



UNIVERSITÀ DEGLI STUDI DI TRIESTE
XXIX CICLO DEL DOTTORATO DI RICERCA IN
BIOMEDICINA MOLECOLARE

FONDO SOCIALE EUROPEO

HMGA1 PROTEINS REGULATE GENE
EXPRESSION IN TRIPLE-NEGATIVE BREAST
CANCER CELLS THROUGH THE MODULATION
OF THE EPIGENETIC CODE

Settore scientifico-disciplinare: **BIO/10**

DOTTORANDA
CARLOTTA PENZO

COORDINATORE
PROF. GERMANA MERONI

SUPERVISORE DI TESI
PROF. GUIDALBERTO MANFIOLETTI

CO-SUPERVISORE DI TESI
RICCARDO SGARRA, PhD

ANNO ACCADEMICO 2015/2016



UNIVERSITÀ DEGLI STUDI DI TRIESTE
XXIX CICLO DEL DOTTORATO DI RICERCA IN
BIOMEDICINA MOLECOLARE

FONDO SOCIALE EUROPEO

HMGA1 PROTEINS REGULATE GENE
EXPRESSION IN TRIPLE-NEGATIVE BREAST
CANCER CELLS THROUGH THE MODULATION
OF THE EPIGENETIC CODE

Settore scientifico-disciplinare: **BIO/10**

DOTTORANDA
CARLOTTA PENZO

COORDINATORE
PROF. GERMANA MERONI

SUPERVISORE DI TESI
PROF. GUIDALBERTO MANFIOLETTI

CO-SUPERVISORE DI TESI
RICCARDO SGARRA, PhD

ANNO ACCADEMICO 2015/2016

To my family

ABSTRACT

Triple negative breast cancers (TNBCs) are high-grade invasive ductal carcinoma which lack expression of hormone receptors and which do not benefit from either endocrine or anti-human epidermal growth factor receptor 2 therapies leading to poor outcome. This heterogeneous group of cancers encompass different molecular subtypes reflecting treatment responsiveness. Among others, mesenchymal-like TNBC subtype is characterized by the enrichment of proteins involved in cell motility and Epithelial to Mesenchymal transition (EMT). High Mobility Group A1 (HMGA1) is a family of oncofetal proteins identified as inducers of this phenotypic transition that have been extensively associated to cancer onset and progression. Despite their well-described role in chromatin architecture and gene expression regulation, their potential role in setting histones post-translational dynamics has never been taken into account. In this thesis, we provide evidences of HMGA1 involvement in histone post-translational modifications modulation by sustaining the RAS/RAF/MEK/ERK/RSK2 pathway in mesenchymal-like TNBC cells. In particular, we suggest that HMGA1 proteins make chromatin prone to epigenetic modifiers actions sustaining both H3 serine 10 phosphorylation by RSK2 and H2B lysine 5 acetylation by CBP and modulating the expression of genes involved in tumor progression and EMT. We suggest that aside being considered as drug target, HMGA1 could be used as biomarker predicting responsiveness to epigenetic therapies in triple negative breast cancers.

RIASSUNTO

Il cancro al seno “Triplo Negativo” è un carcinoma duttale invasivo di grado elevato in cui le cellule tumorali non esprimono i recettori ormonali rendendo inefficaci sia la terapia convenzionale endocrina che quella anticorpale anti-HER2. Questi tumori sono di solito caratterizzati da una prognosi infausta. Questa tipologia di carcinoma della mammella comprende un gruppo estremamente eterogeneo di tumori che vengono classificati in sottotipi molecolari distinti e che permettono spesso di determinare la risposta ad un trattamento farmacologico. Tra questi, il cancro al seno triplo negativo simil-mesenchimale è caratterizzato dall’espressione elevata di proteine coinvolte nella motilità cellulare e nella transizione epitelio-mesenchimale. Le proteine oncofetali appartenenti alla famiglia delle *High Mobility Group A1* (HMGA1) sono state identificate come induttori di questa transizione fenotipica e la loro espressione è stata ampiamente correlata a trasformazione e progressione tumorale. Sebbene il ruolo di queste proteine sia stato ampiamente investigato nel contesto della regolazione dell’espressione genica e della strutturazione dell’architettura cromatinica, l’influenza che esse possono avere nello scenario dinamico delle modifiche post-traduzionali degli istoni non è stato ancora presa in considerazione.

In questa tesi sono fornite evidenze del coinvolgimento delle proteine HMGA1 nella regolazione delle modifiche post-traduzionali degli istoni attraverso la modulazione della via di trasduzione del segnale RAS/RAF/MEK/ERK/RSK2 in cellule di cancro al seno triplo negativo. In particolare, viene dimostrato come queste proteine possano intervenire predisponendo la cromatina all’azione di modificatori epigenetici, quali la chinasi RSK2 e l’acetiltransferasi CBP, favorendo rispettivamente la fosforilazione della serina 10 dell’istone H3 e l’acetilazione della lisina 5 dell’istone H2B, e come questo sia associato all’attivazione trascrizionale di geni coinvolti nella progressione tumorale e nella transizione epitelio-mesenchimale. Si suggerisce quindi che le proteine HMGA1 possano essere considerate non solo come bersaglio terapeutico ma anche come bio-marcatori in grado di prevedere la responsività a terapie epigenetiche per il cancro al seno di triplo negativo.

TABLE OF CONTENTS

| | |
|--|-----------|
| INTRODUCTION | 1 |
| BREAST CANCER..... | 1 |
| <i>Breast cancer classifications.....</i> | <i>1</i> |
| <i>The molecular portrait of breast cancer</i> | <i>3</i> |
| <i>Triple Negative Breast Cancer (TNBC).....</i> | <i>5</i> |
| <i>Current treatment.....</i> | <i>6</i> |
| EPITHELIAL TO MESENCHYMAL TRANSITION (EMT) | 9 |
| <i>The controversy role of EMT in tumor metastasis</i> | <i>10</i> |
| <i>High Mobility Group Proteins drive EMT in cancer cells.....</i> | <i>12</i> |
| HIGH MOBILITY GROUP PROTEINS | 15 |
| <i>High Mobility Group A proteins (HMGA)</i> | <i>15</i> |
| <i>HMGA1 dynamic interplay with chromatin.....</i> | <i>16</i> |
| <i>Basal and oncogenic factors which converge in HMGA1 expression or modulate its function....</i> | <i>18</i> |
| <i>The causative role of HMGA1 in cancer.....</i> | <i>20</i> |
| <i>HMGA1 is a master regulator of breast cancer.....</i> | <i>21</i> |
| <i>Targeting HMGA1: the challenge of specificity</i> | <i>23</i> |
| BREAST CANCER EPIGENETICS..... | 24 |
| <i>Chromatin landscape of breast cancer cells.....</i> | <i>25</i> |
| <i>Brest cancer epigenetics marks in clinical managements of disease.....</i> | <i>28</i> |
| AIM OF THE STUDY | 31 |
| MATERIALS AND METHODS | 32 |
| <i>Breast cancer cell lines</i> | <i>32</i> |
| <i>Cellular cultures and treatments.....</i> | <i>32</i> |
| <i>Cell counting.....</i> | <i>32</i> |
| <i>Small interfering RNA (siRNA)</i> | <i>33</i> |
| siRNA | 33 |
| <i>Metabolic activity assay: MTS assay</i> | <i>33</i> |
| <i>Lysates pre-fractionation: Histone Extraction Methods.....</i> | <i>34</i> |
| Strong acid histone extraction (modified Schechter D et al., 2007)..... | 34 |
| Histone purification kit (<i>Active Motif</i>)..... | 34 |
| <i>Total proteins extraction for Western Blot analysis</i> | <i>34</i> |
| <i>SDS- PolyAcrylamide Gel Electrophoresis (SDS-PAGE).....</i> | <i>35</i> |
| <i>Blue Coomassie staining - Total Protein samples quantification and normalization</i> | <i>35</i> |
| <i>Bicinchoninic Acid (BCA) method for protein quantification.....</i> | <i>36</i> |

| | |
|--|-----------|
| <i>Semi-quantitative Western Blot (WB)</i> | 36 |
| Antibodies | 37 |
| <i>Multiple enzymes digestion Filter Aided Sample Purification (MED-FASP) and Stage Tips purification</i> | 37 |
| <i>Orbitrap Q exactive</i> | 38 |
| <i>RNA extraction - DNase I digestion - Clean Up</i> | 39 |
| <i>RNA analysis on denaturing agarose gel electrophoresis and quantification of extracted RNA</i> .. | 40 |
| <i>Reverse Transcriptase PCR (RT-PCR)</i> | 40 |
| <i>Real-time quantitative PCR (qPCR)</i> | 40 |
| Primers | 41 |
| <i>Co-ImmunoPrecipitation (Co-IP)</i> | 42 |
| <i>Immunofluorescence (IF)</i> | 43 |
| Antibodies | 43 |
| <i>Statistical analysis</i> | 43 |
| RESULTS | 45 |
| <i>HMGA1 expression level affects epigenetic modifications</i> | 45 |
| <i>HMGA1 silencing induces downregulation of histone H2B acetylation</i> | 48 |
| <i>CBP/p300 acetyltransferases are responsible for H2B acetylation in MDA-MB-231</i> | 51 |
| <i>CBP is the acetyltransferase responsible for transcriptional regulation of genes belonging to HMGA1-signature in MDA-MB-231</i> | 54 |
| <i>CBP histone acetylation is functional coupled with RSK2 phosphorylation of H3S10 in MDA-MB-231</i> | 57 |
| <i>HMGA1 modulate RSK2 and CBP activities without interfering with CREB phosphorylation</i> | 59 |
| <i>HMGA1 regulates epigenetic modifications acting downstream RSK2 activation</i> | 60 |
| <i>Different HMGA1 abundances reflect different signalling regulation of histone H2B acetylation pattern</i> | 64 |
| DISCUSSION | 69 |
| CONCLUSIONS AND FUTURE PERSPECTIVES | 74 |
| PRELIMINARY DATA | 76 |
| BIBLIOGRAPHY | 79 |
| ACKNOWLEDGEMENTS | 89 |

TABLE OF FIGURES

| | |
|---|----|
| FIGURE 1.1 BREAST CANCER IS MAINLY ADENOCARCINOMA ARISING FROM EPITHELIA OF THE LOBULES AND DUCTS. | 3 |
| FIGURE 1.2 TNBC SUBTYPES ARE ENRICHED FOR DISTINCT GENE ONTOLOGIES. | 7 |
| FIGURE 1.3 EMT AND MET ARE MECHANISMS THAT TRIGGER CANCER EVOLUTION TOWARD METASTASIS..... | 10 |
| FIGURE 1.4 MOLECULAR BIOMARKERS OF EMT USED IN ONGOING CANCER CLINICAL TRIALS FOR PROGNOSIS AND THERAPEUTIC INTERVENTION..... | 12 |
| FIGURE 1.5. THE STEREOSPECIFIC COMPLEX OF ENHANCEOSOME. | 17 |
| FIGURE 1.6 FUNCTIONAL CLASSIFICATION OF GENES MUTATED IN CANCER DEVELOPMENT. | 24 |
| TABLE 1. LIST OF siRNA SEQUENCES AND THEIR CONCENTRATION USED TO TARGET SPECIFIED MRNA. | 33 |
| TABLE 2. LIST OF PRIMARY AND SECONDARY ANTIBODIES USED WITH THEIR DILUTION FACTORS IN WESTERN BLOT ANALYSIS. | 37 |
| TABLE 3. LIST OF PRIMER SEQUENCES (FORWARD AND REVERSE) USED FOR RT-QPCR. | 41 |
| TABLE 4. LIST OF PRIMARY AND SECONDARY ANTIBODIES USED WITH THEIR DILUTION FACTORS IF ANALYSES..... | 43 |
| FIGURE 3.1 HMGA1 INTERACTS WITH CORE HISTONES AND INFLUENCES THEIR PTMS IN MDA-MB-231. | 47 |
| FIGURE 3.2 HMGA1 LEVELS INFLUENCE ACETYLATION OF HISTONE H2B IN MDA-MB-231. | 51 |
| FIGURE 3.3 CBP/p300 ACETYLTRANSFERASES ACETYLATE HISTONE H2B IN MDA-MB-231 CELLS PRODUCING TRANSCRIPTIONAL ALTERATIONS OVERLAPPING TO HMGA1-SIGNATURE. | 54 |
| FIGURE 3.4 SILENCING OF CBP MIRRORS HMGA1 SILENCING EFFECTS ON HISTONE H2B ACETYLATION AND GENE EXPRESSION ALTERATION IN MDA-MB-231 CELLS..... | 55 |
| FIGURE 3.5 SILENCING OF P300 AFFECT HISTONE H2B ACETYLATION IN MDA-MB-231 CELLS BUT LEAVES UNCHANGED THE EXPRESSION OF THE MAJORITY OF HMGA1-SIGNATURE GENES. | 56 |
| FIGURE 3.6 KINASE AND HAT ACTIVITIES OF CBP AND RSK2 ARE FUNCTIONAL COUPLED IN MODULATING GENE EXPRESSION OF HMGA1-SIGNATURE IN MDA-MB-231. | 59 |
| FIGURE 3.7 HMGA1 INTERPLAY WITH CBP AND RSK2 DOES NOT INVOLVE CREB PHOSPHORYLATION..... | 60 |
| FIGURE 3.8 HMGA1 MODULATE HISTONES PTMS ACTING DOWNSTREAM RSK2 PHOSPHORYLATION. | 63 |
| FIGURE 3.9 DIFFERENT MOLECULAR SUBTYPES OF BREAST CANCER DISPLAY DISTINCT SIGNALLING MECHANISM OF HISTONE H2B ACETYLATION WHICH MIRROR HMGA1 ABUNDANCES..... | 67 |
| FIGURE P1 PRELIMINARY DATA..... | 78 |

LIST OF ABBREVIATIONS

| | |
|-------|--|
| ac | acetylation |
| AKT | RAC serine/threonine-protein kinase (protein kinase B) |
| AI | Aromatase Inhibitors |
| AIOM | Associazione Italiana Oncologia Medica |
| AP1 | Activator protein 1 |
| AR | Androgen Receptor |
| ATCC | American Type Culture Collection |
| ATF2 | Activating transcription factor 2 |
| AURKB | Aurora Kinase B |
| BCA | Bicinchoninic acid |
| BCS | Breast Conserving Surgery |
| bHLH | Basic Helix-loop-Helix |
| BRCA1 | Breast Cancer type 1 susceptibility protein |
| BSA | Bovine serum Albumin |
| BSCS | Breast Cancer Stem Cells |
| c-Jun | Jun Proto-Oncogene, AP-1 Transcription Factor Subunit |
| c-Myc | v-myc avian myelocytomatosis viral oncogene homolog |
| c-Myb | V-Myb Avian Myeloblastosis Viral Oncogene Homolog |
| CBP | CREB binding protein |
| CBX | Chromobox 7 |
| CCND2 | Cyclin D2 |
| Cdc2 | Cell-cycle dependent kinase |
| CD | Cluster of differentiation |
| CDH3 | Cadherin 3 |
| CENPF | Centromere Protein F |
| CDH1 | Cadherin-1 (E-cadherin) |
| ceRNA | Competitive endogenous RNA |
| ChIP | Chromatin Immunoprecipitation |
| CHK1 | Checkpoint-Kinase 1 |
| CYP33 | Cyclophilin 33 |
| CK2 | casein kinase 2 |
| COX2 | Cyclooxygenase-2 |
| CpG | Cytosine-Guanine dinucleotide |
| CREB | cAMP response element-binding protein |
| CTC | Circulating Tumor Cells |
| CTCF | CCCTC-binding factor |
| DCIS | Ductal Carcinoma <i>in situ</i> |
| DFS | Disease Free Survival |

