals, avoiding antigrowth and apoptotic signals and gaining a limitless replicative potential (Hanahan and Weinberg, 2011).

To enable the uncontrolled cell proliferation and growth, cancer cells have to adapt their metabolism. Currently, reprogramming energy metabolism appears a fundamental cancer hallmark to assist cell growth and division. Otto Warburg in 1930 for the first time observed an aberrant use of energy in cancer cell metabolism (Warburg and Dickens, 1930). Indeed, cancer cells consume more glucose than normal cells and they switch from mitochondrial metabolism to glycolysis, which produces less energy but faster in oxygen low condition. This shift in the metabolism state, associated to an acid environment, is a common trait of cancer biology (Jones and Thompson, 2009; Warburg and Dickens, 1930).

1.2 GENOMIC INSTABILITY: THE TRIGGER THAT SUSTAINS TUMOR PROGRESSION

Interestingly, all the properties of the cancer cells are sustained in large part on a succession of alterations in the genome of neoplastic cells. Normal cells have a surprising ability to detect and resolve defects in the DNA through a complex genome maintenance system. In cancer cells the mutation rate is increased because there is a higher sensitivity to mutagenic agents, through a breakdown in one or several components of the genomic maintenance machinery (Schneider and Kulesz-Martin, 2004). This leads to general genome instability, which sustains the neoplastic transformation (Hanahan and Weinberg, 2011).

In fact, once cells acquire the ability to replicate limitless, division after division they accumulate mutations, which can confer new skills to cancer cells.

1.3 METASTATIC TRAIT AS A CHALLENGE FOR TUMOR ERADICATION

1.3.1 MIGRATION AND INVASION PROPERTIES

One of the prominent capacities that cancer cells gain is the acquisition of migration and invasion ability that ultimately falls in metastatic development. In this process a bulk of mutated cells lose their polarity and cell adhesion and spread from the primary tumor to

distant sites. The development of metastasis is a multistep program that begins with local invasion and then goes on the intravasation of cancer cells into nearby vessels and the transition through the lymphatic and blood systems. Whence cancer cells extravasate from the lumina of such vessels into the parenchyma of distant tissues where they form metastasis, colonizing the new tissue (Talmadge and Fidler, 2010). To do this, epithelial cells undergo first to Epithelial to Mesenchymal transition (EMT), acquiring the properties to migrate and invade, then to Mesenchymal to Epithelial transition, forming new tumors in distant locations (Tsai and Yang, 2013). The EMT program controls a particular type of invasiveness that has been termed ''mesenchymal''. Noteworthy, two other distinct modes of cell invasion have been identified: collective and ameboid modes (Friedl and Wolf, 2008, 2010). ''Collective invasion'' involves nodules of cancer cells that advance as a mass into adjacent tissues, for example this form of invasion is characteristic of squamous cell carcinomas and it is mainly sustained by carcinoma associated fibroblast (Gaggioli, 2008). In ''amoeboid'' form of invasion, cancer cells show morphological plasticity, enabling them to slither through existing interstices in the extracellular matrix (Fig. 2) (Madsen and Sahai, 2010; Sabeh et al., 2009).
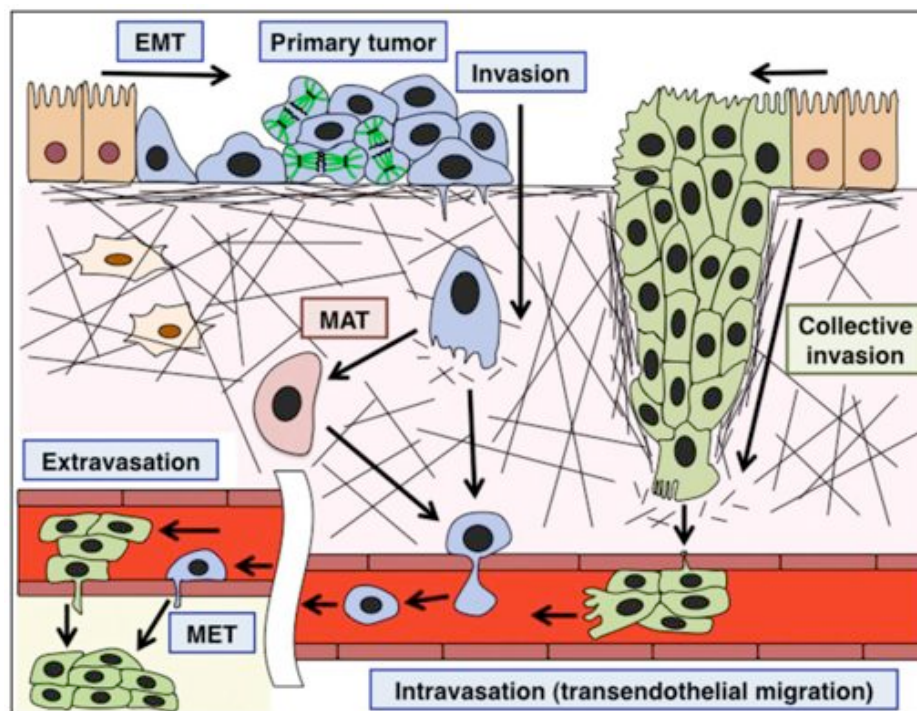
**Figure 2. EMT-mediated invasion, collective invasion and ameboid invasion in cancer metastasis.** Normal epithelial cells (orange cells) undergo EMT and form a primary tumor (blue cells). Some primary tumor cells invade and migrate into blood circulation as a multicellular strand (green cells) or in a mesenchymal fashion or undergo to MAT (Mesenchymal to Ameboid Transition) (pink cells) (Takeshi Kawauchi, 2012).

1.3.2 EMT AND CANCER STEM CELLS (CSCS): THE KEYS OF METASTATIC SUCCESS

Cancer progression is associated with the loss of epithelial differentiation and a switch toward mesenchymal phenotype, which is accompanied by increased cell motility and invasion (Lamouille et al., 2014).

During EMT cells lose their cell-cell contact, undergo actin reorganization and reprogram their molecular networks up-regulating several mesenchymal markers such as vimentin. Meanwhile, there is a loss of epithelial markers such as E-cadherin and the translocation of β-catenin from cell-cell junctions to nucleus (Thiery, 2002). In addition, also Notch and Wnt pathways are involved in this transition. Very recently it was shown that EMT is strongly linked to chemo-resistance *in vivo*. In this work, it has been demonstrated that cells that undergo EMT are protected from chemotherapy and show relapse in metastatic disease (Fischer et al., 2015).

Recently, Mani and his colleagues presented an interesting connection between CSCs and EMT, demonstrating that CSCs are the main actors in the metastatic process. Indeed, through specific stimulations (such as TGF-β), epithelial cells undergo to EMT acquiring many of the properties of self-renewing stem cells, and become able to invade new tissues and give rise to an entire tumor. Therefore, CSCs are defined as those cells within a tumor that can self-renew, drive tumorigenesis and give rise to metastases (Mani et al., 2008). CSCs act as a reservoir of cancer cells resistant to chemotherapy and this could in part explain why metastases are so difficult to eradicate and why they aggravate patient prognosis. CSCs are characterized by the presence of specific cell surface markers such as CD44 (high)/CD24 (low)/ALDH-positive, whereby it is possible to identify them. Moreover, other works underline that the CSCs pool is maintained after radiotherapy or chemotherapy because of the involvement of several molecular pathways such as the up-regulation of ABC transporters, the DNA repair system and the induction of Notch, Wnt and Hedgehog pathways (Elliott et al., 2010; Sampieri and Fodde, 2012). Dissecting the molecular pathways, that underline stemness maintenance, is fundamental for finding new targets in cancer therapy. This is a challenge for researchers because CSCs can be quiescent for years and then cause cancer relapse. Thus, destroying this nice of cells could eradicate cancer relapse.

1.4 TUMOR IS SUSTAINED BY ANGIOGENESIS

Until some years ago the metastatic process was seen as the last step of carcinogenesis. Recently this view is changing and metastases seem to occur at any stage of tumor growth,

moreover, metastases can be sustained by other hallmarks such as angiogenesis. Indeed, through the formation of new vessels cancer cells could easily escape from the primary tumor to the blood flow. Moreover, when a growing tumor exceeds 1 mm of dimension, it requires the access to oxygen and nutrients that are provided by angiogenesis. During this process cancer cells send out signals to the cells of neighbor blood vessels, inducing them to produce new branches in order to feed the tumor growth (Nishida et al., 2006). Therefore, angiogenesis, sustaining both primary tumor growth and metastatic process, encases a fundamental trait of cancer biology.

1.5 INFLAMMATION: THE DUAL ROLE IN CANCER OF NON-TUMORIGENIC CELLS

In many tumors, the tumorigenic signals are mainly sustained by oncogenes such as Ras and Myc, often mutated in cancer cells; whereas in other tumors such inductive signals are produced by immune inflammatory cells. These cells belong to cancer microenvironment, and their crosstalk with the neoplastic lesion is closely connected with cancer development. Immune system provides two distinct roles in tumors. On one side it contributes to many hallmark capabilities. Indeed, the vast majority of cancers shows a high infiltration of immune cells that through the secretion of bioactive molecules can sustain proliferative and proangiogenic signals, limiting cell death and facilitating angiogenesis, invasion, and metastasis (DeNardo et al., 2010; Grivennikov and Karin, 2010). On an opposite side immune system could have the ability to eradicate tumor (Teng et al., 2008). However, cancer cells can acquire the ability to evade immune destruction by disabling components of the immune system that have been dispatched to eliminate them (Mougiakakos et al., 2010; Ostrand-Rosenberg and Sinha, 2009). This is considered as an emerging hallmark of evading immune destruction.

All these hallmarks are common traits of cancer and represent processes that can be targeted to eradicate this disease definitively. Understanding basic mechanisms that control and define them is thus fundamental for the identification of new drug target.

2 BREAST CANCER

Breast cancer is the most frequent neoplastic disease and the leading cause of cancer death among women worldwide. Globally, 1.4 million of new breast cancer cases were diagnosed in 2008 and this reflect the number of new cases per year. Approximately, one-third of patient dies for the disease (Fig. 3). The Western world has the highest incidence rate and the lifetime risk of developing breast cancer is estimated to be one in nine women. (Larsen et al., 2014).
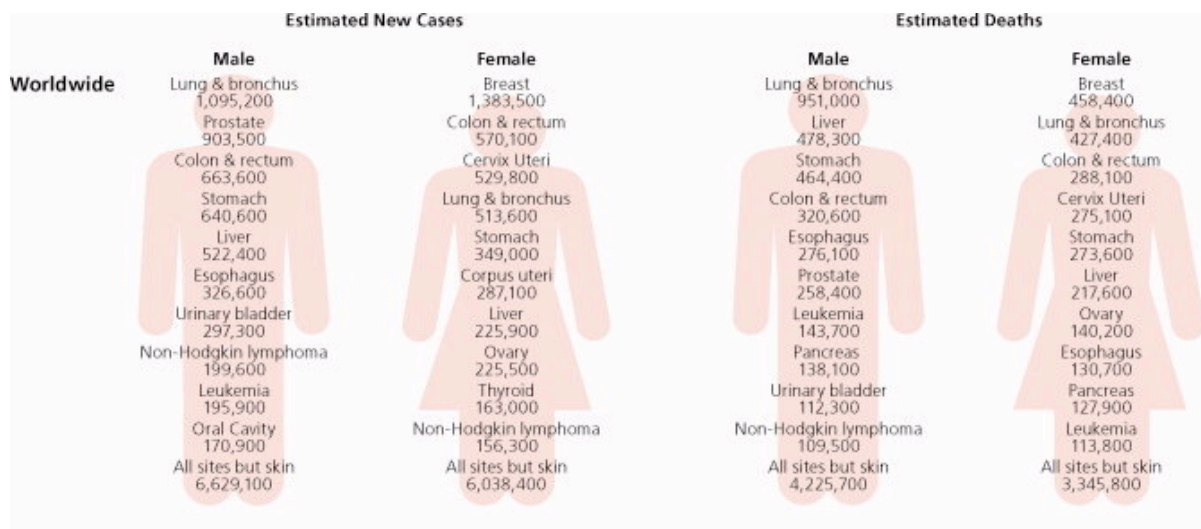
Figure 3. Estimated new cancer cases and deaths worldwide for leading cancer sites in 2008. Source: GLOBOCAN 2008 (A. Jemal 2011).

So, what is breast cancer? Breast cancer is a malignant tumor that starts in the cells of the mammary gland. Breast cancer cells, that grow uncontrollably and abnormally, can invade surrounding tissues and/or metastasize to distant areas of the body, preferentially to lung, brain and bones (Minn et al., 2005; Scully et al., 2012). In the last years the chances to survive have been increased and, thanks to the increase of knowledge of the molecular basis that govern breast cancer development, the treatments have been improved. This is true when the neoplastic lesion is diagnosed at early stage, but when breast cancer is detected later the healing options are very narrow.

2.1 MAMMARY GLAND

The mammary gland or breast is an exocrine tubuloalveolar gland part of the skin located in the breast lying on the top of the pectoralis muscles that is mainly involved in women milk production after childbirth. The functional units are alveoli and ductules, which respectively produce and transport milk through lactiferous ducts to lactiferous sinuses out to papilla mammae (nipple). Both structures lie in a stroma consisting of adipocytes and fibroblasts which is infiltrated by the blood and lymph vasculature.

Each structure described above is composed of a bi-layered epithelium consisting of luminal cells surrounded by contractile myoepithelial cells which contribute to milk ejection during lactation (Oakes et al., 2014). These two cellular lineages derived from multipotent Mammary Stem Cells (MaSCs) give rise to committed progenitors. MaSCs, that are present as niches in all mammary epithelial tree, are specialized cells with the capacity to both self-renew and generate daughter cells that can differentiate in several lineages. Moreover, these cells contribute to breast development and to support numerous rounds of proliferation, differentiation, and apoptosis that take place during and after pregnancy (Macias and Hinck, 2012).

Maturation of mammary gland occurs through ductal and alveolar morphogenesis. The first begin at birth when the mammary epithelium emerges from the mammary placode and continues growing isometrically until puberty. During this period the epithelium extends the ducts throughout the fat pad resulting in the growth of the breast. Instead the alveolar morphogenesis starts with pregnancy. Initial alveolar epithelium proliferates until mid-pregnancy when it differentiates into milk-secreting alveoli. This step is called lactogenesis I or secretory initiation and it is followed by lactogenesis II or secretory activation in which the final phase of differentiation results in milk and lipid movement into the alveolar lumens until weaning. Afterwards, breast undergoes rapid involution with the apoptosis of mammary epithelial cells that takes mammary gland back to puberty (Oakes et al., 2014). A second involution occurs in women life and it is menopausal age-related. This process involves the replacement of glandular epithelium and connective tissues with fat. The whole development is mainly coordinated by hormones: a) Growth Hormone (GH), that is fundamental in ductal morphogenesis and directs global postnatal development through Insuline Growth Factor 1 (IGF1); b) estrogens which promote branching differentiation at birth and during puberty; c) progesterone that regulates lobuloalveolar formation, maintains pregnancy and inhibits milk production; d) and prolactin which secretion induces milk formation and oxytocin, that acting

on myoepithelial layer, cause milk ejection (Macias and Hinck, 2012; Sternlicht, 2006). Moreover, the same hormones impact on many pathways that mediate cell-cell and cell-stroma crosstalk, whose finely regulation marks the transition between healthy and disease. Indeed, their de-regulation could provoke neoplastic transformation, ultimately inducing breast cancer. In this context, hormone receptors, and particularly estrogen and progesterone receptors (ER and PR), take a crucial role. In fact, the large percentage of breast cancer is ER/PR positive and this means that tumor is fueled by hormones. To stop this vicious loop, several molecules, blocking ER/PR (such as tamoxifen) or hormone production (aromatase inhibitors) were developed. The presence of these receptors or others like Her, as well as understanding the molecular profile, the stage, and the type of breast cancer, is fundamental whether for treatment of the disease that for the patient prognosis.

2.2 INVASIVE BREAST CANCER CLASSIFICATION

In order to clarify this point, the International Agency for Research on Cancer (IARC) has published in 2012 the WHO *Classification of Tumours of the Breast*, in which invasive breast cancer classification is based on tumor biomarkers expression, tumor grading, tumor staging according to the primary Tumor, regional lymph Nodes, distant Metastases (classification TNM), molecular expression profiling and histological tumor typing (Gannon et al., 2013).

2.2.1 BIOMARKER EXPRESSION

The main biomarkers considered in breast cancer are: Estrogen Receptor (ER), Progesteron Receptor (PR), HER2 and Ki67. As said before, hormones play a crucial role in breast cancer development and treatment, therefore their deregulation is a major point in defining prognosis and clinical therapy. Historically, the introduction of Immuno HistoChemistry (IHC) and in situ hybridization in standard practice allowed to identify biomarkers which expression is able to stratify patients and give important information about prognosis and treatment of the disease (Weigel and Dowsett, 2010).

2.2.1.1 ER/PR

ER and PR are nuclear transcription factors stimulated respectively by estrogen and progesteron. They promote the growth of normal and neoplastic breast epithelium. Approximately 70% of invasive breast carcinomas express ER and PR. These tumors are responsive to hormone therapies including tamoxifen, aromatase inhibitors and ovarian ablation.

2.2.1.2 HER2

The HER2 gene encodes for a HER2 receptor, a member of the human epidermal growth factor receptor family, which also includes EGFR (alias HER1), HER3, and HER4. It is expressed on the surface of normal breast epithelial cells and amplified in 15% of invasive breast cancer. Its expression correlates with poor prognosis. Nevertheless, tumors that have HER2 amplification well respond to target therapy, including trastuzumab and lapatinib.

2.2.1.3 Ki67

Ki67 antigen is a nuclear protein closed related with active phase of cell cycle. It is present in proliferating cells and absent from resting cells. Ki-67 labeling index can stratify patient outcome and treatment with hormone receptor presence (Scholzen and Gerdes, 2000). Unlike ER/PR and HER2 biomarkers, Ki67 evaluation is less used in breast cancer management, because there is no international score system.

2.2.2   TUMOR GRADING

Tumor grading defines the aggressiveness of breast cancer and classifies it, in accordance with Nottingham Histologic Score, in order to identify its differentiation status. It is based on microscopic estimation of three different parameters: tubule formation, nuclear pleomorphism and mitotic activity (Elston and Ellis, 1991). In the first, the formation of tubular organization is evaluated counting structures with a central lumen. Nuclear pleomorphism is estimated according to the size, the shape, the presence and prominence of nucleoli and the character of the nuclear chromatin of the tumor cell nuclei. Lastly, mitotic activity that depends on field diameter of microscope is assessed counting a minimum of 10 fields at the periphery of the tumor. A score from 1 to 3 is assigned at each parameter and the sum of scores of the three parameters permits to assign the tumor grade. Lower is the grade less aggressive is the tumor.

2.2.3   TUMOR STAGING

Tumor staging or TNM system, outlined by the American Joint Committee on Cancer / Union for International Cancer Control, classifies information about the size of the primary tumor (T), the status of the regional axillary lymph nodes (N) and the presence of distant metastases (M) at diagnosis into one of five stages (0, I, II, III, IV) (Webber et al., 2014). As for tumor grading, lower is the stage less aggressive is the tumor. The size of the primary invasive tumor is assessed first through radiological and clinical estimation than through histological analysis, calculating the dimension of the tumor or the largest one in case of multiple lesions. The status of the regional axillary lymph nodes considers the extent of regional lymph node

involvement and the size of lymph node metastasis. It is assessed mainly through the identification of sentinel lymph nodes, that are the most proximally lymph nodes present surrounding the tumor. The presence or absence of metastasis is the third parameter taken into consideration by the TNM system and the evaluation goes through patient symptomatology, clinical examination and radiological explorations.

2.2.4 HISTOLOGICAL SUBTYPING

Histopathologic evaluation of a breast cancer is necessary to provide the diagnosis of the tumor, to determine a patient's prognosis, and to understand the nature of breast cancer. From the histopathological point of view breast cancer is divided in two main groups: in situ carcinoma and invasive (infiltrating) carcinoma. The first is further sub-classified in ductal and lobular breast carcinoma of which ductal carcinoma in situ (DCIS) is more common and heterogeneous than the lobular one. On the basis of architectural features, it is possible to recognize five different DCSIs: Comedo, Cribiform, Micropapillary, Papillary and Solid. Similar to in situ carcinoma, invasive ones are heterogeneous and sub-classified in at least six subtypes among which the more common is the infiltrating ductal subtype. The others are the tubular, ductal lobular, invasive lobular, mucinous and medullary (Malhotra et al., 2010).

2.2.5 MOLECULAR PROFILING

The advent of high throughput technologies, to analyze gene expression profiles, provides the basis for an improved molecular taxonomy of breast cancer. Starting from microarray-based gene expression profiling studies on individual breast cancer patients, among large cohorts of breast cancer samples, it is possible to simplify the complexity of this disease and to stratify patients in molecular clusters, that reflect a specific prognosis and treatment response (Perou et al., 2000; Sorlie et al., 2001, van 't Veer, 2002). These molecular clusters, also defined as subtypes, are at least five: luminal A, luminal B, HER2-like, normal-like and basal-like (Perou et al., 2000).

2.2.5.1 LUMINAL A and LUMINAL B

Luminal A and B breast cancer subtypes are the most common, characterized by the presence of ER expression, a good prognosis and a fairly clear treatment management. Luminal B breast cancer subtype expresses also HER2 and, for this reason, it has a higher histological grade and a worse prognosis than luminal A.

2.2.5.2 HER2-LIKE

HER2-like breast cancer subtype is associated with HER2 gene amplification and/or HER2 protein overexpression. It is characterized by poor clinical outcomes, but is also predictive of positive therapeutic responses to anti-HER2 drugs such as trastuzumab.

2.2.5.3 NORMAL-LIKE

The normal-like breast cancer cluster is closely related with normal breast epithelium in microarray studies. It is not clear if this is a distinct molecular subtype of breast cancer or simply a group of breast cancers in which there is a large contamination of normal epithelium.

2.2.5.4 BASAL-LIKE

Basal-like breast cancer is the most aggressive subtype characterized by high histological grade, poor recurrence-free and overall survival outcomes. Because of the lack of ER, PR and HER2 it is not suitable for hormone and target therapy and the only option is chemotherapy, although patients have a worse outcome after chemotherapy than patients with other breast cancer subtype.

2.3 BASAL-LIKE BREAST CANCER: FEATURES AND AGGRESSIVENESS

Basal-like subtype comprises a heterogeneous group of breast cancer that account for up to 15% of all breast cancers. It is more common in African-American young women and in patients with BRCA1 mutation. It is characterized by a weakened correlation between the size of primary tumor and the probability of survival (Foulkes et al., 2010). This subtype expresses genes and proteins present in breast myoepithelial cells including cytokeratins 5/6, 14 and 17. Moreover, the p53 proto-oncogene is often mutated and alterations of the pRB and p16 G1/S cell-cycle checkpoint are remarkably prevalent in these cancers (Badve et al., 2011). The origin of these tumors is still controversial. Indeed, even if the majority of basal-like breast cancers takes origin from the myoepithelial cells of breast a subgroup may originate from luminal progenitor because of the expression of cytokeratins 8/18 (Lim et al., 2009). Moreover, basal-like breast cancers display features consistent with those of breast stem cells but it is still unclear whether basal-like breast cancers originate from breast cancer stem cells or through EMT produced cells with stem cell properties. This is crucial for the correct classification and therefore for patient management.

Notably, the majority of basal-like breast cancers is also classified as Triple Negative Breast Cancer (TNBC). Indeed, they are usually negative for ER, PR and HER2 expression and for

this reason conventional therapies are not suitable for these kinds of breast cancers (Foulkes et al., 2010).

TNBC comprises different molecular subgroups of breast cancer. Although approximately 80% of TNBC are basal-like, the rest include claudin-low subgroup of breast tumors, which are reported to be enriched with cells that have features of epithelial-to-mesenchymal transition, and the interferon-rich subgroup, which incorporate tumors with a significantly better prognosis.

All these classifications underline the complexity of breast cancer and highlight the necessity of a more personalized therapy that needs to find new targets and new biomarkers that can stratify patients.

2.4   BREAST CANCER GENE-SIGNATURE AS CLINICAL TOOL

The ability to assess the gene-expression profile, introduces in medicine and in research the concept of gene signature. A 'gene-signature' can be defined as a single or a combined genetic alteration with specificity in terms of diagnosis, prognosis or prediction of therapeutic response (Chibon, 2013). It is possible to recognize two mainly types of gene signature: the prognostic and the predictive. In general, a prognostic factor is a clinical or a biological feature, measurable, that provides information on the outcome of the cancer disease. Instead, a predictive factor gives notions on the benefit for a treatment stratifying patients who will response better to a specific therapy. For example, the molecular profiling presented by Perou and colleagues in 2000, that subdivides breast cancers in the subtypes described in chapter 2.2.5. (Colombo et al., 2002; Perou et al., 2000), could be considered not only diagnostic but also prognostic because it provides information on the breast cancer outcome in terms of risk of relapse and death.

Focusing on prognostic gene-signature, there are three mainly rules that defined it: (i) it has to be specifically associate with clinical outcome, (ii) it has to be reproducible in an independent group of patients and (iii) it has to be an independent prognostic value from other standard factors in a multivariate analysis. Two opposite approaches can be used for identify a prognostic gene signature. The first is the "top-down" supervised approach in which the expression profile associates with the occurrence of the event investigated, while in the second strategy, called "bottom-up", a genic expression is firstly associated with a biological phenotype, than it is tested weather it correlates with patient outcome (Chibon, 2013).

Gene-signatures are currently used to predict in a more accurate way the outcome, and to improve the selection of the adjuvant systemic treatment for individual patients with invasive breast cancer. Indeed, starting from 2000, several independent groups have conducted gene-expression profiling studies in breast cancer patients to develop prognostic gene-signatures (such as MammaPrint, Oncotype DX and 76-gene signature) (Gingras et al., 2015; Sotiriou and Piccart, 2007). Among these gene-signatures, two are currently used as diagnostic assay in clinical management to estimate, in ER positive breast cancer, a woman's risk of recurrence of early-stage breast cancer and those women that will be cured even if they do not receive adjuvant chemotherapy (Arranz et al., 2012; Sotiriou and Piccart, 2007). The first is based on RT-PCR assay performed on paraffin sections that score the recurrence of 21 genes (Oncotype DX) (Kaklamani, 2006) and the second one is the 70-gene signature (MammaPrint) (Glas et al., 2006) and it is a microarray based assay performed on formalin-fixed, paraffin embedded tissue. Currently, many works take in consideration TNBC in order to find gene-signature that stratified patients on a survival and on a specific treatment point of view. Due to the extreme complexity of this disease, TNBC gene signatures are still far to be applied in clinical management and they have to be tested prospectively (Arranz et al., 2012).

3 MOLECULAR FACTORS THAT INFLUENCE BREAST CANCER AGGRESSIVENESS

Breast cancer is a heterogeneous disease characterized by dysregulation of multiple cellular pathways, which confer to each cancer specific characteristics and different aggressiveness degrees. Beside this, there is not a standard way to describe breast cancer aggressiveness, considering that factors associated with this peculiar condition have a considerable impact on patient management. Aggressiveness is closely linked with metastasis formation and breast cancer recurrence, regarding which the causes remain barely unknown. Many molecular factors and signaling pathways are described to be linked with a more aggressive phenotype (Ahmad, 2013). In this thesis section I will summarize the well-known molecular signals that explain, at least in part, the metastatic relapsing cancer behavior.

3.1 NOTCH SIGNALING PATHWAY

Notch signaling pathway activation is considered a main feature in breast cancer development and recurrence, correlating with poor prognosis (Yuan et al., 2015). In humans, there are four

Notch receptors named Notch-1, Notch-2, Notch-3, Notch-4 and several ligands such as Jagged-1, Jagged-2, Delta-like 1, Delta-like 3 and Delta-like 4. Mechanistically, after ligand interaction with the receptor, Notch receptor undergos a proteolytic cleavage that releases its intracellular portion, which translocates to the nucleus and regulates several target genes (Kopan and Ilagan, 2009). Notch pathway is closely linked to aggressiveness under many points of view, affecting multiple processes including proliferation, stem cell maintenance and migration. Indeed, Jagged-1 is upregulated both at mRNA and protein levels, in poor prognosis breast cancer patients, particularly in TNBC compared with ER/HER2 positive breast cancers (Reedijk et al., 2008). In TNBC, Jagged-1 regulates through the interaction with Notch-1 receptor Cyclin D1 expression, impacting cell cycle progression (Cohen et al., 2010). Moreover, Notch pathway controls CSCs potential in TNBC and in HER2 positive breast cancer. Firstly, it has been found an increased activity of Notch-1 and Notch-4, in breast cancer cells enriched for CSC markers, suggesting an involvement of this cell signaling in breast CSC sub-population (Harrison et al., 2010). Moreover, Notch-3 has been suggested to have a role in CSC quiescence helping cells to evade cytotoxic effects of anticancer therapy (Kent et al., 2011). Finally, it has been demonstrated that it is possible to kill CSCs both *in vitro* and *in vivo* inhibiting Notch pathway (Grudzien et al., 2010). All these data suggest that Notch signaling pathway has a fundamental role in breast cancer pathology and therefore it could be a good target for evading breast cancer aggressiveness.

3.2   WNT SIGNALING PATHWAY

Wnt signaling pathway begins with the interaction between Wnt glicoproteins family with the seven trans membrane passage receptor Frizzled that activates the cytoplasmic phosphoprotein Disheveled. Then, the signaling can follow three pathways: the canonical Wnt pathway, the noncanonical planar cell polarity pathway, and the noncanonical Wnt/calcium pathway (Komiya and Habas, 2008). The best studied is the canonical Wnt pathway, in which the main actor is the β-catenin. In the absence of Wnt ligands, β-catenin is located into a destruction complex composed of APC and axin, which leads to ubiquitylation and proteasomal degradation of β-catenin. After Wnt activation, β-catenin accumulates in the cytoplasm and translocates to the nucleus, where it forms a complex with TCF/LEF transcription factors and regulates the transcription of Wnt-target genes (Widelitz, 2005). Several studies highlighted the involvement of Wnt pathway in breast cancer aggressiveness. In particular, it is emerged that activation of Wnt correlates with poor prognosis and

metastatic behavior of breast cancer, regulating crucial processes such as EMT and stemness. In fact, it has been demonstrated that in breast cancer cells, canonical Wnt signaling activates EMT gene expression program (Wang et al., 2013). Some evidence rise from mutations or inactivation of Wnt inhibitors such as HBP1 whose inactivation leads to increased invasive potential, moreover its inactivation correlates with invasive breast cancer in clinical breast cancer samples (Paulson et al., 2007). Moreover, it has been demonstrated that Wnt pathway is upregulated in basal-like subtype and in triple negative breast cancer compared to normal patient samples (Debies et al., 2008). The Wnt signaling pathway is largely involved in stem-like cell properties, in fact, the inhibition of Wnt signaling pathway by cisplatin plus TRAIL reduces CSCs (Debies et al., 2008). In addiction to this, the over-expression of the downstream Wnt target LEF1 is present in breast CSCs (Lamb et al., 2013). Very recently, it has been shown that in mouse Wnt pathway is fundamental for CSCs maintenance and pro-metastatic properties (Jang et al., 2015).

3.3   HEDGEHOG SIGNALING PATHWAY
Hedgehog signaling pathway is involved in breast cancer recurrence when altered, but in physiological condition it is implicated in organ development. In mammals, there are three ligands that induce this signaling pathway, Sonic Hedgehog (Shh), Desert Hedgehog (Dhh), and Indian Hedgehog (Ihh). The best characterized is Shh which binds the receptor Patched resulting in activation and translocation into the nucleus of Gli1 that activates Hh-target genes transcription (Ingham and McMahon, 2001). The up-regulation of activity of Hh pathway is a strong determinant of increased risk of metastasis in invasive breast ductal carcinoma (O'Toole et al., 2011). Moreover, Shh has been shown to increase angiogenesis in a VEGF-independent pathway and to induce bone metastasis (Harris et al., 2012). Indeed, blocking Hh pathway could impair bone metastasis of breast cancer. In addition, Gli1 is overexpressed in basal-like breast cancer cells and negatively correlates with ER presence and with survival (Tao et al., 2011).

3.4   HIPPO SIGNALING PATHWAY
Hippo signaling pathway is a master regulator of organ size control and a tumor suppressor pathway frequently deregulated in cancer (Gomez et al., 2014). The canonical cascade is composed by two core kinases family, MST1/2 and LATS1/2 kinases, that act in concert with SAV1 and MOB1 co-factors, and once activated phosphorylate the proto-oncogenes YAP and

TAZ which then are inactivated through cytoplasm retention and/or degraded via proteasome (Huang et al., 2005). Instead, when Hippo pathway is inactive, YAP and TAZ move into the nucleus, acting as transcriptional co-activators of several transcription factors, such as TEAD1-4, SMADs, TBX5 RUNX1/2, and p73 and then inducing gene expression (Johnson and Halder, 2014). YAP and TAZ activity is upregulated in various human cancers, supporting the idea that they act as proto-oncogenes. Indeed, they play an important role in several cancer-related cellular processes, increasing cellular proliferation, inhibiting apoptosis and deregulating cellular differentiation (Moroishi et al., 2015). However, their role in breast cancer aggressiveness is under debate. Indeed, several works demonstrate that overexpression of YAP and TAZ can induce anchorage independent growth, EMT and invasion properties of immortalized mammary epithelial cells *in vitro* (Lamar et al., 2012; Pan, 2010). Likewise, a down-regulation of YAP reduces breast cancer cell proliferation. Moreover, TAZ is overexpressed in high-grade breast cancers (Zhang et al., 2015). Beside its role as oncogenes, YAP acts also as tumor suppressor potentiating p-73-mediated apoptosis (Strano et al., 2001). In contrast, TAZ confers chemoresistance to breast cancer stem cells (Pan, 2010).

4   THE HIGH MOBILITY GROUP A (HMGA)

"High Mobility Group" (HMG) proteins are chromatin architectural factors that constitute critical hubs in the chromatin network, modulating chromatin structure and orchestrating the participation of other proteins in nuclear activities (Cleynen and Van de Ven, 2008). They are characterized by a small molar masses of less than 30 kDa, a relatively high mobility in acidic (pH 2.4) polyacrylamide gel electrophoresis system compared to other nuclear protein (Reeves and Wolffe, 1996), the presence of a long negatively charged carboxy-terminal tail, and the involvement in some common biological processes such as embryonic development, regulation of transcription, recombination and modulation of DNA repair (Reeves, 2015). HMG proteins are divided in three families that differ each other for the presence of a typical functional sequence motif. Through their respective functional domains, HMG proteins are able to recognize specific structure in the DNA or chromatin rather particular nucleotide sequence; in this way they differentially affect chromatin structure, nuclear metabolism and cellular phenotypes. The three families are composed by:

- HMGA or HMG-AT-hook family characterized by the presence of three positively charged AT-hooks able to bind DNA minor groove rich in Adenine and Thymine (A/T),

- HMGB or HMG-Box proteins containing two alpha-helix HMG boxes which can bind with high affinity non-B-form DNA structures,

- HMGN or HMG-Nucleosome Binding proteins that contain a positively charged nucleosome binding domain through which they associate with nucleosomes.

My thesis will be focused on HMGA proteins family.