| | |
|----------------|---|
| DMEM | Dulbecco's Modified Eagle Medium |
| DNMT3A | DNA methyltransferase 3A |
| dNTPs | Deoxynucleotide triphosphates |
| DTC | Disseminating Tumor Cells |
| EGF | Epidermal Growth Factor |
| EGFR | Epidermal Growth Factor Receptor |
| ECL | Enhanced Chemiluminescence |
| ECM | Extracellular Matrix |
| EMT | Epithelial to Mesenchymal Transition |
| ER | Estrogen Receptor |
| ERB-B2 | proto-oncogene erb-b2 receptor tyrosine kinase 2 |
| ERK | Extracellular signal-Regulated Kinases |
| EZH2 | Zeste homolog 2 |
| FBS | Fetal Bovine Serum |
| FOXC2 | Forkhead box protein C2 |
| FOXM1 | Forkhead box protein M1 |
| FOXO3a | Forkhead box O3 |
| Fra1 | Fos-related antigen 1 |
| GAPDH | Glyceraldehyde-3-phosphate dehydrogenase |
| GemOE | Geminin Overexpressing |
| GCN | General Control Non-derepressible |
| HAND1 | Heart And Neural Crest Derivatives Expressed 1 |
| HAT | Histone Acetyltransferase |
| HDAC | Histone Deacetylase |
| HER2/neu | Human Epidermal growth factor receptor 2 |
| hESC | Human Embryonic Stem Cells |
| HIPK2 | homeodomain-interacting protein kinase 2 |
| HMG | High Mobility Group proteins |
| HMGA1P | High Mobility Group Proteins A 1 pseudogene |
| HP1 | Heterochromatin Protein 1 |
| ID | Intrinsically disordered proteins |
| IDC | Invasive Ductal Carcinoma |
| IFN | Interferon |
| IHC | Immunohistochemistry |
| IL | Interleukin |
| IL-2R α | Interleukin 2 receptor α |
| ILC | Invasive Lobular Carcinoma |
| K | Lysine |
| KDM3A/JMJD1a | JmjC Domain-Containing Histone Demethylation Protein 2A |
| KIF4A | Kinesin Family Member 4A |
| KIF23 | kinesin family member 23 |

| | |
|------------------------------|---|
| KITLG | KIT Ligand |
| iNOS | Inducible Nitric Oxid Synthase |
| IF | Immunofluorescence |
| IP | Immunoprecipitation |
| iPSC | Induced pluripotent stem cells |
| IRF | Interferon regulatory factor |
| ISH | <i>In situ</i> hybridization |
| LADs | Lamina associated domains |
| LAR | Luminal Androgen Receptor |
| LCIS | Lobular Carcinoma <i>in situ</i> |
| let-7 | lethal-7 microRNA |
| LOCK | Large organized chromatin K9 modification |
| LREA | Long-range epigenetic activation |
| LRRC59 | Leucine rich repeat containing 59 |
| LSD1 | Lysine-Specific histone Demethylase homolog |
| MAPK | Mitogen-Activated Protein Kinase |
| MBC | Metaplastic Breast Cancer |
| me | Methylation |
| MED-FASP | Multi-Enzyme Digestion Filter Aided Sample Purification |
| MEK | Mitogen-activated protein/Extracellular signal regulated Kinase |
| MET | Mesenchymal to Epithelial Transition |
| MH | MAD homology domains |
| miRNA | micro-RNA |
| MMP | Matrix MetalloProteinase |
| NANOG | Homeobox protein NANOG |
| NF-κB | Nuclear Factor kappa-light-chain-enhancer of activated B cells |
| NOS | Not otherwise specified |
| NST | No Special type |
| Nurf55 | 55 kDa nucleosome remodelling factor protein |
| O/N | Over night |
| OCT4 | Octamer-binding transcription factor 4 |
| OS | Overall Survival |
| OSM | Oncostatin M |
| p16 ^{ink4a} /CDKN2A | Cyclin-dependent kinase inhibitor 2A |
| p300 | E1A binding protein p300 |
| PARP | Poly-ADP-Ribose Polymerase |
| PBS | Phosphate buffered saline |
| PCAF | P300/CBP-associated factor |
| pCR | Pathological Complete Response |
| PEV | Position effect variegation |
| ph | Phosphorylation |

| | |
|----------------|--|
| PI3K | Phosphoinositide 3 Kinase |
| PKC | Protein kinase C |
| PPAR- γ | Peroxisome proliferator-activated receptor gamma |
| PPP2R2B | Phosphatase 2 regulatory subunit B |
| PRC | Polycomb Repressive Complex |
| PR | Progesterone Receptor |
| PRMT | protein arginine methyltransferases |
| PTMs | Post-Translational Modifications |
| R | Arginine |
| Raf | Rapidly Accelerated Fibrosarcoma kinase |
| RAGE | Receptor for Advanced Glycation Endproducts |
| RAR- β | Retinoic acid receptor beta |
| Ras | Rat sarcoma GTPase |
| RASSF1A | Ras association domain family 1 isoform A |
| RER | Regional epigenetic regulation |
| RP-HPLC | Reversed-Phase High Pressure Liquid chromatography |
| RSK2 | Ribosomal protein S6 kinase, 90kDa, polypeptide 3 |
| RT | Room temperature |
| RT-qPCR | Reverse Transcriptase quantitative Polymerase Chain Reaction |
| RTS | Rubinstein-Taybi syndrome |
| SAHF | Senescence-associated heterochromatin foci |
| S | Serine |
| SAR | Scaffold Associated Regions |
| SC | Stem cells |
| SD | Standard deviation |
| SERM | Selective Estrogen Receptor Modulators |
| Shc | Src homology 2 domain containing) transforming protein |
| shRNA | Short hairpin RNA |
| Slug | Zinc Finger protein SNAI2 |
| Smad | Small Mother Against Decapentaplegic |
| Snail | Zinc Finger Protein SNAI1 |
| SDS-PAGE | Sodium Dodecyl Sulphate - PolyAcrylamide Gel Electrophoresis |
| SOX2 | SRY (sex determining region Y)-box 2 |
| SP1 | Specificity Protein 1 |
| STAT3 | Signal Transducer And Activator Of Transcription 3 |
| Suv39h1 | Suppressor of variegation 3-9 homolog 1 |
| SUZ12 | Suppressor of zeste 12 |
| SWI/SNF | SWItch/Sucrose Non-Fermentable |
| T | Threonine |
| TAR | transactivating response element |
| TBP | TATA binding Protein |

| | |
|--------------|--|
| TEMED | N,N,N',N'-Tetramethylethylenediamine |
| TF | Transcription factor |
| TGF β | Transforming Growth Factor-beta |
| TIC | Tumor Initiating Cells |
| TLR | Toll-like receptor |
| TNBC | Triple Negative Breast Cancer |
| TNF α | Tumor Necrosis Factor α |
| TNM | Tumor Node Metastasis |
| TPA | 12-O-Tetradecanoylphorbol-13-acetate |
| TSS | Transcription start site |
| TWIST1 | Twist-related protein 1 |
| TP53 | Tumor Protein P53 |
| UTR | Untranslated Region |
| ZEB | Zinc finger E-box-binding |
| WB | Western Blot |
| Wnt | Wingless/integrated signalling molecules |

INTRODUCTION

Breast Cancer

Breast cancer is the second most common cancer worldwide, accounting for the 12% of the new cancer cases estimated in 2012 and ranks second as leading cause of cancer death among women in developed countries (Ferlay et al. 2015). Since the 1977, in which tamoxifen has been approved as the gold standard method to treat ER+ breast cancer tumors (roughly 75% of all tumors), and the 1990's with trastuzumab treatment for Her2/neu+ tumors, breast cancer mortality rate has decreased in many countries according to cultural and economic barriers involved in women lifestyles (Servick 2014). Breast cancer is considered a Western women disease because it has the higher incidence rates in the higher income countries; Northern America, Northern and Western Europe and Australia are indeed at the top of the list even if the mortality rate in most cases seems to have the opposite trend. This has been associated with different geographical distribution of risk factors along with early detection with regular screening and the possibility to afford state-of-the-art treatment (DeSantis et al. 2015). However, the increasing breast cancer incidence in developing and low-income countries (i.e. Asia, Africa and South America) could mirror westernization process in terms of urbanization and economic development that are in strict relationship with life styles changes promoting obesity, physical inactivity, alcohol consumption, and delayed childbearing (Torre et al. 2015). The spread of bad costumes and the growth and aging of population worldwide will amplify the numbers of new cases of breast cancer suggesting that the understanding of molecular basis of malignant processes is a pressing need to fight against breast cancer.

Breast cancer classifications

Mammary gland consists of two compartments, epithelia and stroma, which responding to local and endocrine signals are responsible for milk secretion upon pregnancy. Two different lineages of mammary gland epithelial cells exist: luminal cells, outlining the ductal lumen and secreting milk in the lobular-alveolar compartment, and the basal/myoepithelial cells placed below the former and allowing cells contraction for milk release. Stem cells (SC) localize inside the duct and can differentiate in all breast cellular types through asymmetric division during normal breast development or to continue to cycle. Stem cells maintain self-renewal

capacity and are hypothesized to accumulate genetic changes and become breast cancer stem cells (BCSC), which are thought to be able to generate all cancer cell populations (Ellsworth et al. 2016). As other solid tumors, breast cancer looks like a heterogeneous population of cells with different replicative potential depending on access to oxygen and nutrients in the microenvironment. In fact, necrotic regions of tumors are characterized by the presence of quiescent cells that can grow slowly or not grow at all. Tumors usually grow in hypoxic conditions and the surrounding stroma provides vascularization elements (angiogenesis) that provide nutrients for tumor growth (Carels et al. 2016). Moreover, breast cancer owns genomic variability within primary tumors, metastases, circulating tumor cells (CTC; the ones which migrate from the primary sites to the vascular system), disseminating tumor cells (DTC; which migrate from the primary sites to the bone marrow), and cell-free DNA released by apoptosis/necrosis/macrophages phagocytosis (Ellsworth et al. 2016).

Rather than a single disease, breast cancer seems to be a heterogeneous collection of diseases reflecting morphological, molecular, and clinical differences. Despite the great heterogeneity, histopathological profiles of breast cancer are unique and predict disease courses and outcomes. Breast cancer arises generally from cells of the inner epithelial lining of the duct or the lobules; indeed 95% of breast malignancies are adenocarcinomas. Terms as lobular or ductal generally refer to the site in which cancer originates whereas the denominations invasive/infiltrating concern the penetration of cancer cells into the stroma through the ductal/lobular wall. Ductal carcinoma in situ (DCIS) is a non-malignant intraductal proliferation that can potentially give rise to invasive ductal carcinoma (IDC) (**figure 1.1**). Both types of ductal carcinoma originate from the same terminal duct lobular unit but in situ carcinoma presents usually myoepithelial cells of the outer ductal layer preserved. Lobular carcinoma in situ (LCIS) is an intralobular proliferation which represents likewise a risk factor and non-obligatory precursor of its invasive form, the invasive lobular carcinoma (ILC; **figure 1.1**). Some IDCs can be classified accordingly to their distinctive morphological features into well-characterized histological subtypes; however, 75% of them exhibit wide range of morphological and clinical behaviours and are defined as Not Otherwise Specified (NOS) or No Special Type (NST; Makki 2015). Microscopic morphological differences between cancer and normal breast cells correlate with the degree of malignancies, and in particular semi-quantitative evaluations of tubules, nuclear pleomorphism and mitotic counts provide a grading system that aid in classifying tumors with respect to differentiation level and in giving a certain prognosis (Elston & Ellis 1991).

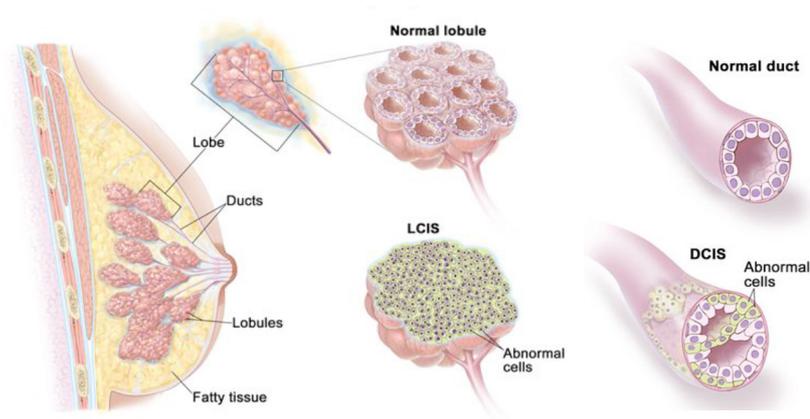


Figure 1.1 Breast cancer is mainly adenocarcinoma arising from epithelia of the lobules and ducts. Both lobular and ductal carcinoma in situ are non-invasive abnormal proliferation of the lining epithelium of the breast, respectively in lobules and ducts, which can give rise to their invasive forms (PDQ Adult Treatment Editorial Board 2016, modified).

Since 1950's the International Union against cancer provides a staging system for solid cancer classification in order to deal with therapeutic decision, the Tumor Node and Metastasis system (TNM). This system describes the extent of the disease at the time of primary treatment, with Tumor size, involvement of regional lymph Nodes, and distant Metastasis as indicators of prognosis. TNM classification was born to discriminate between operable breast cancers from metastatic ones or best candidates for breast conserving therapy, but it suffers of its own limitations even if it has improved during these years coming now to its 8th edition. Besides misleading terminology (i.e. terms as carcinoma used for non-malignant intraductal proliferation), cancers belonging to same categories usually differ from prognostic point of view. For instance, tumors that have few millimetres and tumors with 2 centimetres of dimeters, fall in the same T1 group even if they can display huge differences in prognosis. Categories should be dynamic and flexible and, moreover, they should include more precise prognostic indicators, such as gene expression profiles and others molecular information, making the way straight for personalized medicine (Veronesi et al. 2006; Arnone et al. 2010).

The molecular portrait of breast cancer

Among the main hallmarks of breast cancer is the “proliferative sustaining signalling” through hormonal and growth receptors with estrogen receptors (ER) representing the most important biomarker since 1970s. ER α and ER β are two isoforms of nuclear transcription factors activated by steroids (i.e. estrogen), both expressed in normal mammary gland tissues developing breast cancers but with prevalence of ER α isoform in the pathogenesis of cancer

(De Abreu et al. 2014). 75% of all breast cancer cases are ER+ and positively respond to endocrine therapy with better outcome after surgery.

Progesterone is a steroid hormone which drives growth and differentiation of breast during pubertal expansion and has been associated with the replicative potential of mammary gland. Progesterone receptor (PR) is induced upon ER signalling activation and rarely PR+ tumors are ER-. On the other hand, 40% of ER+ cancers are PR- indicating less responsiveness to endocrine therapy and suggesting the importance of PR expression assessment on breast cancer classification (Dai et al. 2016).

Human Epidermal growth factor receptor 2 (HER2; HER2/neu) or proto-oncogene erb-b2 receptor tyrosine kinase 2 (ERB-B2) is a membrane receptor belonging to a family of tyrosine kinase receptors involved in cell proliferation and survival. Its overexpression in more aggressive breast cancers is linked to gene amplification and has been identified in 15-20% of all breast cancers. Assessment of HER2 status is performed by validated immunohistochemical (IHC) and *in situ* hybridization (ISH) approaches (Nitta et al. 2016). HER2+ is linked to relative endocrine resistance consistent with the inverted correlation between ER/PR and HER2 expression (Dai et al. 2016).

A better stratification of breast cancer patients and therefore better prognostic and/or predictive indications could be achieved considering other distinctive characteristics, such as the activation of invasion/metastasis programs, evasion from the immune system surveillance, deregulation of cellular energetics, genome instability/mutations, and resistance to cell death (Carels et al. 2016).

The development of technologies and genetic knowledge allowed the set-up of the intrinsic molecular subtypes classification which helped to shed light on the tumor heterogeneity. Perou and colleagues in early 2000 outlined four different breast cancer subtypes based on their gene expression profiles (Perou et al. 2000). Among the great diversification of ER+ and ER- subgroups, the former seems to account for luminal breast phenotype, while the latter groups three different subtypes, the basal like, the *Erb-B2* overexpressing, and the normal breast (with adipose and basal-like markers expression and low detection of luminal markers). Luminal subtype reminds luminal ductal/epithelial cells of the breast concerning low molecular weight cytokeratins expression (cytokeratins 8 and 18) whereas basal phenotype expression pattern is more similar to basal epithelial cells or myoepithelial cells of the breast (expressing cytokeratins 5/6 and 17). In 2001, the same group has divided luminal subtypes in at least two more different subgroups, Luminal A and B according to the expression of ER, GATA binding protein 3, X-box binding protein 1, trefoil factor 3, hepatocytes nuclear factor 3, and estrogen

regulated LIV-1 (Sørli et al. 2001). Luminal A subtype comprises usually low-grade tumors with good prognosis while luminal B cancers express higher levels of proliferative markers, such as Ki-67, have higher histological grade and poorer prognosis (Makki 2015). Basal like subtype is defined by a gene expression profile similar to the one of basal/myoepithelial cells of normal breast even if the debate on its cellular ancestor is still opened. The majority of tumors belonging to this heterogeneous subtype is made by cells lacking expression of ER, HER2 and PR, the so-called Triple Negative Breast Cancer (TNBC) phenotype, to the point that they were thought to be synonymous (Kreike et al. 2007). However, 18-40% of basal-like breast cancers analysed by immunohistochemistry are not TNBCs. On the other hand, even if 80% of TNBCs belongs to basal-like subtype, this category encompasses also other molecular subtypes. Together, basal-like and TNBCs account for 15% of all invasive breast cancers and present common features, such as high histological grade, prevalence on specific racial/ethnic groups, BRCA1 susceptibility and a surface marker profile similar to the one hold by stem cells (CD44+CD24-;Foulkes, Smith, and Reis-Filho 2010).