4.1   HMGA FAMILY OF PROTEINS

In humans, the HMGA family is composed by two functional members: the HMGA1 gene, located at the chromosome band 6p21, and the HMGA2 gene, located at the chromosome band 12q13–15. They encode for HMGA1a (107 amino-acid residues) and HMGA1b (96 amino-acid residues) proteins produced through alternative splicing from HMGA1 gene and for HMGA2 protein (108 amino-acid residues) from HMGA2 gene (Fig. 4) (Fusco and Fedele, 2007). As mentioned before, HMGA proteins contain three highly conserved palindromic basic amino acid motifs (proline-arginine-glycine-arginine-proline) defined as "AT-hooks" that bind to the minor groove rich in A/T of B-form of DNA. Moreover, they are characterized by the presence of an acidic carboxy-terminal tail, highly post-translational modified, probably involved in the regulation of functional activities of HMGA proteins (Sgarra et al., 2004).
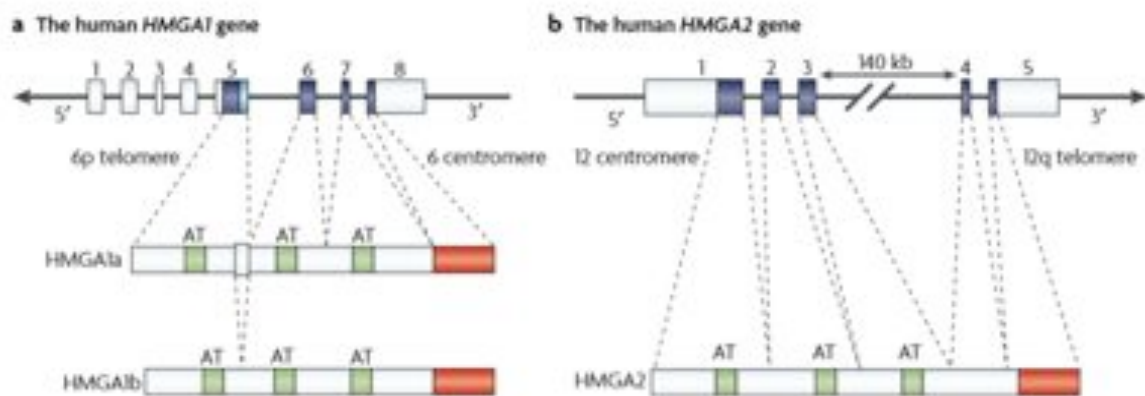

Figure 4. Gene structure and organization and schematic protein product of HMGA1 and HMGA2. In blue are presented coding exons, in green AT-hooks and in orange C-terminal tail (Fedele and Fusco 2007).

Peculiarly, as free molecules, they do not show any secondary or tertiary structure and so they are considered as a paradigm of disordered random coiled proteins. HMGA proteins assume a distinct secondary structure after binding to DNA or other substrates (Reeves, 2001). This feature could explain the extreme plasticity of these proteins in participating in a wide variety of biological processes, including cell growth, proliferation, differentiation, DNA repair and cell death. Notably, they are expressed at high level during embryogenesis and generally absent in adult cells, but they are re-expressed in cancer (Fusco and Fedele, 2007).

4.2 HMGA REGULATION

In general, gene expression is regulated at several layers that start with transcriptional control, passing through post-transcriptional and translational regulation, concluding with post-translational modifications. HMGA proteins expression is not exempt from this fine regulation. Several works demonstrate that HMGA gene expression can be induced in response to a plethora of cellular stimuli such as many growth factors (EGF, TGF-alpha, PDGF and FGF) (Cleynen et al., 2007; Wood et al., 2000) and mediators of inflammation (IL-1beta and endotoxin) (Pellacani et al., 1999), and can be repressed in different cellular contexts by retinoic acid (Giannini et al., 2005). How induction of HMGA proteins occurs during cell transformation is not completely clear. It is emerged that HMGA1 and HMGA2 are regulated at different levels. Indeed, HMGA1 is principally deregulated at the transcriptional level, due to the extreme complexity of its promoter structure, that presents many transcriptional start sites and several cis-acting elements as RAS responsive region and AP1 and SP1 binding sites (Cleynen et al., 2007) and C-MYC target region (Wood et al., 2000). Instead HMGA2, beyond the transcriptional control, is mainly regulated at a post-transcriptional level by miRNA machinery. In particular, independent groups showed that HMGA2 transcript is strictly regulated by let-7, interestingly, disruption or loss of 3'UTR of HMGA2 cause its protein overexpression in different tumor diseases (Lee and Dutta, 2007). In the last years, also post-transcriptional regulation of HMGA1 is emerging, principally through miR-34b, miR-548c-3p (D'Angelo et al., 2012; Kaddar et al., 2009; Liu et al., 2015) Once HMGA mRNAs are translated in proteins, their molecular function is finely regulated by post-translational modifications. Both HMGA1 and HMGA2 protein are highly modified by different enzymes. The best known enzymes responsible for HMGA protein modifications are: cdc2, that phosphorylates HMGA1a Thr-52 and Thr-77 (Schwanbeck et al., 2000), CK2

which phosphorylates the C-terminal domain of both HMGA1 and HMGA2, regulating DNA binding ability (Palvimo and Linnala-Kankkunen, 1989), CBP and PCAF that acetylate HMGA1 on Lys64 and Lys70 coordinating the transcriptional switch of INF-beta (Munshi et al., 2001) and PRMT6 that methylates HMGA1 on Arg57 and Arg59 (Sgarra et al., 2006).

4.3 MOLECULAR MECHANISMS OF ACTION OF HMGA

HMGA proteins contribute to many biological processes through their ability to bind both DNA and proteins, assembling and/or modulating macromolecular complexes. They are mainly involved in gene expression even if they do not have transcriptional activity per se. Indeed, they regulate the transcriptional activity of several genes, interacting with the transcription machinery and altering chromatin structure. There are three principal mechanisms of action through which HMGA proteins influence transcription: they can interact specifically with (i) DNA and/or (ii) transcription factors or, more in general, they can act on (iii) chromatin level.

(i)   HMGA proteins bind and bend DNA at enhancer level forming stereospecific macromolecular complexes called "enhanceosome". The best studied is the regulation of Interferon-Beta (INF-ß) which is a paradigm of the action of HMGA proteins on DNA. In fact, before the interaction of HMGA1a to the DNA sequence of the INF- ß enhancer, the binding affinity of transcription factors (c-JUN, ATF2 and NF-kB) to the enhancer is very low. The binding of HMGA1a to the enhancer unbends the DNA and increases transcription factors binding affinity with consequently positive regulation of INF-ß expression (Yie et al., 1999) (Fig. 5 panel a).

(ii)   HMGA proteins can interact directly with transcription factors by modifying their conformation or availability and modulating their affinity to DNA. One of the first examples is given by the work of M T Chin (Chin et al., 1998) and colleagues, in which they show that HMGA1a protein binds directly Serum-Response Factor (SRF) and enhances SRF-DNA interaction on c-FOS promoter, potentiating c-FOS expression. In addiction, among the best studied regulation of transcriptional activity by HMGA interactions with transcription factors, there are: the binding to p120E4F (E4F transcription factor 1) that induces Cyclin A expression (Tessari et al., 2003), the binding with pRB that permits the release of E2F1 and the transcriptional activation of E2F1-target genes (Fedele et al., 2006b) and the inhibition of p53 oligomerization

interfering with p53-mediated gene transcription blocking the p53-dependent DNA damage response (Fig. 5 panel b) (Frasca et al., 2006).

(iii)   More in general, HMGA proteins are able to alter chromatin structure. In particular, HMGA1 proteins are associated with matrix and scaffold-associated region (MAR/SAR) that are segments of genomic DNA with high affinity for the nuclear matrix, enriched in A/T sequences, whereby form loops. Moreover, HMGA1 protein, competing with histone H1 for DNA binding, de-represses the histone H1-mediated inhibition of SAR transcription (Fig. 5 panel c) (Zhao et al., 1993).
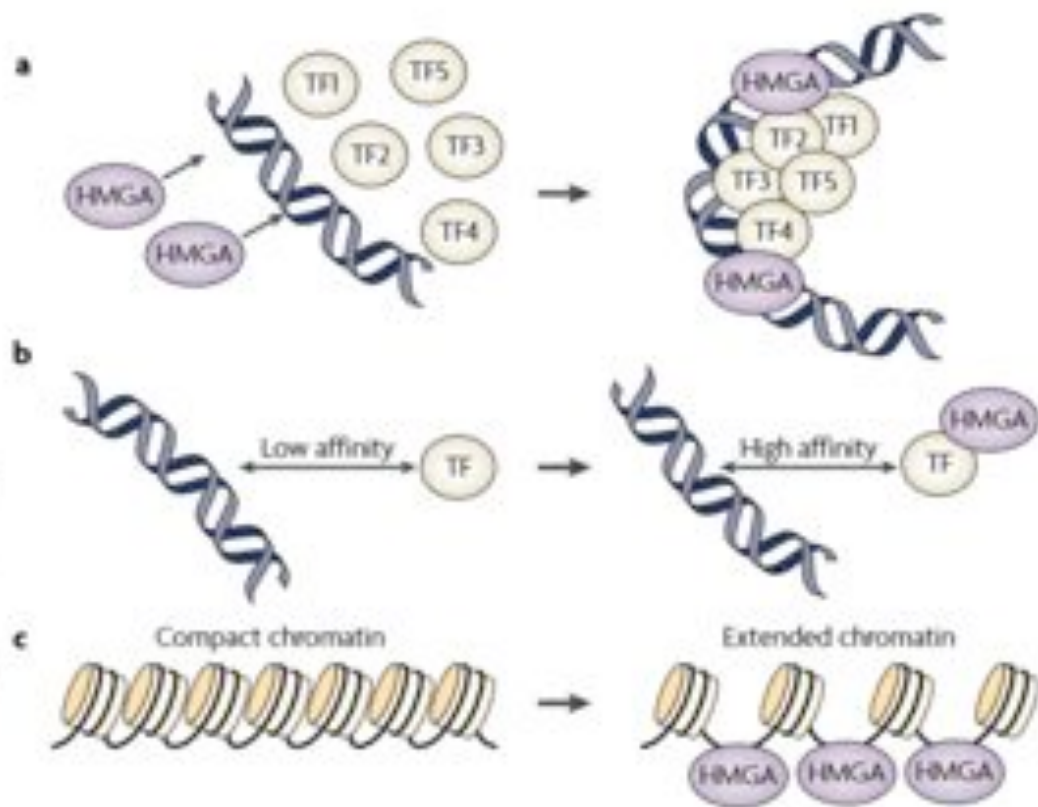
Figure 5. Ways of action of HMGA in gene expression regulation. In a. HMGA proteins directly bind to the DNA, modifying its conformation and consequently facilitating the binding of a group of transcriptional factors. In b. HMGA proteins directly interact with other proteins modifying their conformation and enhancing the affinity of its binding to DNA. In b. HMGA proteins globally alter chromatin structure.

Beside the role of HMGA proteins in the regulation of gene transcription, their involvement in other molecular processes is emerging, such as DNA repair (Reeves, 2015), RNA processing (Manabe et al., 2003), and modifying enzymes cooperation (Sgarra et al., 2010).

4.4 BIOLOGICAL FUNCTIONS OF HMGA PROTEINS AS ONCOFETAL PROTEIN

HMGA proteins are involved in a wide range of biological processes, both physiological and pathological, thanks to their biochemical properties and the molecular mechanisms described above. They are highly expressed during embryogenesis but their expression levels drop following cellular differentiation, being almost undetectable in the vast majority of adult tissue. Nevertheless, they are re-expressed in tumors correlating with grade (Flohr et al., 2003). For these features, HMGA proteins are considered oncofetal proteins. Their role in development was demonstrated *in vivo* in mice through knockout (KO) of Hmga1, Hmga2 and the double KO that give origin to different phenotypes. Indeed, while Hmga1-/- mice are normal in size, exhibiting cardiac hypertrophy, insulin resistant type 2 diabetes and myelo-lymphoproliferative disorders (Fedele et al., 2006a), Hmga2-/- mice exhibit pygmy phenotype, showing reduction in the amount of fat tissue and block of spermatogenesis (Zhou et al., 1995). Instead, double Hmga1/Hmga2 KO mice display a strong embryonic lethality, and the few mice alive exhibit a "super-pigmy" phenotype and a reduced vitality (Federico et al., 2014).

The over-expression of HMGA proteins is a common feature of cancer development and their expression often correlates with tumor grade. Since their discovery in HeLa S3 cells, their involvement in malignancy has been hypothesized (Lund et al., 1983). Nowadays, several papers show a causal role of HMGA proteins in cell transformation mainly, but not exclusively, through their ability to regulate gene expression. In fact, it has been demonstrated that, in cancer they positively modulate the expression of malignant phenotype inducing for example genes such as VEGF and matrix-degrading proteases type-I collagenase, cooperating with the transcription factors AP1 (Vallone et al., 1997). On the contrary, HMGA proteins downregulate the potential tumor-suppressor HAND1 in human thyroid carcinoma-derived cell lines (Martinez Hoyos et al., 2009). Therefore, it is widely accepted that HMGA proteins control different biological processes, involved in cancer hallmarks, mainly acting on the regulation of gene transcription. In fact, HMGA proteins act on multiple fronts on cell proliferation. On one side HMGA2 stimulates cell cycle through the activation of E2F1. Indeed, HMGA2 is able to bind pRB, displacing HDAC1 from pRB/E2F1 complex and then enhances acetylation of E2F1 and histones, inducing cell cycle-gene transcription (Fedele et al., 2006b). Moreover, HMGA2 promotes cyclin A gene expression, it associates with the transcriptional repressor p120E4F and this way it interferes with p120E4F binding to cyclin A promoter (Tessari et al., 2003). On the other side, HMGA1 inhibits p53-dependent apoptosis

blocking its homodimerisation and so its activity as "guardian of genome" (Frasca et al., 2006).

Both HMGA1 and HMGA2 are involved in affecting genome integrity, promoting its instability. For example, several works show that HMGA proteins negatively regulate a number of DNA repair genes, for example XPA and ERCC1, which are involved in single strand DNA damage response and are repressed respectively by HMGA1 and HMGA2 (Adair et al., 2005; Borrmann et al., 2003). Moreover, HMGA proteins regulate also double strand DNA damage response. In fact, HMGA2 impairs DNA-PK activity inhibiting non-homologous end joining (NHEJ) (Li et al., 2009), while HMGA1 negatively regulates BRCA1 gene expression in breast carcinoma (Baldassarre et al., 2003). In addition, HMGA proteins are inlaid in ATM pathway, promoting both the genomic instability and chromosomal rearrangements. It has been demonstrated that these alterations are often observed and associated with HMGA overexpressing breast cancer cells (Palmieri et al., 2011).

Importantly, in the last fifteen years the role of both HMGA1 and HMGA2 in epithelial-mesenchymal transition emerged. In 2001 Reeves and colleagues suggested, for the first time, that the over-expression of HMGA1 in human mammary epithelial cell line induces EMT (Reeves et al., 2001). More recently, it was confirmed the involvement of HMGA1 in this process, in fact, it has been demonstrated that the down-regulation of HMGA1 in breast cancer mesenchymal cells induces a strong mesenchymal to epithelial transition and a concomitant impairment of cancer stem cell properties and migration ability closely linked with EMT process (Shah et al., 2013). Meanwhile, several studies demonstrated that HMGA2 regulates EMT in different cancer cells intersecting many pathways involved in this process. In one of the best described work, it is shown that in mammary mouse epithelial cell line, TGF-beta mediates EMT by inducing HMGA2 via Smad pathway. In this context, the silencing of HMGA2 leads to E-cadherin up-regulation restoring cell-cell contact (Thuault et al., 2006).

Another cancer hallmark, in which HMGA proteins are involved probably in cooperation with NF-kB, is inflammation. Indeed, in hypoxia condition, HMGA1 induces the up-regulation of COX-2, which is thought to elicit a number of cellular pathways involved in neoplastic transformation, including proliferation, angiogenesis, metastasis, and inhibition of apoptosis (Tesfaye et al., 2007). Moreover, HMGA1a is able to upregulate STAT3, a key mediator of

inflammatory signals and molecular pathways that contribute to cancer initiation and progression (Hillion et al., 2008).

Therefore, HMGA proteins have a double role in embryonic development and in cancer. In this thesis I will focus on HMGA1 protein in cancer and in particular its role in breast cancer aggressiveness.

4.5   HMGA1 IN BREAST CANCER

HMGA1 protein was discovered for the first time in the human cell line HeLa (Lund et al., 1983) and in thyroid rat cell lines FRTL5-C12 and PC-C13 transformed with viral oncogenes (Giancotti et al., 1987). As described above, its expression is almost undetectable in adult tissue but it is over-expressed in several epithelial solid tumors such as thyroid, uterine cervix, prostate, colon, lung, pancreas, stomach and breast cancer. A number of works have demonstrated a positive correlation between cellular levels of HMGA1 mRNA or protein and metastatic potential in different tumors (Bussemakers et al., 1991; Chiappetta et al., 1996; Fedele et al., 1996; Leman et al., 2003; Tamimi et al., 1996). Furthermore, transgenic mice overexpressing HMGA1a develop lymphomas and uterine tumors (Xu et al., 2004). These data underline not only an association between HMGA1 expression and neoplastic transformation, but also a direct involvement.

A specific contribution of HMGA1 protein in neoplastic development in breast cancer was underlined by a series of studies. In first works, HMGA1 mRNA levels were evaluated in a range of mouse mammary cell lines with different degree of aggressiveness from pre-neoplastic to highly metastatic, derived from the same hyperplastic parent cell population. Interestingly, the levels of mRNA positively correlate with the relative degree of the transformation and metastatic grade of mouse mammary cells (Ram et al., 1993). A similar observation was done in human breast neoplastic cell lines in which it has been shown that HMGA1 expression is increased compared with normal breast epithelial cells (Nacht et al., 1999). Interestingly, these two works suggested a possible role of HMGA1 in breast cancer tumorigenesis. Meanwhile, other groups have demonstrated that HMGA1 expression levels are regulated by metastatic cell signals, such as stimulation by growth factors and activity of metalloproteinase. Indeed, their expression was increased by several fold in metastatic breast cancer cell lines upon EGF treatment, that leads to the induction of signaling pathways correlated with breast cancer progression to invasion and metastasis (Holth et al., 1997). Moreover, it has been shown that HMGA1 is induced in non-metastatic breast cancer cell

lines after heregulin (HRG) overexpression, inducing metastatic ability. On the other hand, in metastatic breast cancer cell lines it has been observed a decrease of HMGA1 expression after blocking matrix metalloproteinase-9 (MMP-9) that induces the loss of metastatic ability (Liu et al., 1999). These results suggest a strong connection between HMGA1 levels and metastatic abilities (Evans et al., 2004). Furthermore, the first data showing the direct involvement of HMGA1 protein in breast carcinogenesis result from experiments done on breast cancer cell lines in which HMGA1 protein has been overexpressed or downregulated, through antisense or riboenzime. Reeves and colleagues in 2001 published a work that provides strong evidence on a causal role of HMGA1 in breast cancer aggressiveness and EMT. Indeed, they repressed HMGA1 expression in two breast cancer cell lines MCF7 and Hs578T, achieving a strong decrease of cell proliferation and anchorage-independent cell growth. In parallel, after overexpression of HMGA1a and HMGA1b in MCF7 they reached an opposite behavior, obtaining a more aggressive phenotype. In fact, MCF7 cells over-expressed HMGA1b injected in BALBc mice, were more aggressive in promoting tumor growth and metastasis (Reeves et al., 2001). Moreover, analyzing gene expression changes induced by HMGA1 over-expression in MCF7 through cDNA microarray, they found 85 genes differentially expressed involved in many biological function, among which several markers of EMT suggesting that HMGA1 could directly regulates this process (Reeves et al., 2001). Dolde and colleagues reached similar results, hinting that HMGA1 is an important oncogene in breast cancer pathogenesis (Dolde et al., 2002). Even though several works have been done studying HMGA1 in breast cancer, how HMGA1 protein exerts their role in this context is not totally clear and still under study. Few molecular mechanisms that sustain aggressive phenotype were highlighted. Among these, it has been shown that HMGA1 intersects signaling pathway such as Ras/ERK, p53 and BRCA1. Ras/ERK pathway is involved in cell growth and invasion (Santen et al., 2002). Using HMGA1 overexpressing MCF7, Treff and colleagues demonstrated that HMGA1 regulates negatively caveolin 1 and 2 and proteins involved in cholesterol biosynthesis, and positively KIT ligand, notably, all these genes are under the control of Ras/ERK pathway (Treff et al., 2004). p53 controls apoptosis after DNA damage regulating gene transcription. In this context, HMGA1 proteins are able to counteract p53 action inducing Bcl-2 transcription and contributing to the escape from apoptosis (Esposito et al., 2010). As described before, DNA integrity is fundamental in tumor prevention. One of the major proteins involved in DNA integrity maintenance is BRCA1, which is often mutated or downregulated in familiarly breast cancer causing genome

instability. HMGA1 protein negatively regulates BRCA1 expression by binding its promoter, moreover, their expression inverse correlate in 10 different cell lines and in 14 chirurgic sample of breast cancer (Baldassarre et al., 2003). Finally, very recently it has been demonstrated that HMGA1 has a role in cancer stem cell. On a molecular point of view Shah and colleagues define an HMGA1-gene signature correlated with pluripotent stem cells and they demonstrated that HMGA1 silencing in MDA-MB 231 and Hs578T breast cancer cell lines strongly impaired cancer stem cell properties (Shah et al., 2013).

Considering all these information HMGA1 is involved in breast cancer pathogenesis but the molecular role of it and the pathways that it intersect in breast cancer progression and aggressiveness are still unclear.

Aim Of The Thesis

Breast cancer is a woman disease difficult to eradicate due to the extremely heterogeneity and the tendency to metastasize. The study of the molecular alterations that underline breast cancer aggressiveness allowed to develop more specific treatments that can block breast cancer progression. The aim of my Thesis is to study the role of HMGA1 in conferring aggressive traits to a particular breast cancer subtype, the basal-like. In particular, we want to unravel the cellular and molecular alterations orchestrated by HMGA1 in this tumor type. For this purpose, we have taken into consideration different basal-like breast cancer cell lines that express high levels of HMGA1 and studied the involvement of HMGA1 in several hallmarks of cancer. In order to deepen the role of HMGA1 in aggressiveness, we have evaluated which pathways HMGA1 intersects and dissected them.

Materials And Methods

1 CELL LINES

MDA-MB-231, MDA-MB-157 and HEK-293 cell lines were cultured in DMEM (Dulbecco's Modified Essential Medium, Euroclone) with high glucose content, while MDA-MB-468 cell line was grown in RPMI (Roswell Park Memorial Institute, Euroclone) 1640, both containing 10% tetracycline-free FBS (Fetal Bovine Serum, Euroclone), 2 mM L-glutamine (Euroclone), 100 U/ml penicillin and 100 μg/ml streptomycin (Euroclone). For the expansion cells were dissociated by trypsin-EDTA (0.05% trypsin, 0.02% EDTA in PBS, Euroclone). For long term storage cells were maintained in FBS 10% (v/v) DMSO (AnalaR BDH) in liquid nitrogen. MDA-MB-231 stably transfected with the vector pcDNA6/TRTM (Invitrogen) and pSUPERIOR.neo (OligoEngine) shCTRL, pSUPERIOR.neo shA1_1 or pSUPERIOR.neo shA1_3, previously present in the laboratory in which I performed the work of my thesis, were grown in DMEM containing 10% tetracycline-free FBS, 2 mM L-glutamine, 100 U/ml penicillin and 100 μg/ml streptomycin, 2 μg/ml Blasticidin (Sigma) and 200 μg/ml G418 (Sigma). For shRNA induction 1 μg/ml of doxycycline were added to medium. MDA-MB-231 stably infected with pBABE-FLAG or pBABE-FLAG-YAP-5SA were a kind gift of Prof. G. Del Sal.

2 REAGENTS AND PLASMIDS

All siRNA used in this thesis were designed using an Invitrogen tool, except siCCNE2a, a kind gift of Prof. G. Grassi and siLATS1/2, a kind gift of Prof. G. Del Sal. Target gene symbol is used as name throughout the text (siRNA), and the sequences were reported in the following table.

Target Gene	siRNA	Sequence
	siCTRL	ACAGUCGCGUUUGCGACUG
HMGA1a/b	siA1_1	GACAAGGCUAACAUCCCAC
HMGA1a/b	siHMGA1 or siA1_3	ACUGGAGAAGGAGGAAGAG
KIF4A	siKIF4A	CCCGUCUACUGUUUGGGAA
MLF1IP	siMLF1IP	GGUCCCAGGUAUGAGCUAU
DEPDC1	siDEPDC1	CCGUAGUCUAAGAUAACUA

AURKB	siAURKB	AAUCUGCUCUUAGGGCUCA
GTSEI	siGTSEI	UGUGAAGCCUCAGUCACUC
KIF23	siKIF23	CCUGUAGACAAGGCAAUAU
CENPF	siCENPF	CCAUCUGUGUUCACAGCAA
CCNE2	siCCNE2b	GAAAGCCUCAGGUUUGGAG
MST1	siMST1	GCAGGUCAACUUACAGAUA
MST2	siMST2	CCACAAGUACAAAGACCAU
CDK2	siCDK2_1	GCUUGGCCUUGGGCUAUUU
CDK2	siCDK2_2	GCCUUCCUACACGUUAGAU

We used the following plamids:

- pcDNA3HA, pcDNA3HA-HMGA1a, pEGFP-N1, pEGFP-N1 HMGA1a, pRL-CMV Renilla (Promega) and pGL4.11 (Promega) were already present in the laboratory.

- pcDNA3HA-CCNE2: for this plasmid construction, the CCNE2 coding sequence was amplified from reverse-transcribed total RNA of MDA-MB-231 cells using the following primers: forward -5' TTAACGGAATTCATGTCAAGACGAAGTAGC 3'- and reverse -5' TCTCCTCGAGTTATTAGTGTTTTCCTGGTGG 3'- and the resulting fragment was cloned into EcoRI and XhoI restriction sites of pcDNA3HA. The plasmid DNA was checked by DNA sequencing.

- pGL4-CCNE2 (kindly given by Jay A. Nelson laboratory) and pGL4-ΔCCNE2 that was obtained by amplifying a fragment of 223nt from pGL4-CCNE2 using the following primers: forward -5' AATCTCGAGGTGCGGGGCGGGAC 3'- and reverse 5'- ACCCAAGCTTACGGAACGCGGGAACCCA 3'- and cloned in HindIII and XhoI restriction sites of pGL4.11 (Promega).

CDK inhibitor screening were performed treating cells with PD03329911, SNS-032, ROSCOVITINE, PHA793887, AT7519, BS181HCl, BMS265246, AZD5438,

FLAVOPIRIDOL, PHA767491, PHA848125, DINACICLIB or JNJ7706621 (Sellekchem) already resuspended in DMSO.

3   CELL TRANSFECTIONS AND TREATMENTS

For transfection of siRNA, from $10.4*10^3$ to $20.8*10^3$ cells/cm^2 of all cell lines were seeded and the day after transfected with Lipofectamine^{TM} RNAiMAX reagent (Invitrogen) in antibiotics-free medium for 72 hours, if not specifically specified, according to manufacturer instructions.

Plasmid transfections were performed using either FuGENEHD (Promega) or Lipofectamine 3000 (Invitrogen) according to manufacturer instructions or standard Calcium Phosphate method.

For functional-rescue experiments, $20.8*10^3$ cells/cm^2 of MDA-MB-231 were seeded the day after were transfected with siRNA; 24 h later, the cells were transfected with 2 μg of either pcDNA3HA or pcDNA3HA-HMGA1a or pcDNA3HA-CCNE2 plasmids using FuGENEHD (Promega). Experiments were done 48 hours after plasmid transfection.

For overexpression experiment, $36.5*10^3$ cells/cm^2 of MDA-MB-231 cells were seeded. The day after, cells were transfected with 2.5 μg of pEGFP-N1 plasmid or with 625 ng pEGFP-N1 HMGA1a plasmid compensate to 2.5 μg with PEGFP-N1 plasmid, using Lipofectamine 3000 (Invitrogen), experiment was done 30 h after transfection.

For luciferase experiments $36.5*10^3$ cells/cm^2 of HEK293 cells were seeded in 35 mm diameter well. The day after cells were transfected with 500 ng of the reporter construct, 1 μg of pcDNA3HA or pcDNA3HA-HMGA1a, and 50 ng of pRL-CMV Renilla luciferase expression vector (Promega) to normalize for transfection efficiencies and processed 46.5 hours after standard calcium phosphate transfection.

For LATS1/2 and MST1/2 silencing, the cells were initially transfected with siLATS1/2 or siMST1/2; after 24 h, the cells were transfected with siHMGA1 and siCCNE2b for 48 hours.

For CDK inhibitor screening, $15.6*10^3$ cells/cm^2 cells were treated with PD03329911, SNS-032, ROSCOVITINE, PHA793887, AT7519, BS181HCl, BMS265246, AZD5438, FLAVOPIRIDOL, PHA767491, PHA848125, DINACICLIB or JNJ7706621 (Sellekchem) for 24 hours. These inhibitors were applied at 1 μM, except for SNS-032 and DINACICLIB, which were applied at 0.1 μM. For other experiments cells were seeded from $15.6*10^3$ cells/cm^2 to $46.9*10^3$ cells/cm^2.

The number of cells seeded depend from each experiment.

4   CELL GROWTH IN 3D

Cells were seeded at $10.0*10^3$ cells/cm^2 in 24-well plates pre-coated with Laminin-rich extracellular matrix, Matrigel, (BD Biosciences) as previously described (Kenny et al., 2007). Briefly, 200µl of ice-cold Matrigel were dispensed in 24-well plates and allowed to solidified for 30 minutes at 37°C. Then, cells were resuspended in appropriate medium containing 2% Matrigel and seeded over. Cells were grown for 11 days. The colonies were photographed with a digital Canon PowerShot A630 camera.

5   MAMMOSPHERE CULTURE

Mammosphere Assay was performed as previously described with modifications (Mani et al., 2008). $0.52*10^3$ cells/cm^2 were seeded in 6-well ultra-low adhesion plates (Corning) in MEGM (Mammary Epithelial Cell Growth Medium, Lonza), containing 2% methylcellulose supplemented with 20 ng/ml EGF (Stem Cell Technologies), 10 ng/ml bFGF (Orfgenetics), and B27 (GIBCO). Cells were grown for 12 days. After this time, mammospheres were counted at optical microscope, dissociated by trypsinization and plated at $0.52*10^3$ cells/cm^2 in new 6-well ultra-low adhesion plates. We maintained cells until tertiary sphere formation.

6   GROWTH IN SEMISOLID MEDIUM

Soft agar colony formation assay was performed dispensing 3 ml of 0.6% low melting temperature agarose (Sigma) in DMEM and waiting the solidification for 15 minutes at RT. Than $10*10^3$ cells were seeded on the top in 2 ml of 0.3% low melting temperature agarose waiting the solidification for 20 minutes at RT. Cells were grown in incubator for 3 weeks. Then, colonies were stained with 0.5 mg/ml MTT (Sigma) incubating for 2 hours at 37°C, 5% CO2 and counted.

7   MIGRATION AND INVASION ASSAY

For wound healing assays, cells were cultured to 90% confluence on 35-mm plates. The cells were then scraped with a 200-µl tip, and wound closure was followed for 7-8 hours. Images of same area were taken for each plate with digital Canon PowerShot A630 camera and wound closure was analyzed using ImageJ software. For transwell migration and invasion assays, 24-well PET inserts were used (8.0 µm pore size, Falcon) without or with Matrigel-coated filters

for invasion; 40,000 and 100,000 cells were seeded, respectively. Migrated cells were fixed in PFA 4% (w/v) (paraformaldehyde) in PBS (137 mM NaCl, 27 mM KCl, 4.3 mM Na_2HPO_4, 1.4 mM KH_2PO_4, pH 7.4) for 20 minutes and stained with Crystal Violet 0.5% (Sigma) for 20 minutes. After several washes in H2O, a minimum of 4 images to a maximum of 9 images per well, were taken with digital Canon PowerShot A630 camera. The time at which we fixed migrated cell depends on the cell line, but no at more than 24 hours. Cells were counted using ImageJ software.

8   CELL DISPERSION ASSAY

For cell dispersion assay 25 cells were seeded in 10 cm dishes and cultured for 20 days. Afterwards photos were taken with digital Canon PowerShot A630 camera.

9   *IN VIVO* EXPERIMENTS

The *in vivo* experiments to test tumor growth and metastasis formation in presence and in absence of HMGA1 were done by Doctor A. Rosato and Doctor R. Sommaggio from Department of Surgery, Oncology and Gastroenterology, University of Padova and Istituto Oncologico Veneto IRCCS, Padova, Italy. Below are reported protocols for animal treatment.

9.1   CELLULAR TRANSDUCTION

For in (pHR'tripCMV-luc2-IRES-tNGFR-SIN), envelope (hCMV-G) and packaging plasmids (p8.74) as previously described vivo metastasis assays, MDA-MB-231 cells were co-transduced with a lentiviral vector coding for the Firefly Luciferase reporter gene. The vector was produced in 293T cells by transient co-transfection of the transfer (Breckpot et al., 2003). Cells were cultured and expanded.

9.2   MOUSE STRAIN AND ANIMAL CARE

We used SCID female mice (Charles River Laboratories, Lecco, Italy) aged 7 weeks for *in vivo* studies. Mice were administered drinking water supplemented with 4% sucrose plus 2 mg/ml Doxycycline to induce shRNA expression (shCTRL and shA1_1). Doxycycline-supplemented water was changed every 2 days. Procedures involving animals and their care were in conformity with the institutional guidelines (D.L. 116/92 and subsequent complementing circulars), and all experimental protocols were approved by the Ethical Committee of the University of Padua (CEASA).

9.3   TAIL VEIN AND FAT PAD INJECTION

For the intravenous injection, 100,000 cells were resuspended in 200 µl of DMEM for each mouse, whereas for the fat pad injection, 1 million of cells were resuspended in 100 µl of DMEM. We performed *in vivo* imaging at 20 and 26 days after i.v. injection or at 14, 21, 28, 34 and 40 days after fat pad injection. Anesthetized animals (1-3% isoflurane, Merial Italia S.p.A, Italy) were given the substrate D-Luciferin (Biosynth AG, Switzerland) by intraperitoneal injection at 150 mg/kg in PBS (Sigma). Imaging times ranged from 15 seconds to 5 minutes, depending on the tumor model and time point. The light emitted from the bioluminescent tumors or metastasis was detected using a cooled charge-coupled device camera mounted on a light-tight specimen box (IVIS Lumina II Imaging System; Caliper Life Sciences, Alameda, CA). Regions of interest from the displayed images were identified around the tumor sites or metastasis regions, such as the lymph node and lungs, and quantified as total photon counts (photon/s) using Living Image® software (Xenogen). In some experiments, the lower portion of each animal was shielded before reimaging to minimize the bioluminescence from the primary tumor to ensure that the signals from the metastatic regions could be observed *in vivo*. For ex vivo imaging, 150 mg/kg of D-Luciferin was injected into the mice just before necropsy. The lungs were excised and imaged for 5 minutes.

10 CELL PROLIFERATION ANALYSIS

To assess cell proliferation, we use two different methods. A direct and an indirect methods. In the direct method $10.4*10^3$ cells/cm^2 cells were seeded in 6-well plates. Then, cell proliferation was examined every 24 hours by cell counts through Burker chamber. Cell death was highlighted diluting cells resuspension 1:1 in 0.5% Trypan Blue (Euroclone) in PBS. In the indirect method, $15.6*10^3$ cells/cm^2 were seeded in 96-well dishes. Every 24 hours, cell metabolic activity was measured via CellTiter 96® AQueous One Solution Cell Proliferation Assay (Promega) according to the manufacturer's instructions with some modifications. At each time point, the medium in each well was replaced with a solution composed of 100 µl of PBS containing 4.5 g/L glucose (Sigma) and 20 µl of CellTiter 96® AQueous One Solution. Cells were incubated for 3 hours at 37°C 5%CO_2. Then, the quantity of formazan product, that is directly proportional to the number of living cells, was measured at 492 nm by TECAN microplate reader.

11 ANALYSIS OF DNA/PROTEIN CONTENT (PI/FITC STAINING)

For each sample, 10^6 cells were fixed in 70% ethanol, pre-hydrated in PBS for 6 minutes, washed twice with PBS, and allowed to balance in PBS for 1 h. The cells were stained overnight with 500 μl of a PBS-based solution containing 10 μg of PI and 0.25 ng of FITC (all from Sigma). All flow cytometric measurements were performed using a CYTOMICSTM FC500 (Beckman Coulter Inc. Fullerton, CA, USA) equipped with an Argon laser (488 nm, 5 mV) and configured according to standard parameters for green-filtered (525 nm, FL1) and red-filtered (610 nm, FL3) fluorescent detection (http://dsv.units.it/Dipartimento/fcs_dsv). After acquisition of at least 10,000 events per run, the data, which were stored as list mode files, were analyzed using FCS Express V3 software; alternatively, the saved FL3 histograms were subjected to cell cycle analysis, which was performed using MultiCycle® software.

12 PROTEIN EXTRACTION AND WESTERN BLOT (WB) ANALYSIS

After treatments, cells were washed in chilled PBS twice and lysed using SDS sample buffer (62.5 mM Tris, pH 6.8; 2% SDS; 10% glycerol; 200 mM DTT; 1 mM Na_3Vo_4; 5 mM NaF; and mammalian protease inhibitor cocktail (PIC) (Sigma)). Then, lysates were disrupted with an insulin syringe and boiled for 5 minutes.

The lysates were separated via SDS-PAGE (polyacrylamide gel electrophoresis). SDS-PAGE uses SDS-PAGE gel system composed by Stacking gel and running gel. Stacking gel: 5%: 5% acrylamide/bisacrylamide (29:1), 0.125 M Tris-HCl pH 6.8, 0.1% (w/v) SDS, 0.05% (w/v) APS and 0.05% (v/v) TEMED. Running gel: 10%: 10% acrylamide/bisacrylamide (29:1), 0.5 M Tris-HCl pH 8.8, 0.1% (w/v) SDS, 0.05% (w/v) APS and 0.05% (v/v) TEMED; 15%: 15% acrylamide/bisacrylamide (29:1), 0.375 M Tris-HCl pH 8.8, 0.1% (w/v) SDS, 0.05% (w/v) APS and 0.05% (v/v) TEMED. Gel thickness: 0.75 mm, length of stacking gel: 0.5 cm, length of running gel: 8 cm, gel width: 8 cm. Running buffer. Electrophoretic run: accumulation 50 V for 30 minutes, separation 200 V until complete separation of molecular weight (PRE stand Thermo scientific). After separation gel may be stained by using a methanol/water/acetic acid solution (in a volume ratio of 5/4/1) containing 0.05% (w/v) Coomassie Brilliant Blue R 250 or transferred to nitrocellulose membrane $\varnothing$ 0.2 μm (GE Healthcare, WhatmanTM) using a wet transfer system (Transfer buffer: 25 mM Tris, 200 mM Glycine, 20% methanol) at 4°C for 16 hours at 30 V. Membrane were then stained with Red Ponceau solution (0.2% red

poceau S, 3% trichloroacetic acid, 3% sulfosalicylic acid). Western blot analyses were performed according to standard procedures. To revealed horseradish peroxidase coniugated to secondary antibody on autoradiography films (GE Healthcare) were used ECL-kit (Thermo Scientific). The antibodies used are reported in the table 2 and table 3.

13 IMMUNOFLUORESCENCE (IF)

For immunostaining cells were, firstly, grown on glass slides. At the end of the experiments, cells were washed once with ice cold PBS and fixed with 4% (v/v) PFA. Then cells were incubated in 0.1 M glycine (Sigma) in PBS. After permeabilization with 0.3% Triton X 100 (Sigma) in PBS and saturation in 0.5% BSA (Sigma) in PBS, the cells were incubated in primary antibodies diluted in PBS containing 0.5% BSA. α-YAP primary antibody (Santa Cruz) was diluted in 0.5% BSA and 0.01% Triton. The secondary antibody was applied diluted in PBS. For F-Actin staining, Phalloidin-conjugated Alexa Fluor 488 (Invitrogen) was used at a concentration of 1:40 in PBS. Then, the cells were stained with 0.2 µg/ml Hoechst (Sigma) to detect the nuclei. After that, slides with cells were mounted on microscope slides with Prolong Gold Antifade Reagent (Invitrogen). The images were visualized using a Nikon Eclipse e800 microscope and acquired using Nikon ACT-1 software.

Primary Antibody	Antibody concentration for WB	Antibody concentration for IF
α-HMGA1 (homemade)	1:500	1:50
α-β-actin A2055 (Sigma)	1:2000	/
α-β-catenin given by Prof. Del Sal	/	1:100
α-CCNE2 TA300691(Abcam)	1:500	/
α-HA H9658 (Sigma)	1:1000	1:1000
α-YAP sc-101199 (Santa Cruz) given by Prof. Del Sal	1:1000	1:200
α-P127-YAP EP1675Y (Novus Biological)	1:1000	/
α-CDK2 sc6248 (Santa Cruz)	1:200	/
α-FLAG peroxidase conjugate	1:2000	/

(Sigma) given by Prof. Tell		

Secondary antibody	Antibody concentration for WB	Antibody concentration for IF
α-rabbit IgG peroxidase conjugate A0545 (Sigma)	1:5000	/
α-mouse peroxidase conjugate A9044 (Sigma)	1:5000	/
α-mouse IgG Alexa 488 (Invitrogen)	/	1:750
α-rabbit IgG Alexa 594 (Invitrogen)	/	1:750

14 YAP LOCALIZATION

For the quantification of the number of cells with YAP presents in a mainly nuclear, mainly cytoplasmic or equally distributed localization, at least 100 cells from slide taken in different fields were counted.

15 GENE EXPRESSION ANALYSIS

For RNA collection following the experiments, cells were washed once with ice cold PBS and lysed in 1 ml TRIzol Reagent (Invitrogen). RNA was purified according to the manufacturer's instructions of TRIzol Reagent and finally resuspended in H_2O nuclease free. Afterword, to eliminate DNA, RNA samples were subjected to DNase-I (Deoxyribonuclease I, Amplification Grade, Invitrogen) treatment according to datasheet and subsequently purified either by RNeasy kits (QIAGEN) or standard RNA Phenol/Clorophorm purification resuspended RNA in H_2O nuclease free at the end. Then, RNA samples were quantified by QuBit fluorometer (Invitrogen) using Quant-iT^{TM} RNA Assay kit (Invitrogen) as reported by the datasheet and their quality was controlled either by denaturing RNA electrophoresis in gel MOPS (1% (w/v) agarose (Sigma), 6.67% (v/v) formaldehyde (Sigma) in MOPS buffer (20 mM MOPS, 5 mM sodium acetate, 1 mM EDTA, pH 7), or by BioAnalyzer (Agilent) using multichannel Nanochips according to Agilent protocols (this quality control was done before microarray analysis by Centro di Biologia Molecolare (CBM) facility).

For quantitative RT-PCR (qRT-PCR), mRNA was, firstly, reverse transcribed using either Superscript II (Invitrogen) or Superscript III (Invitrogen) according to the manufacturer's instructions. qRT-PCR was performed using iQ^{TM} SYBR Green Supermix (BIO-RAD) mixed with specific primers (see table 4 for sequences) and sample, in technical duplicate. PCR reaction was performed for 40 cycles (denaturation 95°C for 5 seconds, amplification and elongation at 60°C for 45 seconds) using BIO-RAD CFX instrument and analyzing data with BIO-RAD CFX Manager software. GAPDH was used for normalization. Relative gene expression was calculated by ΔΔCt method.

For microarray analysis, mRNA for each group (siCTRL or siHMGA1) was processed accordingly to Affymetrix standard protocols by CBM facility. Biotinylated cRNA were hybridized on Affymetrix GeneChip Human Genome U133A 2.0 array that allowed to analyze the expression level of 18,400 transcripts and variants, including 14,500 well-characterized human genes. Arrays were than stained, washed (GeneChip Fluidics Station 450) and image was acquired by Scanner GeneChip 3000 7G. The intensity of single spots was calculated using Affymetrix Expression Console.

For Chromatin ImmunoPrecipitation (ChIP) assay coimmunoprecipitated DNA was analyzed by qPCR. Promoter occupancy was calculated as percent of chromatin input immunoprecipitated using the $2^{-\Delta Ct}$ method.

Gene / Primer name	Forward primer (5'-3')	Reverse Primer (5'-3')
GAPDH	TCTCTGCTCCTCCTGTTC	GCCCAATACGACCAAATCC
HMGA1/HMGA-3'UTR	ACCAGCGCCAAATGTTCATCCTCA	AGCCCCTCTTCCCCACAAAGAGT
HMGA1 CDS	CAGCGAAGTGCCAACACCT	CGAGATGCCCTCCTCTTCC
VIM	CCAGCTAACCAACGACAAAGCCCG	TGCGTTCAAGGTCAAGACGTGCCA
JAG1	GGGAAAACGTGCCAGTTAGA	TCACAATTCTGACCCATCCA
HES1	GAAGGCGGACATTCTGGAAA	CGGTACTTCCCCAGCACACT
MLF1IP	TAAGGCAGCCATCGCCACATTT	ACCGAAGCAGTTCATCCTGGACTT

KIF23	CCTGAGGGCTACAGACTC AACCGA	TCTGGGTGGTGTGAGTGCCAA
KIF4A	AAGCCAAACGCCATCTGA ATGACC	TTGACCACGCACTTCAGTAAGG GA
CENPF	TCAGGCAAGAGGCAAAGA TCCAGT	TGGCTCAAACTCAGTACCTTCC GT
AURKB	TGAGGAGGAAGACAATGT GTGGCA	AGGTCTCGTTGTGTGATGCACT CT
GTSEI	ACTGAACCAACAAGGGAG AGCAAC	CCGAGCTACTTCTGTGGCAGTA CTTT
DEPDC1	TGGGTATTATCTGCCATGA AGTGCCT	AGGTTGCAGCAAGCCCAAAAT GT
LEF1	CGAATGTCGTTGCTGAGT GT	GCTGTCTTTCTTTCCGTGCT
SETD8	CGCAACAGAATCGCAAAC TTAC	TCCTTCTTCCTTCCCACTTTCA
CCNE2	TGAGCCGAGCGGTAGCTG GT	GGGCTGGGGCTGCTGCTTAG
ATRR-1 (for ChIP)	AGTCCTGTTCCACCTGCTC AAT	TGATGGATGATAGACGGATGG T
ATRR-2 (for ChIP)	AATTACCTACAGAAGCAG GGGTTG	ACTGAGTGATGGATGCACAGA GATTA
CCNE2 promoter (for ChIP)	TCACTTCCCAGCCACCCTA T	CCAGTTTGCTTTTGCTTCCTC
BIRC6	AGCATTCGTCCGGTTGCG CT	TCGATGGCACGGCGCACTTT
CTGF	AGGAGTGGGTGTGTGACG A	CCAGGCAGTTGGCTCTAATC
CYR61	AGCCTCGCATCCTATACA ACC	TTCTTTCACAAGGCGGCACTC

16 BIOINFORMATICS ANALYSIS

Bioinformatics analysis on microarray, patient datasets and survival analysis were done by Dr. S. Piazza and Dr. Y. Ciani from Laboratorio Nazionale CIB, (LNCIB), Area Science Park, Trieste, Italy. Below are reported all protocols for bioinformatics analysis.

16.1 MICROARRAY ANALYSIS

16.1.1 LOW-LEVEL ANALYSIS

For microarray analysis, biological mRNA replicates for each group (siCTRL or siHMGA1) were hybridized on Affymetrix hgu133plus2 chips. Cell intensity values were computed using the Affymetrix Expression Console. Further data processing was performed in the R Computing Environment version 2.14 (http://www.r-project.org/) with BioConductor packages (http://www.bioconductor.org/). Robust Multi-Array Average (RMA) normalization was applied (Irizarry et al., 2003). Statistical analysis for differentially expressed genes was performed with limma (Smyth, 2004). P-values were adjusted for multiple testing using the Benjamini and Hochberg's method to control the false discovery rate (Hochberg and Benjamini, 1990). Genes with adjusted p-values below 10-4 and fold change greater than 2.6 (log 1.4) or lower than -2.6 (-log 1.4) were considered differentially expressed. Gene annotation was obtained from R-Bioconductor metadata packages, and the probesets were converted in Entrez Gene Id and Symbol Id.

16.1.2 CLUSTER ANALYSIS

Starting from the normalized annotated expression matrix after gene median centering, features that had standard deviation of less than 0.3 were filtered out. Unsupervised hierarchical cluster analysis (average-linkage method) was performed using Cluster software (EisenLab). Cluster results were then visualized using Java TreeView.

16.2 FUNCTIONAL ANALYSIS

Differentially expressed gene lists obtained from low-level procedures were analyzed for functional associations.

- Data were analyzed through DAVID Bioinformatics Resources v6.7 (Dennis et al., 2003) using the suggested standard parameters.
- Data were analyzed through Ingenuity Pathway Analysis (Chiappetta et al.) software. Core analysis was performed, and the top associated networks table was reported.
- Data were analyzed through the Oncomine Pro web tool using suggested standard parameters. Custom concept analysis was performed, and the "Summary view" (adapted) was reported.

16.3 BREAST CANCER PATIENT DATASETS AND SURVIVAL ANALYSIS

16.3.1 BREAST CANCER META-DATA

Data processing was performed using several BioConductor packages (survival, affy and limma; see http://www.bioconductor.org/) in the R Computing Environment version 3.0.2 (http://www.r-project.org/).

Several published gene expression datasets (breast cancer meta-dataset) were collected from the gene expression omnibus (GEO) public gene expression database (GSE1456, GSE4922, GSE5327, GSE6532, GSE7390, GSE11121, GSE12093, GSE2603, GSE16446, GSE19615, GSE20685 and GSE21653). The data were normalized in the R/Bioconductor environment using the RMA normalization method (affy package), generating a breast cancer meta-dataset. Gene annotation was obtained from brainarray custom CDF metadata packages, and the probe sets were converted to Entrez Gene Id and Symbol Id. Each dataset was analyzed separately to avoid platform and signal merging issues, and only the results were combined.

This breast cancer meta-data was used to compare with our HMGA1 dataset and CCNE2 expression.

To verify the correlation of the HMGA1-gene signature and breast cancer clinical data, a Mantel-Haenszel test was applied to the normalized meta-dataset (survival R package), and the Kaplan–Meier survival curve of distant metastasis-free survival (DMFS) of breast cancer patients classified according to the expression of HMGA1 signature was obtained. Moreover, Kaplan–Meier survival curves of DMFS of breast cancer patients were classified according to the expression of CCNE2 or HMGA1 using the survival package.

Moreover, we searched for the distribution of the gene expression intensities of HMGA1 signature across different breast cancer subtypes (stats R package) and the correlation between HMGA1 expression and the HMGA1 signature expression levels. The same were done for CCNE2 expression.

A subset of the meta-dataset was used to compare the gene expression data to the DMFS duration for the single regression Cox gene analysis. In particular, we used only the Affymetrics HGU133A platform data to avoid the platform type as a confounding variable in the models.

16.3.2 BREAST CANCER TCGA DATASET

Gene expression data for the breast cancer samples (TCGA data set) were obtained from the Memorial Sloan Kettering Cancer Genomics Portal (http://www.cbioportal.org/public-portal; last accessed on 6 July 2014). Beginning with the Breast Invasive Carcinoma data set (TCGA,

Provisional, $n = 1037$), each patient was classified as expressing high or low levels of YAP/TAZ profile genes and of the HMGA1 or CCNE2 genes. The genes comprising the YAP/TAZ signature are as follows: KRT34; STXBP6; OLR1; THBS1; INHBA; CTGF; SERTAD4; ANKRD1; HSD3B1; ORC1; CENPM; DAW1; ITGB2; IGFBP3; TGM2; ADAMTS1; BDNF; TMEM171; SERPINE2; PTGS2; CCDC18; PLCB4; DEPDC1B; ZBED2; MATN3; CCNA2; TBXA2R; SERPINE1; SLIT2; BCAR4; ZWINT; RAD51; DIAPH3; MCM10; NAV3; SKA1; SHCBP1; RAD51AP1; DDAH1; RIMS2; RRM2; CDC6; PRR16; DAB2; PLK4; ASF1B; KIF14; FMN2; CDC25C; and GINS2. Statistical independence between the different molecular conditions was calculated using Pearson's Chi-squared contingency table tests in the R/Bioconductor environment.

16.4 GENE EXPRESSION-BASED OUTCOME FOR BREAST CANCER WEB TOOL (GOBO)

To evaluate the correspondence between the HMGA1 expression levels and breast cancer clinical data, we utilized the gene expression-based Outcome for Breast Cancer web tool (GOBO) (Ringner et al., 2011).

16.5 PROTEIN PHOSPHORILATION ANALYSIS

The protein phosphorylation data were obtained from the TCGA breast invasive carcinoma dataset via cBioPortal (http://www.cbioportal.org/public-portal/index.do). In particular, using CCNE2 as the gene of interest in the input form, we selected the proteins displaying differential phosphorylation according to the RPPA data.

17 PROMOTER ANALYSIS

Putative MARs were mapped using MAR-Wiz (http://genomecluster.secs.oakland.edu/cgi-bin/mar-upload.cgi), with default parameters. Putative HMGA1 binding sites were obtained using MatInspector tool (http://www.genomatix.de/cgi-bin//eldorado/main.pl). The sequence analyzed is from +226 to − 11430 of CCNE2 gene.

18 LUCIFERASE ASSAY

The assays were performed Dual-Luciferase® Reporter (DLR™) Assay system (Promega) according to the manufacturer's instruction. Because HMGA1 is not soluble in Passive Lysis Buffer (PLB), given with the kit, 1/6.6 of total cells resuspended in PBS were taken and after 5 minutes of centrifugation at 200 g the pellet containing cells were lysed with 20 µl of SDS sample buffer.

19 CHROMATIN IMMUNOPRECIPITATION

MDA-MB-231 cells were cross-linked in culture medium with 1 % formaldehyde for 10 min, neutralized using 125 mM glycine in PBS for 5 min and washed in PBS. Nuclei were obtained by treatment with hypotonic buffer (5 mM Pipes pH 6.8, 85 mM KCl, 0.5 % NP-40 and protease inhibitors) and centrifugation. The nuclei pellet was resuspended in RIPA 100mM buffer (20 mM Tris-HCl pH 7.5, 100 mM NaCl, 1 mM EDTA, 0.5% NP-40, 0.5% Na-Deoxycholate, 0.1% SDS supplemented with protease inhibitors). Chromatin was sonicated (Digital Sonifier 250, Branson) at 30% potency for 80 seconds (10 seconds ON, 30 seconds OFF on ice for 8 times) to 500-800 bp average fragment size and precleared for 1 h at 4°C with protein A/G PLUS-Agarose (Santa Cruz Biotechnology). Agarose was removed by centrifugation and an aliquot of supernatant was taken as input. Chromatin was immunoprecipitated overnight at 4°C with anti-HMGA1 polyclonal antibody. As a negative control for immunoprecipitation, IgGs purified from rabbit serum were used (Abcam). DNA protein complexes were recovered with protein A/G and washed sequentially with RIPA 100 mM buffer, RIPA 250 mM buffer (20 mM Tris-HCl pH 7.5, 250 mM NaCl, 1 mM EDTA, 0.5% NP-40, 0.5% Na-Deoxycholate and 0.1% SDS), LiCl solution (10 mM Tris-HCl pH 8.0, 1 mM EDTA, 250 mM LiCl, 0.5% NP-40 and 0.5% Na-Deoxycholate) and TE. To reverse cross-linking, samples were incubated overnight at 65°C and after that RNase A (35 ng/μl) and proteinase K (90ng/μl) treatment was performed for 2 h at 55°C (0.2%, SDS, 35mM Tris HCl pH6.5, 8.8 mM EDTA pH 8). In parallel, input was treated in the same way. After phenol/chloroform extraction and ethanol precipitation samples were resuspended in H_2O.

20 STATISTICAL ANALYSIS

The data were analyzed via two-tailed Student's t tests, and the results were considered significant at a $P < 0.05$. The results are presented as the mean.

Results

1 HMGA1 EXPRESSION IN PRIMARY BREAST TUMORS

We started asking whether HMGA1 expression was enriched in a particular breast cancer subtype in primary tumors. Literature data reported that HMGA1 is overexpressed in 60% of sporadic ductal carcinomas (Chiappetta et al., 2004) and its protein expression positively correlates with a more advanced tumor grade (Flohr et al., 2003). Then, to analyze deeply the relation between HMGA1 expression and breast cancer, we have taken advantage from high-throughput analysis of breast cancer specimens, and thanks to the collaboration with the Bioinformatics Unit of LNCIB, we have evaluated the abundance of HMGA1 mRNA by performing a bioinformatics analysis of a primary breast cancer public microarray data collection composed by 1881 different samples. As shown in figure 1 panel A, HMGA1 mRNA levels were higher in the basal-like and HER2+ subtypes, which are more aggressive than luminal A and B and normal-like breast cancer subtypes. In addition, HMGA1 mRNA expression is strongly associated with oestrogen receptor-negative group of breast cancers, which is a characteristic status of basal-like and HER2+ subtypes (Fig. 1B). Finally, in agreement with literature data we have observed a positive association of HMGA1 mRNA with more advanced tumor grade (Fig. 1C). All these data suggest that HMGA1 is expressed more abundantly in the most aggressive breast cancer specimen.
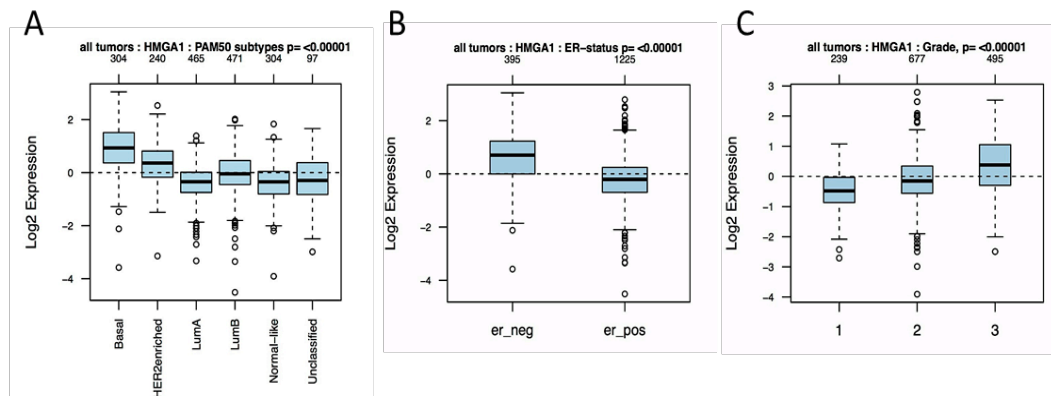

Figure 1. Association of HMGA1 mRNA Levels with Subtype and Tumor Grade in Human Breast Cancers.
Boxplots of the distribution of the gene expression intensities of HMGA1 mRNA across different:
(A) breast cancer subtypes,
(B) ER-statuses and
(C) tumor grades.

2 HMGA1 SILENCING IN BASAL-LIKE BREAST CANCER CELL LINES REVERTS THE BREAST CANCER CELL TRANSFORMATION PHENOTYPE

Basal-like breast cancer subtype is one of the most aggressive and poor prognosis-linked subtypes. In fact, it tends to be highly metastatic, and because of the lack of ER, PR, and HER2 receptor expression, it is not treatable with conventional therapies (Foulkes et al., 2010). Therefore, since we found that HMGA1 mRNA expression was associated with it, we decided to investigate the role of HMGA1 protein in basal-like breast cancer. To do this we choose the highly metastatic oestrogen receptor-negative human basal-like breast cancer cell line MDA-MB-231, usually used as model for this breast cancer subtype, that expresses high levels of HMGA1 protein (Liu et al., 1999). To unravel the role of HMGA1 in breast cancer aggressiveness we decided to silence its expression exploiting the RNA interference (RNAi) system. As shown in figure 2, we efficiently knock-down HMGA1 protein both in the cellular system inducible for HMGA1 silencing based on short hairpin RNA (shRNA) and in the transient silencing based on small interfering RNA (siRNA) (Fig. 2 A and B). Surprisingly, after HMGA1 depletion, we observed strong morphological alterations to a more differentiated phenotype respect to control cells (Fig. 2C). Indeed, HMGA1-silenced cells (ShA1_1 and ShA1_3) displayed a flattened and polygonal morphology and grew as a highly ordered monolayer sheet, whereas control cells (ShCTRL) maintained a spindle-like fibroblastic phenotype (Fig. 2C upper panel). The same changes in morphology were observed also in MDA-MB-231 cells transiently silenced for HMGA1 expression (Fig. 2C lower panel). Consistently with the reversion to a more differentiated epithelial-like phenotype, we observed a weak but significant decrease in proliferation in HMGA1-silenced cells at day 5 after cell seeding when cells reached confluence (Fig. 2D), suggesting a reacquisition of cell-cell contact inhibition properties. To gain further insight into the potential role of HMGA1 in differentiation, we analyzed the effects of HMGA1 depletion in a 3D culture assay on Matrigel that allows appreciating cells organized in three dimensions. Interestingly, in HMGA1-depleted cells, the disorganized morphology, characteristic of tumor cells like MDA-MB-231, changed to acini-like spheroids with hollow lumens, typical of non-malignant breast epithelial cells (Fig. 2E two different magnification of the same field), confirming that HMGA1 may be involved in the tumor cell differentiation process. Epithelial to mesenchymal transition (EMT) is a central process in differentiation and in cancer development, where cells lose their polarity and global organization acquiring a fibroblastic-

like shape and a more motile and invasive behavior (Thiery, 2002). Thus, we investigated the role of HMGA1 expression in EMT in basal-like breast cancer cell lines. After HMGA1 depletion, we observed a drastic reorganization of the actin cytoskeleton of the cells
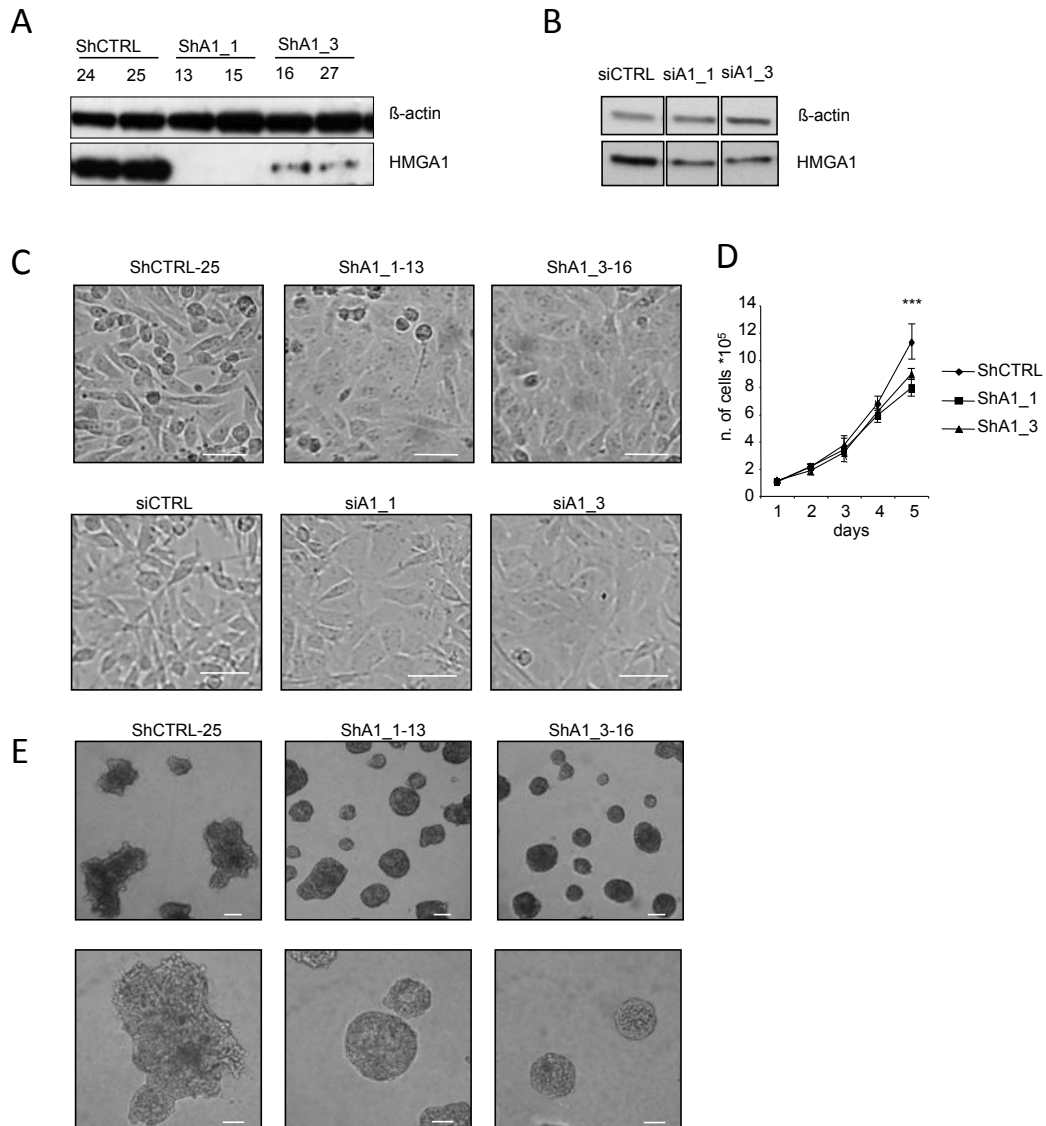
Figure 2. HMGA1 Depletion Induces Phenotypic Changes in MDA-MB-231 Breast Cancer Cell Line.

(A) Western blot analysis of HMGA1 levels for different clones of the MDA-MB-231 inducible ShRNA cell lines after doxycycline addition. Two clones for each shRNA were selected. ShA1_1 and ShA1_3 are different clones obtained using two shRNA molecules targeting a region in the 3'UTR and the coding sequence of HMGA1, respectively, to avoid potential off-target effects. ShCTRL was obtained using a control shRNA.
(B) Western blot analysis of HMGA1 levels after transfection of the MDA-MB-231, with siCTRL, siA1_1 or siHMGA1 siRNA for 72 hours. In (A) and (B) β-actin was used as an internal control.
(C) Representative pictures of 2D morphology of MDA-MB-231 after HMGA1 depletion (ShA1_1 and ShA1_3 upper and siHMGA1 and siA1_1 below) respect to control (ShCTRL upper and siCTRL below). The scale bar represents 50 μm.
(D) Proliferation curves. Cells were counted every 24 hours for 5 days. The data are represented as the means ± SD (n=3). The presented data were obtained from two independent clones for MDA-MB-231 ShCTRL, ShA1_1 and ShA1_3.

(E) Representative phase contrast pictures of 3D growth after 11 days in culture. In this panel are presented two different magnifications of the same field. The scale bar represents in the upper part of panel is 100 μm and in the below panel is 50 μm.

***P < 0.001; two-tailed Student's t-test

analyzed. Indeed, F-actin became more cortical and there was a reduction of stress fibers (Fig. 3 A, C and D), a symptom of a mesenchymal to epithelial transition (Banerjee et al.). In addition, we observed the down-regulation of mRNA encoding the mesenchymal marker vimentin (Palvimo and Linnala-Kankkunen) in MDA-MB-231 and in MDA-MB-468, another breast cancer basal-like cell line (Fig. 3B). EMT is a process induced by many pathways, including Notch and Wnt/β-catenin signaling. We observed that the expression of mRNAs encoding key players in the Notch pathway such as JAG1 and HES1 (Li et al., 2013) in HMGA1-depleted cells were downregulated at different extent in the cell lines analyzed (Fig. 3B). Moreover, concomitantly with the F-actin reorganization, there was a translocation of β-catenin to cell-cell contacts (Fig. 3 C and D), indicating a possible participation of HMGA1 in the Wnt canonical signaling and supporting the hypothesis of its involvement in MET.

In 2008 Mani and colleagues demonstrated a strong connection between EMT and cancer stem cell phenotype. Indeed, cells that undergo EMT may acquire stem cell properties showing also tumor initiating capability (Mani et al., 2008). Thus, given that we showed that HMGA1 is involved in MET, we asked whether HMGA1 influenced cancer stem cell population and then its self-renewal capacity. One of the most used assays to test stem cell properties of breast cancer cells is mammosphere-formation assay in which the pool of cancer stem cells forms mammospheres under non-adherent condition (Shaw et al., 2012). Mammospheres appear as spheres in which single cells are difficult to detect, forming a very compact structure. MDA-MB-231 cells are known to possess a sub-population of CSCs that forms mammospheres in this type of assay (Wang et al., 2014). After HMGA1 depletion in MDA-MB-231, we observed a strong reduction in mammosphere formation and dimension (Fig 3E). Self-renewal is one of the main features of stem-cells and to evaluate it, it is necessary to dissociate the primary mammospheres into single cells and reseed them to assess their ability to form secondary mammospheres (Shaw et al., 2012). HMGA1 depletion maintained mammosphere inhibition in the subsequent passages, demonstrating that HMGA1 decreases the self-renewal capacity of mammosphere-forming cells. Instead the control cells maintained the self-renewal potential at least until third passage (Fig. 3F). These data suggest

that HMGA1 is involved in the growth and self-renewal capacity of breast cancer stem cells in a process linked to EMT.
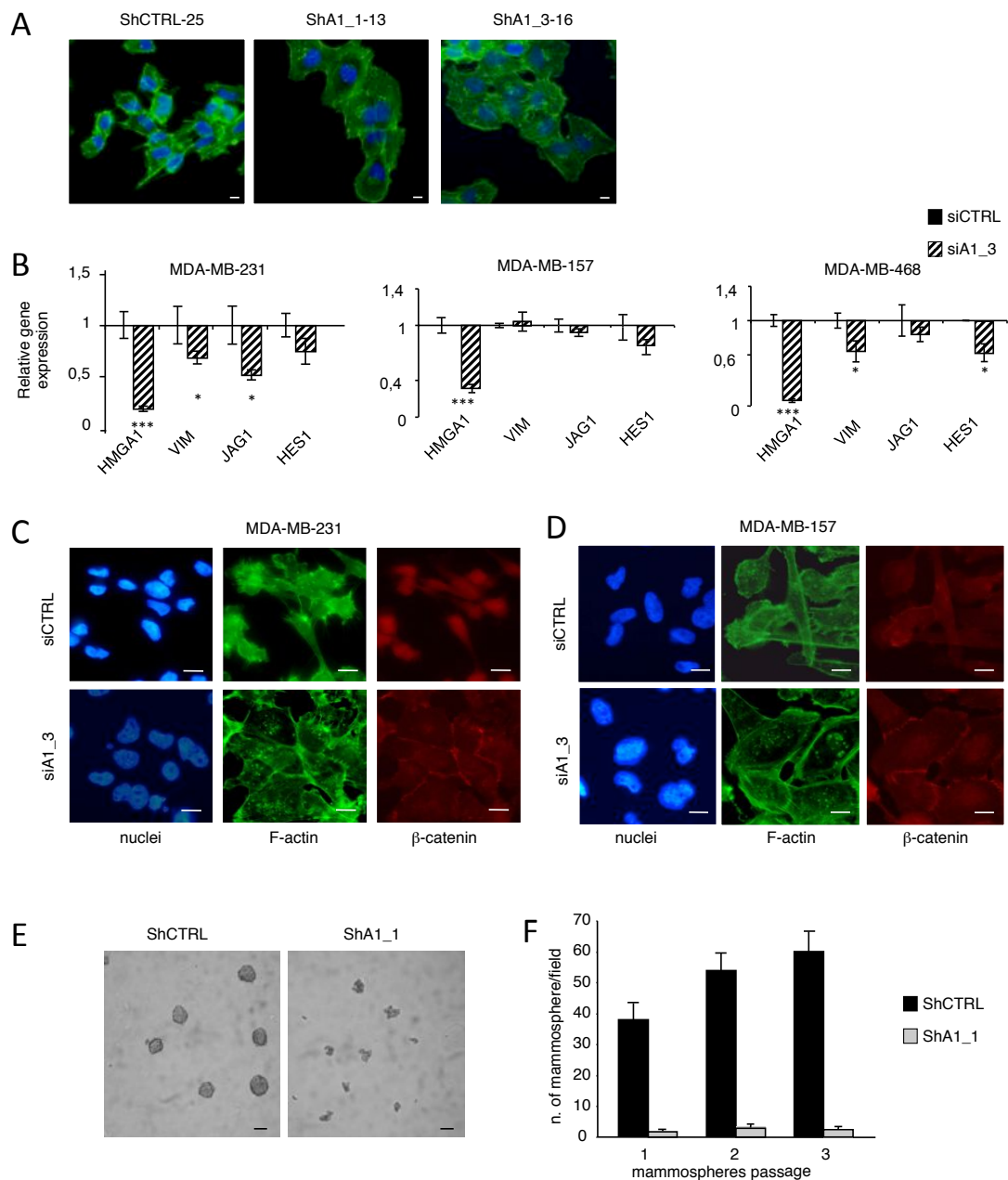
Figure 3. HMGA1 Modulates EMT in Basal-Like Breast Cancer Cell Lines and Stem Cell Properties in MDA-MB-231.

(A) Representative images of MDA-MB-231 induced for ShCTRL, ShA1_1 and ShA1_3 were stained with phalloidin (green) to visualize F-actin. Nuclei were stained with Hoechst (blue). Images were taken at X 60 magnification.