Triple Negative Breast Cancer (TNBC)

Hormone-receptors negative breast cancers account for 12-17% of all breast cancers and stand for their aggressiveness and poorer outcomes. Triple negative phenotype is often diagnosed in young women, spreads among black ethnicities, and is detected between two mammographies (interval cancer), with 18% of tumors occult in initial mammography. It has an unfavourable histopathological profile (high grade, high tumor size, and lymph node positivity) and tends to form metastases at lungs and brain (Kumar & Aggarwal 2016). Survival curves demonstrated that patients with TNBC are likely to die within 3-5 years after diagnosis even if after this period the risk of distant relapse is lower compared to other subtypes (Foulkes et al. 2010). 60-80% of breast cancers displaying the germline mutation of the susceptibility gene BRCA1 possess TNBC phenotype indicating that BRCA1 pathway dysfunction could be involved in TNBC biology (Reis-Filho & Tutt 2008).

TNBC is a heterogeneous group of orphan breast cancers that exhibit different clinical and morphological features with dissimilar response to therapy strategies. They have been classified in different molecular subtypes according to their molecular profiles: i) the basal-like 1 and 2, with high levels of cell-cycle and DNA-damage response pathway elements activated; ii) the mesenchymal/mesenchymal-like, with epithelial to mesenchymal transition (EMT), cell motility, and growth signalling pathway genes highly expressed; iii) the immune

enriched, with genes involved in immunological processes triggered; iv) the luminal androgen receptor (AR), with high level of androgen signalling (Ahn et al. 2016; **figure 1.2**). The “claudin-low” subtype converges towards the mesenchymal phenotype and it is characterized by low expression of genes coding for tight junction or cell-cell adhesion proteins, such as Claudin 3, 4, and 7, Occludin, and E-cadherin (Herschkowitz et al. 2007). Related to this particular subtype are the Metaplastic Breast Cancers (MBCs), which are triple negative and belong to mesenchymal/sarcomatoid/squamous metaplasia of breast carcinoma: these tumors are unique for their DNA copy alterations and share with the claudin-low subtype stem cell and chemoresistance properties (Hennessy et al. 2009).

Current treatment

Surgery is the first line treatment of breast cancer and consists mainly of two approaches: breast conserving surgery (BCS), when tumor is localized or regional; mastectomy, in presence of larger or multiple tumors. Most of women diagnosed in early stages undergo BCS with adjuvant therapy (chemotherapy, radiotherapy, or hormone therapy) whereas same percentage (59%) of women with high grade breast cancer face mastectomy (Desantis et al. 2014). Adjuvant therapy consists of a series of treatments that are given after the main surgery in order to prevent distant recurrences or the occurrence of contra-lateral breast cancers (Salmon 2016). Radiotherapy and chemotherapy hit actively dividing cells through DNA damage or production of oxygen free radicals that activate apoptosis mechanisms. Widely used chemotherapeutic agents are methotrexate, 5-fluoruracil, taxane, cyclophosphamide, and anthracycline, i.e. doxorubicin (Davis et al. 2014). Chemotherapy is generally given as a combination of drugs that are more effective than monotherapies in certain subtypes of breast cancer, i.e. node positive, inflammatory, and advanced cancers. However, the single chemotherapeutic agent has fewer side effects and is preferred for women who no longer respond to other treatments (Carels et al. 2016). For both hormone receptor expressing and TNBC early tumors, adjuvant cytotoxic chemotherapy is based primarily on anthracyclines and taxanes; in particular, BRCA1/2 and TP53 mutated cancers well respond to platinum agents, which cause double strand breaks and prevent breast cells proliferation (Miller et al. 2014). Nowadays, personalized prognostic and predictive approaches to manage breast cancer are available. Immunohistochemical detection of expression of ER, PR, and HER2 currently drive adjuvant therapeutical approaches based on endocrine, anti-HER2 or chemotherapy strategies (De Abreu et al. 2014).



Figure 1.2 TNBC subtypes are enriched for distinct gene ontologies. The public available Lehmann classification of triple negative breast cancers emerges from k-means clustering of the most differentially expressed genes of a training set of TNBCs and is used to predict the subtypes of a validation set. By gene set enrichment analysis results shown on the right, each subtype displays unique gene ontologies specific for top canonical pathways (Lehmann et al. 2011).

ER expression is one of the most important features taken into consideration for the selection of an endocrine therapy. Currently, two are the main strategies to fight ER expressing breast cancers: the use of selective estrogen receptor modulators (SERMs, such as tamoxifen, raloxifene and toremifene) or aromatase inhibitors (AIs, non-steroidal [letrozole and anastrozole] and steroidal [exemestane]). SERMs competitively bind to ER interfering with

DNA synthesis and cell cycle progression towards G1 from G0. AIs instead block testosterone conversion to estradiol (E2) and androstenedione to estrone by aromatase which is the main source of estradiol in post-menopausal women (Miller et al. 2014). HER2 overexpression is indicative of cancers responsiveness to monoclonal antibodies blocking the receptor activation, such as trastuzumab and pertuzumab, or small molecules inhibitors, such as lapatinib (De Abreu et al. 2014).

Neoadjuvant treatments arise as methods that allow breast conservative strategies or surgery where it was technically complicated (Salmon 2016). To this end, therapies based on endocrine drugs are potentially harmful because of their slow response rate which could delay surgery. In fact, most of neoadjuvant treatments consist of chemotherapy both in receptor positive or negative cancers even if in TNBC has shown controversial results: TNBCs have a better chemo-sensitivity compared to patients affected by other breast cancer subtypes in terms of pathological complete response (pCR) but after neo-adjuvant chemotherapy still exhibit worse outcomes in terms of disease free survival (DFS-period of time from diagnosis to evidence of disease) and/or overall survival (OS-time lapse from diagnosis to death; Miller et al. 2014). Historical therapeutical trials for TNBC patients are still lacking.

Current breast cancer treatment trends address non-toxic potent targeted therapies and the advent of the *-omic* era has provided personalized prognostic and predictive approaches to manage tumors, even if with gross limitations: anti-angiogenetic drugs, poly-ADP-ribose polymerase (PARP), epidermal growth factor receptors (EGFR), androgen receptor (AR), phosphoinositide 3 kinase (PI3K), mitogen-activated protein/extracellular signal regulated kinase (MEK), checkpoint-kinase 1 (CHK1), and histone deacetylase (HDAC) signalling inhibitors are used in pre-clinical and clinical studies to treat advanced TNBC in place of cytotoxic treatment (Palma et al. 2015). Given the high heterogeneity of breast cancer and TNBC in particular, the prospective is to deeply understand the biology of these cancers in terms of pathways activations in order to define more feasible TNBC subtypes and to develop drugs targeting/affecting those specifically altered molecular mechanisms.

Epithelial to Mesenchymal Transition (EMT)

Epithelial cells are apico/basal-polarized units of regular shape which lean on a basal lamina keeping contact with other cells by adherens junction, desmosomes and tight junction. E-cadherin (epithelial calcium dependent adhesion molecule) is a transmembrane glycoprotein which composes adherens junction through homophilic interaction in a Ca^{2+} dependent way. Dependent on the tissue type, this cadherin is expressed in the entire circumference of the cells (non-polarized) or in peripheral intercellular borders (Eidelman et al. 1989). E-cadherin interacts by its cytoplasmic domain with β -catenin which in turn anchor cadherin complex to cytoskeletal actin. In contrast, mesenchymal cells are spindle-shaped cells which are loosely organized in extracellular matrix arranging connective tissue. Tissues containing these cells (i.e. connective tissue, lymphoid organ, cardiovascular system) are devoid of E-cadherin expression (Eidelman et al. 1989) with the consequent disappearance of specialized junction. Along with changes in surface proteins, mesenchymal cells exhibit a specific organization of cytoskeleton components, differential transcriptional factor expression, and organelle distribution (Zeisberg & Neilson 2009).

Epithelial to Mesenchymal transition is of utmost importance during morphogenesis, a process in which cells de-differentiate from polarized epithelial cells to mesenchymal phenotype in order to migrate and define the complex body plan in multicellular organism. Gastrulation and neural crest delamination are two primary EMT events occurring early in embryonal development, followed possibly by secondary and tertiary EMTs in specific organ development (Thiery et al. 2009). During this functional transition, cells lose polarity and epithelial marker expression, switch cadherins, rearrange actin-stress fibers, increase synthesis of cytoskeletal proteins, change transcriptional factors expression, relocate on fibrillar extracellular matrix (ECM), and acquire spindle shape and migratory capabilities (Zeisberg & Neilson 2009).

The plasticity of phenotypes acquired by this cellular transdifferentiation is underlined by the reversibility of the transition. Indeed, cells can undergo mesenchymal to epithelial transition (MET) once reached their final destination. The term transition is therefore preferred to transdifferentiation since it underlies the transient nature and reversibility of the process (Zeisberg & Neilson 2009).

Three main types of EMT have been defined: i) the first is related to implantation, embryo formation and organ development; ii) the second deals with wound closure, tissue regeneration

and organ fibrosis and iii) the third type characterizes neoplastic cells which are enabled to invade and metastasize during the final stages of cancer progression (Kalluri & Weinberg 2009). Concerning the latter case, the observation that cells relying on the invasive front of primary tumors acquire a mesenchymal phenotype leads authors to consider EMT as part of the invasion-metastasis cascade and to be responsible for the escape from the primary tumor mass, intravasation, transport in the circulating system, extravasation, micrometastasis formation, and colonization (Kalluri & Weinberg 2009).

The role of EMT in tumor metastasis

The vast majority of deaths related to solid cancer is due to occult metastasis occurring in lung, liver, brain, and bone marrow. However, metastatic process is for its nature very ineffective: in one cubic centimeter of solid tumor mass containing one billion of cells only <0.1% of tumor cells successfully develop distal metastasis (Mack & Marshall 2010). As reviewed by van Zijl F and co-authors, the origin of metastasis is still a debated question. Two are the main hypothesis: the “linear progression model”, suggesting that dissemination occurs after primary tumor growth, and the “parallel progression model” claiming an early migration of disseminating cells from a relative small primary tumor of 1-4 mm size. Cells can invade mainly by two different mechanisms: as epithelial sheet, maintaining cell-cell contacts and leading/tip cells giving a front-rear polarity and interacting with microenvironmental stimuli, or as single mesenchymal/amoeboid cell (Van Zijl et al. 2011). EMT has been thought as an early process in cancer progression that enables the subsequent colonization of distant sites and the formation of micro- and macrometastasis for cells coming from the primary tumor through the acquisition of self-renewal capability (Mani et al. 2008). Both EMT and stemness represent indeed cornerstones which enable primary carcinoma cells to invade and to form the bulk of a second tumor mass (**figure 1.3**).

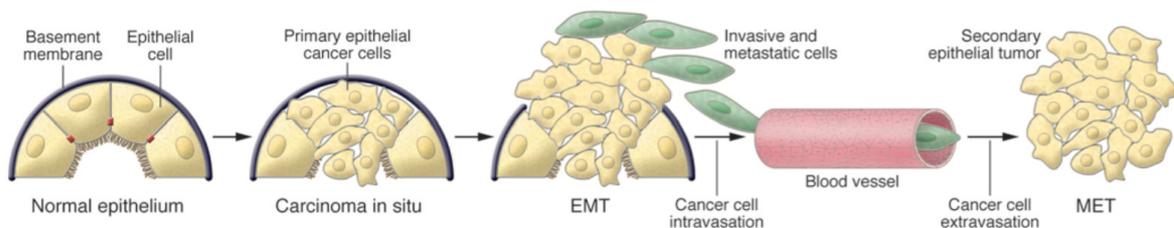


Figure 1.3 EMT and MET are mechanisms that trigger cancer evolution toward metastasis. Cancer is thought as a multistep process in which cells de-differentiate losing cell polarity and cell-cell contacts, detach from the basement membrane, acquire plasticity and motility modifying cytoskeletal proteins and acquiring new sensory extensions through an epithelial to mesenchymal transition, intravasate in the blood stream (or lymph stream) where they disseminate and through emboli or contact with endothelial cells they extravasate. The

formation of micro- and subsequent macro-metastasis require that cells undergo a new transition toward epithelial phenotype (MET) and proliferate defining a new metaplastic lesion. The tumorigenic and self-renewal properties of cells allow proliferation, resistance to apoptosis and senescence stimuli and suppression of immune system reactions in the macrometastasis development (Kalluri & Weinberg 2009).

Once cells acquire proliferation potential, resistance to apoptosis, and have a triggered metabolism, they firstly need to lose polarity and detach from the basement membrane in order to proceed in the metastatic process (Mack & Marshall 2010). Evidences that E-cadherin expression correlates with invasive breast carcinoma phenotype arise in 1997 along with the awareness that in invasive ductal carcinoma its loss might contribute to tumor dedifferentiation and hormone insensitivity (Siitonen et al. 1996). Lombaerts et al. demonstrated that inactivation of E-cadherin expression (encoded by *CDH1*) by promoter methylation is a part of the EMT program (Lombaerts et al. 2006). The relationship between cancer cells undergoing EMT and Extracellular Matrix components (ECM, i.e. collagen, fibronectin, laminin, and elastin) is another important point in the invasion step: cells gradually change integrin expression repertoire and up-regulate matrix metalloproteinases (i.e. MMP2 and MMP9) enabling ECM degradation and evasion from primary site towards lymphatic and blood vessels, in a process called intravasation (Lamouille et al. 2014; Mack & Marshall 2010). Among the well-known master regulators of EMT are transcriptional factors involved in E-cadherin repression during mesenchymal differentiation. Some of these factors bind directly E-boxes present in the promoter of *CDH1*, such as the zinc fingers Snail, Slug, ZEB, and the bHLH E47, some others act indirectly, without binding E-cadherin promoter, such as the bHLH Twist1, the forkhead FOXC2, and the homeobox Gooseoid (Yang & Weinberg 2008). The main signalling pathways involved in the expression of these transcriptional factors are TGF β , Notch, and Wnt pathways that can cooperate with other pathways (i.e. MAPK and PI3K pathways) and signals coming from tumor microenvironment (i.e. Hypoxia). The best studied model is TGF β -dependent activation of Smad2 and/or Smad3 that form a trimeric complex with Smad4. This complex migrates in the nucleus where it activates transcription of proteins involved in EMT (Lamouille et al. 2014).

EMT molecular profile is currently used to try to stratify patients based on predicted outcome. Meta-analysis on gene expression signatures shows indeed a correlation between EMT-core gene list and pathological complete response (pCR) in breast cancer patients (Gröger et al. 2012). In 2015 more than 100 clinical trials had EMT as key word and try to define a clear EMT expression pattern in different kind of cancers and tumor stages in order to use them as prognostic factors to tailor treatments (**figure 1.4**; Pasquier et al. 2015).