(B) The down-regulation of selected genes after HMGA1 silencing (gray bar) in MDA-MB-231, MDA-MB-157 and MDA-MB-468 was measured using qRT-PCR. Expression was normalized to the levels in cells that had been transfected with control siRNA. GAPDH was used as an internal normalizer. The data are represented as the means ± SD (n=3).

(C) and (D) Representative images of MDA-MB-231 (C) and MDA-MB-157 (D) cells that have been transfected with control (siCTRL) and HMGA1 (siHMGA1) siRNA and that were stained with phalloidin (green) to

56

visualize F-actin. Immunofluorescence for β-catenin was shown (red). Nuclei were stained with Hoechst (blue). Images were taken at X 60 magnification.

(E) Representative phase contrast pictures of mammosphere growth after 12 days in culture. The scale bar represents 100 μm.

(F) *In vitro* quantification of mammospheres that were formed by the cells described in (E).

*P < 0.05, **P < 0.01, ***P < 0.001; two-tailed Student's t-test.

3   SILENCING OF HMGA1 REDUCES PRO-METASTATIC FEATUREs OF HUMAN BREAST CANCER CELLS

The ability of cancer cells to metastasize distant organs is one of the hallmarks that characterize carcinomas and it is the major cause of death for this disease being responsible for more than 90 percent of the cancer related mortality (Hanahan and Weinberg, 2000). How a cancer cell acquires the ability to metastasize is still unclear. It is widely accepted that EMT play a critical role in this context sustaining invasion and conferring migration skills to cells (Kalluri and Neilson, 2003). In addition, given that EMT and CSCs are interconnected and CSCs have tumor initiating potential, we asked whether HMGA1 depletion impact on cell metastatic potential.

First we investigated if HMGA1 silencing impairs migration in several basal-like breast cancer cell lines such as MDA-MB-231, MDA-MB-157 and MDA-MB-468. To assess its involvement, we performed trans-well migration assay in which cells that have motile properties pass through a porous membrane. After HMGA1 depletion all the three cell lines showed a strong reduction, more than 50% compared to controls, in trans-well migration (Fig. 4A). Moreover, we confirmed the inhibitory effect of HMGA1 knockdown on migration in MDA-MD-231 in wound healing-induced migration assay (Fig. 4B) and we demonstrated that this effect was strictly dependent on HMGA1 because the reintroduction of a siRNA-resistant HMGA1 construct almost entirely rescued cell migration (Fig. 4C). This assay allows following cell migration in 2D confluent cell culture after scratching it with a tip, and moreover, it permits to highlight how cells migrate. As described in introduction, cells may migrate in a collective, mesenchymal or amoeboid ways. Polarized epithelial cells, which have intact cell-cell contacts, move collectively in sheet, while mesenchymal cells migrate individually (Friedl and Wolf, 2008). Indeed, through high magnification, we observed that HMGA1-depleted MDA-MB-231 cells migrated collectively to the wound center in wound healing assay, a finding that is consistent with the re-acquisition of a polarized epithelial phenotype, whereas control cells moved individually (Fig. 4D). To deepen this aspect, we used a cell dispersion assay to simulate in 2D the metastatic spreading of cells from the

primary tumor to surroundings. This assay underlined that HMGA1-depleted cells grow in a monolayer as well-defined colonies, while control cells formed not-defined colonies in which cells break the colony boundaries (Fig. 4E).
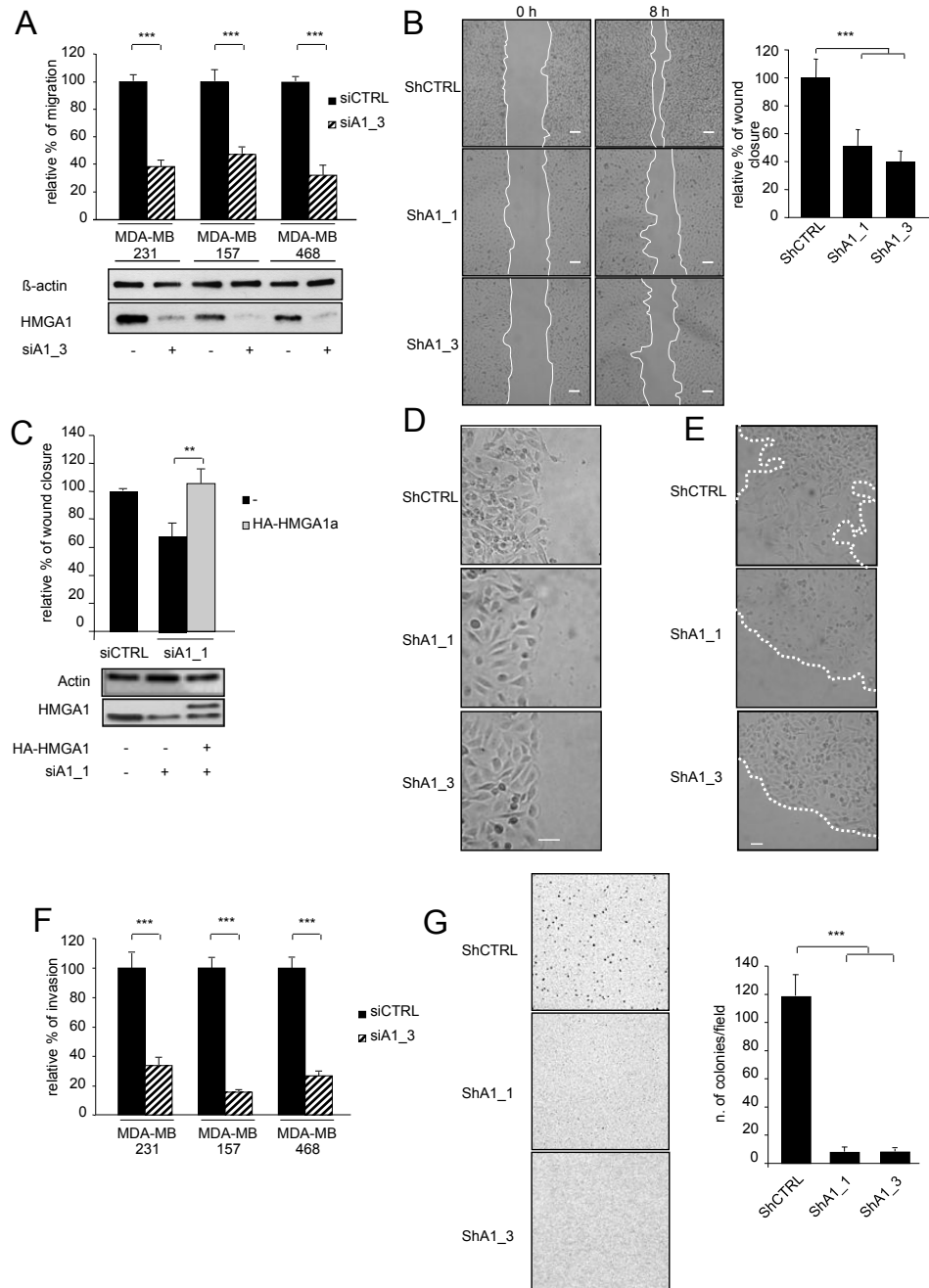
Figure 4: HMGA1 Depletion Inhibits Pro-Metastatic Properties of Breast Cancer Cells *in vitro*.
(A) Transwell migration assays in MDA-MB-231, MDA-MB-157 and MDA-MB-468 cells transfected with control (siCTRL) and HMGA1 (siHMGA1) siRNA. Cells migrated through the porous membrane were fixed after 16 hours from the seed. The data are represented as the means of the percentage of the number of cells relative to control ± SD (n=3). Under the graph the western blot analysis of the HMGA1 levels after transfection of the MDA-MB-231, MDA-MB-157 and MDA-MB-468 cells with siCTRL or siHMGA1. Actin was used as an internal control.

(B) Representative pictures from the scratch wound-healing assay, performed with HMGA1-depleted (ShA1_1 and ShA1_3) and control (shCTRL) cells. The scale bar represents 100 μm. Quantification of the scratch wound-healing assay is shown. The data are represented as the means of the percentage of wound closure at 8 hours relative to control ± SD (n=3).

(C) The scratch wound-healing assay in MDA-MB-231 cells that have been cotransfected with control (siCTRL) or HMGA1 (siA1_1) siRNA and vectors expressing siRNA-resistant HA-HMGA1a were presented. The empty vector (-) was used as a negative control. The data are represented as the means of the percentage of wound closure relative to control ± SD (n=3). Under the graph the western blot analysis of the endogenous and transfected (HA-HMGA1) HMGA1 levels in the MDA-MB-231 cells used in the scratch experiment was reported. Actin was used as an internal control.

(D) Representative pictures illustrating that the MDA-MB-231 cells in which HMGA1 (ShA1_1 and ShA1_3) was depleted move as a coherent group. The scale bar represents 50 μm.

(E) Representative pictures of the cell dispersion assay. The scale bar represents 50 μm.

(F) Transwell invasion assays in the cells shown in (A). Cells migrated through the Matrigel and the porous membrane, were fixed after 24 hours from the seed.

(G) Quantification of anchorage-independent growth in soft-agar. The data are represented as the means ± SD (n=3). On the right representative pictures of soft-agar assay.

*P < 0.05, **P < 0.01, ***P < 0.001; two-tailed Student's t-test.

The second main question about the involvement of HMGA1 in metastatic potential was about the impact of HMGA1 on invasion and survival skills out of the primary tumor. Using Matrigel-coated trans-well assay we tested the ability of HMGA1-silenced MDA-MB-231, MDA-MB-157, and MDA-MB-468 to invade, since cells have to degrade the matrix to cross the porous membrane. The depletion of HMGA1 conferred a strong inhibition of this property in all cell lines tested (Fig. 4F). Cell survival outside of the primary tumor depends on the process of anchorage-independent growth, which is a key aspect of the tumor phenotype. Through *in vitro* soft agar assay analysis, we observed that inhibition of HMGA1 expression dramatically suppressed the ability of MDA-MB-231 cells to form colonies in soft agar (Fig. 4G), demonstrating that HMGA1 is required for the survival and proliferation of breast cancer cells in the absence of external stimuli.

All these data, obtained through *in vitro* assay, suggest an involvement of HMGA1 in metastatic processes. Then, we asked whether HMGA1 supports this role in *in vivo* mouse models. The experiments were done using Severe Combined ImmunoDeficiency (SCID) mice, that are widely used as model in malignant disease transplant study because of the impairment of immune-system. We took advantage of MDA-MB-231 cell lines inducible for the expression of HMGA1 specific shRNAs (ShA1_1) and with controls (shCTRL) that were genetically engineered to express the firefly luciferase reporter gene. Once injected in mice these cells could be easily detect after D-luciferin administration, through *in vivo*

bioluminescence imaging (BLI), monitoring tumor growth and metastasis formation. All the following experiments were done in collaboration with Dr. Rosato at the Department of Surgery, Oncology and Gastroenterology, University of Padova.
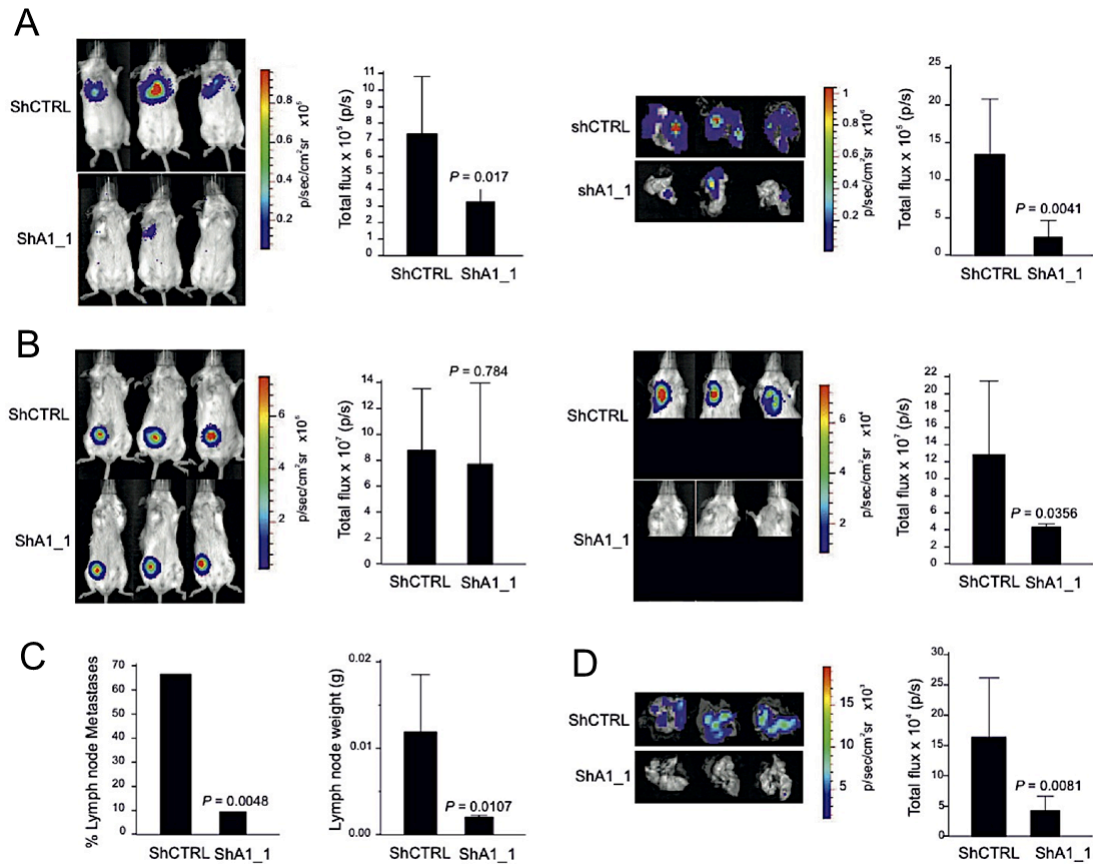
Figure 5. HMGA1 Depletion Suppresses Metastasis *in vivo* in a Mouse Xenograft Model.
(A) MDA-MB-231 ShCTRL and ShA1_1 cells carrying the firefly luciferase reporter gene were injected into the tail veins and mice were analyzed for metastasis using bioluminescence. Three representative mice from each group exhibiting metastasis are shown in the left panel with the corresponding averages of the total flux analyses of 12 mice for each group. Representative ex vivo images of the excised lungs and corresponding averages of the total flux analyses are shown on the right. Analyses were performed at day 26 after cell injection. The data are represented as the means ± SD (n=12 for each group of mice).
(B) MDA-MB-231 ShCTRL and ShA1_1 cells carrying the firefly luciferase reporter gene were injected into the fat pad. Three representative mice showing primary tumors detected *in vivo* and the corresponding averages of the total flux analyses are shown on the left. Shown on the right are the metastases detected *in vivo* and the corresponding averages of the total flux analyses. Here, the lower portion of each animal was shielded before reimaging to minimize the bioluminescent signal from the primary tumor.
(C) The number of metastatic lymph nodes (left) and the average weights of the lymph nodes (right) excised from mice injected with cells in the fat pads. The lymph node weight data are represented as the means ± SD (n=21 for ShCTRL mice and n=21 for ShA1_1 mice).
(D) Representative ex vivo images of the lungs excised from mice that had been injected with cells in the fat pad and the corresponding averages of the total flux analyses are shown. For (B) and (D), the analyses were performed at day 40 after cell injection. The data are represented as the means ± SD (n=19 for ShCTRL mice and n=21 for ShA1_1 mice).
*P < 0.05, **P < 0.01, ***P < 0.001.

Firstly, we examined the impact of HMGA1 on metastatic dissemination after tail vein xenograft in mice. After two weeks we observed a strong reduction in BLI signal in the lung regions of animals that had been injected with HMGA1-depleted cells (Fig. 5A, left) with respect to control cells. These data were confirmed via bioluminescence analysis of explanted lungs (Fig. 5A, right). These results suggested an active regulation of the metastatic process in breast cancer by HMGA1. However, we considered these results as preliminary data, because intravenous injection of tumor cells bypasses several critical steps of the metastatic cascade, including invasion of the tumor border and intravasation into the vasculature. Then, we asked whether HMGA1 silencing impaired metastasis in a more physiological experimental model that summarize different steps of the metastatic process, from tumor growth to dissemination to distal metastasis formation. To this aim, cells were injected subcutaneously into the fat pad of SCID mice. Results indicated that HMGA1 knockdown did not affect tumor growth (Fig. 5B, left). Conversely, we observed a significant reduction in the BLI in the region associated with tumor colonization of the regional homolateral axillary lymph nodes (Fig. 5B, right), the primary site of metastatic dissemination (Nathanson et al., 2010). Indeed, although 14 of 21 control mice displayed lymph node metastasis at this site, only 2 of 21 HMGA1-depleted mice were positive (Fig. 5C, left). Consistently with this observation, we detected a strong reduction of lymph node weight (FIG. 5C, right). Accordingly, we observed a dramatic reduction in lung colonization in mice injected with cells silenced for HMGA1 respect to control mice. Indeed, *ex vivo* BLI analysis of the lungs confirmed these data (Fig. 5D).

Both *in vitro* and *in vivo* data underline a clear involvement of HMGA1 in breast cancer metastatic processes.

4 DEFINITION OF THE PROGNOSTIC HMGA1-GENE SIGNATURE

HMGA1 is a chromatin architectural factor able to bind DNA and to regulate gene transcription (Reeves, 2001). Taken advantage of microarray technology, that allows to compare the expression level of thousands of genes at the same time in different conditions, we examined the functional involvement of HMGA1 in breast cancer malignancy. Therefore, we asked whether HMGA1 altered the transcriptional program in MDA-MB-231 cell line comparing the transcriptional profile of these cells in presence and after silencing of HMGA1. Thanks to the collaboration with Bioinformatics Unit of LNCIB, we were able to highlight

two cluster of genes which change after HMGA1 depletion in MDA-MB-231: a small group of 38 genes that were most upregulated and a larger group containing 130 most down
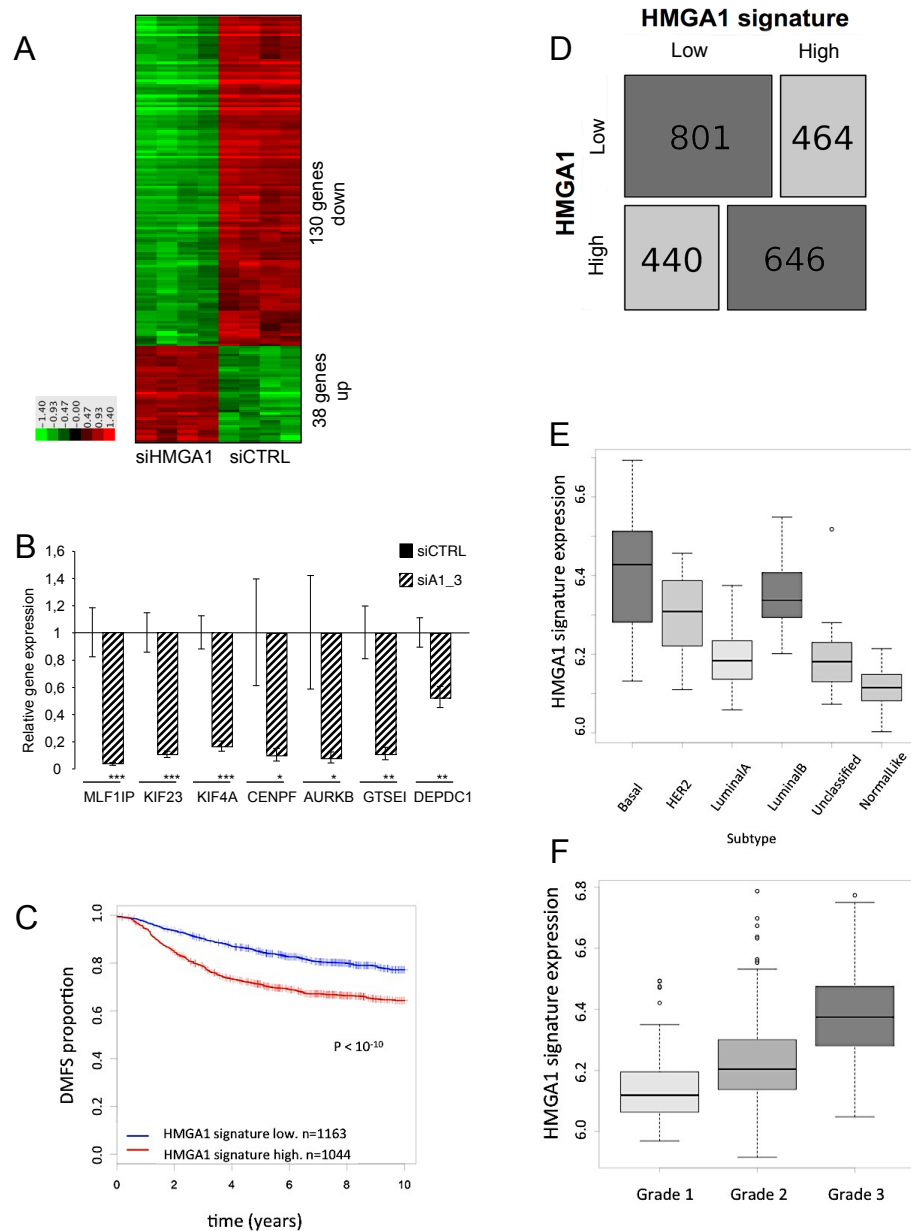
Figure 6. HMGA1-Gene Signature Correlates with Breast Cancer Aggressiveness in Breast Cancer Meta-dataset.

(A) The results of the cluster analysis for the genes obtained from differential genes analysis from microarray analysis of MDA-MB-231 breast cancer cells after HMGA1 depletion for 72 hours are displayed.

(B) Downregulation of selected genes after HMGA1 silencing was measured by qRTPCR. Expression was normalized to the level in MDA-MB-231 cells transfected with siCTRL; GAPDH was used as a normalizer. Data are presented as the mean ± SD (n=3).

(C) Kaplan–Meier survival curve of time to distant metastasis (TDM) for breast cancer patients who were classified according to HMGA1 expression. Red line: cases with high HMGA1 expression; blue line: cases with low HMGA1 expression.

(D) Correlation between HMGA1 and HMGA1-gene signature expression. We classified all of the breast cancer samples (from the public microarray datasets used in previous analyses) using the expressions of HMGA1 or corresponding signature obtained by our microarray experiments to evaluate their association. The mosaic plot shows the proportion of the four possible groups (High_HMGA1-High_HMGA1_signature, High_HMGA1-Low_HMGA1_signature, Low_HMGA1-High_HMGA1_signature, and High_HMGA1-Low_HMGA1_signature) along with the number of samples in each group. Statistical analysis (using the Pearson's Chi-squared test) showed a significant correlation ($p < 2.2e^{-16}$).

(E and F) Boxplots of the distribution of the gene expression intensities of the HMGA1-gene signature across different breast cancer subtypes (intrinsic subtypes or Grades 1, 2 or 3).

*P < 0.05, **P < 0.01, ***P < 0.001; two-tailed Student's t-test.

regulated genes (Fig. 6A, Table.1 upregulated genes, Table. 2 downregulated genes, in Appendix). The last cluster contains genes that could be positively regulated by HMGA1 and the fact that this is the largest group is consistent with the predominantly function of HMGA1 as a transcriptional activator (Reeves, 2001). We selected a pool of genes among the most downregulated and confirmed through qRT-PCR their decrease in expression after HMGA1 depletion (Fig. 6B).

Microarray experiments produce an extensive quantity of data that have to be managed in order to identify groups of genes with a similar function, which could describe a specific phenotype. We used two different tools: first, the Database for Annotation, Visualization and Integrated Discovery / Expression Analysis Systematic Explorer (DAVID/EASE) that provides a comprehensive set of functional annotation tools to handle large lists of genes. The second tool considered was Ingenuity Pathway Analysis (Chiappetta et al.), which allows interactive pathway analysis of complex 'omics' data. Both tool outputs suggest that HMGA1 depletion affected genes involved in cell cycle, cancer, cellular movement and metabolism (Table 3 in Appendix).

Because we were interested in the analysis of the transcriptional program induced by HMGA1 we focused our attention on downregulated genes after HMGA1 depletion and we asked whether these genes were linked to tumor aggressiveness. The 130 downregulated genes were analyzed using the Oncomine web tool, which allows to investigate cancer transcriptome data and in particular if there are any associations between the list of gene of interest and the gene expression profiles of different cancer types present in the database (Rhodes and Chinnaiyan, 2004). Interestingly, this analysis revealed a higher expression of the 130 genes in tumor tissue vs. normal tissue (ratio 89/6), and in bad vs. good clinical outcome (ratio 55/4), primarily in breast cancer (ratio 31/0) (Table 4 in Appendix). These results suggest a possible clinical correlation for the genes downregulated after HMGA1 silencing. Hence, to deepen

this aspect, we performed Kaplan-Meier survival analysis on the 130 HMGA1-downregulated genes in a breast cancer meta-dataset consisted of more than 2000 patients for which were available gene expression data and clinical value. We observed a significant correlation with clinical outcome in terms of Time to Distant Metastasis (TDM). Indeed, patients expressing high levels of these genes displayed a shorter TDM (Fig. 6C). Moreover, the higher expression of 130 genes correlated with HMGA1 mRNA levels (Fig. 6D) and associated with basal-like subtypes and high-grade breast cancer as HMGA1 expression (Fig. 6 E and F). All together these results show that HMGA1 orchestrates a specific transcriptional program in basal-like breast cancer that confers high malignant characteristics to the tumor, defining the 130 genes downregulated after HMGA1 depletion as an HMGA1-gene signature in breast cancer. Therefore, we asked whether this signature could be an independent predictor of clinical outcome. Cox multivariate analysis was applied showing that the HMGA1-gene signature was a significant ($p<0.05$) independent prognostic factor together with ER status, lymph node impairment and tumor size in a cohort of 586 patients (Table 5 in Appendix). Because of the small number of clinical variables in this cohort, we decided to investigate a second independent cohort of patients (n=115) that had more clinical variables available. Concordantly with the first analysis, the HMGA1-gene signature added prognostic information to clinical variables commonly used in practice (Table 6 in Appendix). These data suggest that HMGA1 plays a central role in breast cancer aggressiveness by modulating a specific gene network that behaves as a prognostic signature.

5   FUNCTIONAL ROLES OF HMGA1-GENE SIGNATURE

All results, previously showed, demonstrate the involvement of HMGA1 in the reversion of the malignant phenotype to a more differentiated and epithelial phenotype supporting its participation in EMT and in breast cancer stem cell promotion. The contribution of HMGA1 in these processes ultimately leads to regulation of migration and invasion, key properties of the metastatic disease. Moreover, HMGA1 modulates a gene network that can be used to stratified patients who develop metastasis earlier and whose expression is prominent in high grade metastatic breast cancers. Given these data, we asked the functional roles of HMGA1-gene signature, analyzing more in detail its genes. Firstly, we observed the presence of several genes involved in EMT and cancer stem cell, as for example CD24 (Ponti et al., 2005), IL1R1 (Franco-Barraza et al., 2010), SERPINE1 (Hogan et al., 2013), LIFR (Pitman et al., 2004), LEF1 and SET8. Of particular interest, LEF1 and its partner SET8 cooperate in complex with

β-catenin in the Wnt canonical pathway sustaining EMT and stem cell properties. We confirmed the down-regulation of LEF1 and SETD8 mRNA using qRT-PCR in MDA-MB-231 and MDA-MB-157 after HMGA1 depletion (Fig. 7A) whereas β-catenin mRNA levels did not change (Fig. 7A, upper). However, as previously shown we observed β-catenin re-localization from nucleus to cell-cell contacts, suggesting a clear involvement in this pathway.
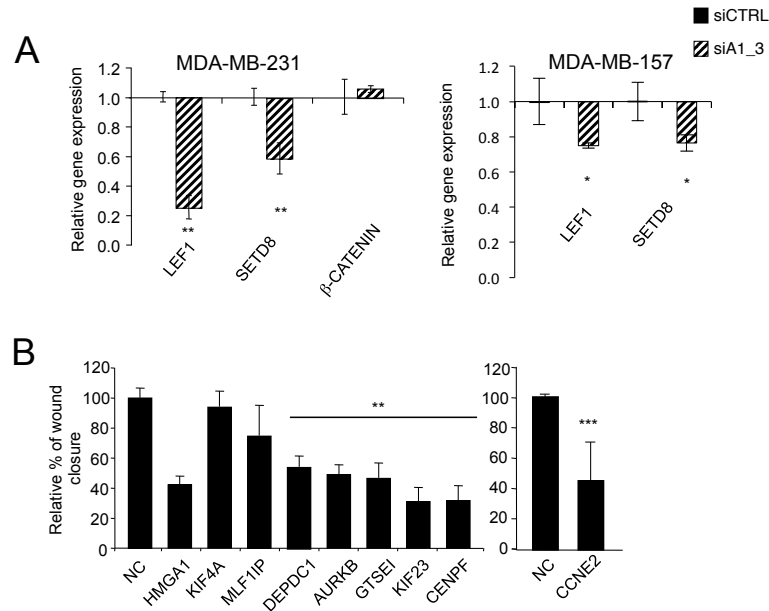
Figure 7. HMGA1 Promotes a Molecular Program Linked to EMT, Stem Cell and Migration Properties.
(A) Down-regulation of Wnt-related genes after HMGA1 silencing was measured using qPCR in MDA-MB-231 and MDA-MB-157. Expression was normalized to the levels in control siRNA cells. GAPDH was used for normalization. The data are represented as the means ± SD (n=3).
(B) The scratch wound-healing assay in MDA-MB-231 cells transfected with control siRNA (negative control, NC) or siRNA specific for each analysed gene. The data are represented as the means of the percentage of wound closure relative to control ± SD (n=4).
*P < 0.05, **P < 0.01, ***P < 0.001; two-tailed Student's t-test.

Then, we investigated the involvement of HMGA1-gene signature in the regulation of migration abilities. In order to select a pool of genes to be tested for motility properties we ordered the genes present in the HMGA1-gene signature considering their correlation between their expression levels and the clinical data concerning patient poor prognosis. Thus, we constructed corresponding Cox proportional hazards model using public datasets of breast cancer gene expression. Therefore, we selected eight of the most differentially expressed genes that were also most correlated with clinical data outcomes (i.e., GTSEI, AURKB, CENPF, MLF1IP, DEPDC1, KIF23, KIF4A and CCNE2) and we explored their role in cell motility silencing their expression (via siRNA) in MDA-MB-231 and performing wound

healing assay. These genes are associated with the microtubule network, cytoskeletal organization and cell cycle but less is known about their involvement in cell movement. Beside their already known function, we observed that silencing of six of the eight genes (GTSEI, AURKB, CENPF, DEPDC1, KIF23 and CCNE2) significantly reduced cell migration (Fig. 7B). These results sustain previous observations that place HMGA1 as a center node in a molecular network that sustains the invasiveness of breast cancer cells.

6 HMGA1 REGULATES CCNE2 IN BREAST CANCER CELL LINES

CCNE2 is the gene that codified for Cyclin E2, the second member of the E-type cyclin family, crucial proteins involved in cell cycle that act through the association and activation of Cyclin Dependent Kinase (CDK). CCNE2 expression depends on the cell cycle stage and globally peaks at late G1 until the end of S phase concurrently with DNA replication differently to CCNE1 which expression decreases during S phase (Caldon et al., 2013). In breast cancer, CCNE2 is frequently overexpressed, leading to genomic instability without affecting mitotic progression (Caldon et al., 2013). Of interest, besides its known role in cell cycle, CCNE2 takes part in several prognostic signatures that predict a shorter metastasis-free survival or relapse-free interval in breast cancer patients (Sotiriou et al., 2006; van 't Veer et al., 2002; Wang et al., 2005). Therefore, CCNE2 is involved in breast cancer progression and aggressiveness but how it contributes to the disease is still unknown. We found that CCNE2 is present in our HMGA1-gene signature. It is one of the most regulated genes by HMGA1 and it is one of the most highly correlated genes with poor clinical outcome (Fig. 8A). Thus, we decided to deepen the relation between HMGA1 and CCNE2 and their role on breast cancer aggressiveness. Firstly, we investigated whether HMGA1 regulates directly or un-directly CCNE2. We confirmed the down-regulation of CCNE2 expression after HMGA1 silencing in MDA-MB-231 and in MDA-MB-157 breast cancer cell lines. In both cell lines, depletion of HMGA1 caused a significant decrease of CCNE2 expression at mRNA (Fig. 8B) and protein level (Fig. 8C). This observation was further established using a second independent siRNA targeting a different region of HMGA1 mRNA in MDA-MB-231 cell line (Fig. 8C). Moreover, the overexpression of HMGA1a tagged with Green Fluorescence Protein (GFP) was able to upregulate the expression of endogenous CCNE2 mRNA in the same cell line (Fig. 8D). Subsequent, because HMGA1 is a chromatin architectural factor and works predominantly binding DNA and controlling gene transcription, we asked whether HMGA1

regulates directly CCNE2 gene expression by binding to its promoter. To answer this question, we performed Chromatin Immuno Precipitation (ChIP) assay of HMGA1 in MDA-MB-231 cells and we designed primers to detect three different regions spanning from 9.0 kb before CCNE2 transcriptional start site (TSS) to CCNE2 TSS. The first region includes CCNE2 proximal promoter (primers designed to detect a region from -368 to -246), the second and the third are AT-Rich Region (named AT-RR), whose primers were designed

A

Symbol	Description	ProbeID	Fold Change	P Value	Score
NCAPG	non-SMC condensin I complex, subunit G	218662_s_at	-4.07	2.5E-10	11.28
MLF1IP	MLF1 interacting protein	218883_s_at	-6.64	6.8E-12	9.98
CCNE2	cyclin E2	205034_at	-4.86	1.8E-09	9.81
CENPF	centromere protein F, 350/400kDa (mitosin)	207331_at	-4.14	7.3E-09	8.63
KIF23	kinesin family member 23	204709_s_at	-5.01	2.4E-11	8.57
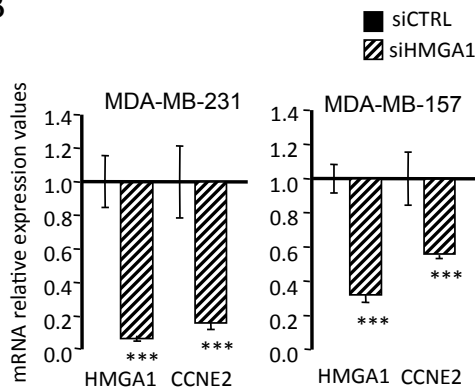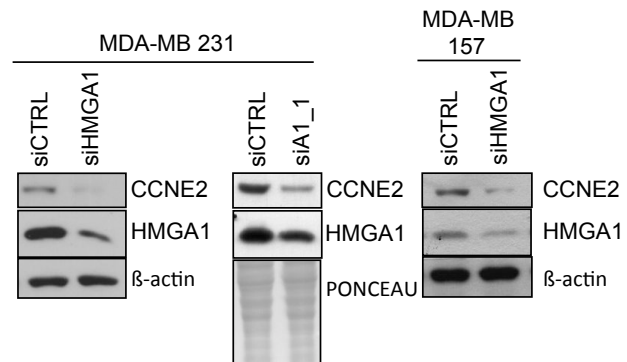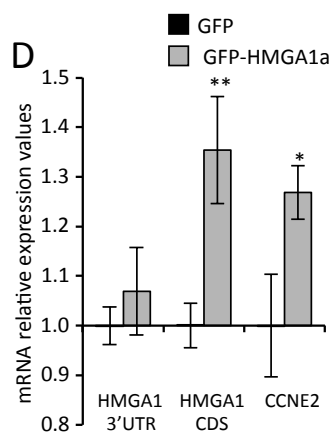
Figure 8. HMGA1 Mediates CCNE2 Expression.
(A) 130 HMGA1-gene signature genes were ranked in order to the overall survival. The top five genes are shown.
(B) CCNE2 mRNA expression was measured after 72 h of HMGA1 silencing (siHMGA1) in MDA-MB-231 and MDA-MB-157 cells using qRT-PCR. The level of CCNE2 expression in HMGA1 silenced cells was compared to that in cells transfected with control siRNA (siCTRL). GAPDH was used for normalization. The data are represented as the mean±SD (n = 3).
(C) CCNE2 protein expression was analyzed after 72 h of HMGA1 silencing (siHMGA1 or siA1_1) in MDA-MB-231 and (siHMGA1) in MDA-MB-157 cells. A representative Western blot is shown. ß-actin was used as a loading control.

(D) HMGA1 3'UTR, CDS and CCNE2 mRNA expression was measured after 30 h of GFP-HMGA1a overexpression in MDA-MB-231. 3'UTR amplification detects only endogenous HMGA1 while CDS amplification detects both endogenous and overexpressed HMGA1 mRNA because pEGFP-N1 HMGA1a vector contains only the HMGA1a coding sequence without 3'UTR. Levels of each mRNA were compared to that in cells transfected with control vector (GFP). GAPDH was used for normalization. The data are represented as the mean±SD (n = 3).
*P < 0.05, **P < 0.01, ***P < 0.001; two-tailed Student's t-test.

to detect a region from -2833 to -2714 (AT-RR1) and from -8606 to -8469 (AT-RR2). We chose these two regions because of their unusual AT-rich composition, which is of 62% for AT-RR1 and 55% for AT-RR2. HMGA1 classically binds the minor groove of DNA in AT-rich region. In particular, it is known that HMGA1 is able to bind Scaffold/Matrix-Attachment-Region (S/MAR) regulating gene expression (Chavali et al., 2011). Using the MAR-Wiz tool, we analyzed the presence of S/MAR throughout the 11.0 kb before the TSS of CCNE2 finding that AT-RR2 is calculated as a S/MAR (Fig. 9A). Moreover, using MatInspector analysis, we highlighted the presence of several HMGA1 binding sites in this region and in AT-RR1 (Fig. 9A red arrows). The results of the two bioinformatics tools suggested that AT-RR1 and AT-RR2 could be putative regulatory elements in the regulation of CCNE2 by HMGA1. ChIP experiments showed that HMGA1 bond AT-RR1 and AT-RR2, suggesting that it could act through these regulatory regions (Fig. 9B). Moreover, HMGA1 was able to bind also the proximal promoter of CCNE2 (Fig. 9B). ChIP experiments allow researchers to identify whether a DNA binding protein binds a specific DNA region *in vivo*. To test if the binding of HMGA1 on CCNE2 proximal promoter regulates gene transcription, we performed luciferase assay overexpressing HMGA1a in Hek-293 cells, together with a luciferase reporter gene construct containing the promoter region of CCNE2 (from −580 to +226 nt) cloned upstream the firefly luciferase gene. The overexpression of HMGA1a was able to increase the activity of the promoter while a deletion mutant ΔCCNE2 (containing the region spanning from +1 to + 223) had a much lower basal activity and it was weakly induced by HMGA1a (Fig. 9C). Consistently, the silencing of endogenous HMGA1 expression was able to decrease the promoter activity of the reporter gene (Fig. 9D).

Next, we asked whether the relationship between HMGA1 and CCNE2 observed in MDA-MB-231 and MDA-MB-157 was confirmed in breast cancer patients in terms of co-expression correlation with respect to different breast cancer grades and subtypes, and analyzing the ability of two genes to stratify patients with distant metastasis free survival (DMFS). These analyses were done in collaboration with Bioinformatics Unit of LNCIB.
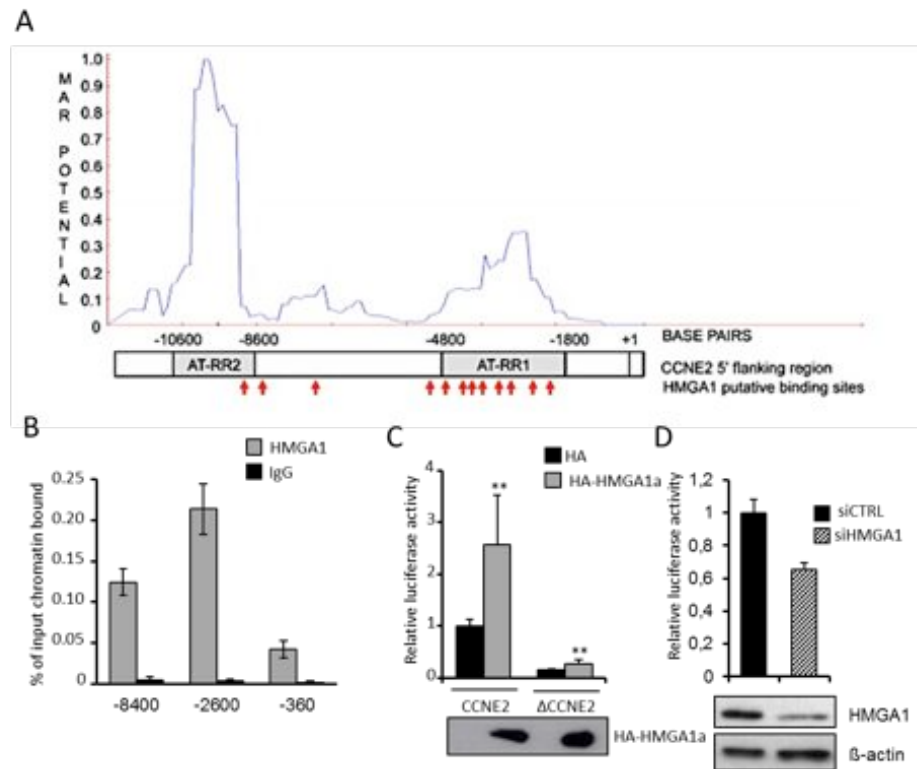
Figure 9. HMGA1 Regulates CCNE2 Expression by Binding its Promoter.
(A) MAR potential analysis by MAR-Wiz tool of sequence from -11430 to +226 of CCNE2 5'flanking region. Gray boxes correspond to AT-rich regions. Red arrows indicate the HMGA1 putative binding sites calculated with MatInspector tool.
(B) Chromatin immunoprecipitation analysis of HMGA1 binding to CCNE2 promoter from MDA-MB-231 cells. Chromatin was immunoprecipitated with anti-HMGA1 antibody or rabbit purified IgG as negative control. Promoter occupancy was analyzed by qPCR amplifying three different regions: AT-RR2 (from -8606 to -8469), AT-RR1 (from -2833 to -2714) and CCNE2 promoter (from -368 to -246) and calculated as percentage of input chromatin bound (n = 3).
(C) HEK293 cells were transiently cotransfected with the luciferase reporter plasmid CCNE2 or ΔCCNE2 in combination with the expression plasmid pcDNA3HA or pcDNA3HA-HMGA1a. pRL-CMV Renilla luciferase expression vector was included to normalize for transfection efficiencies. Values are reported as relative luciferase activity comparing to cells transfected with the reporter vector CCNE2 and pcDNA3HA. The data are represented as the mean±SD (n > 5). Below the graph is reported the Western blot of HA-HMGA1a.
(D) Luciferase assay on HEK293 silenced for HMGA1 and transfected with luciferase reporter plasmid CCNE2. Below the graph is reported the Western blot of HA-HMGA1a.
*P < 0.05, **P < 0.01, ***P < 0.001; two-tailed Student's t-test.

First, breast cancer samples from the meta-dataset or from TCGA dataset were stratified according to the relative expression levels of HMGA1 and CCNE2. As shown in the two contingency tables there was a significant difference in patient distribution in both datasets (meta-dataset: chi-square, $P < 10^{-15}$, TCGA-dataset: chi-square, $P < 10^{-8}$) (Fig. 10A). Moreover, when patients present in meta-dataset were classified according to grade and molecular subtype, high co-expression of CCNE2 and HMGA1 was found in Grade 3 breast cancer and in the more aggressive breast cancer subtypes such as luminal B and basal-like

(Fig. 10 B and C). Lastly, we performed a Kaplan-Meier survival analysis considering the probability of developing distant metastasis (DMFS) in the case of high or low expression of HMGA1 and CCNE2 alone or together. We observed that patients expressing high levels of both genes displayed a significantly higher probability ($P < 10^{-15}$) of developing distant metastasis than patients expressing low levels of both genes (Fig. 10D, red line vs. green line), suggesting that the HMGA1-CCNE2 axis confers an aggressive phenotype to the tumor. Take alones HMGA1 was highly expressed in patients with a better DMFS when CCNE2 expression levels were low (orange line), whereas CCNE2 alone correlated with a poor prognosis when HMGA1 expression levels were low (blue line). Therefore, CCNE2 participated to breast cancer aggressiveness even alone.

A

	CCNE2		
HMGA1	**High**	**Low**	**Difference**
High	54.7%	40.6%	14.0%
Low	45.3%	59.4%	-14.0%
	100%	100%	

BC Meta-dataset

	CCNE2		
HMGA1	**High**	**Low**	**Difference**
High	43.3%	25.7%	17.6%
Low	56.7%	74.3%	-17.6%
	100%	100%	

BC (TCGA)

D

HMGA1 low, CCNE2 low, N= 787
HMGA1 low, CCNE2 high, N= 399
HMGA1 high, CCNE2 low, N= 543
HMGA1 high, CCNE2 high, N= 478

B

		Grade		
HMGA1	CCNE2	G1	G2	G3
High	High	7	95	202
	Low	49	173	121
Low	High	23	103	122
	Low	163	258	75

C

|		Intrinsic Subtype (PAM50)				
CCNE2	HMGA1	Basal (B)	Her2 (H)	LumA (LA)	LumB (LB)	Normal-like (NL)
High	High	166	67	55	210	10
	Low	65	35	88	228	5
Low	High	84	37	295	124	38
	Low	41	27	631	120	25

Figure 10. CCNE2 Expression Correlates with HMGA1 Expression in Breast Cancer Sample of Patient and Tumor Aggressiveness.
(A) Contingency table frequencies of breast cancer (BC) samples (meta-dataset (up) and TCGA dataset (down)) classified as having high or low levels of HMGA1 and/or CCNE2.
(B) Mosaic plot showing the expression of HMGA1 and CCNE2 in breast cancer samples. The samples were stratified based on tumor grade and the expression of these two genes (higher or lower that the mean expression of the given gene in the meta-dataset). The table shows the number of samples for each category.
(C) Mosaic plot showing the expression levels of the two genes in the breast cancer Meta-dataset. Patients are stratified based on the cancer subtype. The table shows the number of samples for each category.
(D) Kaplan–Meier survival curves showing the relevance of CCNE2 and HMGA1 expression to clinical outcome, specifically DMFS. The patients were stratified based on the expression of these two genes.

Indeed, there is a subset of patients expressing high CCNE2 and low HMGA1 levels, in which CCNE2 may be regulated by other still unknown transcriptional regulators.

Thus, these results clearly show that HMGA1 regulates CCNE2 in breast cancer cell lines by binding to the CCNE2 promoter and their co-expression may mediate oncogenic properties of breast cancer conferring a shorter DMFS in breast cancer patients.

7   CCNE2 MEDIATES CELL MOTILITY OF BREAST CANCER CELLS DOWNSTREAM OF HMGA1 WITHOUT AFFECTING CELL GROWTH

CCNE2 belongs to E-cyclin family and its conventional roles include cell cycle regulation through CDK2 activation. For this reason, we initially investigated whether CCNE2 affected cell proliferation in basal-like breast cancer cell lines MDA-MB-231 and MDA-MB-157. To assess the involvement of CCNE2 in this process we efficiently silenced its expression using two siRNA targeting two different regions of its mRNA (Fig. 11A) and then, we evaluated cell proliferation through MTS assay. CCNE2 depletion did not significantly alter proliferation both in MDA-MB-231 and MDA-MB-157 cells (Fig. 11B), suggesting that CCNE2 mediates breast cancer aggressiveness in an unconventional way. In agreement with these data, cell cycle analysis did not show strong changes in cell cycle, revealing only a slight accumulation of cells in G0/G1 phase only in MDA-MB-157 cell line (Fig. 11C). Thus, CCNE2 involvement in cell cycle progression seems to be very limited in basal-like breast cancer cell lines. As said above, CCNE2 is expressed in several prognostic gene expression profiles that predict a shorter metastasis-free survival, indeed is one of the six genes for the molecular detection of circulating tumor cells in the blood of female cancer patients and previously data suggest its possible involvement in migration (Girardini et al., 2011).

Moreover, we demonstrated with the previous experiments that HMGA1 acts principally in regulating cell motility and invasiveness in breast cancer cells. Thus, we asked whether CCNE2 had a functional importance in cell motility and invasiveness in breast cancer cells and if it acts downstream of HMGA1. Therefore, we silenced CCNE2 expression in MDA-MB-231 and in MDA-MB-157 and we performed wound-healing assay and trans-well assay.
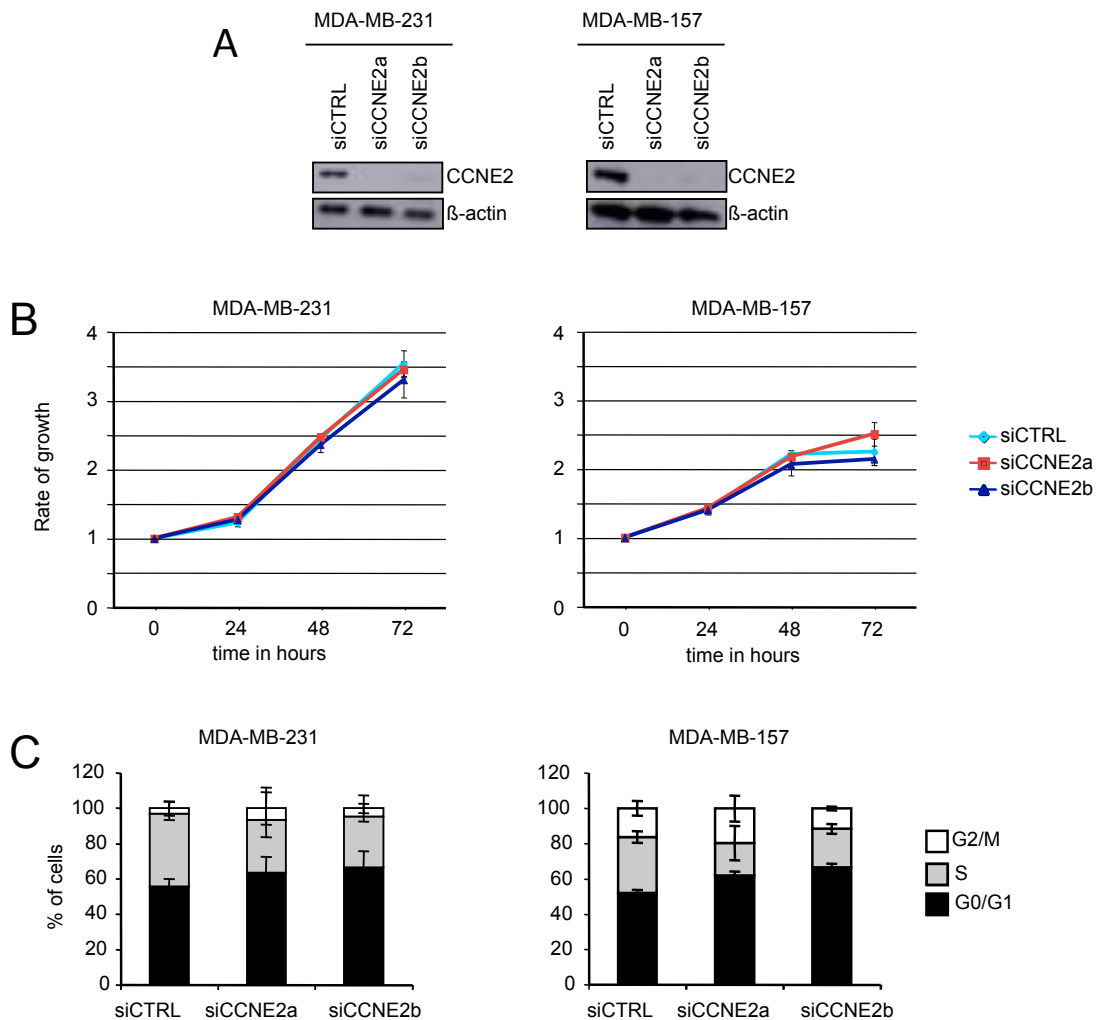
Figure 11. CCNE2 Depletion does not Impaired Cell Growth.
(A) Western blot analysis of CCNE2 levels after 72 hours from transfection of the MDA-MB-231 (left) and MDA-MB-157 (right), with siCTRL, siCCNE2a or siCCNE2b. In (A) and (B) β-actin was used as an internal control.
(B) MTS assay in MDA-MB-231 and MDA-MB-157 silenced for CCNE2 with siCCNE2a and siCCNE2b. Metabolic activity was followed every 24 hours until 72 hours. Data are presented as means ± SD (n≥4).
(C) Cell cycle was analyzed for DNA content by propidium iodide staining in MDA-MB-231 and MDA-MB-157 after 72 hours from CCNE2 silencing with siCCNE2a and siCCNE2b. Data show the means of the percentage of cells in the different cell cycle phases ± SD (n=5).

Interestingly, depletion of CCNE2 significantly impaired motility ability of these cells both in 2D and in 3D migration (Fig. 12 A and B). Moreover, using Matrigel-coated inserts, all cell lines tested showed a dramatic decrease of invasion properties after CCNE2 knock down (Fig. 12C). These results clearly demonstrate the involvement of CCNE2 in pro-metastatic skills of breast cancer cells. Next, we tested whether CCNE2 acts as a downstream effector of HMGA1-induced cell migration. We reintroduced CCNE2 expression using an expressing vector in which we cloned a tagged form of it (HA-CCNE2) transfected in MDA-MB-231 cells previously depleted for HMGA1 expression and we performed wound-healing assay. We observed a rescue of the migration ability of the cells silenced for HMGA1 to levels similar to control cells (Fig. 12D). All these results demonstrate for the first time that CCNE2 plays a role in promoting cell migration and invasion in highly aggressive basal-like breast cancer cell lines downstream of HMGA1.
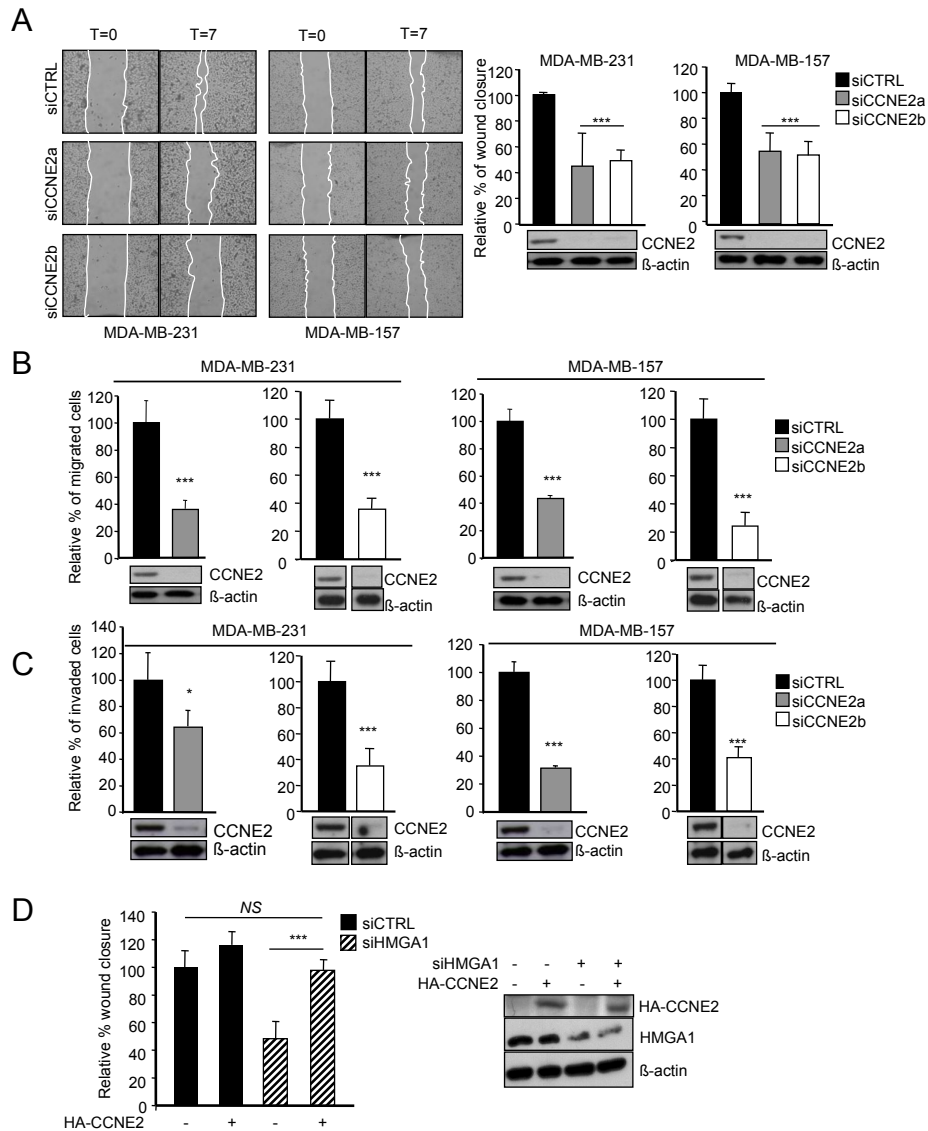
Figure 12. CCNE2 Silencing impairs Migration and Invasion of Basal-like Breast Cancer Cells.
(A) Left, representative images of a wound-healing assay in which MDA-MB-231 and MDA-MB-157 cells were transfected with control siRNA (siCTRL) or two different siRNAs against CCNE2 mRNA (siCCNE2a or siCCNE2b) are shown. Confluent cell cultures were scratched, and wound closure was analyzed after 7 h with respect to time zero. Right, quantification of the wound-healing assay is presented as the means of the percentage of wound closure relative to the control±SD (n > 3).
(B) and (C) quantification of transwell migration (B) and invasion (C) assays performed on MDA-MB-231 and MDA-MB-157 cells transfected with siCTRL, siCCNE2a or siCCNE2b. The data are presented as the mean of the percentage of the number of cells relative to the control±SD (n > 3).
(D) Analysis of a wound-healing assay in MDA-MB-231 cells co-transfected with siCTRL or siHMGA1 and a vector expressing HA-CCNE2. The empty vector (-) was used as a negative control. The data are presented as the means of the percentage of wound closure relative to the control±SD (n = 4). Right, a representative Western blot analysis of the cell lysates is presented. ß-actin was used as a loading control.
*P<0.05, ***P < 0.001; two-tailed Student's t-test.

8 HOW DOES CCNE2 REGULATE CELL MIGRATION?

The notion that cyclins can exert a critical role in pro-metastatic properties is fairly recent. Indeed, it is already known that Cyclin D1 and Cyclin A2 regulate motility independently to cell cycle regulation. Cyclin D1 positively acts on migration through CDK4/CDK6 or by binding p27^{kip1} in a CDK independent manner, whereas Cyclin A2 inhibits cell motility, activating RhoA (Casimiro et al., 2012). Thus, Cyclins can unconventionally control this feature via canonical or not-canonical partner interaction. Therefore, the next question concerns the mechanism by which CCNE2 mediates cell migration and invasion. To unravel which pathway could be regulated by CCNE2 in this context, we adopted an unbiased strategy analyzing a cohort of 408 breast cancer patients samples from TCGA dataset for which expression data (from RNA-seq) and phospho-proteomics data (from Reverse Phase Protein Array (RPPA)) were available, allowing us to explore whether alteration in CCNE2 mRNA expression correlates with some protein phosphorylations. The top five most significant phosphorylations found to be correlated with CCNE2 expression were the modification at Ser127 of YAP, at Tyr705 of STAT3, at Thr70 and Ser65 of EIF4EBP1 and at Thr198 of CDKN1B (Fig. 13A). Interestingly, excluding the phosphorylations on EIF4EBP1, which is involved in protein translation, the phosphorylations on YAP, STAT3 and CDKN1B are linked with cell pro-metastatic ability regulation (Kamran et al., 2013; Lamar et al., 2012; Larrea et al., 2009). On the whole, we focused our attention on YAP, finding an inverse correlation between CCNE2 expression and the phosphorylation of YAP at Ser127 (Fig. 13 A and B). This particular modification induces YAP translocation from the nucleus to the cytoplasm, repressing its activity (Piccolo et al., 2013). Moreover, Lemar and colleagues demonstrated in an elegant work, that the mutated active form YAP^{S127A} rendered the non-metastatic NMuMG, 67NR and A375 cells highly metastatic, enhancing invasion and suggesting a critical role of YAP and of Ser127 phosphorylation in this process (Lamar et al., 2012). We hypothesized that CCNE2, induced by HMGA1 in basal-like breast cancer, could regulate migration and invasion modulating YAP modification status, localization and consequently its activity.

Protein	Residue	RPPA score difference	p-value
YAP1	pS127	-0.48	1.01E-05
STAT3	pY705	-0.44	4.12E-07
EIF4EBP1	pT70	0.62	1.05E-06
EIF4EBP1	pS65	0.53	1.18E-05
CDKN1B	pT198	0.51	7.73E-07
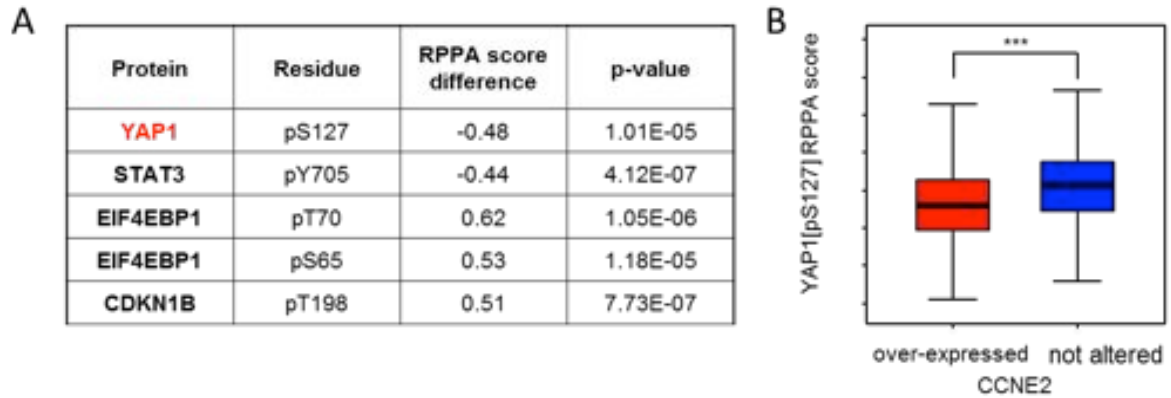
Figure 13. CCNE2 Overexpression in Patient Breast Cancer Sample Correlates with Phospho-modifications mainly Involved in Migration.
(A) Table reporting the first five position of the correlation analysis between RPPA data and mRNA expression of CCNE2 from TCGA dataset.
(B) The boxplot shows the YAP protein phosphorylation level in breast cancer samples. The samples were stratified based on the expression level of CCNE2.

9 HMGA1 AND CCNE2 AFFECT THE NUCLEAR LOCALIZATION AND ACTIVITY OF YAP THROUGH THE MODULATION OF HIPPO PATHWAY IN BREAST CANCER CELL LINES

To unravel the molecular contribution of HMGA1-CCNE2 axis on YAP regulation we evaluated YAP localization via immunofluorescence analysis in MDA-MB-231 and MDA-MB-157 after HMGA1 or CCNE2 silencing. These two cell lines are characterized by a predominantly YAP nuclear localization when maintained at low confluence as shown in figure 14A (see images and values of siCTRL). Depletion of HMGA1 or CCNE2 induced delocalization of YAP that displayed a more cytoplasmic cell distribution, losing its predominantly nuclear localization in both cell lines (Fig. 14A). YAP compartmentalization is finely regulated by post-translational modifications. In particular, the phosphorylation status of YAP at Ser127 is crucial for its localization and then for its activation (Piccolo et al., 2013). Moreover, as described in the chapter 8, an inverse correlation between CCNE2 expression and this modification was observed in invasive breast cancer patient dataset from TCGA. Therefore, we tested whether the depletion of HMGA1 or CCNE2 affected the
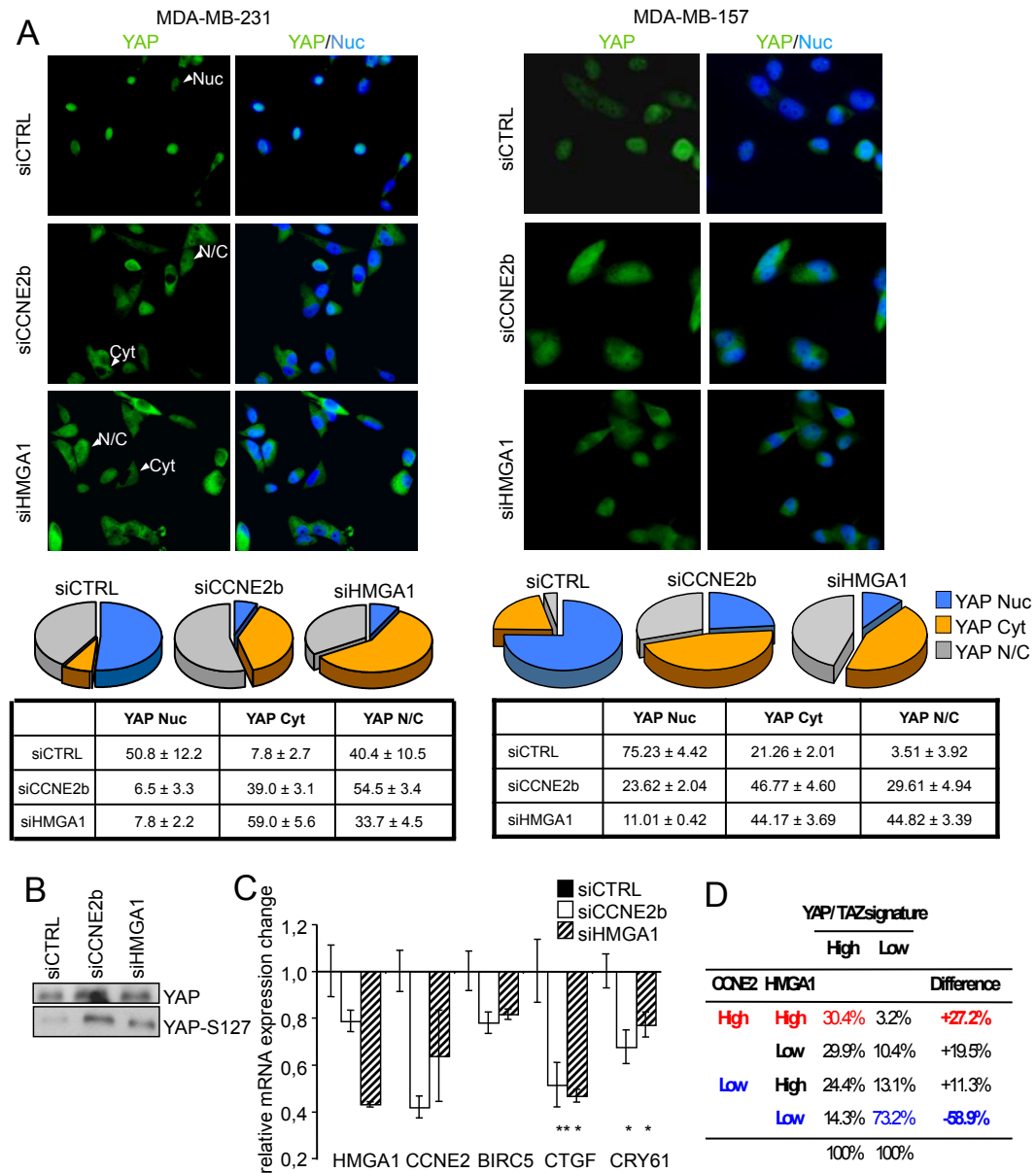
Figure 14. HMGA1 and CCNE2 Affect YAP Localization and Activity.

(A) Representative immunofluorescence images of YAP in MDA-MB-231 cells (on the left) and in MDA-MB-157 (on the right) after HMGA1 and CCNE2 silencing are shown. Representative cells are indicated by arrowheads with abbreviations (Nuc, primarily nuclear; Cyt, primarily cytoplasmic; N/C, diffuse in the nucleus and the cytoplasm). Images were taken at X 60 magnification. Under the representative images, the quantification of YAP localization in MDA-MB-231 and in MDA-MB-157 silenced with siCTRL, siCCNE2 and siHMGA1. Cells from at least 4 fields for slide (n=3) were counted considering YAP nuclear, cytoplasmic or both localization.

(B) Western blot analysis of total YAP and YAP phosphorylated at Ser127 (YAP-S127) in protein lysates of MDA-MB-231 cells transfected with siCTRL, siCCNE2b or siHMGA1. A representative image was reported (n ≥ 5).

(C) qRT-PCR analysis of YAP target genes (CTGF, BIRC5, CYR61) in MDA-MB-231 cells after HMGA1 and CCNE2 silencing. GAPDH was used for normalization. The data are presented as the mean±SEM (n = 4).

(E) Contingency table frequencies of breast cancer samples (TCGA dataset) classified as expressing high or low levels of YAP/TAZ profile genes or displaying high or low HMGA1 and CCNE2 expression levels.

*P < 0.05, **P < 0.01, ***P < 0.001; two-tailed Student's t-test.

phosphorylation of YAP at Ser127 in MDA-MB-231. Silencing both HMGA1 and CCNE2 increased the phosphorylation levels of YAP at Ser127 (Fig. 14B), also confirming in an *in vitro* model the inverse correlation between CCNE2 expression and YAP Ser127 phosphorylation that was found in the tumor samples.

When YAP is nuclear, it can bind several transcription factors regulating transcription of many genes involved in proliferation, apoptosis, metastatic process, EMT, drug resistance and cancer stem cell (Johnson and Halder, 2014). Some of the well-established endogenous YAP target genes are BIRC5, CYR61 and CTGF, so we monitored their expression in MDA-MB-231 after HMGA1 and CCNE2 knockdown. We found that CTGF and CYR61 were significantly downregulated upon HMGA1 and CCNE2 silencing, whereas BIRC5 appeared to be downregulated, although this result was not significant (Fig. 14C). Moreover, in collaboration with the Bioinformatics Unit of LNCIB we observed that the differential activation of the YAP/TAZ signature is associated with the expression levels of both the CCNE2 and HMGA1 genes (Fig. 4D, Chi-square, $P < 10^{-15}$). Furthermore, as HMGA1-CCNE2 axis is mainly implicated in cell migration regulation as described above, we investigated whether YAP is a downstream mediator of this biological function. To study this point, we used a constitutively active form of YAP carrying serine-to-alanine substitutions at the primary YAP phosphorylation sites (YAP-5SA), including Ser127. YAP-5SA expression in MDA-MB-231 rescued the effect of HMGA1 or CCNE2 depletion on cell migration (Fig. 15A). All these data underline a close relationship between HMGA1-CCNE2 axis and YAP, in pro-metastatic properties regulation that was confirmed even in breast cancer patients (TCGA breast cancer dataset).