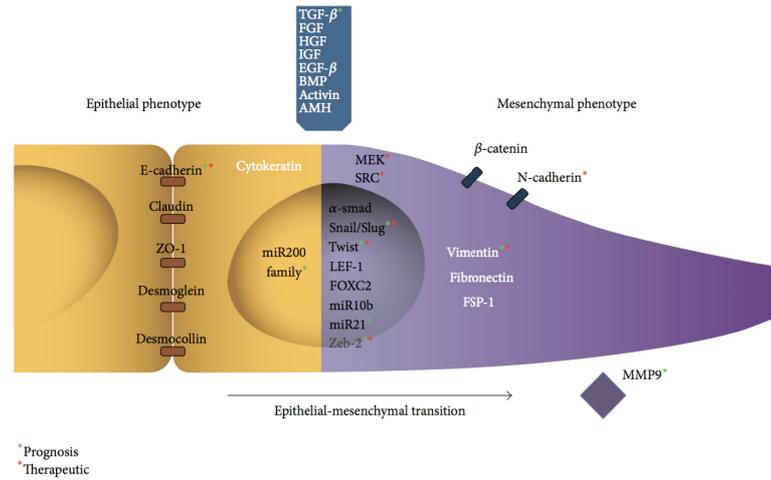


Figure 1.4 Molecular biomarkers of EMT used in ongoing cancer clinical trials for prognosis and therapeutic intervention. The figure summarizes EMT inducers and effectors marked with red or green asterisk if studied respectively in therapeutic or prognostic field (Pasquier et al. 2015).

However, one of the main problems concerning EMT process in cancer pathogenesis is the difficult to distinguish between cell undergoing EMT and stromal cells or tumor associated fibroblasts together with the lack of convincing evidences in clinical samples (Thiery et al. 2009). Often cancer cells do not need a complete EMT process to escape from the primary tumors and once they reach the remote site they undergo MET to colonize and proliferate. In fact, one of the main reasons of the controversy in EMT-based treatments relies on assumption that EMT is a permanent irreversible process whereas it seems not to be the case (Yang & Weinberg 2008). Another critical issue for the use of EMT markers in clinical trials is tumor heterogeneity: their expression widely varies among tumor sections, increasing mesenchymal markers in tumor periphery, and it is very difficult to establish cut offs in order to meet the stringent parameters of robustness and reproducibility required for clinical prognostic markers to be used as adjuvant therapy. Most of the ongoing clinical trials are trying to target inducing signals rather than EMT effectors and regulators (Pasquier et al. 2015) but none of them has translated into routine clinical practice yet. However, the understanding of precise combination of tumor microenvironmental signals and cellular molecular profiles in each step of cancer evolution could lead to the definition of an appropriate therapeutic options for patients.

High Mobility Group Proteins drive EMT in cancer cells

Among the inducers of pathological EMT is the aberrant activation of signalling pathways induced by inflammatory cytokines such as TGFβ, TNFα and IL-6. The inflammatory

response appears to elicit the tumor progression associated with High Mobility Group (HMG) proteins overexpression. HMGB1 protein could indeed act as extracellular inflammatory mediator and therefore enhances tumor development supporting chronic innate immune system inflammation. HMGB1 which is passively released from necrotic cells or actively secreted by inflammatory cells bind different types of receptor including RAGE, TLR2, TLR4, TLR9 and CD24 to promote inflammation (Tang et al. 2010). Recently, it has been suggested that in the necrotic and hypoxic breast tumor core, geminin overexpressing (GemOE) cells survival is connected to the presence of acetylated HMGB1 in the microenvironment. This oncoprotein is able to induce the recruitment of mesenchymal stem cells that in turn crosstalk with and enhance tumor initiating cells (TIC), stimulate the acquisition of a basal phenotype and induce an EMT on the necrotic GemOE tumor core (Ananthula et al. 2016).

The chromosomal HMGN proteins are involved in a TGF β -Smad driven EMT that seems to be enhanced by the pro-inflammatory cytokine TNF α in lung cancer cell lines. In this model, miR-23a is a target of TGF β pathway and in turn regulates HMGN2 expression (Saito et al. 2013). From recently published data, also HMGN5 (NSBP1) 3'UTR is targeted by miR-186, in a process that culminate with cadherins switch and vimentin down-regulation (Yao et al. 2015).

Tumor overexpressing HMGA proteins have been clearly correlated to highly aggressive and metastatic phenotype. The expression of these proteins has been demonstrated to be linked with both EMT and chemoresistance. In colon-rectal and breast cancer, HMGA2 enhances TGF β signalling and correlates with TGF β RII expression, β -catenin nuclear localization and loss of cortical E-cadherin in invasive front of primary tumors and metastatic lesions (Morishita et al. 2013). Also in prostate cancer, HMGA2 was highlighted as a driver of TGF β -induced EMT, and linked to metalloprotease MMP2 and MMP9 expression (Shi et al. 2016). Moreover, in a feed-forward mechanism, Smad enhances HMGA2 expression, which in turn controls other key EMT regulators in mammary epithelial NMuMG cells (Thuault et al. 2006), such as Snail and Twist (Thuault et al. 2008; Tan et al. 2012). In Snail induction, the AT hooks seem to allow HMGA2 binding to Snail promoter and the contact with MH1 and MH2 domains of Smad3, whereas C-terminal acidic tail appear as the transactivating domain for Snail expression (Thuault et al. 2008). Besides TGF β , HMGA2 support Ras-induced EMT in pancreatic cancer cells: HMGA2 is able to restore Snail expression affected by the inhibition of MEK1/2 activation of ERK1/2 by U0126 (Watanabe et al. 2009).

As previously shown for HMGN proteins, EMT is also affected by cytokines through mechanism involving miRNA. let7d miRNA has demonstrated to affect HMGA2 and mesenchymal markers expressions induced by TGF β in lung fibroblasts (Huleihel et al. 2014). In breast cancer model, down-regulation of let7 and miR-200 by oncostatin M (OSM, a member of IL-6) converge in HMGA2 expression and phenotype transition (Guo et al. 2013). In hepatocellular carcinoma another member of the let7 family, let7g seems to affect malignancy through the K-Ras/HMGA2A/Snail axis (Chen et al. 2014). Other epigenetic mechanisms are also involved in mediating TGF β -HMGA2 induction of EMT such as post-translational modification of DNA or histones (methylation of histone H3 in K9 or K27). HMGA2 are shown to be able to affect E-cadherin (*CDH1*) promoter DNA methylation by interacting with DNA methyltransferase 3A (DNMT3A) and favouring the detachment of CCCTC-binding factor (CTCF) a protein responsible for tumor suppressor genes insulation in mammary epithelial cells (Tan et al. 2015)

Reeves and colleagues early highlighted HMGA1 proteins connection with the EMT process in MCF7 in which HMGA1 was overexpressed. Once inoculated directly into the fat pats of nude mice, these cells were able to induce secondary tumors. Among genes that have been found regulated in epithelial breast cells overexpressing HMGA1, collagens, vimentin, cytokeratins, and MMP were up-regulated whereas EGF receptor and TGF β receptor III were down-regulated. Three years later, the same group discovered that HMGA1a protein regulates malignant metastatic progression of breast cancer cells triggering the Ras/Raf/MEK/ERK mitogenic pathway (Treff et al. 2004). Other evidences arose in poorly differentiated metastatic colon cancer cells where upon HMGA1 silencing Twist1 and vimentin were down-regulated and, on the contrary, E-cadherin was up-regulated (Belton et al. 2012).

Recently published data coming from our laboratory confirmed HMGA1 proteins role in breast cancer cell EMT process and stemness (Pegoraro et al. 2013). In fact, HMGA1 depletion in MDA-MB-231 cells causes a drastical change in cell morphology with a redistribution of F-actin in a cortical pattern and the reduction of stress fibers. The Wnt/ β -catenin and Notch pathways have been suggested as main players of this transition. Moreover, *in vivo* results highlighted HMGA1 active role in metastasis formation rather than in tumor growth. Among the most differentially expressed genes between controls and HMGA1-depleted cells, AURKB turned out to be one of the genes that mostly correlate with clinical outcome and cell migration. AURK pan-inhibitor, Danusertib, has recently shown to interferes with EMT process in malignant MCF7 and MDA-MB-231 breast cell lines (Li et al. 2015).

High Mobility Group proteins

The superfamily of High Mobility Group (HMG) proteins includes chromatin architectural factors that exhibit similar physical and chemical features and are able to tune DNA-related processes by binding to DNA or other chromatin proteins, first among others histones.

Three families, namely HMGA, HMGB, and HMGN proteins, which differ for their functional motifs and therefore for the specificity of their binding sites, are members of this superfamily.

HMGA proteins lack secondary or tertiary structures when free in solution and are characterized by modular sequence organization with three positively charged “AT-hooks” engaging minor groove of AT-rich DNA regions (Sgarra et al. 2004).

Conversely, HMGB proteins possess two homologous L-shaped domains, called HMG boxes, composed by three alpha-helices domains. These proteins mainly localize into the nucleus even if they can also be secreted activating immune response or inflammatory reactions (Müller et al. 2001).

The family of HMGN proteins binds nucleosomes by their positively charged domain, the nucleosomal binding domain, relaxing higher chromatin structure and enabling processes such as replication and transcription (Catez et al. 2002).

HMG proteins lack binding specificity towards specific genomic sequences although they prefer certain DNA conformations. The kinetics of HMG proteins binding to DNA molecules have shown dynamicity, with a sort of “stop and go” mechanism ensuring high proteins turnover and competition with other family members and histone H1 (Hock et al. 2007). In this way, HMG proteins cooperate in maintaining chromatin plasticity with dynamic global and local transitions occurring during major cellular events.

High Mobility Group A proteins (HMGA)

HMGA proteins were discovered about 30 years ago, by inducing transformation in fully differentiated Fischer rat thyroid epithelial cells with oncogenic Kirsten murine sarcoma virus. The tumorigenesis associated with this retroviral infection is due to expression of viral v-ras, which in turn activates the production of secondary elements, i.e. growth factors, able to sustain the transformation. HMGA proteins were induced as a consequence of the v-ras induced neoplastic transformation. These proteins were afterward identified in actively dividing cells but were almost absent in adult non proliferative tissues (Giancotti et al. 1985; Goodwin et al. 1985). By sequencing information available in 1991, it became clear that the new family

of proteins was composed of three closely related but different proteins, HMGA1a (previously known as HMGI), HMGA1b (previously known as HMGY) and HMGA2 (previously known as HMGI-C; Manfioletti et al. 1991). HMGA1a and HMGA1b are isoforms produced by alternative splicing of the same gene located at chromosome 6p21, which generate protein A1a of 107 aa and protein A1b of 96 aa (Johnson et al. 1989). For clearness purpose they will be referred together as HMGA1 protein throughout the text. The HMGA2 coding gene is longer than the one coding for HMGA1, it is located on chromosome 12q13-15 and encodes for a 109 aa long protein (Chau et al. 1995).

Besides neoplastic transformation, HMGA1 gene was found to be highly expressed in embryonic and fetal tissue with some quantitative differences between tissues and was barely detectable if not absent in adult rat and human tissues (Chiappetta et al. 1996).

Human embryonic stem cells (hESC) hold high levels of HMGA1 while it is turned off during differentiation along with OCT4, SOX2 and NANOG, all pluripotency factors. Furthermore, HMGA1 drives somatic cells de-differentiation into induced pluripotent stem cells (iPSC). Specifically, HMGA1 is shown to bind promoters of cMYC, SOX2 and LYN28 pluripotency/stem cell associated factors triggering their expression (Shah et al. 2012). Their absence or mutation lead to aberrant developmental phenotypes such as cardiac hypertrophy and type 2 diabetes in HMGA1 null mice and “pigmy” phenotypes (reduction till 60% of weight) in HMGA2 null mice (reviewed by Pallante et al. 2015). The double HMGA1 and HMGA2 knock-out synergistically resulted in the “superpigmy” mice phenotype, with about 80% of weight loss (Federico et al. 2014). Although severe phenotypic alterations caused by loss of HMGA proteins, their presence is not critical for vitality being functional redundant and acting in compensatory fashion with other HMG proteins (Reeves 2010).

HMGA1 dynamic interplay with chromatin

HMGA1 is a prototype of intrinsically disordered proteins (IDs) that adopts stable conformations once bound to its DNA target or other substrates (Huth et al. 1997). In this way HMGA1 is able to bind to a wide spectrum of factors that are in turn involved in a great number of biological processes. Through contacts with DNA and proteins, HMGA1 affects cellular growth, proliferation, differentiation, and apoptosis by acting locally or globally on the chromatin landscape (Sgarra et al. 2004).

An extensively studied model of locally regulated event by HMGA1 is the expression of Type I interferons (IFN- α and IFN- β), which are proteins secreted by immune system upon viral

infection and that interfere with viral replication. In the 65 bp region of DNA involved in IFN enhanceosome formation, HMGA1 has found to enhance the binding of several transcription factors, i.e. NF- κ B, IRFs, ATF-2/c-Jun by inducing conformational changes onto DNA stretches and only later promoting assembly by protein-protein interaction (Yie et al. 1999). The enhanceosome surface contacts two histone acetyltransferase (HAT) GCN5/PCAF and CBP/p300 which in turn acetylates both histone's tails and HMGA1, respectively on K71 and K65. These transcriptional activators work synergistically to recruit basal machinery and amplify gene expression (Merika et al. 1998; **figure 1.5**).

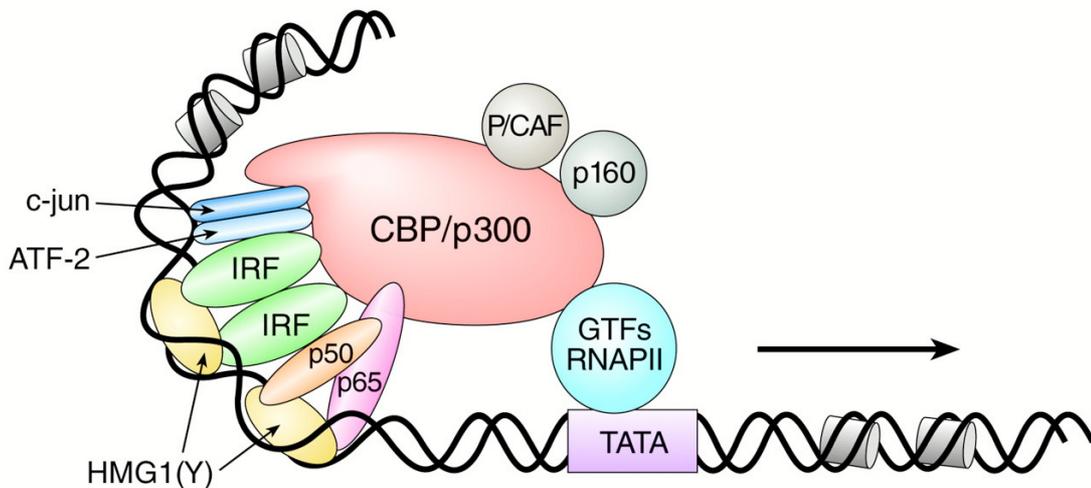


Figure 1.5. The stereospecific complex of enhanceosome. The illustration represents the enhanceosome formation through co-activators and RNA PolII recruitments and DNA conformational changes on well-studied model of IFN- β . HMGA proteins, indicated as HMGA1(Y), interact with IRF, ATF2/c-jun and NF- κ B subunits p50 and p65 and imparts DNA bending (Vo & Goodman 2001).

HMGA1 seems to play with a “hit and run” mechanism rather than being a stable component of the enhanceosome and its modifications, i.e. acetylation, act as molecular switch of its function (Reeves 2001).

Regulatory regions of genes could be studded with occluding nucleosomes that mask the binding sites of transcription factors. Sometimes AT-rich regions are exposed on nucleosomal DNA or they flank core nucleosomes and could overlap the binding sites of other activators. In all these cases, HMGA1 acts mainly enhancing chromatin remodelling complexes displacing nucleosome and forming enhanceosome (Reeves 2010). Among other genes, the inducible expression of cytokine IL-2 that sustains the T-cell autocrine loop growing, is tuned by HMGA1 which triggers the binding of transcription factors to its weak promoter lacking their consensus motif (Himes et al. 2000). IL-2 high affinity receptor, IL-2R α , is likewise regulated by HMGA1 with formation of stereospecific complexes (John et al. 1995).

Aside from canonical binding to B-form DNA and genomic sequences wrapped around nucleosome core particles, HMGA1 is able to recognize unusual structures, such as four-way junction (Cleynen & Van de Ven 2008). Four-way junction DNA mimics Holliday-type recombination crossover structures and the linker DNA close by the entrance and the exit of nucleosomes; HMGA1 has shown the ability to displace histone H1 through the competition for the same binding sites in these structures (Hill & Reeves 1997). Similar evidences of antagonism with repressive histone H1 has emerged from scaffold-associated region (SAR) involved in chromatin loops delimitation. HMGA proteins are thought to act similar to distamycin, an antibiotic which blocks the interaction of proteins with AT rich sequences of SAR by amide-base hydrogen bonds and van der Waals interactions; distamycin, Hoechst 33258 and netropsin structural motives are very similar to the AT-hooks and compete with H1 for the binding to SARs. Moreover, HMGA have shown to lead to histone H1 outplacement in non SAR-DNA regions (Zhao et al. 1993).