Next, we asked whether the Hippo pathway, which is the major way of YAP regulation, mediates the effect of HMGA1 and CCNE2 on YAP nuclear localization. Hippo pathway is composed of a core kinase constituted by MST1/2 and LATS1/2, which are primarily responsible for YAP phosphorylation and inactivation. MST1/2 bind SAV1 forming an active complex that can directly phosphorylate LATS1/2 and MOB1 enhancing their interaction and inducing the formation of LATS1/2-MOB1 complex. Activated LATS1/2 phosphorylate YAP at Ser127 promoting 14-3-3 binding and inhibiting its translocation into the nucleus and thus its activation (Halder et al., 2012). To unravel the involvement of the Hippo core kinases in YAP localization following HMGA1 and CCNE2 depletion, we concomitantly silenced MST1/2 or LATS1/2, via siRNAs, in MDA-MB-231 and MDA-MB-157 cells and we
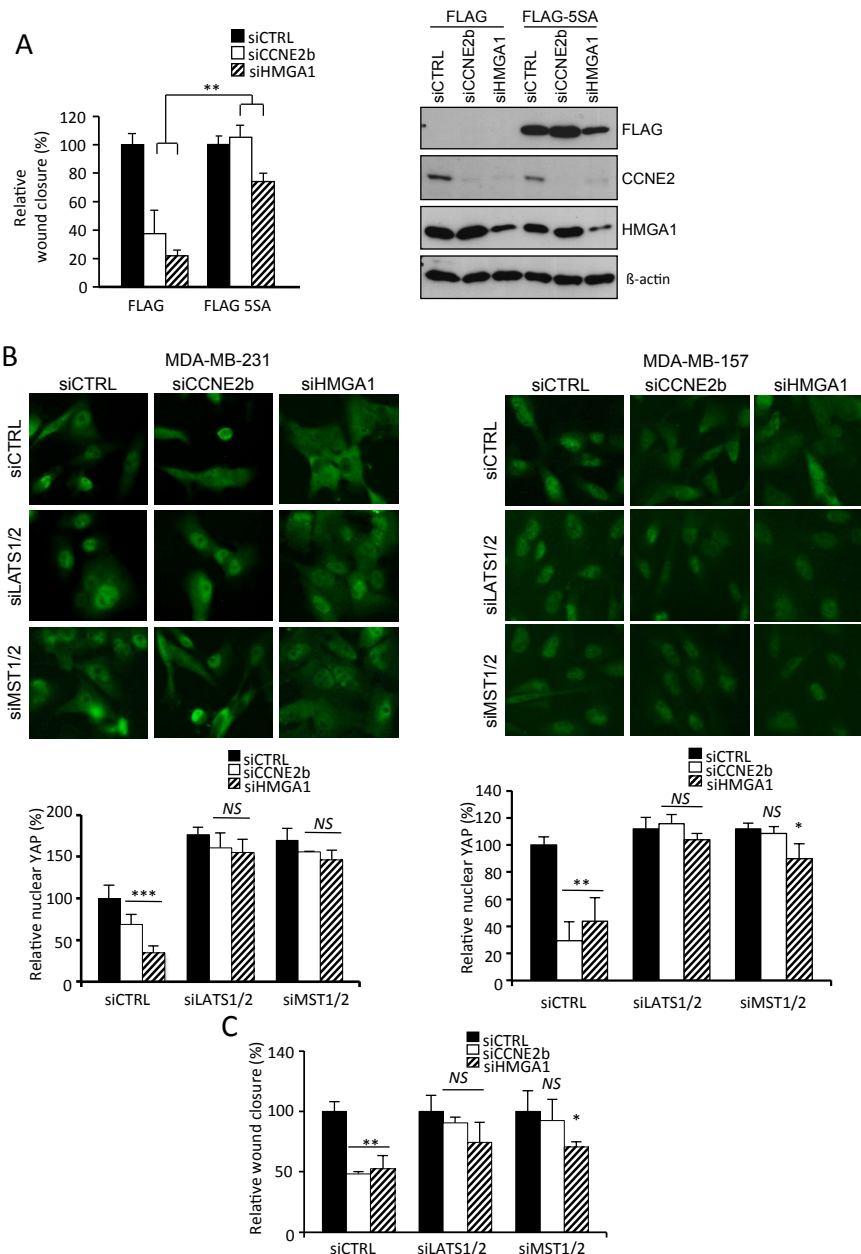
Figure 15. HMGA1 and CCNE2 Regulate YAP Localization in a MST1/2 and LATS1/2 Kinase-dependent manner in Breast Cancer Cells.
(A) Quantification of wound-healing assay using MDA-MB-231 cells stably infected with pBABE-FLAG (FLAG) or pBABE-FLAG-YAP-5SA (FLAG-5SA) after transfection with siCTRL, siCCNE2b or siHMGA1 for 72 h. The data are presented as the means of the percentage of wound closure relative to each control (siCTRL)±SD (n = 3). On the right, representative western blot analysis for FLAG, CCNE2 and HMGA1, using ß-actin as loading control.
(B) Immunofluorescence images of YAP in MDA-MB-231 cells (on the left) and MDA-MB-157 (on the right), indicating the effect of LATS1/2 and MST1/2 silencing on cells in which HMGA1 and CCNE2 were depleted. In the graphs below the images, the ratio of nuclear YAP relative to siCTRL ±SD was presented (n = 3). Images were taken at X 40 magnification.
(C) Wound-healing assay using MDA-MB-231 cells treated as in panel (B).
*P < 0.05, **P < 0.01, ***P < 0.001, NS = not significant; two-tailed Student's t-test.

observed an increase of YAP nuclear localization and an almost completely rescue of the effect of HMGA1 and CCNE2 on YAP localization (Fig. 15B representative images and below the graph of quantification). These results suggest that MST1/2 and LATS1/2 are necessary in HMGA1-CCNE2-induced YAP translocation. Therefore, HMGA1 and CCNE2 regulate YAP upstream to MST1/2 and LATS1/2. Finally, we investigated whether LATS1/2 and MST1/2 kinases were key regulators for HMGA1-CCNE2-YAP mediated effect on cell migration. In agreement with our observations, both Hippo core kinases rescued almost completely the effect of HMGA1 and CCNE2 depletion on cell motility (Fig. 15C). Altogether these results demonstrate that HMGA1 and CCNE2 affect the nuclear localization and activity of YAP via the canonical Hippo pathway.

10 CKD INHIBITORS IMPAIR THE NUCLEAR LOCALIZATION OF YAP AND CELL MIGRATION IN BREAST CANCER CELLS

CDK2 is the major partner of CCNE2 through which it exerts its functions, regulating many nuclear substrates through phosphorylation. We investigated whether CDK2 silencing altered cell motility and YAP localization and thus whether CDK2 is a downstream effector of HMGA1-CCNE2-YAP axis. We efficiently depleted CDK2 using a pool of siRNA in MDA-MB-231 (Fig. 16A) and we tested cell migration using wound healing assay. Following CDK2 silencing we observed a clear impairment of cell motility (Fig. 16B). Moreover, we checked also YAP localization in the same condition and, as shown in figure 16C, CDK2 knockdown reduced YAP nuclear localization. These data suggest that HMGA1-CCNE2-YAP axis employs CDK2 to ensure its function.

The activity of cyclins and CDKs is frequently deregulated in cancers and this leaded researchers to develop small molecules, targeting cyclin-CDK complexes in malignancies. Accumulating evidence underline that inhibiting CDKs may be effective, even if the efficacy on inhibition of proliferation in clinical trial is limited to hematological malignancies (Sanchez-Martinez et al., 2015). Globally these trials consider only tumor growth but not metastatic process producing a limited knowledge. Considering all data presented in this

thesis   we   hypothesized   a   use   of   these   inhibitors   in
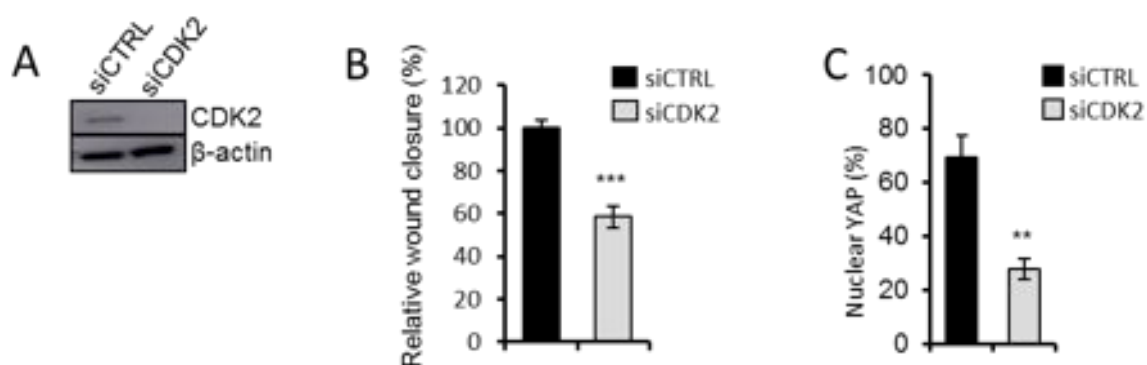
Figure 16. CDK2 silencing Impairs Migration and YAP Nuclear Localization in MDA-MB-231 Cell Line.
(A) Western blot analysis of CDK2 levels after transfection of the MDA-MB-231, with siCTRL, or a pool of siRNAs against CDK2. β-actin was used as an internal control.
(B) Quantification of wound-healing assay in MDA-MB-231 cells transfected with control siRNA (siCTRL) or pool of siRNAs against CDK2, data is presented as the means of the percentage of wound closure relative to the control±SD (n > 3).
(C) Immunofluorescence analysis of YAP localization in MDA-MB-231 treated as in (B). Data are presented as percentage of cells with nuclear YAP in each condition.
**P < 0.01, ***P < 0.001, two-tailed Student's t-test.

impairing migration properties mediated by HMGA1-CCNE2-YAP axis in breast cancer cells with the final aim to break this axis. Therefore, we performed a screening using a panel of CDK inhibitors on MDA-MB-231. CDK inhibitors chosen are considered broad range, in fact as presented in table 7, present in the Appendix section, they target more than one CDK. This feature permits to eliminate CDK redundancy obtaining, at least in theory, a strongest effect. In fact, CDK2 action inhibition could be compensate by CDK1 (Bashir and Pagano, 2005). To test whether CDK inhibitors could be exploited as potential inhibitors of YAP activity, we firstly assessed YAP localization upon treatment. Interestingly, most but not all of the compounds tested affected YAP nuclear localization (Fig. 17A, on the right representative image of the YAP localization upon treatment with the most effective compounds). The three most effective inhibitors (AZD5438, JNJ7706621 and PHA793887) were also tested in MDA-MB-157 obtaining similar results (Fig. 17B on the right representative image of the YAP localization). Moreover, we found that they induced an increase of YAP phosphorylation at Ser127 (Fig. 17D) and that they were able to downregulate the expression of the YAP target genes CTFG and CYR61 (Fig. 17E).

Finally, we tested whether they are able to also impair migration in MDA-MB-231 cells. We performed both wound healing and transwell assay after 24 hours of CDK inhibitor treatment.

We chose 24 hours because at this time point cell proliferation was not affected by treatments (Fig. 18A). All three inhibitors dramatically altered the migratory ability of MDA-MB-231
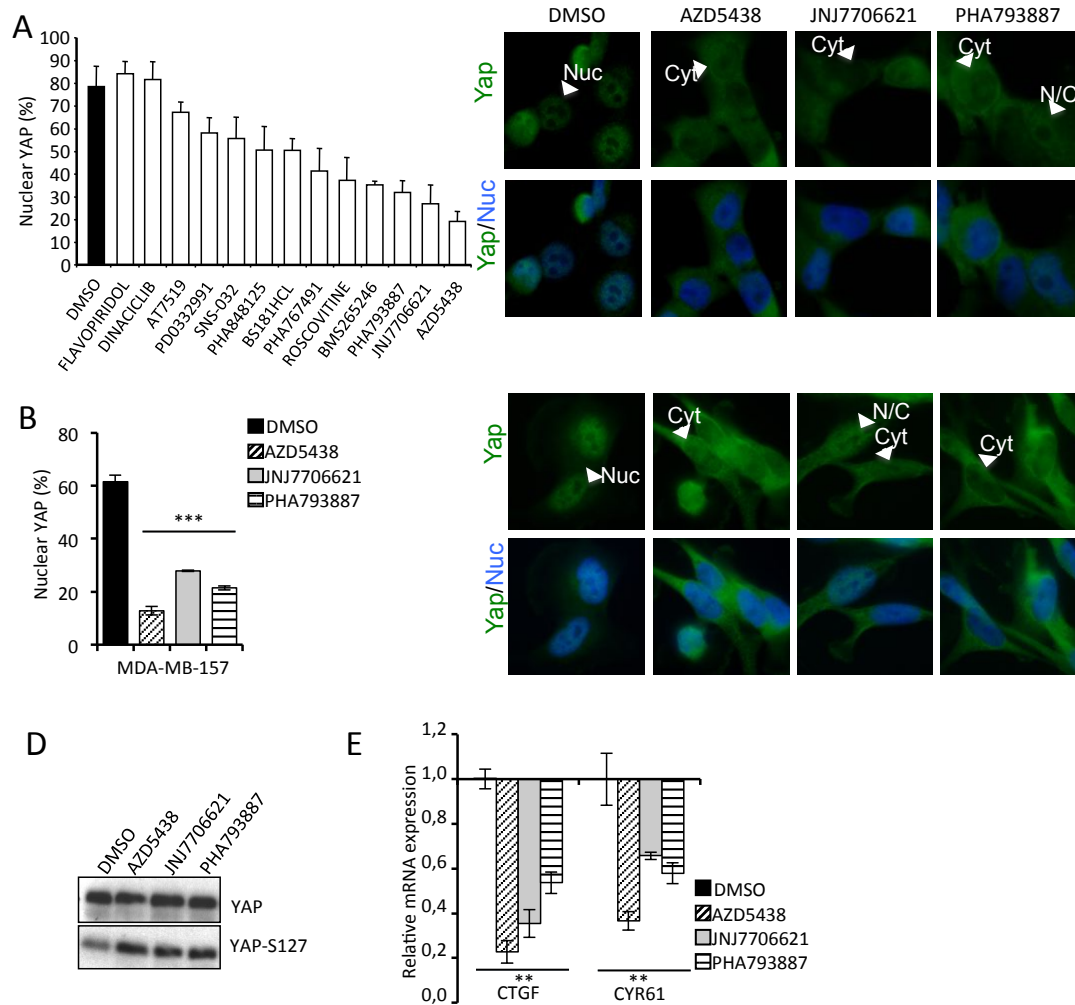
Figure 17. CDK Inhibitors Affect YAP Nuclear Localization in Breast Cancer Cell Lines.
(A) On the left, results of the CDK inhibitor library screening evaluating YAP localization. MDA-MB-231 cells were treated for 24 h with 1 mM of inhibitor, except for SNS-032 and DINACICLIB, which were applied at 0.1 mM. The data are presented as the means of percentage of cells with nuclear YAP±range between replicates (n = 2). On the right, the representative images of YAP (green) localization on cells treated with the most effective compounds from CDK inhibitor library. Nuclei were stained with Hoechst (blu). Images were taken at X 60 magnification.
(B) Immunofluorescence analysis of YAP localization in MDA-MB-157 treated with the most effective CDK inhibitors (AZD5438, JNJ7706621 or PHA793887). Data is presented as percentage of cells with nuclear YAP in each condition (n = 3). On the right, the representative images of YAP (green) localization on cells treated with the most effective compounds from CDK inhibitor library. Nuclei were stained with Hoechst (blu). Images were taken at X 60 magnification.
(C) Western blot of total YAP and phosphorylated YAP at Ser127 (YAP-S127) in protein lysates of MDA-MB-231 cells treated with CDK inhibitors (n = 3).
(D) qRT-PCR analysis of YAP target genes (CTGF and CYR61) in MDA-MB-231 cells treated with CDK inhibitors. GAPDH was used for normalization. The data are presented as the mean ± SD (n = 3).
*P < 0.05, **P < 0.01, ***P < 0.001; two-tailed Student's t-test.

(Fig. 18 B and C, on the left representative images of the assay, on the right the quantification). Moreover, this effect was mediated by YAP; indeed, the overexpression of the active form of YAP, YAP-5SA, rescued cell migration (Fig. 18 B and C).
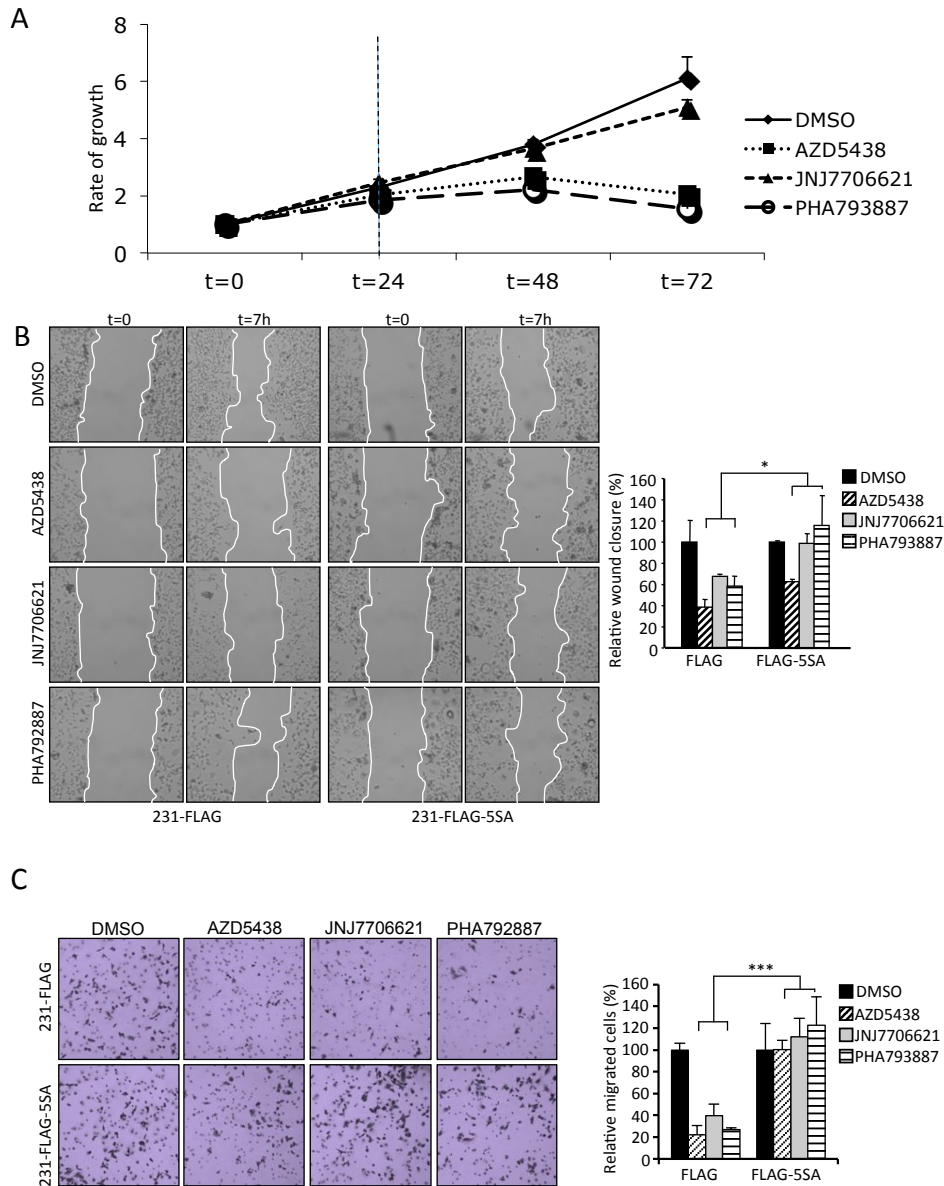
Figure 18. CDK Inhibitors Affect the Motility of Breast Cancer Cell Lines Through YAP.
(A) MTS assay in MDA-MB-231 treated with DMSO, AZD5438, JNJ7706621 and PHA793887. Metabolic activity was followed every 24 hours until 72 hours. Data are presented as means ± SD (n=5).
(B) and (C) on the right representative images of wound-healing (B) and transwell assay (C) using MDA-MB-231 cells stably infected with pBABE-FLAG (FLAG) or pBABE-FLAG-YAP-5SA (FLAG-5SA) after the treatment with AZD5438, JNJ7706621 or PHA793887 CDK inhibitors, on the left is reported the relative assay quantification. The data are presented as the means of the percentage relative to each control (DMSO)±SD (n = 3).
*P < 0.05, ***P < 0.001; two-tailed Student's t-test.

Altogether these results indicate that CDK-inhibitors may block HMGA1-CCNE2-YAP axis modulating YAP cellular localization. Moreover, they suggest a different use of these inhibitors to impair metastatic process beside their primary use in tumor growth.

Discussion And Conclusions

The aim of my Thesis was to unravel the role of HMGA1 in breast cancer aggressiveness focusing on discovering molecular pathways regulated by HMGA1. Previous works suggested an involvement of HMGA1 in cancer both on cell proliferation and/or metastatic properties, depending on the molecular context or tumor type (Cleynen and Van de Ven, 2008; Reeves, 2010; Shah et al., 2013) but the role of HMGA1 in breast cancer was still unclear. Here, we found that HMGA1 mRNA level is enriched in basal-like breast cancer, the poorest prognosis breast tumor subtype (Pegoraro et al., 2013). Among breast cancers, basal-like subtype is less responsive to adjuvant chemotherapeutics and completely insensitive to hormone therapy (Foulkes et al., 2010), we focused our study on the involvement of HMGA1 in this breast cancer subtype in order to evaluate the participation of HMGA1 in its aggressiveness and to identify pathways regulated by HMGA1 that, finally, could be blocked with small molecules. Firstly, we demonstrated that HMGA1 plays a pivotal role in different hallmarks of cancer, regulating invasive processes and determining poor prognostic outcomes in breast cancer by sustaining the mesenchymal phenotype and stem cell properties. Indeed, the silencing of HMGA1 in basal-like breast cancer cell lines impaired mesenchymal fibroblastic features, inducing a phenotypic reversion and a mesenchymal to epithelial transition (Pegoraro et al., 2013). EMT is strongly linked to the acquisition of metastatic properties, in particular cells gain migration and invasion skills (Kalluri and Neilson, 2003). In accordance, our data underline that basal-like breast cancer cells depleted for HMGA1 expression acquire a more differentiated and epithelial behavior, strongly decreasing migration and invasion ability both *in vitro* and *in vivo*, although tumor growth is not impaired. Moreover, it is widely accepted that EMT program induces and sustains stem cell capabilities that are mainly important in cancer relapse and new tissue colonization, forming metastasis (Mani et al., 2008). Interestingly, we observed that cell lines silenced for HMGA1 lose the ability to form mammosphere in culture and lack self-renewal capability. Emerging evidences further indicates that HMGA1 is a master regulator of stem-cell properties in basal-like breast cancer (Shah et al., 2013). In this work, the gene expression program induced by HMGA1 has been compared with that of stem cells demonstrating the presence of a stem-cell gene signature regulated by HMGA1. All these data suggest a specific mechanism by which HMGA1 controls the progression of tumor cells to an aggressive and invasive phenotype: aberrant expression of HMGA1 in tumor cells promotes metastasis by inducing EMT- and stemness-related processes, which in turn may enhance the ability of breast cancer cells to migrate and grow at secondary sites.

The first evidence linking HMGA1 to EMT came from a study in 2001 in epithelial MCF-7 breast cancer cells, which demonstrated that over-expression of HMGA1 cause metastatic progression and EMT (Reeves et al., 2001). Now, we added new relevant connection between HMGA1, EMT, and stemness, deeply analyzing the HMGA1-molecular pathways involved. One of the critical pathways that regulate stem cell maintenance and acquisition of EMT characteristics during tumorigenesis is the Wnt/β-catenin pathway (Micalizzi et al., 2010). A connection between Wnt/β-catenin pathway and HMGA1 was already described in colon cancer in two papers even if the results are conflicting. Indeed, one group described HMGA1 as downstream effector of Wnt/β-catenin signaling (Bush et al., 2013), instead the other identified HMGA1 as a partner of β-catenin that cooperates in the regulation of β-catenin-target genes (Xing et al., 2014). In our system, we found that HMGA1 silencing induces a dramatic relocalisation of β-catenin to cell-cell contact points and likely modulates the expression of two known β-catenin coactivators, LEF1 and SETD8, that in complex with β-catenin act as mediators of Wnt signaling (Li et al., 2011). Therefore, it is reasonable that HMGA1 can contribute to EMT and stem cell skills regulating Wnt canonical pathway. One of the main genes regulated by this pathway is JAG1, which is a top player of the Notch pathway, interestingly, it is known that Wnt and Notch pathways work together during carcinogenesis to induce self-renewal. Moreover, the Notch pathway regulates the EMT in both physiological and pathological conditions (Creighton et al., 2010). Here, we discovered that HMGA1 silencing downregulates JAG1 and its downstream effector HES1, allowing a crosstalk between HMGA1 and the Wnt and Notch pathways (Pegoraro et al., 2013). These findings collocate HMGA1 as a critical hub in the regulation of relevant pathways that promote the EMT and stemness in breast cancer.

In the last years, gene signatures are imposed as a new tool for clinical management that can predict prognosis or better treatment for a patient starting from the gene expression profile of the primary tumor (Arranz et al., 2012). For example, Mammaprint is actually used in clinic to predict the risk that a primary breast cancer will metastasize to other organs (Mook et al., 2010). Starting from microarray analysis of gene expression in cells depleted for HMGA1, we defined a HMGA1-gene signature composed by 130 genes positively correlated with HMGA1 expression, with the most aggressive and undifferentiated basal-like subtype, with high relapse rates, and with poor patient survival (Pegoraro et al., 2013). Moreover, some of the genes in the HMGA1-gene signature such as CENPF, CENPA, CCNE2, BUB1 and PSMD2

are also part of the Mammaprint 70 gene signature indicating a possible connection with metastatic advancement. Intriguingly, our data indicate that the HMGA1 gene signature is a strong independent factor in the prediction of disease outcome. In particular, it is associated with poorer clinical outcomes that are related to a short time to distant metastasis. This functional analysis suggests that HMGA1 may directly regulate the expression of a number of genes that stratify for metastasis development. We also demonstrated that 6 out of the 8 genes tested were directly involved in promoting breast cancer cell migration. Some of these are involved in microtubule regulation such as Kif4A (Mazumdar et al., 2004), Kif23 (White and Glotzer, 2012) during metaphase and cytokinesis and GTSE1 in plus-end tracking. Interestingly, GTSE1 has emerged as a microtubule-associated protein that is correlated with tumor metastasis in breast cancer. GTSE1 promotes migration via focal adhesion turnover (Scolz et al., 2012), suggesting a possible role for HMGA1 in controlling migration through the regulation of microtubule pathways. In addition, we found a significant overlap between our HMGA1-gene signature and pathways known to be involved in migration and invasion, for example, the recently defined Pin1/mutant p53 gene signature (Girardini et al., 2011). In agreement with previous findings indicating that DEPDC1, a direct downstream target of the mutant p53 pathway, is relevant to migration and invasion, we found a functional link between HMGA1 and DEPDC1. DEPDC1 regulates migration downstream of HMGA1, suggesting that HMGA1 may cooperate with the mutant p53 pathway in modulating breast cancer aggressiveness (Pegoraro et al., 2013). Among the common genes with Pin1/mutant p53 gene signature there was also CCNE2 (Girardini et al., 2011), for which we deeply analyzed its role.

Intriguingly, beside the known role of CCNE2 in cell cycle progression, growing evidences suggest an involvement of CCNE2 in breast cancer cell migration (Caldon and Musgrove, 2010; Desmedt et al., 2006). Indeed, CCNE2 is a component of three prognostic gene expression signatures that predict shorter metastasis-free survival (Sotiriou et al., 2006; van 't Veer et al., 2002; Wang et al., 2005). Moreover, in the last years it has become evident how cyclins play important cell cycle independent roles. In particular, cyclin D1 and cyclin A2 show a strong and opposite, involvement in migration through the regulation of RhoA pathway (Bendris et al., 2012; Li et al., 2006a; Li et al., 2006b). The present work extends the current understanding of the alternative roles of cyclins in migration. Indeed, the silencing of CCNE2 strongly impaired cell migration and invasion in basal-like breast cancer cells downstream to HMGA1. We demonstrated for the first time that HMGA1 regulates CCNE2

and together negatively modulate the Hippo pathway inducing the activation of the oncogene YAP identifying an HMGA1-CCNE2-YAP axis (Pegoraro et al., 2015).

Our results show that HMGA1 binds *in vivo* to CCNE2 promoter and activates CCNE2 transcription. Because of the nature of HMGA1, that has not *per se* transcriptional activity, we speculate that HMGA1 could activate transcription of CCNE2 in concert with others factors, such as E2F1. Notably, we found that HMGA1-gene signature is enriched for E2F-target genes. Indeed, previous works show that promoter region of CCNE2 contains *in vivo* binding sites for E2F and it can to be activated by E2F1 (Caldon et al., 2009). Moreover, it was demonstrated that HMGA1, binding pRb, activates E2F target genes (Ueda et al., 2007). In addition, we provide evidence that HMGA1 can bind to two AT-rich regions, AT-RR1 and AT-RR2, upstream the CCNE2 promoter, one of which, AT-RR2, is a potential S/MAR. HMGA1 is a nuclear architectural factor that can organize local chromatin structures regulating gene expression by binding to long AT-rich stretch and to S/MAR (Chavali et al., 2011; Kishi et al., 2012). Therefore, it is possible that AT-RR1 and AT-RR2 distal elements could regulate CCNE2 transcription by physically interacting with the CCNE2 proximal promoter through HMGA1-mediated chromatin looping.

Our results demonstrated that the inhibition of HMGA1 and CCNE2 induced YAP inactivation by promoting its cytoplasmic localization, and we showed that this effect is mediated by MST1/2 and LATS1/2, the core kinases of the Hippo pathway (Pegoraro et al., 2015). Hippo pathway is often deregulated in cancer and this causes a hyper activation of YAP with a consequent gain of aggressive properties (Piccolo et al., 2013). Intriguingly, we found a functional correlation between HMGA1 and CCNE2 expression and a specific post-translational modification on YAP: the phosphorylation on serine 127 (Pegoraro et al., 2015). It was demonstrated that this particular phosphorylation is strongly connected with metastatic potential. Indeed, cells mutated in this amino acid residue acquired the ability to metastasize *in vivo* once injected in mice (Lamar et al., 2012). We demonstrated that the migration abilities, impaired after HMGA1 and CCNE2 silencing, were rescued using an active form of YAP (YAP-5SA) that contain also the mutated form of S127. This result clearly underline that YAP lies downstream of HMGA1 and CCNE2. How these two proteins regulate YAP status is an open question. One of the major regulators of the Hippo pathway is represented by the cytoskeletal pathway, which acts through mechanotransduction processes dependent on the RHO/ROCK signaling (Low et al., 2014). We showed that HMGA1 has a strong effect on

the modulation of the actin cytoskeleton, the shape, and the size of breast cancer cells (Pegoraro et al., 2013). Therefore, we hypothesize that HMGA1 acts on YAP through the regulation of cytoskeletal pathway. Future studies will be done to unravel this point.

Several studies revealed that pharmacological inhibition of YAP activity represents an appealing anticancer strategy. However, to date, few small molecules that target the Hippo pathway have been discovered. Among these inhibitors, statins have very recently been shown to display a strong potential in targeting the malignant effects of YAP in cancer cells, repositioning this drugs from the cholesterol levels control to cancer treatment (Sorrentino et al., 2014). The uncovering of the molecular mechanisms underlying metastatic process is determinant for developing novel therapeutic agents. Here, we found a novel HMGA1/CCNE2/YAP axis that could be potentially targeted, thus, we decided to block YAP activity inhibiting CDKs. To avoid compensatory activity (Bashir and Pagano, 2005), we used CDK pan-inhibitors evaluating YAP localization and activity in basal-like breast cancer cells. Thus, our work highlights that several CDK inhibitors effectively inhibit YAP, decreasing its nuclear localization and activity and impairing cell migration even beside their primary activity in proliferation inhibition. An intensive search is on-going for possible therapeutic applications of CDK inhibitors; in fact, several CDK inhibitors are under investigation in clinical trials but the success of monotherapy on inhibition of solid tumor growth in human is not completely satisfactory (Cicenas et al., 2014; Criscitiello et al., 2014; Phelps et al., 2009). So, we suggest to investigate the efficacy of these compounds for their ability to reduce metastasis *in vivo*.

In conclusion, we demonstrated that (Fig. 1):

1. HMGA1 plays a critical role in breast cancer aggressiveness particularly in basal-like breast tumors modulating EMT, stem cell ability, migration, and invasion properties both *in vitro* and *in vivo.*

2. HMGA1 controls breast cancer aggressiveness by coordinating a complex gene expression program that include many genes involved in pathways that sustain malignant phenotype, such as Wnt, Notch, Pin1/mutant p53, and Hippo.

3. HMGA1-gene signature has a prognostic value that inversely correlates with distant metastasis free survival.

4. HMGA1-CCNE2-YAP axis regulates cell migration in basal-like breast cancer cells, ultimately sustaining metastatic properties.

5. CDK inhibitors target the Hippo pathway inactivating YAP activity and impairing cell migration.
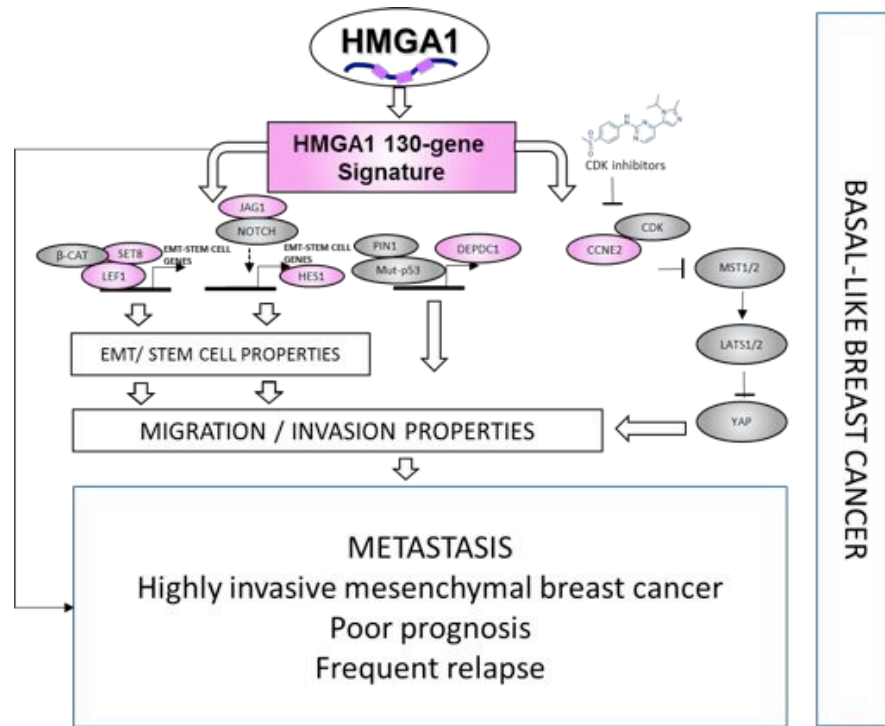
Figure 1. Graphical Abstract of HMGA1 Action in Basal-Like Breast Cancer Cells.