During interphase HMGA1a distribution overlaps the one of histone H1 preferentially localizing on heterochromatin (Harrer M et al., 2004). In contrast to their activating functions, HMGA proteins are also structural components of senescence-associated heterochromatic foci (SAHS) along with heterochromatin protein 1 (HP1) and its docking site, the histone H3 trimethylated on lysine 9 (H3K9me3). The senescence program supplies cell cycle arrest with the repression of genes related to proliferation. In striking contrast with evidences that link HMGA1 to open chromatin configurations sustaining tumorigenic factors expression, these studies suggest an anti-proliferative action of HMGA proteins triggering senescence, a process that possess clearly tumor-suppressive features (Narita et al. 2006). HMGA are thought to act in mechanisms that clearly depend on environmental and intracellular contexts and owing the natural plasticity required to adapt to a great number of target it is able to mediate even opposite functions.

Basal and oncogenic factors which converge in HMGA1 expression or modulate its function

Several signalling pathways impair HMGA1 expression in response to extracellular stimuli. From the analysis of the 2 kb region upstream the transcription start site of HMGA1, three functional regulatory regions have been characterized, one of them being a distal region while two of them being proximal to the TSS. Although basal transcription involves the cooperation of specificity protein 1 (SP1) and activator protein 1 (AP1) in the proximal regions, the

overexpression of HMGA1 stimulated by oncogenic Ras through mitogen activated protein kinase (MAPK) cascade involve the distal or enhancer region (Cleynen & Van de Ven 2008). Transforming growth factor β (TGF- β) promotes HMGA1 expression in breast cancer cell lines by phosphorylation of SP1 affecting its occupancy on HMGA1 promoter. LY294402 and wortmannin, two potent inhibitors of phosphoinositide 3 kinases (PI3Ks), suggest that TGF- β acts mainly through non-Smad signalling pathway to activate HMGA1 expression (Zu et al. 2015).

Likewise, the canonical Wnt/ β catenin pathway has shown the ability to regulate HMGA1 expression through c-myc effector; moreover, HMGA1 positively correlates with nuclear localization of β -catenin in gastric cancer tissues (Akaboshi et al. 2009).

Endogenous microRNA targeting HMGA1 also modulate its translation owing a tumor-suppressive feature; among others, miR-296 (Wei et al. 2011), miR-26a (Zhao et al. 2015), miR-142-3p (Xu et al. 2014), let7a (Liu et al. 2015), miR16 (Kaddar et al. 2009) miR-548c-3p and miR-34b (D'Angelo et al. 2012) are found down-regulated in cancer tissues compared to normal ones and most of them considerably affect proliferation and motility. HMGA1 action is modulated by RNA interaction: the nuclear non coding 7SK RNA is a factor which negatively affect Polymerase II elongation and influence HMGA1 biological functions by competing for the first AT-hook binding with DNA (Benecke & Eilebrecht 2015).

Other epigenetics mechanisms affect HMGA1 expression: by bioinformatics studies, eight pseudogenes, otherwise called “genomic fossils”, which belong to ancestor genes but lost their primitive functions through the accumulation of mutations are found high homologous to functional gene. HMGA1Ps (pseudogenes) are still able to affect gene expression in several ways: some could originate different proteins/peptides with alterations on post-translational modifications patterns and compete with HMGA1 for partners binding; others act as competitive endogenous RNA (ceRNA) which sequester miRNA regulating their action or modulate the binding of RNA stability factors (De Martino et al. 2016).

Last but not least, in addition to its turnover, localization, and natively-disordered structure, also post-translational modifications modulate HMGA1 activities (Sgarra et al. 2010). As review by Zhang Q and Wang Y, HMGA1 is the most phosphorylated protein in the nucleus and its phosphorylation interfere with DNA binding capabilities and interactions with other proteins (Zhang & Wang 2008).

Different levels of HMGA1 regulation, from transcriptional to post-translation mechanisms, let us understand how HMGA1 is a hub of chromatin biological functions, responding to extracellular environmental changes and orchestrating a myriad of intracellular events.

The causative role of HMGA1 in cancer

The pivotal role played by HMGA1 in cancer onset emerged since its discovery in early 1985 and subsequently many experiments confirmed a causative association between HMGA1 and different types of cancer rather than a marginal event of transformation. In 2001, Reeves and colleagues demonstrated that nude mice injected with human breast epithelial cells overexpressing HMGA1 developed primary tumors and metastasis and that HMGA1 expression in tumor cells is responsible for the induction of factors critical for cell signalling, proliferation, tumor initiation, invasion, migration, angiogenesis, and colonization (Reeves et al. 2001). Several other *in vitro* evidences supported this causal role of HMGA1 in cancer: HMGA1 antisense construct prevents the transformation of rat thyroid cells by Kirsten murine sarcoma virus carrying v-Ras-Ki oncogene (Berlingieri et al. 2002); adenovirus harbouring HMGA1 antisense cDNA has cytotoxic effects on lung, breast, colon, and thyroid neoplastic cells but no effects were reported on normal cells (Scala et al. 2000); HMGA1 ectopic expression in Rat1 fibroblasts enables them to form colonies in soft agar and provides tumorigenic potential when fibroblasts are introduced in athymic nude mice (Wood et al. 2000); in pancreatic cell lines, HMGA1 overexpression made cells unsusceptible for apoptosis mediated by cell-matrix interaction also known as anoikis (Liau et al. 2007). Furthermore, the advent of transgenic mice as laboratory tools allows to assess the oncogenicity of HMGA1 *in vivo*: HMGA1 knockout mice upon chemically induced skin neoplasia developed fewer tumors in respect to wild-type mice (Visone et al. 2008) and overexpression of HMGA1 under control of cytomegalovirus promoter led to pituitary adenomas and natural killer (NK)-T/NK cell lymphomas (Fedele et al. 2005).

Among benign tumors, especially of mesenchymal types, chromosomal rearrangements involving both genes coding for HMGA1 and HMGA2 have been observed (Fusco & Fedele 2007).

Since 1991, by differential hybridization analyses of steady-state mRNA belonging to prostatic tumors associated with different grade of malignancy, HMGA1 emerged as prognostic marker for advanced solid cancer (Bussemakers et al. 1991). Aside mesenchymal-originated tumors, HMGA1 is also overexpressed in human leukaemia of diverse origin and correlates with transformed state (Pierantoni et al. 2003).

As in-deep reviewed by Fusco and Fedele, HMGA promotes cell transformation by inferring with basic biological processes in cells, mainly through the regulation of the expression of tumor suppressor genes or transcriptional complex factors. Indeed, HMGA1 promotes AP-1

action, a complex that modulates expression of genes promoting malignant phenotype, by the induction of the member Fra1, triggering cyclin A expression and affecting cell-cycle progression. HMGA1 fine-tunes the expression of several other proteins involved in cancer progression such as HAND1, a transcription factor which play major role during differentiation and thyroid carcinogenesis, and STAT3, a key signal transducer. Inflammation is also modulated by transcriptional regulation exerted by HMGA1 on inducible nitric oxide synthase (iNOS), cyclooxygenase 2 (COX2), E-selectin, Immunoglobulin E, IL2, and IL4 genes; all these inflammatory players expression in turn results in suppression of apoptosis and boost of proliferation, angiogenesis, and invasion. Moreover, HMGA1 decreases expression of elements which enable DNA-damage recognition and repair. HMGA1 impacts also epigenetic mechanisms controlling gene expression; i.e. HMGA1 regulates miR-181 which in turn negatively affect expression of critical downstream elements, such as Polycomb gene CBX7 (Fedele & Fusco 2010 and references therein). Furthermore, as previously reported, HMGA overexpression is associated with up- or down-regulation of key markers of epithelial to mesenchymal transition, an event characteristic of advanced epithelial tumor cells promoting invasion and metastasis. HMGA2 have shown to sustains EMT though the association with DNMT3A and the modulation of DNA methylation at *CDH1* promoter which result in E-cadherin suppression (Tan et al. 2015).

Although HMGA1 works mainly through modulation of gene expression by well-known mechanism involving both DNA and protein interactions, it is also able to control biological events by interacting with crucial factors and modulating their action. The “guilty by association principle” led to the discovery of a wide range of functional involvement of HMGA1 beyond transcription (Sgarra et al. 2010). One of the most striking examples is the ways by which HMGA1 influences p53-mediated apoptosis: HMGA1 interacts with p53 altering downstream p53 effectors expression and regulates cytoplasmic localization of HIPK2, a kinase responsible for p53 activation through its phosphorylation (Fusco & Fedele 2007).

HMGA1 is a master regulator of breast cancer

Among others, also HMGA1 emerges as biomarker for prognosis and diagnosis of breast cancer: Chiappetta and colleagues demonstrated by immunohistochemistry that HMGA1 staining was higher both in lobular and ductal carcinoma compared to normal tissues, and in particular 40% of hyperplastic lesions without cellular atypia and 60% of ductal carcinoma

display high HMGA1 intensity staining (Chiappetta et al. 2004). Later on, studies confirmed that HMGA1 expression is higher both in term of mRNA and protein levels in breast cancer tissues compared with normal ones (Huang et al. 2015). Moreover, higher expression of HMGA1 was found among ductal carcinoma with respect to lobular breast cancer types (Zu et al. 2015). A recent study analysed HMGA1 expression on a tissue microarray with 1338 resection of breast tumors, mainly ductal carcinoma variants. Results confirmed HMGA1 absence on normal tissue whereas in breast cancer tissues its expression positively correlated with HER2/neu and progesterone expression and negatively correlated with estrogen receptor but none association was observed among histotypes, staging, and grading of tumors (Sepe et al. 2016). Conversely another study suggested that HMGA1 expression is higher in patients with high histological grade, worst clinical stage, larger tumor size, greater lymph node involvement, presence of distant metastasis, and having triple-negative status of hormone receptor expression and that their presence inversely correlates with overall survival of these patients (Huang et al. 2015). In fact, HMGA1 gene expression profile significantly crosses over the ones of others critical factors inferring with poor prognosis and high risk of distal metastasis (Pegoraro et al. 2013). Overlapping proteomics and gene expression data obtained from MDA-MB-231 cell lines transiently silenced for HMGA1 expression and control cells evidenced a HMGA1 signature with prognostic value in terms of overall-, relapse free-, and distant metastasis free-survival in breast cancer (Maurizio et al. 2016). Moreover, the tumor suppressor microRNA Let7a is found to be downregulated in breast cancer tissues and cells and is thought to act primarily by targeting HMGA1 expression which in turn results in migration and growth impairment (Liu et al. 2015).

In familial breast cancer, showing autosomal dominant transmission pattern, BRCA1 and BRCA2 are often found mutated and BRCA1 patients exhibit particular tumor histopathological features. Baldassare and colleagues demonstrated that overexpression of HMGA1 protein leads to downregulation of BRCA1 promoter activity in breast cancer, with impairment of DNA repair mechanism and additional genomic alteration (Baldassarre et al. 2003). By modulating BRCA1 expression, HMGA1 is able to affect cisplatin sensitivity and increased DNA damage-induced cell death (Baldassarre et al. 2005).

Other molecular implications have been found for HMGA1 expression in breast cancer. In the mammary adenocarcinoma cell line MCF7, HMGA1a overexpression promotes tumor progression triggering sensitivity to Ras-extracellular signal related kinase (Ras/ERK) mitogenic signalling pathway. In particular, HMGA1a promotes expression of KITLG and its receptor c-Kit whereas downregulates both caveolins 1 and 2 and cholesterol biosynthesis

genes, all involved in modulating Ras/ERK pathway activation. HMGA1 modulates also the expression of downstream elements of this pathway, i.e. Shc, and its targets, i.e. Stat3, ER and PPAR- γ (Treff et al. 2004). An emerging role of HMGA1 in breast cancer progression came out in TNBC cellular models, in which lentiviral mediated delivery of shRNA targeting HMGA1 mRNA led to impairment of invasion and migration along with changes in morphology and depletion of tumor initiators or cancer stem-like cells (Shah et al. 2012). In this light, it becomes clear the importance of understanding the metastatic molecular mechanisms driven by HMGA1 in order to design drugs able to target these oncoproteins and also their downstream effectors.

Targeting HMGA1: the challenge of specificity

It is clear that HMGA1 could be a promising target for the treatment of highly aggressive cancers and efforts have been addressed towards the design of drugs targeting directly or indirectly HMGA1.

Such treatment strategies could be efficiently grouped into four categories: i) compounds that competitively bind to AT-rich regions in the minor groove of DNA interfering with HMGA1 action, i.e. antibiotics (distamycin and netropsin); ii) oligonucleotides decoy/sponges which prevent HMGA1 binding to DNA, i.e. spiegelmers and aptamers; iii) molecules which among others antitumor effects are also able to repress HMGA1 expression, i.e. doxorubicin and flavopiridol; and iv) inhibitors of HMGA1 downstream oncogenic factors, i.e. COX2, STAT3 and MMP2/MMP9. Last but not least, higher HMGA1 expression is indicative of responsiveness to DNA-damaging agents; in fact, it downregulates expression of factors critical for DNA repair mechanism (BRCA1; Baldassarre et al. 2003). On the other hand, we recently demonstrated HMGA1 involvement in Non Homologous End Joining DNA repair mechanism by enhancing Ligase IV activity and resulting in faster recovery upon induction of dsDNA breaks; we therefore suggest an opposite oncogenic mechanism by which HMGA1 could promote progressive accumulation of DNA lesions that could provide selective advantages to cancer cells compared to normal ones (Pellarin et al. 2016) .

Most of the above mentioned HMGA1-interfering strategies often lack the specificity required to avoid undesirable effects (Huso & Resar 2014). Specificity could be gained by delivering specific shRNA/siRNA by nanoparticles or by therapeutical approaches which address downstream molecular mechanisms altered by HMGA1 expression.

Breast Cancer Epigenetics

In breast cancer onset and progression, the misregulation of epigenetic mechanisms is a critical early event that affects chromatin dynamics resulting in gene expression and phenotype alterations. These mechanisms consist of hereditary molecular alterations that do not directly affect DNA sequences but covalently modify chromatin components: DNA methylation, histone's post translational modifications, small non coding and antisense RNAs, and alterations in chromatin protein complexes are all involved in gene transcription regulation. Although cancer has been essentially defined as a genetically disease with major focus on transcriptional factors (TFs) dysregulation, TFs targeting is challenging because they are not enzymes and addressing protein-protein interactions remains tricky. On the other hand, enzymes responsible for chromatin modifications represent ideal targets for drug design: writers, readers and erasers of modifications are often mutated with the consequence of a disrupted epigenetic homeostasis (Jones et al. 2016).

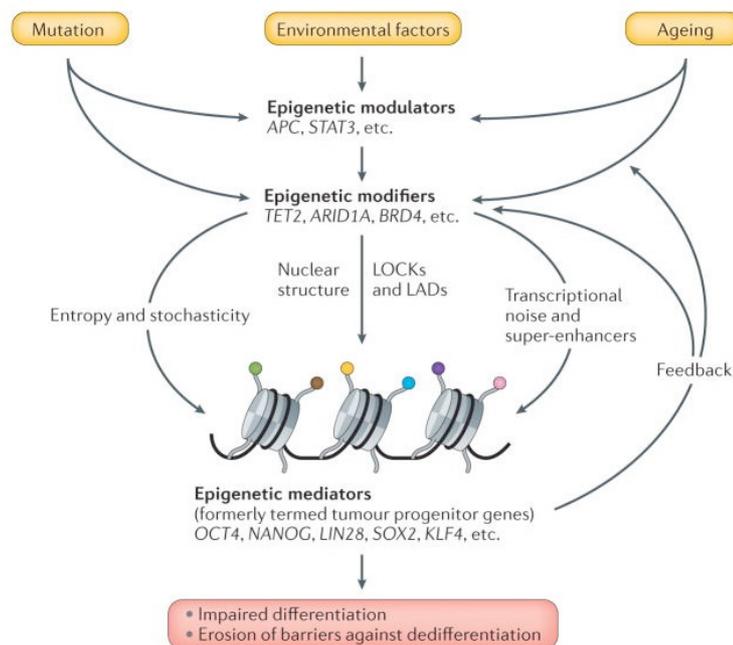


Figure 1.6 Functional classification of genes mutated in cancer development. In induced pluripotent stem cells, developmental potential is ruled by pluripotency factors, i.e. OCT4 and NANOG, genes which are frequently dysregulated in cancer. These factors are collectively referred to as epigenetic mediators and are prone to epigenetic regulation played by modifiers, another class of genes which could be mutated or fine-tuned by cellular signalling pathways elements (epigenetic modulators). In parallel to mediators' expression, phenotypic plasticity is gained by large organized chromatin K9 modification (LOCK; block of repressive DNA and histone methylation) and lamina associated domains (LADs) which contribute to differentiation prevention and dedifferentiation promotion. To stabilize transformation, increasing transcriptional noise at differentiation-promoting genes and re-localization of super-enhancer on oncogenes occur. Tumor heterogeneity is accomplished by fluctuations in environmental pressure and stochastic changes into unstable chromatin states and result in increased cellular entropy (Feinberg et al. 2016).