Appendix

Table 1. Up-regulated genes after HMGA1-silencing

Official Symbol	Affyprobe	Description	EntrezID	logFC	P.Value	adj.P.Val
ERBB4	214053_at	v-erb-a erythroblastic leukemia viral oncogene homolog 4 (avian)	2066	2.380185287	1.06E-08	9.92E-07
NAALAD2	1554507_at	N-acetylated alpha-linked acidic dipeptidase 2	10003	2.035759847	1.08E-07	4.31E-06
PTGS1	238669_at	prostaglandin-endoperoxide synthase 1 (prostaglandin G/H synthase and cyclooxygenase)	5742	1.984579405	1.64E-07	5.85E-06
CAPN8	229030_at	Calcium-regulated non-lysosomal thiol-protease	388743	1.940939582	2.06E-07	6.68E-06
LGR5	213880_at	leucine-rich repeat containing G protein-coupled receptor 5	8549	1.90338254	1.72E-08	1.29E-06
COL3A1	215076_s_at	collagen, type III, alpha 1	1281	1.866503099	1.91E-10	1.09E-07
SH2D3A	219513_s_at	SH2 domain containing 3A	10045	1.842985613	8.35E-10	2.17E-07
FABP4	203980_at	fatty acid binding protein 4, adipocyte	2167	1.806973448	1.03E-09	2.39E-07
TMEM156	241844_x_at	transmembrane protein 156	80008	1.806631177	5.72E-10	1.91E-07
DNAH12	243802_at	dynein, axonemal, heavy chain 12	201625	1.756020634	1.15E-10	8.68E-08
BCCIP	227896_at	BRCA2 and CDKN1A interacting protein	56647	1.736916494	1.99E-10	1.09E-07
GDPD1	238681_at	In multiple Geneids		1.704464339	1.36E-08	1.15E-06
KBTBD8	239835_at	kelch repeat and BTB (POZ) domain containing 8	84541	1.678145504	1.16E-09	2.56E-07
DPYD	1554536_at	dihydropyrimidine dehydrogenase	1806	1.674129559	4.23E-10	1.55E-07
ABCG2	209735_at	ATP-binding cassette, sub-family G (WHITE), member 2	9429	1.656780431	1.38E-11	3.91E-08
GKN2	238222_at	gastrokine 2	200504	1.642175935	1.01E-08	9.66E-07
AGXT2L1	221008_s_at	alanine-glyoxylate aminotransferase 2-like 1	64850	1.62274403	1.32E-07	4.95E-06
AKR1B10	206561_s_at	aldo-keto reductase family 1, member B10 (aldose reductase)	57016	1.604670181	1.09E-09	2.48E-07
PFKFB2	209992_at	6-phosphofructo-2-kinase/fructose-2,6-biphosphatase 2	5208	1.575324808	4.55E-09	6.01E-07
C9orf95	219147_s_at	chromosome 9 open reading frame 95	54981	1.558257031	5.19E-08	2.66E-06
LOC100505989	238103_at	uncharacterized LOC100505989	100505989	1.547521028	1.28E-09	2.64E-07
CHMP4C	226803_at	charged multivesicular body protein 4C	92421	1.528872554	6.88E-09	7.61E-07
TXNRD3	59631_at	thioredoxin reductase 3	114112	1.527761015	4.21E-08	2.31E-06
CBLB	227900_at	Cbl proto-oncogene, E3 ubiquitin protein ligase B	868	1.496649681	1.68E-09	3.19E-07
MAP7	202890_at	microtubule-associated protein 7	9053	1.493606946	4.41E-08	2.38E-06
SAMD5	228653_at	sterile alpha motif domain containing 5	389432	1.491671828	7.42E-09	7.94E-07
MAL2	224650_at	mal, T-cell differentiation protein 2 (gene/pseudogene)	114569	1.47734711	3.85E-08	2.19E-06
RELN	205923_at	reelin	5649	1.470064128	2.65E-08	1.67E-06
FAT3	236029_at	FAT tumor suppressor homolog 3 (Drosophila)	120114	1.456911606	6.49E-10	1.97E-07
PGAP1	244321_at	post-GPI attachment to proteins 1	80055	1.446829261	1.38E-08	1.15E-06
ESRP2	229223_at	epithelial splicing regulatory protein 2	80004	1.443871702	2.03E-07	6.61E-06
PPM1H	212686_at	protein phosphatase, Mg2+/Mn2+ dependent, 1H	57460	1.423044087	3.95E-09	5.48E-07
FAM83H	226129_at	family with sequence similarity 83, member H	286077	1.422178809	2.44E-10	1.14E-07
LOC100506713	229833_at	uncharacterized LOC100506713	100506713	1.422014448	5.82E-09	6.77E-07
PCDH7	205535_s_at	protocadherin 7	5099	1.404865074	2.87E-09	4.51E-07
ST20	217104_at	suppressor of tumorigenicity 20	400410	1.404617709	2.41E-07	7.51E-06
TMEM45A	219410_at	transmembrane protein 45A	55076	1.401958859	4.33E-08	2.35E-06
PPP2R1B	222351_at	protein phosphatase 2, regulatory subunit A, beta	5519	1.40007178	2.16E-09	3.79E-07

Table 2. Down-regulated genes after HMGA1-silencing

Official Symbol	Affyprobe	Description	EntrezID	logFC	P.Value	adj.P.Val
SERPINE1	202627_s_at	serpin peptidase inhibitor, clade E (nexin, plasminogen activator inhibitor type 1), member 1	5054	-2.92745006	1.85E-13	3.66E-09
WIPF1	202664_at	WAS/WASL interacting protein family, member 1	7456	-2.833101111	1.24E-09	2.64E-07
MLF1IP	218883_s_at	MLF1 interacting protein	79682	-2.73153912	6.77E-12	2.85E-08
TDRKH	223530_at	tudor and KH domain containing	11022	-2.633805277	1.61E-11	3.99E-08
MEST	202016_at	mesoderm specific transcript homolog (mouse)	4232	-2.532520446	2.40E-12	1.58E-08
CCDC34	226287_at	coiled-coil domain containing 34	91057	-2.509641199	2.48E-10	1.14E-07
HMGA1	206074_s_at	high mobility group AT-hook 1	3159	-2.417068055	5.07E-11	5.58E-08
LRRC59	222231_s_at	leucine rich repeat containing 59	55379	-2.339161281	1.48E-12	1.47E-08
KIF23	204709_s_at	kinesin family member 23	9493	-2.323486186	2.43E-11	3.99E-08
CCNE2	205034_at	cyclin E2	9134	-2.282315274	1.81E-09	3.32E-07
ADAM19	209765_at	ADAM metallopeptidase domain 19	8728	-2.280492217	8.64E-12	2.85E-08
RRP1B	212846_at	ribosomal RNA processing 1 homolog B (S. cerevisiae)	23076	-2.198079565	1.65E-10	1.04E-07
GXYLT1	226868_at	glucoside xylosyltransferase 1	283464	-2.170273821	8.02E-11	6.35E-08
MAGEC1	206609_at	melanoma antigen family C, 1	9947	-2.136955751	8.03E-12	2.85E-08
EIF5B	201025_at	eukaryotic translation initiation factor 5B	9669	-2.128508957	1.19E-10	8.68E-08
ZNF828	226194_at	zinc finger protein 828	283489	-2.125439949	1.28E-09	2.64E-07
MDC1	203062_s_at	mediator of DNA-damage checkpoint 1	9656	-2.120524504	3.50E-11	4.27E-08
CASP7	207181_s_at	caspase 7, apoptosis-related cysteine peptidase	840	-2.100628211	1.81E-10	1.09E-07
CENPF	207331_at	centromere protein F, 350/400kDa (mitosin)	1063	-2.050562616	7.31E-09	7.94E-07
ZBTB47	226484_at	zinc finger and BTB domain containing 47	92999	-2.023959077	7.04E-08	3.26E-06
NCAPG	218662_s_at	non-SMC condensin I complex, subunit G	64151	-2.023764249	2.53E-10	1.14E-07
SOAT1	228479_at	sterol O-acyltransferase 1	6646	-1.983926725	7.48E-11	6.32E-08
GAS2L3	238756_at	growth arrest-specific 2 like 3	283431	-1.974461913	1.16E-09	2.56E-07
AFAP1L1	226955_at	actin filament associated protein 1-like 1	134265	-1.965387551	2.10E-10	1.09E-07
RASSF4	226436_at	Ras association (RalGDS/AF-6) domain family member 4	83937	-1.942457697	3.04E-10	1.31E-07
SPC25	209891_at	SPC25, NDC80 kinetochore complex component, homolog (S. cerevisiae)	57405	-1.931358459	3.16E-09	4.78E-07
C19orf2	211563_s_at	chromosome 19 open reading frame 2	8725	-1.912984591	2.90E-11	3.99E-08
CENPB	212437_at	centromere protein B, 80kDa	1059	-1.897336736	6.55E-11	6.32E-08
SUPT16H	217815_at	suppressor of Ty 16 homolog (S. cerevisiae)	11198	-1.880214345	2.49E-11	3.99E-08
E2F8	219990_at	E2F transcription factor 8	79733	-1.879394048	1.13E-08	1.04E-06
WHSC1	222778_s_at	Wolf-Hirschhorn syndrome candidate 1	7468	-1.878342105	1.68E-10	1.04E-07
NES	218678_at	nestin	10763	-1.872610019	2.30E-09	3.99E-07
FAM54A	228069_at	family with sequence similarity 54, member A	113115	-1.868300214	5.00E-09	6.22E-07
TSPAN12	219274_at	tetraspanin 12	23554	-1.840765524	7.23E-07	1.70E-05
ANP32E	229128_s_at	acidic (leucine-rich) nuclear phosphoprotein 32 family, member E	81611	-1.83951562	4.74E-09	6.12E-07
C1orf115	218546_at	chromosome 1 open reading frame 115	79762	-1.825841333	2.34E-11	3.99E-08
TIMELESS	203046_s_at	timeless homolog (Drosophila)	8914	-1.820356525	3.12E-10	1.32E-07
FAM35A	220547_s_at	family with sequence similarity 35, member A	54537	-1.815722601	2.25E-11	3.99E-08
IL11	206924_at	interleukin 11	3589	-1.807289156	3.02E-11	3.99E-08
COPS2	209838_at	COP9 constitutive photomorphogenic homolog subunit 2 (Arabidopsis)	9318	-1.788078896	6.31E-10	1.97E-07
SSX2IP	203018_s_at	synovial sarcoma, X breakpoint 2 interacting protein	117178	-1.778722235	2.75E-09	4.41E-07
PIGK	227639_at	phosphatidylinositol glycan anchor biosynthesis, class K	10026	-1.770005578	1.44E-10	9.84E-08
FAM177A1	227029_at	family with sequence similarity 177, member A1	283635	-1.769033421	1.06E-09	2.43E-07
GFPT1	226886_at	glutamine--fructose-6-phosphate transaminase 1	2673	-1.767972576	2.37E-10	1.14E-07
POLR3GL	223269_at	polymerase (RNA) III (DNA directed) polypeptide G (32kD)-like	84265	-1.767564038	9.61E-10	2.35E-07
GNG11	204115_at	guanine nucleotide binding protein (G protein), gamma 11	2791	-1.766879065	7.35E-11	6.32E-08
TCP11L1	205796_at	t-complex 11 (mouse)-like 1	55346	-1.738701014	2.62E-09	4.32E-07
MAP4K4	222547_at	mitogen-activated protein kinase kinase kinase kinase 4	9448	-1.735678363	6.06E-11	6.31E-08
NOLC1	211951_at	nucleolar and coiled-body phosphoprotein 1	9221	-1.717742452	1.51E-10	9.96E-08
BRCA1	204531_s_at	breast cancer 1, early onset	672	-1.704112966	5.64E-10	1.91E-07
TCF19	223274_at	transcription factor 19	6941	-1.702485639	3.72E-10	1.46E-07
LYRM7	228841_at	Lyrm7 homolog (mouse)	90624	-1.682442844	2.77E-10	1.22E-07
DTL	218585_s_at	denticleless homolog (Drosophila)	51514	-1.682420851	1.05E-08	9.88E-07
MOCS2	218212_s_at	molybdenum cofactor synthesis 2	4338	-1.67785212	1.64E-09	3.19E-07
TGFBI	201506_at	transforming growth factor, beta-induced, 68kDa	7045	-1.676222196	9.83E-10	2.37E-07
LOC100289610	230466_s_at	similar to mesenchymal stem cell protein DSC96	100289610	-1.672985848	4.57E-07	1.19E-05
LOC729082	225225_at	hypothetical LOC729082	729082	-1.664537325	2.70E-11	3.99E-08
FBXO43	236852_at	F-box protein 43	286151	-1.661287561	9.34E-10	2.31E-07
CCDC43	225607_at	coiled-coil domain containing 43	124808	-1.660218814	5.27E-10	1.86E-07
TMED8	225343_at	transmembrane emp24 protein transport domain containing 8	283578	-1.658631963	2.17E-10	1.10E-07
LEF1	221558_s_at	lymphoid enhancer-binding factor 1	51176	-1.65648691	2.32E-10	1.14E-07
KIF4A	218355_at	kinesin family member 4A	24137	-1.653868232	7.66E-11	6.32E-08
C2orf47	219176_at	chromosome 2 open reading frame 47	79568	-1.6495774	7.59E-11	6.32E-08
DDB2	203409_at	damage-specific DNA binding protein 2, 48kDa	1643	-1.640760877	3.29E-10	1.36E-07
PSMC3IP	213951_s_at	PSMC3 interacting protein	29893	-1.637393952	6.45E-09	7.23E-07
CYP24A1	206504_at	cytochrome P450, family 24, subfamily A, polypeptide 1	1591	-1.637082646	6.76E-10	1.97E-07
LIFR	225571_at	leukemia inhibitory factor receptor alpha	3977	-1.630985327	4.11E-07	1.11E-05
FAM122B	225361_x_at	family with sequence similarity 122B	159090	-1.623960448	1.09E-08	9.87E-07
PSMD2	200830_at	proteasome (prosome, macropain) 26S subunit, non-ATPase, 2	5708	-1.618085913	3.66E-11	4.27E-08
MKI67	212020_s_at	antigen identified by monoclonal antibody Ki-67	4288	-1.617640945	8.77E-08	3.76E-06
AURKB	209464_at	aurora kinase B	9212	-1.603190384	6.49E-08	3.09E-06
CAMK4	241871_at	calcium/calmodulin-dependent protein kinase IV	814	-1.594704257	1.87E-07	6.25E-06
SRSF8	213152_s_at	serine/arginine-rich splicing factor 8	10929	-1.59341108	3.83E-10	1.46E-07
LRR1	235113_at	leucine rich repeat protein 1	122769	-1.578999285	3.47E-10	1.40E-07
GTSE1	204318_s_at	G-2 and S-phase expressed 1	51512	-1.574911458	5.38E-09	6.55E-07
MNS1	219703_at	meiosis-specific nuclear structural 1	55329	-1.571351138	8.23E-09	8.36E-07
ZC3HAV1L	228280_at	zinc finger CCCH-type, antiviral 1-like	92092	-1.568719968	4.19E-09	5.68E-07
TIMP4	206243_at	TIMP metallopeptidase inhibitor 4	7079	-1.563456414	3.44E-09	5.04E-07
EXO1	204603_at	exonuclease 1	9156	-1.560775177	1.77E-09	3.30E-07
ZCCHC10	221193_s_at	zinc finger, CCHC domain containing 10	54819	-1.558270648	3.68E-08	2.12E-06
NCAPG2	219588_s_at	non-SMC condensin II complex, subunit G2	54892	-1.558117721	1.69E-09	3.19E-07
ADPGK	220980_s_at	ADP-dependent glucokinase	83440	-1.55774154	1.23E-10	8.68E-08
ZNF367	229551_x_at	zinc finger protein 367	195828	-1.553632794	1.38E-08	1.15E-06
CDCA7	224428_s_at	cell division cycle associated 7	83879	-1.551535689	4.82E-09	6.16E-07
NEGR1	229461_x_at	neuronal growth regulator 1	257194	-1.55126315	3.15E-08	1.89E-06
C9orf100	231892_at	chromosome 9 open reading frame 100	84904	-1.551221833	1.96E-07	6.47E-06
PPP1R3E	227409_at	protein phosphatase 1, regulatory (inhibitor) subunit 3E	90673	-1.550742385	1.71E-08	1.29E-06
WDR76	239680_at	WD repeat domain 76	79968	-1.545411663	4.80E-07	1.21E-05
WDHD1	204727_at	WD repeat and HMG-box DNA binding protein 1	11169	-1.537321579	2.26E-07	7.13E-06
SKA3	227165_at	spindle and kinetochore associated complex subunit 3	221150	-1.532760644	5.50E-09	6.64E-07
KIAA1715	229173_at	KIAA1715	80856	-1.531927195	4.69E-09	6.11E-07
MYBL2	201710_at	v-myb myeloblastosis viral oncogene homolog (avian)-like 2	4605	-1.526612311	6.60E-10	1.97E-07
VEZT	223089_at	vezatin, adherens junctions transmembrane protein	55591	-1.517905739	5.83E-08	2.87E-06
CISD3	226727_at	CDGSH iron sulfur domain 3	284106	-1.507657423	2.35E-09	4.05E-07
DNAJC21	235032_at	DnaJ (Hsp40) homolog, subfamily C, member 21	134218	-1.507652709	5.16E-08	2.66E-06
CALR	200935_at	calreticulin	811	-1.50737424	1.06E-06	2.28E-05
TCEA3	226388_at	transcription elongation factor A (SII), 3	6920	-1.505630403	8.67E-10	2.20E-07
HIPK2	225368_at	homeodomain interacting protein kinase 2	28996	-1.504954063	6.35E-10	1.97E-07
GPR17	206190_at	G protein-coupled receptor 17	2840	-1.498596832	4.65E-08	2.46E-06
APOBEC3B	206632_s_at	apolipoprotein B mRNA editing enzyme, catalytic polypeptide-like 3B	9582	-1.496119423	3.42E-09	5.04E-07
KIAA1549	223575_at	KIAA1549	57670	-1.49581465	5.29E-08	2.69E-06
STK17A	202695_s_at	serine/threonine kinase 17a	9263	-1.482091367	1.61E-09	3.19E-07
ATXN1	203231_s_at	ataxin 1	6310	-1.477916494	3.72E-07	1.03E-05
PNMAL1	218824_at	PNMA-like 1	55228	-1.4776635	2.28E-08	1.52E-06
ILF2	200052_s_at	interleukin enhancer binding factor 2, 45kDa	3608	-1.475782696	3.46E-09	5.04E-07
ABLIM3	205730_s_at	actin binding LIM protein family, member 3	22885	-1.474044603	2.41E-09	4.09E-07
HBEGF	203821_at	heparin-binding EGF-like growth factor	1839	-1.473764956	2.77E-09	4.41E-07
PELP1	215354_s_at	proline, glutamate and leucine rich protein 1	27043	-1.47040906	1.61E-08	1.25E-06

Official Symbol	Affyprobe	Description	EntrezID	logFC	P.Value	adj.P.Val
GPS2	209350_s_at	G protein pathway suppressor 2	2874	-1.455556427	1.59E-07	5.74E-06
TOP2A	201292_at	topoisomerase (DNA) II alpha 170kDa	7153	-1.452439515	1.66E-09	3.19E-07
PROM2	1552797_s_at	prominin 2	150696	-1.451831812	1.94E-07	6.41E-06
CDCA8	221520_s_at	cell division cycle associated 8	55143	-1.448846288	3.83E-10	1.46E-07
RNPS1	200060_s_at	RNA binding protein S1, serine-rich domain	10921	-1.444471478	1.57E-08	1.22E-06
PGPEP1	219891_at	pyroglutamyl-peptidase I	54858	-1.443639258	4.76E-09	6.12E-07
ATP11B	1554557_at	ATPase, class VI, type 11B	23200	-1.443549784	2.27E-08	1.52E-06
DEPDC1	235545_at	DEP domain containing 1	55635	-1.441381887	1.56E-08	1.22E-06
NUP210	213947_s_at	nucleoporin 210kDa	23225	-1.439888558	1.27E-08	1.11E-06
TMEM194A	212621_at	transmembrane protein 194A	23306	-1.438861369	2.08E-09	3.71E-07
AVEN	219366_at	apoptosis, caspase activation inhibitor	57099	-1.433457027	4.16E-10	1.55E-07
UHRF1	225655_at	ubiquitin-like with PHD and ring finger domains 1	29128	-1.433282926	1.82E-09	3.32E-07
SPTBN1	226342_at	spectrin, beta, non-erythrocytic 1	6711	-1.425729549	1.54E-09	3.13E-07
GNA12	224681_at	guanine nucleotide binding protein (G protein) alpha 12	2768	-1.425275633	7.17E-10	2.00E-07
TOR3A	218459_at	torsin family 3, member A	64222	-1.423937706	4.95E-09	6.22E-07
RNF213	232155_at	ring finger protein 213	57674	-1.418960672	2.85E-08	1.76E-06
C1orf85	1558692_at	chromosome 1 open reading frame 85	112770	-1.415335639	8.46E-09	8.50E-07
BARD1	227545_at	BRCA1 associated RING domain 1	580	-1.410797549	3.46E-09	5.04E-07
MYEF2	229464_at	myelin expression factor 2	50804	-1.407647644	6.67E-08	3.16E-06
TPX2	210052_s_at	TPX2, microtubule-associated, homolog (Xenopus laevis)	22974	-1.40700666	3.46E-07	9.73E-06
PLRG1	227246_at	pleiotropic regulator 1 (PRL1 homolog, Arabidopsis)	5356	-1.405257857	8.06E-10	2.16E-07
C16orf53	218300_at	chromosome 16 open reading frame 53	79447	-1.404487743	2.68E-08	1.69E-06

Table3. Functional annotation of HMGA1 modulated-genes.

DOWN regulated genes upon siHMGA1 Functional Analysis	
DAVID	p-value
cell cycle	8.49E-12
M phase	6.40E-09
condensed chromosome	1.57E-06
nuclear lumen	2.10E-06
nuclear division	1.23E-05
mitosis	1.23E-05
Ingenuity Pathway Analysis	p-value
Cancer	4.56E-08 - 3.84E-02
Reproductive System Disease	2.25E-04 - 3.84E-02
Cell Cycle	3.48E-13 - 3.68E-02
DNA replication and Repair	1.45E-08 - 3.30E-02
Embryonic Development	2.25E-05 - 3.84E-02
Hematological System Development	3.04E-04 - 3.30E-02

UP regulated genes upon siHMGA1 Functional Analysis	
DAVID	p-value
regulation of kinase activity	0.031
domain:Cadherin 7	0.034
regulation of transferase activity	0.034
FAD-dependent pyridine nucleotidedisulphide oxidoreductase	0.038
EGF	0.043
Ingenuity Pathway Analysis	p-value
Cancer	3.92E-06 - 1.73E-02
Gastrointestinal Disease	3.92E-06 - 1.73E-02
Cellular Movement	3.32E-05 - 1.73E-02
Cellular Growth and Proliferation	4.88E-05 - 1.73E-02
Lipid Metabolism	1.71E-04 - 1.73E-02
Cell-Mediated Immune Response	3.32E-05 - 1.16E-02

Table 4. Oncomine analysis of HMGA1 regulated genes

Cancer Type	Cancer vs. Normal		Clinical Outcome	
	Over-expression or Copy Gain	Under-expression or Copy Loss	Over-expression or Copy Gain	Under-expression or Copy Loss
Bladder Cancer	2		1	
Brain and CNS Cancer	5	1	4	
Breast Cancer	8	1	31	
Cervical Cancer	2			
Colorectal Cancer	14		2	1
Esophageal Cancer	3			
Gastric Cancer	4			
Head and Neck Cancer	10			
Kidney Cancer			2	
Leukemia	1	3		
Liver Cancer	4			
Lung Cancer	13		4	
Lymphoma			3	3
Melanoma	1			
Myeloma			5	
Other Cancer	5	1	2	
Ovarian Cancer	4			
Pancreatic Cancer	2			
Prostate Cancer	2		1	
Sarcoma	10			
Significant Unique Concepts	89	6	55	4
Ratio	14.83 (p<0.0007)		13.75(p<0.009)	

Table 5. Multivariate analysis 1

Clinical variable	HR	se(coef)	z	p-value
ER status (negative)	1.457	0.215	1.749	0.015
Linfonode (negative)	0.624	0.197	-2.394	0.0011
Size	3.447	0.25	4.945	2^{-09}
Age	1.034	0.104	0.322	0.57
HMGA1 signature	1.574	0.195	2.326	0.019

Table 6. Multivariate analysis 2

Figure Supplementary 3: Multivariate analysis

Clinical variable	HR	se(coef)	z	p-value
ER status (negative)	1.457	0.215	1.749	0.015
Linfonode (negative)	0.624	0.197	-2.394	0.0011
Size	3.447	0.25	4.945	2^{-09}
Age	1.034	0.104	0.322	0.57

Table 7. CDK inhibitors selected from Selleckchem company

Compound	Other Name	CDK high	CDK low	FDA
PD-0332991	Pfizer's Palbociclib	4, 5		Breakthrough Therapy designation
SNS-032	BMS-387032	9, 7, 2	1, 4, 6	TRIAL
Roscovitine	seliciclib	1, 2, 5		TRIAL
PHA-793887		2, 5, 7	1, 4, 9	TRIAL
AT7519		1, 2, 4, 6, 9	3, 7	TRIAL
BS-181 HCL		7,,	1, 2, 4, 5, 6, 9	NO
BMS-265246		1, 2	4,,	NO
AZD5438		1, 2, 9	5, 6, GSK2B	TRIAL
Flavopiridol hydrochloride		1, 2, 4, 6		TRIAL
PHA-767491		7, 9	1, 2, GSK2B	NO
PHA-848125	Milciclib	2, CCNA,	7, 4, 5 TRKA	TRIAL
Dinaciclib	SCH727965	2, 5, 1, 9		TRIAL
JNJ-7706621		1, 2, AURK A/B		NO

References

Adair, J.E., Kwon, Y., Dement, G.A., Smerdon, M.J., and Reeves, R. (2005). Inhibition of nucleotide excision repair by high mobility group protein HMGA1. J Biol Chem *280*, 32184-32192.

Ahmad, A. (2013). Pathways to breast cancer recurrence. ISRN Oncol *2013*, 290568.

Arranz, E.E., Vara, J.A., Gamez-Pozo, A., and Zamora, P. (2012). Gene signatures in breast cancer: current and future uses. Transl Oncol *5*, 398-403.

Badve, S., Dabbs, D.J., Schnitt, S.J., Baehner, F.L., Decker, T., Eusebi, V., Fox, S.B., Ichihara, S., Jacquemier, J., Lakhani, S.R.*, et al.* (2011). Basal-like and triple-negative breast cancers: a critical review with an emphasis on the implications for pathologists and oncologists. Modern pathology : an official journal of the United States and Canadian Academy of Pathology, Inc *24*, 157-167.

Baldassarre, G., Battista, S., Belletti, B., Thakur, S., Pentimalli, F., Trapasso, F., Fedele, M., Pierantoni, G., Croce, C.M., and Fusco, A. (2003). Negative regulation of BRCA1 gene expression by HMGA1 proteins accounts for the reduced BRCA1 protein levels in sporadic breast carcinoma. Mol Cell Biol *23*, 2225-2238.

Banerjee, R., Russo, N., Liu, M., Basrur, V., Bellile, E., Palanisamy, N., Scanlon, C.S., van Tubergen, E., Inglehart, R.C., Metwally, T.*, et al.* (2014). TRIP13 promotes error-prone nonhomologous end joining and induces chemoresistance in head and neck cancer. Nature communications *5*, 4527.

Bashir, T., and Pagano, M. (2005). Cdk1: the dominant sibling of Cdk2. Nat Cell Biol *7*, 779-781.

Bendris, N., Arsic, N., Lemmers, B., and Blanchard, J.M. (2012). Cyclin A2, Rho GTPases and EMT. Small GTPases *3*, 225-228.

Borrmann, L., Schwanbeck, R., Heyduk, T., Seebeck, B., Rogalla, P., Bullerdiek, J., and Wisniewski, J.R. (2003). High mobility group A2 protein and its derivatives bind a specific region of the promoter of DNA repair gene ERCC1 and modulate its activity. Nucleic Acids Res *31*, 6841-6851.

Breckpot, K., Dullaers, M., Bonehill, A., van Meirvenne, S., Heirman, C., de Greef, C., van der Bruggen, P., and Thielemans, K. (2003). Lentivirally transduced dendritic cells as a tool for cancer immunotherapy. J Gene Med *5*, 654-667.

Bush, B.M., Brock, A.T., Deng, J.A., Nelson, R.A., and Sumter, T.F. (2013). The Wnt/beta-catenin/T-cell factor 4 pathway up-regulates high-mobility group A1 expression in colon cancer. Cell Biochem Funct *31*, 228-236.

Bussemakers, M.J., van de Ven, W.J., Debruyne, F.M., and Schalken, J.A. (1991). Identification of high mobility group protein I(Y) as potential progression marker for prostate cancer by differential hybridization analysis. Cancer Res *51*, 606-611.

Caldon, C.E., and Musgrove, E.A. (2010). Distinct and redundant functions of cyclin E1 and cyclin E2 in development and cancer. Cell Div *5*, 2.

Caldon, C.E., Sergio, C.M., Schutte, J., Boersma, M.N., Sutherland, R.L., Carroll, J.S., and Musgrove, E.A. (2009). Estrogen regulation of cyclin E2 requires cyclin D1 but not c-Myc. Mol Cell Biol *29*, 4623-4639.

Caldon, C.E., Sergio, C.M., Sutherland, R.L., and Musgrove, E.A. (2013). Differences in degradation lead to asynchronous expression of cyclin E1 and cyclin E2 in cancer cells. Cell Cycle *12*, 596-605.

Casimiro, M.C., Crosariol, M., Loro, E., Li, Z., and Pestell, R.G. (2012). Cyclins and cell cycle control in cancer and disease. Genes & cancer *3*, 649-657.

Chavali, P.L., Funa, K., and Chavali, S. (2011). Cis-regulation of microRNA expression by scaffold/matrix-attachment regions. Nucleic Acids Res *39*, 6908-6918.

Chiappetta, G., Avantaggiato, V., Visconti, R., Fedele, M., Battista, S., Trapasso, F., Merciai, B.M., Fidanza, V., Giancotti, V., Santoro, M.*, et al.* (1996). High level expression of the HMGI (Y) gene during embryonic development. Oncogene *13*, 2439-2446.

Chiappetta, G., Botti, G., Monaco, M., Pasquinelli, R., Pentimalli, F., Di Bonito, M., D'Aiuto, G., Fedele, M., Iuliano, R., Palmieri, E.A.*, et al.* (2004). HMGA1 protein overexpression in human breast carcinomas: correlation with ErbB2 expression. Clin Cancer Res *10*, 7637-7644.

Chiappetta, G., Ottaiano, A., Vuttariello, E., Monaco, M., Galdiero, F., Gallipoli, A., Pilotti, S., Jodice, G., Siranoush, M., Colombo, M.*, et al.* (2010). HMGA1 protein expression in familial breast carcinoma patients. Eur J Cancer *46*, 332-339.

Chibon, F. (2013). Cancer gene expression signatures - the rise and fall? Eur J Cancer *49*, 2000-2009.

Chin, M.T., Pellacani, A., Wang, H., Lin, S.S., Jain, M.K., Perrella, M.A., and Lee, M.E. (1998). Enhancement of serum-response factor-dependent transcription and DNA binding by the architectural transcription factor HMG-I(Y). J Biol Chem *273*, 9755-9760.

Cicenas, J., Kalyan, K., Sorokinas, A., Jatulyte, A., Valiunas, D., Kaupinis, A., and Valius, M. (2014). Highlights of the Latest Advances in Research on CDK Inhibitors. Cancers (Basel) *6*, 2224-2242.

Cleynen, I., Huysmans, C., Sasazuki, T., Shirasawa, S., Van de Ven, W., and Peeters, K. (2007). Transcriptional control of the human high mobility group A1 gene: basal and oncogenic Ras-regulated expression. Cancer Res *67*, 4620-4629.

Cleynen, I., and Van de Ven, W.J. (2008). The HMGA proteins: a myriad of functions (Review). Int J Oncol *32*, 289-305.

Cohen, B., Shimizu, M., Izrailit, J., Ng, N.F., Buchman, Y., Pan, J.G., Dering, J., and Reedijk, M. (2010). Cyclin D1 is a direct target of JAG1-mediated Notch signaling in breast cancer. Breast Cancer Res Treat *123*, 113-124.

Colombo, E., Marine, J.C., Danovi, D., Falini, B., and Pelicci, P.G. (2002). Nucleophosmin regulates the stability and transcriptional activity of p53. Nat Cell Biol *4*, 529-533.

Cooper, G.M. (2000). The cell : a molecular approach, 2nd edn (Washington, D.C.Sunderland, Mass.: ASM Press ;Sinauer Associates).

Creighton, C.J., Chang, J.C., and Rosen, J.M. (2010). Epithelial-mesenchymal transition (EMT) in tumor-initiating cells and its clinical implications in breast cancer. J Mammary Gland Biol Neoplasia *15*, 253-260.

Criscitiello, C., Viale, G., Esposito, A., and Curigliano, G. (2014). Dinaciclib for the treatment of breast cancer. Expert Opin Investig Drugs *23*, 1305-1312.

D'Angelo, D., Palmieri, D., Mussnich, P., Roche, M., Wierinckx, A., Raverot, G., Fedele, M., Croce, C.M., Trouillas, J., and Fusco, A. (2012). Altered microRNA expression profile in human pituitary GH adenomas: down-regulation of miRNA targeting HMGA1, HMGA2, and E2F1. The Journal of clinical endocrinology and metabolism *97*, E1128-1138.

Debies, M.T., Gestl, S.A., Mathers, J.L., Mikse, O.R., Leonard, T.L., Moody, S.E., Chodosh, L.A., Cardiff, R.D., and Gunther, E.J. (2008). Tumor escape in a Wnt1-dependent mouse breast cancer model is enabled by p19Arf/p53 pathway lesions but not p16 Ink4a loss. J Clin Invest *118*, 51-63.

DeNardo, D.G., Andreu, P., and Coussens, L.M. (2010). Interactions between lymphocytes and myeloid cells regulate pro- versus anti-tumor immunity. Cancer metastasis reviews *29*, 309-316.

Dennis, G., Jr., Sherman, B.T., Hosack, D.A., Yang, J., Gao, W., Lane, H.C., and Lempicki, R.A. (2003). DAVID: Database for Annotation, Visualization, and Integrated Discovery. Genome Biol *4*, P3.

Desmedt, C., Ouriaghli, F.E., Durbecq, V., Soree, A., Colozza, M.A., Azambuja, E., Paesmans, M., Larsimont, D., Buyse, M., Harris, A*., et al.* (2006). Impact of cyclins E, neutrophil elastase and proteinase 3 expression levels on clinical outcome in primary breast cancer patients. Int J Cancer *119*, 2539-2545.

Dolde, C.E., Mukherjee, M., Cho, C., and Resar, L.M. (2002). HMG-I/Y in human breast cancer cell lines. Breast Cancer Res Treat *71*, 181-191.

Elliott, A., Adams, J., and Al-Hajj, M. (2010). The ABCs of cancer stem cell drug resistance. IDrugs *13*, 632-635.

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

Esposito, F., Tornincasa, M., Chieffi, P., De Martino, I., Pierantoni, G.M., and Fusco, A. (2010). High-mobility group A1 proteins regulate p53-mediated transcription of Bcl-2 gene. Cancer Res *70*, 5379-5388.

Evans, A., Lennard, T.W., and Davies, B.R. (2004). High-mobility group protein 1(Y): metastasis-associated or metastasis-inducing? J Surg Oncol *88*, 86-99.

Fedele, M., Bandiera, A., Chiappetta, G., Battista, S., Viglietto, G., Manfioletti, G., Casamassimi, A., Santoro, M., Giancotti, V., and Fusco, A. (1996). Human colorectal carcinomas express high levels of high mobility group HMGI(Y) proteins. Cancer Res *56*, 1896-1901.

Fedele, M., Fidanza, V., Battista, S., Pentimalli, F., Klein-Szanto, A.J., Visone, R., De Martino, I., Curcio, A., Morisco, C., Del Vecchio, L*., et al.* (2006a). Haploinsufficiency of the Hmga1 gene causes cardiac hypertrophy and myelo-lymphoproliferative disorders in mice. Cancer Res *66*, 2536-2543.

Fedele, M., Visone, R., De Martino, I., Troncone, G., Palmieri, D., Battista, S., Ciarmiello, A., Pallante, P., Arra, C., Melillo, R.M*., et al.* (2006b). HMGA2 induces pituitary tumorigenesis by enhancing E2F1 activity. Cancer Cell *9*, 459-471.

Federico, A., Forzati, F., Esposito, F., Arra, C., Palma, G., Barbieri, A., Palmieri, D., Fedele, M., Pierantoni, G.M., De Martino, I.*, et al.* (2014). Hmga1/Hmga2 double knock-out mice display a "superpygmy" phenotype. Biol Open *3*, 372-378.

Fischer, K.R., Durrans, A., Lee, S., Sheng, J., Li, F., Wong, S.T., Choi, H., El Rayes, T., Ryu, S., Troeger, J.*, et al.* (2015). Epithelial-to-mesenchymal transition is not required for lung metastasis but contributes to chemoresistance. Nature *527*, 472-476.

Flohr, A.M., Rogalla, P., Bonk, U., Puettmann, B., Buerger, H., Gohla, G., Packeisen, J., Wosniok, W., Loeschke, S., and Bullerdiek, J. (2003). High mobility group protein HMGA1 expression in breast cancer reveals a positive correlation with tumour grade. Histol Histopathol *18*, 999-1004.

Foulkes, W.D., Smith, I.E., and Reis-Filho, J.S. (2010). Triple-negative breast cancer. N Engl J Med *363*, 1938-1948.

Franco-Barraza, J., Valdivia-Silva, J.E., Zamudio-Meza, H., Castillo, A., Garcia-Zepeda, E.A., Benitez-Bribiesca, L., and Meza, I. (2010). Actin cytoskeleton participation in the onset of IL-1beta induction of an invasive mesenchymal-like phenotype in epithelial MCF-7 cells. Arch Med Res *41*, 170-181.

Frasca, F., Rustighi, A., Malaguarnera, R., Altamura, S., Vigneri, P., Del Sal, G., Giancotti, V., Pezzino, V., Vigneri, R., and Manfioletti, G. (2006). HMGA1 inhibits the function of p53 family members in thyroid cancer cells. Cancer Res *66*, 2980-2989.