Besides mutations on genes coding for these epigenetic modifiers, which are directly responsible for chromatin modifications, epigenetic mediators and epigenetic modulators are two other functional cancer gene classes dysregulated in cancer: the former are genes whose expression is affected by chromatin modifications made by modifiers and which are generally associated to stem cell programming; the latter are genes whose products act before modifiers and mediators and that mediate crosstalk between environmental stimuli and cellular neoplastic propensity (Feinberg et al. 2016; **figure 1.6**).

Cancer is therefore depicted as the result of an interplay between intracellular and extracellular tumorigenic stimuli which alter the epigenetic state of vulnerable epigenetic mediators through the action of signalling modulators and mediators.

Chromatin landscape of breast cancer cells

Epigenetic marks have been extensively queried to assess their relationship with histological types, clinical parameters and molecular profile of breast cancers. This suggest their potential role as both biomarkers for prognosis or outcome prediction upon specific treatment and therapeutical targets. The genetic information is housed in the nucleus of eukaryotic cells by DNA-histones complexes that are regulated by epigenetics modifiers. In particular, epigenetic modifications addressed in breast cancer war have been alteration in DNA methylation pattern and changes on PTMs which stud nucleosome surfaces and hence influence gene expression.

DNA methylation is a post replicative modification present in the 5th position of pyrimidine ring of cytosine of dinucleotide CpG and is indicative of heterochromatic regions with more than 90% of methyl cytosines found in chromatin silent repetitive elements or transposones. DNA methyltransferases (DNMT) drive context dependent DNA methylation: DNMT3 family imparts de novo methylation while DNMT1 carries out methylation upon replication (Cheng & Blumenthal 2008). In normal cells, CpG methylation is associated with development long range silencing of imprinted genes, germ cells genes and those of inactive X chromosome (Stefansson & Esteller 2013). In cancer development, in context dependent ways, DNA could be globally hypomethylated, leading to up-regulation and overexpression of proto-oncogenes, or locally hypermethylated, repressing tumor suppressor, DNA repair, hormone receptors, and angiogenesis genes (Jovanovic J et al., 2010). DNA of estrogen receptor ER α promoter has found to be hypermethylated in 40% of breast cancer patients, with correlation to tumor size and histological grade (Hervouet et al. 2013). Treatment of ER- breast cancer cells with demethylation agents induces a partial demethylation of CpG islands present in the ER 5'

region, and a production of ER mRNA and functional protein (Ferguson et al. 1995). Among genes that are hypermethylated in breast tumors and cell lines are cyclin D2 (CCND2), p16^{ink4a}/CDKN2A, RAR- β , RASSF1A, HIN1, CDH3, and BRCA1 (Jovanovic et al. 2010). Characteristic DNA methylation profiles are displayed by luminal A, luminal B and basal-like breast cancers with higher levels of methylation gained by luminal B and the lower by basal-like while normal breast and HER2+ intrinsic subtypes spread over above mentioned three methylation patterns groups (Holm et al. 2010). Different models have been proposed for gene expression repression by epigenetic machinery: one suggest that DNA methylation specify further histone's modifications positions in order to maintain the silencing; on the other hand, histone PTMs have been proposed to primarily drive transient repression (Dworkin et al. 2009).

Nucleosomes are composed by protein chromatin pillars around which 147 bp of DNA are wrapped around and they organize genome in regulatory regions by their positioning. These units are composed of eight basic proteins called histones, H2A, H2B, H3 and H4 with two copies for each histone type, whose tails protrude in the nucleoplasm and could be extensively covalently modified. Single or combinations of post translational modifications (PTMs) destabilize histones contacts with DNA, nucleosome contacts, and form/disrupt docking sites for protein complexes able to drive biological responses. Some modifications are predictive of gene expression and in particular their specific localization is fundamental for the progression of gene transcription, i.e. the pre-initiation complex formation, PolII recruitment, initiation and elongation, and transcription termination (Karlić et al. 2010). One of the main effects of histone acetylation and phosphorylation is to disturb the electrostatic interactions between histones and DNA by altering the positive charge of N-terminal tails. This is usually associated with the opening of the chromatin fiber that allows transcriptional factors to bind gene regulatory sequences. In addition to this effect, phosphorylation and acetylations can also create/disturb docking sites at chromatin levels, directly participating in the transcription factor recruitment process.

Differently, methylation does not cause overall chromatin structure perturbation via charge effects. Indeed, its action towards transcription appears dependent on the specific residue modified (Bannister & Kouzarides 2011). H3K4, H3K36, H3K79 methylations distinguish actively transcribed genes whereas H3K9, H3K27, H4K20 methylations are associated with repressed chromatin regions. Trithorax group of methyltransferases is responsible for H3K4me; Polycomb group proteins (4 different complexes PRC1-PRC4) formed by highly conserved enhancer of zeste homolog 2 (EZH2), suppressor of zeste (SUZ12), and Nurf-55 are

linked to H3K27 methylation. EZH2 has found to be overexpressed in breast cancer and is associated with aggressiveness, proliferation, invasion and tumor progression (Yoo & Hennighausen 2011). Lysine specific demethylase 1 (LSD1), which is responsible for demethylation of mono- and di- methyl lysine 4 of histone H3, was also found to be expressed at high levels on ER- tumors (Huang et al. 2011). HP1 recognizes and binds di- and tri-methylated H3K9, hallmarks of silent chromatin, and it is further responsible for the formation of higher order chromatin structures. HP1 displays tumor suppressive functions in breast cancer and is found to be differentially expressed between cancer and normal tissues (Vad-Nielsen & Nielsen 2015). From immunohistochemistry analysis of 880 breast cancer tissues which address histone marks known to be involved in transcriptional activation and maintenance of euchromatin state, H4K16 acetylation appears to be low or absent in 79% of breast cancer samples whereas levels of H3K18ac and H4K20me3 were high in the majority of cases. Correlation with clinic-pathological data such as tumor size and lymph node and vascular invasion was observed for most of the modification observed in this study. Moreover, high levels of post translational modifications were associated to steroid hormone receptor, luminal cytokeratins, E-cadherin and BRCA1 expression and better patient outcomes. In particular, high levels of acetylation and methylation are linked to favourable prognosis and luminal breast cancer subtypes, moderate levels of acetylation (H3K9, H3K18 and H4K12), methylation (H3K4me2, H4K20me3), and arginine methylation (H4R3me2) were indicative of poorest prognosis, basal carcinoma and HER2+ subtypes (Elsheikh et al. 2009).

Epigenetic marks have been also causatively associated with tumorigenesis and cancer progression. An *in vitro* breast epithelial cells model of cancer progression was obtained by the introduction of three genes (large T antigen, telomerase reverse transcriptase and Harvey rat sarcoma viral oncogene homolog v12 mutant RAS V12) to human primary mammary cells (HMC). Authors described a decrement of di- and tri- methylated histone H3K9 along with the inversely matching increasing activities of KDM3A/JMJD1a demethylase (Zhao et al. 2016)

In line with epigenetics role in cancer progression, histone H3 acetylation of Slug promoter through the activation of PKC with phorbol ester (TPA), was indicated as a mechanism driving EMT in breast cancer (Kamiya et al. 2016). In addition, epigenome has been demonstrated to either assist or mainly direct SMAD3 differential binding specificity, mediating context dependent pro-oncogenic or tumor suppressive responses. In particular, in breast tumor initiating cells SMAD3 binding pattern is affected by open configuration state marked by promoter H3K4me3, promoter and enhancer H3K27ac and RNA PolIII binding (Tufegdzcic Vidakovic et al. 2015). In spite of opposite roles played by acetylation and

methylation in H3K27 and H3K9, those of H3K4 are both associated with gene expression activation with acetylation playing a major role in early step of cancer progression and methylation being associated with metastatic potential. H3K4ac mediate expression of targets genes of ESR signaling pathway while H3K4ac and H3K4me3 have been associated with EMT response in MDA MB 231 (Messier et al. 2016). Cell invasion, another important step in tumor metastasis, has been inversely correlated with H3K20me3 in breast cancer cells and its loss is indicative of poor prognosis in patients (Yokoyama et al. 2014).

Aside well studied local epigenetic modification at single gene locus, breast cancer heterogeneity is characterized by other long range epigenetics events which results in higher order chromatin organization. Gene which cluster into regions of regional epigenetic regulation (RER) exhibit coordinate expression without evidences of genetic alterations of their copy numbers. Analysing patterns of gene expression, DNA methylation and histone PTMs, large chromosomal domains are found coordinatively regulated and these RERs segregate with breast cancer intrinsic subtypes. In this study, one region of long-range epigenetic activation (LREA) is shared by both ER+ breast tumors and breast cancer cell lines and estrogen-dependent chromatin relaxation was suggested as underpinning mechanism leading to unfolding of high order chromatin structures and chromosomal radial nuclear re-localization (Rafique et al. 2015).

Another study investigated cis interaction between breast-epithelial enhancers and up to 750 kb away target genes expression in normal breast epithelial cells (HMEC) and breast cancer cells (MDA MB 231). The presence of enhancers was connected to the histone mark H3K4me1 (Choe et al. 2012) and nucleosome-depleted regions were chosen. Results suggested a cell specific enhancer control of nearby gene expression which enrich for genes involved in proteolysis, epidermis development, and cell adhesion processes in normal breast cells and, on the contrary, mitosis, cell cycle, DNA replication, chromatin modification, and cell division processes in invasive breast cancer cell line (Rhie et al. 2014).

Brest cancer epigenetics marks in clinical managements of disease

Two strategies in targeting chromatin modifications are the most widely exploited: targeted therapy in a specific subset of patients (precision medicine) and broad reprogrammers, which affect large scale pattern of modifications leading to gene expression regulation (Jones et al. 2016).

Among broad reprogrammers are drugs which reverse aberrant DNA methylation or histone acetylation patterns on tumor suppressors or genes mediating drug resistance. These therapeutical strategies rely on inhibitors of DNA methyltransferases and histone deacetylases which drive back cancer genes expression. In breast cancer, the analogues 5-azacytine and 5-aza-deoxycytine (decitabine) are the most commonly studied inhibitor of DNMTs which covalently trap the enzyme and stop up methylation of genomic DNA. Another nucleoside analogue is zebularine, which after being phosphorylated and converted to deoxynucleotide is incorporated in DNA where it blocks DNMTs and methylated-CpG binding proteins. RG108, Psammaplins, MG98, hydralazine and (-)-epigallocatechin-3-gallate (EGCG; polyphenol compound of the green tea) are additional compounds which have shown demethylating effects on breast cancer cells without being incorporated to DNA, with promising effects in breast cancer treatment (Cai et al. 2011).

Histones deacetylase inhibitors (HDACi) act mainly promoting selectively apoptosis of tumor cells by affecting DNA-histones complexes interactions and other proteins acetylation status. These inhibitors can be classified according to their substrates or by their chemical structures which define 4 types of drugs: hydroxamates, benzamides, cyclic peptides and aliphatic acids. Vorinostat (SAHA) is one of the most studied HDACi which inhibits cellular growth by cell cycle arrest and apoptosis in several breast cancer cell models. SAHA has been also connected to EMT in TNBC (Damaskos et al. 2017). In fact, HDACi could exhibit opposite effects depending on molecular profile of breast cancer: it has been demonstrated that in ER⁺ breast tumors it leads to transcriptional downregulation, while in ER⁻ cancer it restores gene expression (Jovanovic et al. 2010). Besides direct effects of HDAC inhibition on histone PTMs profile, the use of compound which provide chromatin relaxation could also improve access of topoisomerase inhibitors on cancer cells. More often than not, the combinational use of drugs elicits better responses than those of monotherapies and even epigenetics drugs can be used in combination with other epigenetics therapies or with other treatments, such as immunotherapeutic or chemotherapeutic agents (Jones et al. 2016). The combinational use of inhibitors of HDAC and DNMT synergistically affect disease progress with increased re-expression of ER and tamoxifen responsiveness in breast preclinical studies (Connolly & Stearns 2012). LAQ824, inhibitor of deacetylase, and decitabine, inhibitor of DNA methyltransferase, have shown synergic activities on restoration of tumor suppressor gene expression; vorinostat in combination with chemotherapeutic paclitaxel and anti-angiogenic bevacizumab have shown effects on 50% of metastatic breast cancer treated (Jovanovic et al. 2010).

On the other hand, precision medicine aims to address epigenetic mechanisms which affect (i) genes critical for cancer development that are not mutated or (ii) mutations of genes involved in epigenetic pathways. For instance, the design of drugs targeting hyperactivation of PI3K-AKT signalling should account for epigenetic mechanisms which drive PTEN inactivation in breast cancer PTEN defective cells or CpG methylation of PPP2R2B, a phosphatase and negative regulator of AKT (Muggerud et al. 2010). Moreover, epigenetic modifications could be used to predict cancer risk, onset of disease and therapeutical responsiveness. BRCA1 methylation, present in half of triple negative breast cancer, predict responsiveness to PARP inhibitors and cisplatin (Stefansson OA and Esteller M, 2013); likewise, the development of Tamoxifen resistance (roughly 40% of cases) has been connected to ER α and ER β methylation (Dworkin et al. 2009).

Although considerable progresses have been achieved in the field of epigenetics in breast cancer, in the era of personalized targeted therapy, robust biomarkers that accurately predict responsiveness to drug targeting epigenetic modifiers are still lacking. This thesis has been developed accounting for the pressing need for clearness of chromatin dynamics and their causative roles in breast cancer onset and progression.

AIM OF THE STUDY

HMGA1 is a well-known determinant of breast cancer onset and progression and it occupies a pivotal position in the chromatin landscape by organizing genetic information and regulating DNA accessibility to a myriad of factors; furthermore, HMGA1 has been shown to influence nucleosome positioning at specific promoters or enhancers through direct contacts with core histones. Despite this, efforts have not yet been addressed towards understanding whether HMGA1/nucleosome contacts could be linked to a modulation of covalent modifications of histones tails. Therefore, the primary purpose of this study has been to unravel HMGA1 function on chromatin modifications in triple negative breast cancer where HMGA1 is naturally high expressed.

MATERIALS AND METHODS

Breast cancer cell lines

Breast cancer cell lines used in this study are: adenocarcinoma MDA MB 231 (TNBC, mesenchymal like/claudin low-KRAS mutation); medullary carcinoma MDA MB 157 (TNBC, mesenchymal like/claudin low); adenocarcinoma MDA MB 468 (TNBC, basal like - amplification of EGFR); metastatic carcinoma MDA MB 453 (ER-, PR-, HER2+; LAR); and adenocarcinoma MCF7 (ER+, PR+, HER2-; Luminal A). Classification is reported according to the one proposed by Holliday DL and Speirs V (Holliday & Speirs 2011) and American Type Culture Collection (ATCC).

Cellular cultures and treatments

Cells were cultured in 10 % Fetal Bovine Serum (*Euroclone*), 2 mM L-glutamine (*Euroclone*), 100 U/ml penicillin, 100 µg/ml streptomycin in Dulbecco's modified Essential Medium High Glucose (DMEM, *Euroclone*) at 37 °C in humidified 5% CO₂ incubator. To MCF7 cell line 1X MEM non-essential amino acids (*Sigma*) were added to the medium. Cells were expanded every 48 hours when sub-confluent according to ATCC recommendations. Adherent cells were detached by trypsin-EDTA, (*Euroclone*) incubated 5 minutes at 37 °C and trypsin was neutralized by medium addition. Thereafter, cells were seeded according to cell counting. For long term storage, cells were resuspended in 10% DMSO in FBS and conserved in cryovials in liquid nitrogen.

Cell counting

In order to define cell concentration, a Neubauer chamber was used. Briefly, 10 µL of desirable cell dilutions (about 250.000 cells/mL) are pipetted in the two chamber rooms composed of a grid subdividing 9 big squares. Cells present in four external squares (4x4) are counted excluding cells lying on lower and right margins of the smaller squares. Same procedures are used for the second chamber. Minimal and maximal values are excluded for mean calculation. Since distance between glass cover and the bottom of the chamber is 0.1 mm, the volume of one square is 0,1 µL and the cell concentration is obtained by multiplying mean value to factor 10.000. If prior dilutions have been applied to the sample, they must be considered to undiluted sample concentration determination.

Small interfering RNA (siRNA)

Cells were plated $21 \cdot 10^3/\text{cm}^2$ in medium without antibiotics. After 24 hours, siRNA and lipofectamineTM RNAiMAX (*Invitrogen*) were diluted in medium without antibiotics and FBS, incubated together for 15 minutes and then dropped over cell Petri dishes. Lipofectamine was used 5 μL in a 35-mm plate containing 3 mL of medium.

siRNA

| Target mRNA | siRNA sequences | Final concentration |
|-------------------------|-----------------------|---------------------|
| None in mammalian cells | ACAGTCGCGTTTGC GACTG | 10 nM |
| HMGA1 (A1_1) | GACAAGGCUAACAUCCCACTT | 10 nM |
| HMGA1 (A1_3) | ACTGGAGAAGGAGGAAGAG | 10 nM |
| RSK2 | GCAUUCCACCUAGUGCUAATT | 10 nM |
| CBP (ORF) | CCTTCTAGCACCGGTGTAA | 10 nM |
| EP300 (ORF) | GCAAACAATCGAGCGGAAT | 10 nM |

Table 1. List of siRNA sequences and their concentration used to target specified mRNA.

Metabolic activity assay: MTS assay

In order to evaluate the cytotoxicity of a particular treatment, we performed MTS assay that allows the colorimetric determination of actively proliferating cells in the cellular population. In fact, the compound 3 - (4,5 - dimethylthiazol - 2 - yl) - 5 - (3 - carboxymethoxyphenyl) - 2 - (4 - sulfophenyl) - 2H - tetrazolium is bio-reduced to formazan which is soluble and absorbs at 490 nm. The reaction is proportional to viable cells because the reaction is accomplished by dehydrogenase enzymes acting in metabolic active cells. 5000 cells were plated in a 96-multiwell plate and grown for 72 h before treatment, thereafter every 24 hours medium was replaced with 200 μL of MTS diluted 1:6 in 4.5 g/L glucose in PBS (Huang KT et al., 2004). After an incubation of 2 h and 30 minutes at 37°C 5% CO₂, the absorbance was read at TECAN microplate reader. Plotting values of absorbance versus time points, curves of metabolic activity were obtained upon treatment over a period of time.

Lysates pre-fractionation: Histone Extraction Methods

Strong acid histone extraction (modified Schechter D et al., 2007)

Upon silencing, a 150-mm plate of cells was rinsed twice in PBS, gently scraped in PBS and collected in tubes. Intact cells were precipitated by mild centrifugal force (4°C, 450 g, 5 minutes) and resuspended in 1 mL of cold hypotonic lysis buffer (10 mM Tris/HCl pH 8.0, 1mM KCl, 1.5 mM MgCl₂, 1 mM dithiothreitol DTT) complemented with 1 mM NaVO₃, a grain of sodium butyrate and 1/100 (v/v) protease inhibitor cocktail (PIC, *Sigma*). Otherwise, after centrifugation cells could be stored at -80°C up to 3 months. Cells were shaking incubated in lysis buffer at 4°C for 30 minutes and nuclei precipitated by centrifuge (4°C, 10000 g, 10 minutes). Supernatant was discarded and nuclei resuspended in 0.5 M HCl; strong acid incubation is performed O/N. The day after, samples are centrifuged at 4°C 16000 g for 10 minutes. Proteins soluble in strong acid medium (i.e. histones) are present in the supernatant; therefore, supernatant was incubated O/N at -20°C with 8 volumes of acetone so as proteins could be precipitated by centrifuge at 4°C 3345g for ten minutes. Samples were decanted and pellet resuspended with 200 µL of MS-water (*Sigma*).

Histone purification kit (*Active Motif*)

Histones were extracted from two 150-mm plate cells according to manufacturer instruction. Briefly, cells were washed twice in serum-free medium and scraped with 800 µL of Extraction buffer wherewith cells are shaking incubated O/N at 4°C. Then, cells were centrifuged at 4°C 16000 g and supernatant conserved over weekend (O/W) at -80°C. 200 µL of neutralization buffer was added and pH measured (8). Column was filled with resin, washed, equilibrated and samples loaded. Resin was washed and finally core histones eluted in different fractions.

Total proteins extraction for Western Blot analysis

Upon above-mentioned treatments, cells were rinsed twice in PBS (Phosphate Buffered Saline: 2.68 mM KCl, 1.76 mM KH₂PO₄, 137 mM NaCl, 9.93 mM Na₂HPO₄ pH 7.4 sterilized), and lysed by gently scraper in SDS sample buffer (125 mM Tris/HCl pH 6.8, 4% w/v SDS, 20% glycerol, traces of bromophenol blue and 0.2 M DTT). The stoichiometry of anionic detergent SDS binding to protein is around 1 molecule of SDS per two aminoacids, or 1.4 g SDS per 1 g of proteins. To a confluent 50-mm Petri dish, 100 µL of SDS sample buffer were generally used with large excess of SDS. DNA was mechanically destroyed by ten

sample passages into insulin syringe (29 G needle) and lysates were then boiled for 5 minutes at 96 °C to favor reducing agent action (DTT) and denature proteins. Samples were then stored at -20°C.

SDS- PolyAcrylamide Gel Electrophoresis (SDS-PAGE)

The electrophoretic migration of protein lysates was performed in Laemmli discontinuous and denaturing system in which proteins, conditioned with SDS which impairs negative charges to molecules, migrate primarily based on their size towards the positive electrode and determination of molecular weight is allowed with opportune standards loaded next to samples. The upper phase of discontinuous gel, *Stacking gel* (T=5%; C=3.3%; in 1.15 M Tris/HCl pH 8.45, 0.11% w/v SDS and 1 mg/mL of ammonium persulfate APS), is characterized by larger pores in which proteins migrate faster and accumulate in more resolving *Running gel* (T=15%; C=3.3%; in 1.6 M Tris/HCl pH 8.45, 0.16% w/v SDS and 1 mg/mL of ammonium persulfate APS), with smaller pores and slow migration. Polymerization occurs by the addition of catalyst N-N-N-N Tetramethylethylenediamine or TEMED (2 µL for 1 mL of stacking; 4 µL for 10 mL of running). Gel thickness: 0.75 mm; length stacking: 0.5 cm; length running: 8 cm; gel width: 8 cm. In contrast with original Laemmli gel, ours do not contain SDS which is present in SDS sample buffer and cathode buffer enough to sustain denaturation during electrophoresis separation. Cathode buffer: 0.1 M Tris, 0.1 M Tricine, 0.1% w/v SDS pH 8.25. Anode buffer: 0.1 M Tris/HCl pH 8.9. The accumulation was performed at 20 mA whereas separation at 40 mA.

Blue Coomassie staining - Total Protein samples quantification and normalization

Lysates separated by SDS-PAGE were stained by a shaking incubation with methanol/water/acetic acid in a 5/4/1 volume ratio with 0.05% (w/v) of Coomassie Brilliant Blue R 250 which can be protracted overnight (O/N). Dye excess was eliminated by 10% (v/v) acetic acid washes. Optical densities were obtained using Image Scanner and Image Master 2D Software (*Amersham Pharmacia Biotech*). In order to quantify and normalize samples, a curve of HL60 lysate quantified by Bradford (*Sigma*) was loaded next to samples in the SDS-PAGE. The densitometry values performed with 600 dpi (dots per inch) resolution, were then plotted in a calibration curve versus the amount of lysate (µg). By densitometry values interpolation, the sample concentrations were obtained and samples were normalized.

Bicinchoninic Acid (BCA) method for protein quantification

The BCA protein assay reagent kit (*Pierce*) is a detergent compatible method based on biuret method. Briefly, in alkaline medium Cu^{+2} is reduced at Cu^{+1} which in turns is able to chelate molecules such as BCA forming purple products, soluble in water and absorbing at 562 nm. The intensity is proportional to number of peptides bonds, the macromolecular structure of proteins and presence of cysteine, cysteine, tryptophan and tyrosine. According to datasheet, reactions are performed in a 96-multiwell plates and absorbance was read at TECAN microplate reader. The concentration of samples derived from a calibration curve with bovine serum albumin (BSA) as standard.

Semi-quantitative Western Blot (WB)

Once cell lysates have been separated accordingly to protein molecular weight through SDS-PAGE, total proteins could be transferred to a nitrocellulose membrane \varnothing 0.2 μm (*GE Healthcare*) using a wet transfer system (*Biorad*) in 20% methanol, 25 mM Tris, 200 mM Glycine (blotting solution), for 16 hours at 4°C with liming current (75 mA). The quality of transfer was valued by incubation with ponceau red solution (0.2% red ponceau, 3% trichloroacetic acid, 3% sulfosalicylic acid) for 10 minutes at RT. Images were acquired by Image Scanner and densitometry of lysates obtained by Image Master 2D Software (*Amersham Pharmacia Biotech*) were used to normalized western blot results. Membrane was then saturated 1 h at RT with 5% (w/v) non-fat dry milk 0.1% (v/v) Tween 20 in PBS (blocking solution). The amount of total lysate necessary for antibody detection of target protein was estimated by a curve of lysate and the entire membrane was recognised in order to assess antibody specificity. Membrane was then shaking incubated with indicated dilutions of primary antibody in blocking solution 1 h at RT and then washed three times with blocking solution. Afterword membranes were then incubated with horseradish peroxidase-conjugated secondary antibodies diluted in blocking solution and washed three times in blocking solution and twice in PBS. Peroxidase substrate Enhanced Chemiluminescence (ECL, *Thermo Scientific*) kit was used to detect bands with autoradiography films (*GE Healthcare*).

Antibodies

| Primary Antibody | Dilution used in WB |
|---|----------------------------|
| α -HMGA1 | 1/500 |
| α -H2B (<i>Abcam</i> , #ab52985) | 1/2000 |
| α -H2BK5ac (<i>Abcam</i> , #ab40886) | 1/500-1/1000 |
| α -H2BK16ac (<i>Abcam</i> , #ab177427) | 1/10000 |
| α -H2BK20ac (<i>Abcam</i> , #ab177430) | 1/2000 |
| α -H3 (<i>Abcam</i> , #ab1791) | 1/4000 |
| α -H3S10ph (<i>Abcam</i> , #ab5176) | 1/1000 |
| α -ERK 1/2 (<i>Sigma</i> , #M5670) | 1/1000 |
| α -pERK 1/2 (<i>Sigma</i> , #M8159) | 1/1000 |
| α -RSK2 (<i>Millipore</i> , #06-918) | 1/500 |
| α -RSK2S227ph (<i>Abcam</i> , ab75820) | 1/500 |
| α -CREB (<i>Genetex</i> , #GTX112846) | 1/500 |
| α -CREBS133ph (<i>Sigma</i> , SAB4504375) | 1/1000 |
| α -AURKB (<i>Novus</i> , #NB100-294) | 1/1000 |
| Secondary antibody | |
| α -IgG rabbit (Whole molecule) Peroxidase conjugate (<i>Sigma</i> , #A0545) | 1/5000 |
| α - IgG mouse (Whole molecule) Peroxidase conjugate (<i>Sigma</i> , #A9044) | 1/5000 |

Table 2. List of primary and secondary antibodies used with their dilution factors in western blot analysis.

Multiple enzymes digestion Filter Aided Sample Purification (MED-FASP) and Stage Tips purification

Centrifugal Filter Units Microcon 30k (*Millipore*) were used in order to perform solvent exchange and in solution digestion. 50 μ g of core histones were mixed with 200 μ L of buffer UA (8 M Urea in 100 mM Tris/HCl pH 8.5) and loaded onto filter units placed in a centrifuge tube. Centrifugation was performed at 20°C 10.000 g for 20 minutes. Once discarded the flow-through, filter was washed with 200 μ L of buffer UA and alkylated with 100 μ L of IodoAcetAmide buffer (55 mM IAA in UA) was added and centrifuged at 20°C 10.000g for

15 minutes. This step was followed by two washes with UA buffer and two conditioning washes with Arg-C digestion buffer (DB; 100 mM Tris/HCl pH 7.4, 10 mM CaCl₂). Then, samples were digested with 0.5 µg of Arg-C in DB (100 µL) O/N at 37 °C in activating solution and DB (according to manufacturer instruction; *Roche*). Samples digested were collected in tubes by centrifuge at 20°C 10.000g for 15 minutes; filters were washed twice with 100 µL of digestion buffer. Thereafter, remaining undigested polypeptides were incubated O/N hours with 1 µg Trypsin in DB (50 µL) at 37°C. Elution and washes were performed as for Arg-C digestion. Eluates were then de-salted by stage tips device made by three plugs of Solid Phase extraction disk (*Bioanalytical technologies 3M Company Empore 2215 - C18 Octadecyl 47 mm 20 each*) previously washed with ethanol, Acetonitrile 60% in 1% acetic acid and finally with 1% acetic acid. Samples were loaded in 1% acetic acid, washed with 1% acetic acid and eluted by Acetonitrile 60% in 1% acetic acid with a mild centrifuge at 24 °C 500 g 10 minutes and collected into 0.2 mL 24 well PCR plates (*Thermoscientific*). Samples were finally subjected to a vacuum centrifuge 45°C 30 minutes. We established 6% of the Arg-C samples and 12% of Strong Acid Arg-C samples as the right amount of FASP eluted sample to inject so as to do not overload the column. Trypsin samples for both extraction methods are injected as 12% of the total eluted after the FASP purification/digestion.

Orbitrap Q exactive

Peptides were loaded onto a reversed phase C18 column (20 cm × i.d. 75 µm; Dr. Maisch Reprosil-Pur AQ 1.9 µm) (75 µm inner diameter, in-house packed with Reprosil-Pur C18-AQ 1.8-µm resin - Dr. Maisch GmbH, Ammerbuch-Entringen, Germany) with buffer A (0.5% acetic acid). A 180 minutes gradient of acetonitrile in 0.1% formic acid (5 minutes 2-5% B; 80 minutes 5-30% B; 13 minutes 30-65% B; 10 minutes 65-95% B; 3 minutes at 95% B; 5 minutes 95-2% B; 8 minutes at 2%B) was performed at a flow rate of 250 µL. The analyser Q Exactive was coupled to the LC through a Proxeon nano- electrospray source. A full scan over m/z 300-1650 with resolution 60.000 was followed by the fragmentation of the 10 most intense peaks with charge state $\geq +2$ (Top10). Dynamic exclusion was 20 s and 12 ppm.

Data were analysed by Max Quant software version 1.2.6.20, and MS and MS/MS information were matched with UniProt human proteome sequences. Analysis was performed with the option match between runs (matches identifications across samples within a window of 2 minutes). Carbamidomethylation of cysteines was set as a fixed modification whereas N-

terminal acetylation, oxidation of methionines, acetylation of lysines, phosphorylation of threonines, tyrosines and serines, methylation of lysines and tri-methylation of arginines and lysines were set as variable modifications. Up to 2 missed cleavage were allowed. FDR for peptides, proteins and sites was set at 0.01.