Friedl, P., and Wolf, K. (2008). Tube travel: the role of proteases in individual and collective cancer cell invasion. Cancer Res *68*, 7247-7249.

Friedl, P., and Wolf, K. (2010). Plasticity of cell migration: a multiscale tuning model. J Cell Biol *188*, 11-19.

Fusco, A., and Fedele, M. (2007). Roles of HMGA proteins in cancer. Nat Rev Cancer *7*, 899-910.

Gaggioli, C. (2008). Collective invasion of carcinoma cells: when the fibroblasts take the lead. Cell adhesion & migration *2*, 45-47.

Gannon, L.M., Cotter, M.B., and Quinn, C.M. (2013). The classification of invasive carcinoma of the breast. Expert review of anticancer therapy *13*, 941-954.

Giancotti, V., Pani, B., D'Andrea, P., Berlingieri, M.T., Di Fiore, P.P., Fusco, A., Vecchio, G., Crane-Robinson, C., and Goodwin, G.H. (1987). Histone and nonhistone proteins from normal and virus-transformed rat thyroid epithelial cells. Basic Appl Histochem *31*, 229-238.

Giannini, G., Cerignoli, F., Mellone, M., Massimi, I., Ambrosi, C., Rinaldi, C., and Gulino, A. (2005). Molecular mechanism of HMGA1 deregulation in human neuroblastoma. Cancer Lett *228*, 97-104.

Gingras, I., Desmedt, C., Ignatiadis, M., and Sotiriou, C. (2015). CCR 20th Anniversary Commentary: Gene-Expression Signature in Breast Cancer--Where Did It Start and Where Are We Now? Clin Cancer Res *21*, 4743-4746.

Girardini, J.E., Napoli, M., Piazza, S., Rustighi, A., Marotta, C., Radaelli, E., Capaci, V., Jordan, L., Quinlan, P., Thompson, A.*, et al.* (2011). A Pin1/mutant p53 axis promotes aggressiveness in breast cancer. Cancer Cell *20*, 79-91.

Glas, A.M., Floore, A., Delahaye, L.J., Witteveen, A.T., Pover, R.C., Bakx, N., Lahti-Domenici, J.S., Bruinsma, T.J., Warmoes, M.O., Bernards, R.*, et al.* (2006). Converting a

breast cancer microarray signature into a high-throughput diagnostic test. BMC Genomics *7*, 278.

Gomez, M., Gomez, V., and Hergovich, A. (2014). The Hippo pathway in disease and therapy: cancer and beyond. Clin Transl Med *3*, 22.

Grivennikov, S.I., and Karin, M. (2010). Inflammation and oncogenesis: a vicious connection. Current opinion in genetics & development *20*, 65-71.

Grudzien, P., Lo, S., Albain, K.S., Robinson, P., Rajan, P., Strack, P.R., Golde, T.E., Miele, L., and Foreman, K.E. (2010). Inhibition of Notch signaling reduces the stem-like population of breast cancer cells and prevents mammosphere formation. Anticancer Res *30*, 3853-3867.

Halder, G., Dupont, S., and Piccolo, S. (2012). Transduction of mechanical and cytoskeletal cues by YAP and TAZ. Nature reviews Molecular cell biology *13*, 591-600.

Hanahan, D., and Weinberg, R.A. (2000). The hallmarks of cancer. Cell *100*, 57-70.

Hanahan, D., and Weinberg, R.A. (2011). Hallmarks of cancer: the next generation. Cell *144*, 646-674.

Harris, L.G., Pannell, L.K., Singh, S., Samant, R.S., and Shevde, L.A. (2012). Increased vascularity and spontaneous metastasis of breast cancer by hedgehog signaling mediated upregulation of cyr61. Oncogene *31*, 3370-3380.

Harrison, H., Farnie, G., Howell, S.J., Rock, R.E., Stylianou, S., Brennan, K.R., Bundred, N.J., and Clarke, R.B. (2010). Regulation of breast cancer stem cell activity by signaling through the Notch4 receptor. Cancer Res *70*, 709-718.

Hillion, J., Dhara, S., Sumter, T.F., Mukherjee, M., Di Cello, F., Belton, A., Turkson, J., Jaganathan, S., Cheng, L., Ye, Z., *et al.* (2008). The high-mobility group A1a/signal transducer and activator of transcription-3 axis: an achilles heel for hematopoietic malignancies? Cancer Res *68*, 10121-10127.

Hochberg, Y., and Benjamini, Y. (1990). More powerful procedures for multiple significance testing. Stat Med *9*, 811-818.

Hogan, N.M., Joyce, M.R., Murphy, J.M., Barry, F.P., O'Brien, T., Kerin, M.J., and Dwyer, R.M. (2013). Impact of mesenchymal stem cell secreted PAI-1 on colon cancer cell migration and proliferation. Biochem Biophys Res Commun *435*, 574-579.

Holth, L.T., Thorlacius, A.E., and Reeves, R. (1997). Effects of epidermal growth factor and estrogen on the regulation of the HMG-I/Y gene in human mammary epithelial cell lines. DNA Cell Biol *16*, 1299-1309.

Huang, J., Wu, S., Barrera, J., Matthews, K., and Pan, D. (2005). The Hippo signaling pathway coordinately regulates cell proliferation and apoptosis by inactivating Yorkie, the Drosophila Homolog of YAP. Cell *122*, 421-434.

Ingham, P.W., and McMahon, A.P. (2001). Hedgehog signaling in animal development: paradigms and principles. Genes Dev *15*, 3059-3087.

Irizarry, R.A., Bolstad, B.M., Collin, F., Cope, L.M., Hobbs, B., and Speed, T.P. (2003). Summaries of Affymetrix GeneChip probe level data. Nucleic Acids Res *31*, e15.

Jang, B.G., Lee, B.L., and Kim, W.H. (2015). Prognostic significance of leucine-rich-repeat-containing G-protein-coupled receptor 5, an intestinal stem cell marker, in gastric carcinomas. Gastric Cancer.

Johnson, R., and Halder, G. (2014). The two faces of Hippo: targeting the Hippo pathway for regenerative medicine and cancer treatment. Nat Rev Drug Discov *13*, 63-79.

Jones, R.G., and Thompson, C.B. (2009). Tumor suppressors and cell metabolism: a recipe for cancer growth. Genes Dev *23*, 537-548.

Kaddar, T., Rouault, J.P., Chien, W.W., Chebel, A., Gadoux, M., Salles, G., Ffrench, M., and Magaud, J.P. (2009). Two new miR-16 targets: caprin-1 and HMGA1, proteins implicated in cell proliferation. Biol Cell *101*, 511-524.

Kaklamani, V. (2006). A genetic signature can predict prognosis and response to therapy in breast cancer: Oncotype DX. Expert Rev Mol Diagn *6*, 803-809.

Kalluri, R., and Neilson, E.G. (2003). Epithelial-mesenchymal transition and its implications for fibrosis. J Clin Invest *112*, 1776-1784.

Kamran, M.Z., Patil, P., and Gude, R.P. (2013). Role of STAT3 in cancer metastasis and translational advances. BioMed research international *2013*, 421821.

Kenny, P.A., Lee, G.Y., Myers, C.A., Neve, R.M., Semeiks, J.R., Spellman, P.T., Lorenz, K., Lee, E.H., Barcellos-Hoff, M.H., Petersen, O.W.*, et al.* (2007). The morphologies of breast cancer cell lines in three-dimensional assays correlate with their profiles of gene expression. Mol Oncol *1*, 84-96.

Kent, S., Hutchinson, J., Balboni, A., Decastro, A., Cherukuri, P., and Direnzo, J. (2011). DeltaNp63alpha promotes cellular quiescence via induction and activation of Notch3. Cell Cycle *10*, 3111-3118.

Kishi, Y., Fujii, Y., Hirabayashi, Y., and Gotoh, Y. (2012). HMGA regulates the global chromatin state and neurogenic potential in neocortical precursor cells. Nature neuroscience *15*, 1127-1133.

Komiya, Y., and Habas, R. (2008). Wnt signal transduction pathways. Organogenesis *4*, 68-75.

Kopan, R., and Ilagan, M.X. (2009). The canonical Notch signaling pathway: unfolding the activation mechanism. Cell *137*, 216-233.

Lamar, J.M., Stern, P., Liu, H., Schindler, J.W., Jiang, Z.G., and Hynes, R.O. (2012). The Hippo pathway target, YAP, promotes metastasis through its TEAD-interaction domain. Proc Natl Acad Sci U S A *109*, E2441-2450.

Lamb, R., Ablett, M.P., Spence, K., Landberg, G., Sims, A.H., and Clarke, R.B. (2013). Wnt pathway activity in breast cancer sub-types and stem-like cells. PLoS One *8*, e67811.

Lamouille, S., Xu, J., and Derynck, R. (2014). Molecular mechanisms of epithelial-mesenchymal transition. Nature reviews Molecular cell biology *15*, 178-196.

Larrea, M.D., Wander, S.A., and Slingerland, J.M. (2009). p27 as Jekyll and Hyde: regulation of cell cycle and cell motility. Cell Cycle *8*, 3455-3461.

Larsen, M.J., Thomassen, M., Gerdes, A.M., and Kruse, T.A. (2014). Hereditary breast cancer: clinical, pathological and molecular characteristics. Breast Cancer (Auckl) *8*, 145-155.

Lee, Y.S., and Dutta, A. (2007). The tumor suppressor microRNA let-7 represses the HMGA2 oncogene. Genes Dev *21*, 1025-1030.

Leman, E.S., Madigan, M.C., Brunagel, G., Takaha, N., Coffey, D.S., and Getzenberg, R.H. (2003). Nuclear matrix localization of high mobility group protein I(Y) in a transgenic mouse model for prostate cancer. J Cell Biochem *88*, 599-608.

Li, A.Y., Boo, L.M., Wang, S.Y., Lin, H.H., Wang, C.C., Yen, Y., Chen, B.P., Chen, D.J., and Ann, D.K. (2009). Suppression of nonhomologous end joining repair by overexpression of HMGA2. Cancer Res *69*, 5699-5706.

Li, Y., Ma, J., Qian, X., Wu, Q., Xia, J., Miele, L., Sarkar, F.H., and Wang, Z. (2013). Regulation of EMT by Notch signaling pathway in tumor progression. Curr Cancer Drug Targets *13*, 957-962.

Li, Z., Nie, F., Wang, S., and Li, L. (2011). Histone H4 Lys 20 monomethylation by histone methylase SET8 mediates Wnt target gene activation. Proc Natl Acad Sci U S A *108*, 3116-3123.

Li, Z., Wang, C., Jiao, X., Lu, Y., Fu, M., Quong, A.A., Dye, C., Yang, J., Dai, M., Ju, X.*, et al.* (2006a). Cyclin D1 regulates cellular migration through the inhibition of thrombospondin 1 and ROCK signaling. Mol Cell Biol *26*, 4240-4256.

Li, Z., Wang, C., Prendergast, G.C., and Pestell, R.G. (2006b). Cyclin D1 functions in cell migration. Cell Cycle *5*, 2440-2442.

Lim, E., Vaillant, F., Wu, D., Forrest, N.C., Pal, B., Hart, A.H., Asselin-Labat, M.L., Gyorki, D.E., Ward, T., Partanen, A.*, et al.* (2009). Aberrant luminal progenitors as the candidate target population for basal tumor development in BRCA1 mutation carriers. Nat Med *15*, 907-913.

Liu, K., Zhang, C., Li, T., Ding, Y., Tu, T., Zhou, F., Qi, W., Chen, H., and Sun, X. (2015). Let-7a inhibits growth and migration of breast cancer cells by targeting HMGA1. Int J Oncol *46*, 2526-2534.

Liu, W.M., Guerra-Vladusic, F.K., Kurakata, S., Lupu, R., and Kohwi-Shigematsu, T. (1999). HMG-I(Y) recognizes base-unpairing regions of matrix attachment sequences and its increased expression is directly linked to metastatic breast cancer phenotype. Cancer Res *59*, 5695-5703.

Low, B.C., Pan, C.Q., Shivashankar, G.V., Bershadsky, A., Sudol, M., and Sheetz, M. (2014). YAP/TAZ as mechanosensors and mechanotransducers in regulating organ size and tumor growth. FEBS Lett *588*, 2663-2670.

Lund, T., Holtlund, J., Fredriksen, M., and Laland, S.G. (1983). On the presence of two new high mobility group-like proteins in HeLa S3 cells. FEBS Lett *152*, 163-167.

Macias, H., and Hinck, L. (2012). Mammary gland development. Wiley Interdiscip Rev Dev Biol *1*, 533-557.

Madsen, C.D., and Sahai, E. (2010). Cancer dissemination--lessons from leukocytes. Dev Cell *19*, 13-26.

Malhotra, G.K., Zhao, X., Band, H., and Band, V. (2010). Histological, molecular and functional subtypes of breast cancers. Cancer Biol Ther *10*, 955-960.

Manabe, T., Katayama, T., Sato, N., Gomi, F., Hitomi, J., Yanagita, T., Kudo, T., Honda, A., Mori, Y., Matsuzaki, S.*, et al.* (2003). Induced HMGA1a expression causes aberrant splicing of Presenilin-2 pre-mRNA in sporadic Alzheimer's disease. Cell Death Differ *10*, 698-708.

Mani, S.A., Guo, W., Liao, M.J., Eaton, E.N., Ayyanan, A., Zhou, A.Y., Brooks, M., Reinhard, F., Zhang, C.C., Shipitsin, M., *et al.* (2008). The epithelial-mesenchymal transition generates cells with properties of stem cells. Cell *133*, 704-715.

Martinez Hoyos, J., Ferraro, A., Sacchetti, S., Keller, S., De Martino, I., Borbone, E., Pallante, P., Fedele, M., Montanaro, D., Esposito, F., *et al.* (2009). HAND1 gene expression is negatively regulated by the High Mobility Group A1 proteins and is drastically reduced in human thyroid carcinomas. Oncogene *28*, 876-885.

Mazumdar, M., Sundareshan, S., and Misteli, T. (2004). Human chromokinesin KIF4A functions in chromosome condensation and segregation. J Cell Biol *166*, 613-620.

Micalizzi, D.S., Farabaugh, S.M., and Ford, H.L. (2010). Epithelial-mesenchymal transition in cancer: parallels between normal development and tumor progression. J Mammary Gland Biol Neoplasia *15*, 117-134.

Minn, A.J., Kang, Y., Serganova, I., Gupta, G.P., Giri, D.D., Doubrovin, M., Ponomarev, V., Gerald, W.L., Blasberg, R., and Massague, J. (2005). Distinct organ-specific metastatic potential of individual breast cancer cells and primary tumors. J Clin Invest *115*, 44-55.

Mook, S., Knauer, M., Bueno-de-Mesquita, J.M., Retel, V.P., Wesseling, J., Linn, S.C., Van't Veer, L.J., and Rutgers, E.J. (2010). Metastatic potential of T1 breast cancer can be predicted by the 70-gene MammaPrint signature. Ann Surg Oncol *17*, 1406-1413.

Moroishi, T., Hansen, C.G., and Guan, K.L. (2015). The emerging roles of YAP and TAZ in cancer. Nat Rev Cancer *15*, 73-79.

Mougiakakos, D., Choudhury, A., Lladser, A., Kiessling, R., and Johansson, C.C. (2010). Regulatory T cells in cancer. Adv Cancer Res *107*, 57-117.

Munshi, N., Agalioti, T., Lomvardas, S., Merika, M., Chen, G., and Thanos, D. (2001). Coordination of a transcriptional switch by HMGI(Y) acetylation. Science *293*, 1133-1136.

Nacht, M., Ferguson, A.T., Zhang, W., Petroziello, J.M., Cook, B.P., Gao, Y.H., Maguire, S., Riley, D., Coppola, G., Landes, G.M., *et al.* (1999). Combining serial analysis of gene expression and array technologies to identify genes differentially expressed in breast cancer. Cancer Res *59*, 5464-5470.

Nathanson, S.D., Kwon, D., Kapke, A., Hensley Alford, S., and Chitale, D. (2010). The role of lymph node metastasis in the systemic dissemination of breast cancer. Indian J Surg Oncol *1*, 313-322.

Nishida, N., Yano, H., Nishida, T., Kamura, T., and Kojiro, M. (2006). Angiogenesis in cancer. Vasc Health Risk Manag *2*, 213-219.

O'Toole, S.A., Machalek, D.A., Shearer, R.F., Millar, E.K., Nair, R., Schofield, P., McLeod, D., Cooper, C.L., McNeil, C.M., McFarland, A., *et al.* (2011). Hedgehog overexpression is associated with stromal interactions and predicts for poor outcome in breast cancer. Cancer Res *71*, 4002-4014.

Oakes, S.R., Gallego-Ortega, D., and Ormandy, C.J. (2014). The mammary cellular hierarchy and breast cancer. Cell Mol Life Sci *71*, 4301-4324.

Ostrand-Rosenberg, S., and Sinha, P. (2009). Myeloid-derived suppressor cells: linking inflammation and cancer. J Immunol *182*, 4499-4506.

Palmieri, D., Valentino, T., D'Angelo, D., De Martino, I., Postiglione, I., Pacelli, R., Croce, C.M., Fedele, M., and Fusco, A. (2011). HMGA proteins promote ATM expression and enhance cancer cell resistance to genotoxic agents. Oncogene *30*, 3024-3035.

Palvimo, J., and Linnala-Kankkunen, A. (1989). Identification of sites on chromosomal protein HMG-I phosphorylated by casein kinase II. FEBS Lett *257*, 101-104.

Pan, D. (2010). The hippo signaling pathway in development and cancer. Dev Cell *19*, 491-505.

Paulson, K.E., Rieger-Christ, K., McDevitt, M.A., Kuperwasser, C., Kim, J., Unanue, V.E., Zhang, X., Hu, M., Ruthazer, R., Berasi, S.P.*, et al.* (2007). Alterations of the HBP1 transcriptional repressor are associated with invasive breast cancer. Cancer Res *67*, 6136-6145.

Pegoraro, S., Ros, G., Ciani, Y., Sgarra, R., Piazza, S., and Manfioletti, G. (2015). A novel HMGA1-CCNE2-YAP axis regulates breast cancer aggressiveness. Oncotarget *6*, 19087-19101.

Pegoraro, S., Ros, G., Piazza, S., Sommaggio, R., Ciani, Y., Rosato, A., Sgarra, R., Del Sal, G., and Manfioletti, G. (2013). HMGA1 promotes metastatic processes in basal-like breast cancer regulating EMT and stemness. Oncotarget *4*, 1293-1308.

Pellacani, A., Chin, M.T., Wiesel, P., Ibanez, M., Patel, A., Yet, S.F., Hsieh, C.M., Paulauskis, J.D., Reeves, R., Lee, M.E.*, et al.* (1999). Induction of high mobility group-I(Y) protein by endotoxin and interleukin-1beta in vascular smooth muscle cells. Role in activation of inducible nitric oxide synthase. J Biol Chem *274*, 1525-1532.

Perou, C.M., Sorlie, T., Eisen, M.B., van de Rijn, M., Jeffrey, S.S., Rees, C.A., Pollack, J.R., Ross, D.T., Johnsen, H., Akslen, L.A.*, et al.* (2000). Molecular portraits of human breast tumours. Nature *406*, 747-752.

Phelps, M.A., Lin, T.S., Johnson, A.J., Hurh, E., Rozewski, D.M., Farley, K.L., Wu, D., Blum, K.A., Fischer, B., Mitchell, S.M.*, et al.* (2009). Clinical response and pharmacokinetics from a phase 1 study of an active dosing schedule of flavopiridol in relapsed chronic lymphocytic leukemia. Blood *113*, 2637-2645.

Piccolo, S., Cordenonsi, M., and Dupont, S. (2013). Molecular pathways: YAP and TAZ take center stage in organ growth and tumorigenesis. Clin Cancer Res *19*, 4925-4930.

Pitman, M., Emery, B., Binder, M., Wang, S., Butzkueven, H., and Kilpatrick, T.J. (2004). LIF receptor signaling modulates neural stem cell renewal. Mol Cell Neurosci *27*, 255-266.

Ponti, D., Costa, A., Zaffaroni, N., Pratesi, G., Petrangolini, G., Coradini, D., Pilotti, S., Pierotti, M.A., and Daidone, M.G. (2005). Isolation and in vitro propagation of tumorigenic breast cancer cells with stem/progenitor cell properties. Cancer Res *65*, 5506-5511.

Ram, T.G., Reeves, R., and Hosick, H.L. (1993). Elevated high mobility group-I(Y) gene expression is associated with progressive transformation of mouse mammary epithelial cells. Cancer Res *53*, 2655-2660.

Reedijk, M., Pinnaduwage, D., Dickson, B.C., Mulligan, A.M., Zhang, H., Bull, S.B., O'Malley, F.P., Egan, S.E., and Andrulis, I.L. (2008). JAG1 expression is associated with a basal phenotype and recurrence in lymph node-negative breast cancer. Breast Cancer Res Treat *111*, 439-448.

Reeves, R. (2001). Molecular biology of HMGA proteins: hubs of nuclear function. Gene *277*, 63-81.

Reeves, R. (2010). Nuclear functions of the HMG proteins. Biochim Biophys Acta *1799*, 3-14.

Reeves, R. (2015). High mobility group (HMG) proteins: Modulators of chromatin structure and DNA repair in mammalian cells. DNA repair *36*, 122-136.

Reeves, R., Edberg, D.D., and Li, Y. (2001). Architectural transcription factor HMGI(Y) promotes tumor progression and mesenchymal transition of human epithelial cells. Mol Cell Biol *21*, 575-594.

Reeves, R., and Wolffe, A.P. (1996). Substrate structure influences binding of the non-histone protein HMG-I(Y) to free nucleosomal DNA. Biochemistry *35*, 5063-5074.

Rhodes, D.R., and Chinnaiyan, A.M. (2004). Bioinformatics strategies for translating genome-wide expression analyses into clinically useful cancer markers. Annals of the New York Academy of Sciences *1020*, 32-40.

Ringner, M., Fredlund, E., Hakkinen, J., Borg, A., and Staaf, J. (2011). GOBO: gene expression-based outcome for breast cancer online. PLoS One *6*, e17911.

Sabeh, F., Shimizu-Hirota, R., and Weiss, S.J. (2009). Protease-dependent versus -independent cancer cell invasion programs: three-dimensional amoeboid movement revisited. J Cell Biol *185*, 11-19.

Sampieri, K., and Fodde, R. (2012). Cancer stem cells and metastasis. Seminars in cancer biology *22*, 187-193.

Sanchez-Martinez, C., Gelbert, L.M., Lallena, M.J., and de Dios, A. (2015). Cyclin dependent kinase (CDK) inhibitors as anticancer drugs. Bioorg Med Chem Lett *25*, 3420-3435.

Santen, R.J., Song, R.X., McPherson, R., Kumar, R., Adam, L., Jeng, M.H., and Yue, W. (2002). The role of mitogen-activated protein (MAP) kinase in breast cancer. J Steroid Biochem Mol Biol *80*, 239-256.

Schneider, B.L., and Kulesz-Martin, M. (2004). Destructive cycles: the role of genomic instability and adaptation in carcinogenesis. Carcinogenesis *25*, 2033-2044.

Scholzen, T., and Gerdes, J. (2000). The Ki-67 protein: from the known and the unknown. J Cell Physiol *182*, 311-322.

Schwanbeck, R., Manfioletti, G., and Wisniewski, J.R. (2000). Architecture of high mobility group protein I-C.DNA complex and its perturbation upon phosphorylation by Cdc2 kinase. J Biol Chem *275*, 1793-1801.

Scolz, M., Widlund, P.O., Piazza, S., Bublik, D.R., Reber, S., Peche, L.Y., Ciani, Y., Hubner, N., Isokane, M., Monte, M.*, et al.* (2012). GTSE1 is a microtubule plus-end tracking protein that regulates EB1-dependent cell migration. PLoS One *7*, e51259.

Scully, O.J., Bay, B.H., Yip, G., and Yu, Y. (2012). Breast cancer metastasis. Cancer Genomics Proteomics *9*, 311-320.

Sgarra, R., Lee, J., Tessari, M.A., Altamura, S., Spolaore, B., Giancotti, V., Bedford, M.T., and Manfioletti, G. (2006). The AT-hook of the chromatin architectural transcription factor high mobility group A1a is arginine-methylated by protein arginine methyltransferase 6. J Biol Chem *281*, 3764-3772.

Sgarra, R., Rustighi, A., Tessari, M.A., Di Bernardo, J., Altamura, S., Fusco, A., Manfioletti, G., and Giancotti, V. (2004). Nuclear phosphoproteins HMGA and their relationship with chromatin structure and cancer. FEBS Lett *574*, 1-8.

Sgarra, R., Zammitti, S., Lo Sardo, A., Maurizio, E., Arnoldo, L., Pegoraro, S., Giancotti, V., and Manfioletti, G. (2010). HMGA molecular network: From transcriptional regulation to chromatin remodeling. Biochim Biophys Acta *1799*, 37-47.

Shah, S.N., Cope, L., Poh, W., Belton, A., Roy, S., Talbot, C.C., Jr., Sukumar, S., Huso, D.L., and Resar, L.M. (2013). HMGA1: a master regulator of tumor progression in triple-negative breast cancer cells. PLoS One *8*, e63419.

Shaw, F.L., Harrison, H., Spence, K., Ablett, M.P., Simoes, B.M., Farnie, G., and Clarke, R.B. (2012). A detailed mammosphere assay protocol for the quantification of breast stem cell activity. J Mammary Gland Biol Neoplasia *17*, 111-117.

Smyth, G.K. (2004). Linear models and empirical bayes methods for assessing differential expression in microarray experiments. Stat Appl Genet Mol Biol *3*, Article3.

Sorlie, T., Perou, C.M., Tibshirani, R., Aas, T., Geisler, S., Johnsen, H., Hastie, T., Eisen, M.B., van de Rijn, M., Jeffrey, S.S.*, et al.* (2001). Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. Proc Natl Acad Sci U S A *98*, 10869-10874.

Sorrentino, G., Ruggeri, N., Specchia, V., Cordenonsi, M., Mano, M., Dupont, S., Manfrin, A., Ingallina, E., Sommaggio, R., Piazza, S.*, et al.* (2014). Metabolic control of YAP and TAZ by the mevalonate pathway. Nat Cell Biol *16*, 357-366.

Sotiriou, C., and Piccart, M.J. (2007). Taking gene-expression profiling to the clinic: when will molecular signatures become relevant to patient care? Nat Rev Cancer *7*, 545-553.

Sotiriou, C., Wirapati, P., Loi, S., Harris, A., Fox, S., Smeds, J., Nordgren, H., Farmer, P., Praz, V., Haibe-Kains, B.*, et al.* (2006). Gene expression profiling in breast cancer: understanding the molecular basis of histologic grade to improve prognosis. J Natl Cancer Inst *98*, 262-272.

Sternlicht, M.D. (2006). Key stages in mammary gland development: the cues that regulate ductal branching morphogenesis. Breast Cancer Res *8*, 201.

Strano, S., Munarriz, E., Rossi, M., Castagnoli, L., Shaul, Y., Sacchi, A., Oren, M., Sudol, M., Cesareni, G., and Blandino, G. (2001). Physical interaction with Yes-associated protein enhances p73 transcriptional activity. J Biol Chem *276*, 15164-15173.

Talmadge, J.E., and Fidler, I.J. (2010). AACR centennial series: the biology of cancer metastasis: historical perspective. Cancer Res *70*, 5649-5669.

Tamimi, Y., van der Poel, H.G., Karthaus, H.F., Debruyne, F.M., and Schalken, J.A. (1996). A retrospective study of high mobility group protein I(Y) as progression marker for prostate cancer determined by in situ hybridization. Br J Cancer *74*, 573-578.

Tao, Y., Mao, J., Zhang, Q., and Li, L. (2011). Overexpression of Hedgehog signaling molecules and its involvement in triple-negative breast cancer. Oncol Lett *2*, 995-1001.

Teng, M.W., Swann, J.B., Koebel, C.M., Schreiber, R.D., and Smyth, M.J. (2008). Immune-mediated dormancy: an equilibrium with cancer. J Leukoc Biol *84*, 988-993.

Tesfaye, A., Di Cello, F., Hillion, J., Ronnett, B.M., Elbahloul, O., Ashfaq, R., Dhara, S., Prochownik, E., Tworkoski, K., Reeves, R.*, et al.* (2007). The high-mobility group A1 gene

up-regulates cyclooxygenase 2 expression in uterine tumorigenesis. Cancer Res *67*, 3998-4004.

Tessari, M.A., Gostissa, M., Altamura, S., Sgarra, R., Rustighi, A., Salvagno, C., Caretti, G., Imbriano, C., Mantovani, R., Del Sal, G.*, et al.* (2003). Transcriptional activation of the cyclin A gene by the architectural transcription factor HMGA2. Mol Cell Biol *23*, 9104-9116.

Thiery, J.P. (2002). Epithelial-mesenchymal transitions in tumour progression. Nat Rev Cancer *2*, 442-454.

Thuault, S., Valcourt, U., Petersen, M., Manfioletti, G., Heldin, C.H., and Moustakas, A. (2006). Transforming growth factor-beta employs HMGA2 to elicit epithelial-mesenchymal transition. J Cell Biol *174*, 175-183.

Treff, N.R., Pouchnik, D., Dement, G.A., Britt, R.L., and Reeves, R. (2004). High-mobility group A1a protein regulates Ras/ERK signaling in MCF-7 human breast cancer cells. Oncogene *23*, 777-785.

Tsai, J.H., and Yang, J. (2013). Epithelial-mesenchymal plasticity in carcinoma metastasis. Genes Dev *27*, 2192-2206.

Ueda, Y., Watanabe, S., Tei, S., Saitoh, N., Kuratsu, J., and Nakao, M. (2007). High mobility group protein HMGA1 inhibits retinoblastoma protein-mediated cellular G0 arrest. Cancer Sci *98*, 1893-1901.

Vallone, D., Battista, S., Pierantoni, G.M., Fedele, M., Casalino, L., Santoro, M., Viglietto, G., Fusco, A., and Verde, P. (1997). Neoplastic transformation of rat thyroid cells requires the junB and fra-1 gene induction which is dependent on the HMGI-C gene product. EMBO J *16*, 5310-5321.

van 't Veer, L.J., Dai, H., van de Vijver, M.J., He, Y.D., Hart, A.A., Mao, M., Peterse, H.L., van der Kooy, K., Marton, M.J., Witteveen, A.T.*, et al.* (2002). Gene expression profiling predicts clinical outcome of breast cancer. Nature *415*, 530-536.

Wang, R., Lv, Q., Meng, W., Tan, Q., Zhang, S., Mo, X., and Yang, X. (2014). Comparison of mammosphere formation from breast cancer cell lines and primary breast tumors. J Thorac Dis *6*, 829-837.

Wang, Y., Klijn, J.G., Zhang, Y., Sieuwerts, A.M., Look, M.P., Yang, F., Talantov, D., Timmermans, M., Meijer-van Gelder, M.E., Yu, J.*, et al.* (2005). Gene-expression profiles to predict distant metastasis of lymph-node-negative primary breast cancer. Lancet *365*, 671-679.

Wang, Y., Shi, J., Chai, K., Ying, X., and Zhou, B.P. (2013). The Role of Snail in EMT and Tumorigenesis. Curr Cancer Drug Targets *13*, 963-972.

Warburg, O.H., and Dickens, F. (1930). The metabolism of tumours (London: Constable).

Webber, C., Gospodarowicz, M., Sobin, L.H., Wittekind, C., Greene, F.L., Mason, M.D., Compton, C., Brierley, J., and Groome, P.A. (2014). Improving the TNM classification: findings from a 10-year continuous literature review. Int J Cancer *135*, 371-378.

Weigel, M.T., and Dowsett, M. (2010). Current and emerging biomarkers in breast cancer: prognosis and prediction. Endocr Relat Cancer *17*, R245-262.

White, E.A., and Glotzer, M. (2012). Centralspindlin: at the heart of cytokinesis. Cytoskeleton (Hoboken) *69*, 882-892.

Widelitz, R. (2005). Wnt signaling through canonical and non-canonical pathways: recent progress. Growth Factors *23*, 111-116.

Wood, L.J., Mukherjee, M., Dolde, C.E., Xu, Y., Maher, J.F., Bunton, T.E., Williams, J.B., and Resar, L.M. (2000). HMG-I/Y, a new c-Myc target gene and potential oncogene. Mol Cell Biol *20*, 5490-5502.

Xing, J., Cao, G., and Fu, C. (2014). HMGA1 Interacts with beta-Catenin to Positively Regulate Wnt/beta-Catenin Signaling in Colorectal Cancer Cells. Pathol Oncol Res.

Xu, Y., Sumter, T.F., Bhattacharya, R., Tesfaye, A., Fuchs, E.J., Wood, L.J., Huso, D.L., and Resar, L.M. (2004). The HMG-I oncogene causes highly penetrant, aggressive lymphoid malignancy in transgenic mice and is overexpressed in human leukemia. Cancer Res *64*, 3371-3375.

Yie, J., Merika, M., Munshi, N., Chen, G., and Thanos, D. (1999). The role of HMG I(Y) in the assembly and function of the IFN-beta enhanceosome. EMBO J *18*, 3074-3089.

Yuan, X., Zhang, M., Wu, H., Xu, H., Han, N., Chu, Q., Yu, S., Chen, Y., and Wu, K. (2015). Expression of Notch1 Correlates with Breast Cancer Progression and Prognosis. PLoS One *10*, e0131689.

Zhang, K., Qi, H.X., Hu, Z.M., Chang, Y.N., Shi, Z.M., Han, X.H., Han, Y.W., Zhang, R.X., Zhang, Z., Chen, T.*, et al.* (2015). YAP and TAZ Take Center Stage in Cancer. Biochemistry *54*, 6555-6566.

Zhao, K., Kas, E., Gonzalez, E., and Laemmli, U.K. (1993). SAR-dependent mobilization of histone H1 by HMG-I/Y in vitro: HMG-I/Y is enriched in H1-depleted chromatin. EMBO J *12*, 3237-3247.

Zhou, X., Benson, K.F., Ashar, H.R., and Chada, K. (1995). Mutation responsible for the mouse pygmy phenotype in the developmentally regulated factor HMGI-C. Nature *376*, 771-774.

Publications

During my PhD I worked on the following publications:

Expression and functional characterization of Xhmg-at-hook genes in Xenopus laevis.

Macrì S, Sgarra R, **Ros G**, Maurizio E, Zammitti S, Milani O, Onorati M, Vignali R, Manfioletti G.

High Mobility Group A proteins (HMGA1 and HMGA2) are architectural nuclear factors involved in development, cell differentiation, and cancer formation and progression. Here we report the cloning, developmental expression and functional analysis of a new multi-AT-hook factor in Xenopus laevis (XHMG-AT-hook) that exists in three different isoforms. Xhmg-at-hook1 and 3 isoforms, but not isoform 2, are expressed throughout the entire development of Xenopus, both in the maternal and zygotic phase. Localized transcripts are present in the animal pole in the early maternal phase; during the zygotic phase, mRNA can be detected in the developing central nervous system (CNS), including the eye, and in the neural crest. We show evidence that XHMG-AT-hook proteins differ from typical HMGA proteins in terms of their properties in DNA binding and in protein/protein interaction. Finally, we provide evidence that they are involved in early CNS development and in neural crest differentiation. PLoS One. 2013 Jul 25;8(7):e69866

HMGA1 promotes metastatic processes in basal-like breast cancer regulating EMT and stemness.
Pegoraro S, **Ros G**, Piazza S, Sommaggio R, Ciani Y, Rosato A, Sgarra R, Del Sal G, Manfioletti G.
Oncotarget. 2013 Aug;4(8):1293-308.

A novel HMGA1-CCNE2-YAP axis regulates breast cancer aggressiveness.
Ros G, Pegoraro S, Ciani Y, Sgarra R, Piazza S, Manfioletti G.
Oncotarget. 2015 Aug 7;6(22):19087-101.

Acknowledgements

I thank my family to have supported me through the life, to have given me the freedom to make my choices, to be there with passion and dedication in each moment of my path and I thank Giulia and Lorena to be my friends and to give me the opportunity to see the world from other points of view.

Research is a question of love; life is a question of love. In any case both research and life are a journey into the future and I think to fly away you have to have a good co-pilot with the same horizon. I have found him twelve years ago and I thank the universe for him. My gratitude for Francesco is endless. His patience, his way of thinking outside the box, his delicate and constant presence make me a better person and give me the feeling that everything is possible.