RNA extraction - DNase I digestion - Clean Up

Upon desirable treatments, cells cultured in a 35-mm plate at sub-confluence were rinsed once with room temperature PBS and lysed in 1 mL TRIzol Reagent (*Invitrogen*) incubating two minutes at RT; lysates could be stored at -80°C. Chloroform was added as 1/5 volume of Trizol, samples were mixed by inversion and incubated at RT for 2 minutes. Afterwards, lysates were centrifuged at 4°C 17005 g 10 minutes, the upper aqueous phase containing nucleic acid was recovered and transferred to a new tube. For further precipitation, isopropanol as 1/2 of Trizol volume was added to samples, incubated 10 minutes at RT and centrifuged at 4°C 17005 g 10 minutes. Supernatants were discarded and pellets were washed with 75% (v/v) ethanol and centrifuged at 4°C 17005 g 10 minutes. Supernatants were again discarded and air-dried pellets were resuspended in adequate volume of RNase-free water depending on the amount of precipitated (generally 30 µL). Depending on the amount of volume used, DNase I reaction buffer 10X (200mM Tris/HCl pH 8.4, 20 mM MgCl₂, 500 mM KCl) and 1U/µL DNase I (*Invitrogen*) were added to each sample to digest genomic DNA (generally to 30 µL of samples, 4 µL of reaction buffer and 1 µL of DNase I were added). The digestion was performed for 15 minutes at RT and blocked with EDTA 2 mM at 65°C 10 minutes. The sample was diluted to a final volume of 300 µL with RNase-free water and made 300 mM sodium acetate pH 5.2. An equal volume of acid phenol 5:1 chloroform pH 4.5 (*Ambion*) was added to each sample, mixed by vortex and centrifuged at RT 17005 g 2 minutes. Supernatant was recovered and an equal volume of chloroform was added to each sample, mixed by vortex and centrifuge at RT 17005 g 2 minutes. Once recovered the upper aqueous phase, 2.5 volumes of cold ethanol 100% were added in order to precipitate DNA at -20°C O/N. The day after, DNA were pelleted by centrifuge at 4°C 17005 g 30 minutes, washed with 70% (v/v) ethanol and centrifuged at 4°C 17005 g 20 minutes. Supernatants were discarded, pellets air-dried and resuspended with nuclease free water depending on the amount of precipitated (generally 50 µL).

RNA analysis on denaturing agarose gel electrophoresis and quantification of extracted RNA

In order to check the integrity of extracted RNA and the absence of genomic DNA, 2 μ L of samples were conditioned with 1 μ L of loading buffer (30% v/v glycerol, 1.2% w/v SDS, trace of bromophenol blue and GelRed *Biotium* used as 1000X in MOPS) and loaded into 1% (w/v) agarose, 6.67% (v/v) formaldehyde in MOPS (20 mM MOPS, 5 mM sodium acetate, 1 mM EDTA pH 7) gel. All equipment required was pre-treated with 0.1M NaOH for 30 minutes in order to create a RNase-free environment. Conditioned lysates were denatured by heating at 68°C for 5 minutes and kept at 4°C. Then, the electrophoretic separation was done for 40 minutes at 50V. The quantification of the samples was done on Nanodrop 2000 *Thermoscientific* spectrophotometer.

Reverse Transcriptase PCR (RT-PCR)

1 μ g of RNA was retrotranscribed into complementary DNA (cDNA) by adding 150 ng of random primers (*Invitrogen*), 0.834 mM dNTPs (*Sigma*) in nuclease-free water. Samples were denatured at 65°C for 5 minutes, kept on ice 5 minutes and incubated with First Strand buffer 5X (*Invitrogen*; 50 mM Tris/HCl pH 8.3, 75 mM KCl, 3 mM MgCl₂), 10 mM DTT and 40 U RNase OUT (*Invitrogen*). Samples were then incubated at RT for 2 minutes and 200 U of SuperScript III Reverse Transcriptase (*Invitrogen*) were added to each sample. The reaction was performed one hour at 50°C, the inactivation was done at 70°C for 15 minutes and samples were left at 4°C.

Real-time quantitative PCR (qPCR)

The Real Time PCR system (*Biorad*, CFX) was used to perform quantitative Polymerase Chain Reaction (qPCR) on samples diluted accordingly to gene products abundances. Experiments were performed in biological triplicates and technical duplicates. Each gene was normalized on housekeeping genes expression as internal standard (i.e. GAPDH and CYP33). The reaction mix was made with IQ SYBR Green Supermix 2X (*Biorad*; containing Taq DNA polymerase, dNTPs, MgCl₂, SYBR Green dye, enhancers, stabilizers, fluorescein), 150 mM of forward and 150mM reverse primers, cDNA (generally 1 μ L of a dilution 1/50) up to volume with RNase free water. The thermal cycling protocol provides an initial polymerase activation and DNA denaturation 5 minutes at 95°C, followed by 40 cycles of amplification which

consist of 5 seconds of denaturation at 95°C, 45 seconds of annealing/elongation and fluorescence reading at 60°C; a final step to define melting curves with the increasing of 1 °C ever 10 seconds from 60 to 95°C. Primers were designed using Primer3 Plus with mRNA sequences derived from NCBI and they were validated through Integrated DNA technologies (IDT) and Blast alignments. Data were analysed by Biorad CFX Manager software and relative gene expression was calculated by Pfaffl method which take into account amplification efficiencies of target genes and internal standards, calculated by the slopes of standard curves performed with the reaction mix on sequential diluted samples. Statistical analysis performed on differences in expression between samples were done assuming a normal distribution of our target gene expressions among cell population. Therefore, two-tailed Student's t test was done and p values were considered significant when <0.05 accordingly to literature.

Primers

| TARGET GENES | FORWARD PRIMER (5'-3') | REVERSE PRIMER (5'-3') |
|--------------|--------------------------|----------------------------|
| GAPDH | TCTCTGCTCCTCCTGTTC | GCCCAATACGACCAAATCC |
| CYP33 | CTTCATGCTGCGTTCATTCC | CCTCGGTGCTTTTCTGTTC |
| HMGA1 | ACCAGCGCAAATGTTTCATCCTCA | AGCCCCTCTCCCCACAAAGAGT |
| AURKB | TGAGGAGGAAGACAATGTGTGGCA | AGGTCTCGTTGTGTGTGATGCACTCT |
| CENPF | TCAGGCAAGAGGCAAAGATCCAGT | TGGCTCAAACCTCAGTACCTTCCGT |
| KIF23 | CCTGAGGGCTACAGACTCAACCGA | TCTGGGTGGTGTGAGTGCCAA |
| KIF4A | AAGCCAAACGCCATCTGAATGACC | TTGACCACGCACTTCAGTAAGGGA |
| CBP | GGTTTTGTGTGCGACAACCTG | TTCCAAGTGGTTTCCCAGTC |
| EP300 | CAAACGCCGAGTCTTCTTTC | TTGAGCTGCTGTTGGCATAG |
| RSK2 | GAGAGCGGAAAATGGTCTTC | CAAGCAGCATCATAGCCTTG |
| AURKB -800bp | CAAACAACACGGCAGCTAAC | TCTGACATTCCTGGGAGGAG |
| AURKB TSS | TTGGGTTCCTCATGACTTACG | CCCGCAAACAACCTGAATCTG |
| AURKB intrag | CCAAACTGCTCAGGCATAAC | ATCAGGCGACAGATTGAAGG |

Table 3. List of primer sequences (forward and reverse) used for RT-qPCR.

Co-ImmunoPrecipitation (Co-IP)

A 150-mm plate of cells at sub-confluence, was rinsed twice in cold PBS and collected with scrapers. Cells were pelleted by mild relative centrifugal force (4°C, 450 g, 5 minutes) and washed again in PBS. After centrifugation, supernatant was discarded and cells resuspended in lysis buffer (25 mM Tris/HCl pH 8, 0.5% v/v NP40, 125 mM NaCl, 10% v/v glycerol) supplemented with 1/1000 (v/v) Phenylmethylsulfonyl fluoride (PMSF, saturated solution in isopropanol), 1 mM NaVO₃, 5 mM NaF, 10 mM sodium butyrate and 1/100 (v/v) protease inhibitor cocktail (PIC, *Sigma*) and incubated 15 minutes on ice. Samples were then sonicated at 30% of potency for 60 seconds (10 seconds ON, 30 seconds OFF) by *Branson Digital Sonifier 250*, centrifuged at 4°C, 8385 g for 10 minutes, and supernatant collected and stored at -80°C. 30 µL of Protein G Sepharose (*GE Healthcare*) were used for each co-immunoprecipitation. Beads were washed three times with Tris/HCl 50 mM pH 7 (4°C, 1027g, 1 minute) and incubated for one hour at 4°C with 4 µg of antibody or IgG (Abcam #ab37415 used as negative control) adjusted with Tris/HCl 50 mM pH 7 up to 500 µL. After a mild centrifugation (4°C, 1027g, 1 minute) supernatant was discarded and BSA 1 mg/mL in lysis buffer was added at 4°C for 1 hour in order to saturate unspecific binding sites of our antibodies. Beads were then washed three times in lysis buffer (4°C, 1027g, 1 minute). 500 µg of lysates, quantified by BCA, were adjusted to 1 mL with lysis buffer and shaking incubated with beads for 3 hours. If required, to verify that the interaction between interested proteins was not mediated by DNA, a 500 µg of lysate was digested with 1000 U of DNase I (*Sigma*) adjusted to 500 µL with lysate buffer and DNase digestion buffer 10X (*Invitrogen*) and shaking incubated for 20 minutes at 37°C. An aliquot of 50 µL of lysate either digested or not (control without DNase) was kept and made 1% sarcosile and 25 mM EDTA. The aliquot was digested with 10 µg of RNase A (*Sigma*) for 1 hour at 55°C and later with 23 µg of Proteinase K (*Sigma*) for 1 hour at 55°C (shaking incubation). 10 µL were conditioned with an equal volume of Blue/Orange DNA loading dye (*Promega*) and loaded to 1.5 % (w/v) agarose in TAE buffer (1/25000 GelRed *Biotium*). Samples were subjected to electrophoretic separation for 40 minutes at 50 V. The visualization was done with *Gel Doc* (*Biorad*). After the incubation with lysates, beads were centrifuged at 4°C, 1027 g, 1 minute, supernatants were discarded and beads washed three times in lysis buffer. 20 µL of SDS sample buffer were added to each sample, boiled for 5 minutes at 96°C and analysed by WB.

Immunofluorescence (IF)

Cells were cultured on glass coverslips (16 mm) placed into 35-mm plate. Cells were rinsed twice in PBS and fixed for 20 minutes at RT with 4% paraformaldehyde (PFA) in PBS pH 7.2. Cells were washed three times in PBS and 0.1 M glycine was added 5 minutes at RT in order to decrease auto fluorescence given by PFA. After three washes in PBS, cells were permeabilised with 0.3% Triton X for 5 minutes at RT. Cells were washed three times and unspecific sites were blocked with 0.5% (w/v) BSA in PBS (blocking solution) for 30 minutes at RT. Each coverslip was incubated for 90 minutes in a wet room with desirable concentration of antibody diluted in blocking solution (samples were placed over 20 μ L of antibody dilutions). After being washed three times on PBS, samples were incubated with secondary antibody diluted in blocking solution for 60 minutes in a wet room. Coverslips were washed three times in PBS and incubated 5 minutes with 0.2 μ g/mL Hoechst in PBS (*Sigma*). Images were acquired by Eclipse Ti inverted research microscope (*Nikon*) equipped with a NIS elements software.

Antibodies

| Primary Antibody | Dilution used in WB |
|---|---------------------|
| α -H2BK5ac (<i>Abcam</i> , #ab40886) | 1/100 |
| α -RSK2 (<i>Millipore</i> , #06-918) | 1/25 |
| Secondary antibody | |
| α -IgG rabbit Alexa Fluor 488 (<i>Invitrogen</i> , #A11008) | 1/750 |

Table 4. List of primary and secondary antibodies used with their dilution factors IF analyses.

Statistical analysis

Western blot and RT-qPCR analysis were conducted in triplicates and the significance was conferred by two-tailed Student T test, with p value indicative of a significant difference between two means when lower than 0.05. Data are represented as mean and standard deviation (\pm SD). WB histograms represent densitometry values of bands normalized versus densitometry of total lysates of ponceau-stained membrane. For both WB and real-time data values were shown normalized to levels of control samples. Immunofluorescence data were analysed by Shapiro-Wilk test to check their probability distribution: when p values are less

than 0.05 they were considered as not-normally distributed and non-parametric statistics were used, such as Mann-Whitney test which investigate if probability distributions of two populations are different. Data are presented as box-plot, with median and whiskers from 5th to 95th percentile.

RESULTS

HMGA1 expression level affects epigenetic modifications

Data published by our laboratory underlined HMGA1 involvement in stemness and metastatic progression of breast cancer by chairing epithelial to mesenchymal transition tuning the expression of genes which are predictors of poor clinical outcomes (Pegoraro et al. 2013, Maurizio et al. 2016). In the last few years we collected a series of data demonstrating that HMGA1 binds histones and that its presence influences a set of histones post-translational modifications (see preliminary data section at the end of this thesis obtained before my PhD studies). To have a bird's eye view of histone modifications affected by HMGA1 expression, we knocked down HMGA1 in triple negative breast cancer model, MDA-MB-231. These cells are peculiar since they exhibit mesenchymal phenotype and their molecular profile indicates alterations of the expression of genes involved in cell-cell contacts, denoting that they underwent a phenotype transition (EMT); in fact, they belong to the claudin-low molecular subtype. They were collected 72 hours after HMGA-targeting siRNA (siA1_3) or control siRNA (siCTRL) treatment, because at this time point HMGA1 expression was completely switched off; as shown by **figure 3.1A**, upon HMGA1 silencing cells exhibit a reversion of phenotype, reminiscent of mesenchymal to epithelial transition (MET). In order to assess all possible histone PTMs and their combinations, we extracted histones from both conditions, treated the extracts by proteases, and analysed the peptides by state-of-the-art of mass spectrometry technologies. We decided to extract histones with two different methods (see materials and methods) in order to obtain data from two different histone preparations. We used a histone purification kit provided by *Active Motif* and the gold standard strong acid extraction method (Shechter et al. 2007). Blue Coomassie stained SDS-PAGE reported in **figure 3.1B** are indicative of core histones column elution by kit extraction protocol (left panel) and histones extracted by nuclear proteins treatment with hydrochloric acid (right panel - lanes 7 to 12). The two elutions obtained with the kit were pooled together prior to MS analyses. Normalized amount of proteins was digested by multiple enzymes digestion - filter aided sample purification (MED-FASP) method combining Arg-C and trypsin digestions and peptides were further purified by stage tip devices. We opted for Arg-C proteolytic digestion because of the optimal length of peptides for the retention on the chromatography and the ionization process (Soldi et al. 2013); MED-FASP was chosen for the opportunity to analyse

both histone H1 (less rich in arginine) and other proteins that have been co-purified with histones during the pre-fractionation process. Samples were vacuum centrifuged and we injected 12% of the strong acid samples (Arg-C) and 6% of the kit samples (Arg-C) to optimize the TIC signal of the chromatogram so as the column was not saturated. For trypsin, 12% of the samples was used for both strong acid and kit samples. Proteins intensities have been normalized by the sum of all intensities of the sample. We have considered proteins and modifications that are present at least in duplicate in our analysis, and we took into consideration as down/up-regulated the ones in which the ratio between the mean of samples treated with siCTRL and the mean of samples treated with siHMGA1 (and vice-versa) was > 2 and $< 0,5$ respectively. The percentage of MS/MS spectra identified by MaxQuant software have been 10-15% for Arg-C peptides and 30-45% for trypsin peptides leading to 3700 proteins identified in strong acid samples and 2700 in the kit samples with an overlapping of 2000 proteins between the two extraction methods. We confirmed the down-regulation of HMGA1 proteins that are knocked-down for 70% (data not shown). A global quality control of our data arose assessing downregulation of proteins, i.e. LRRC59, or modifications, i.e. phosphorylation of YAP, that are well-described events following HMGA1 silencing (Maurizio et al. 2016; Pegoraro et al. 2015). **Figure 3.1C** reports volcano plot of the fold change (\log_2) plotted versus p values ($-\log_{10}$) of different intensities among peptides belonging to kit histone extraction (left panel) or histones strong acid extraction (right panel). In red are reported peptides up-regulated upon HMGA1 silencing (epithelial phenotype) whereas blue spots indicated peptides downregulated (abundant in mesenchymal breast cancer control cells). Black dots are informative of other peptides co-purified with histones. However, looking at histone PTMs, only two modifications were shared between the two different pre-fractionation methodologies. These were the acetylation of lysines 17 and 21 of histone H2B (further indicated as K16 and K20 without counting the first methionine). Data concerning these histone H2B modifications were recently published by Abell AN et al. and have been linked to the maintenance of epithelial phenotype in EMT process during both development and breast cancer evolution (Abell et al. 2011).

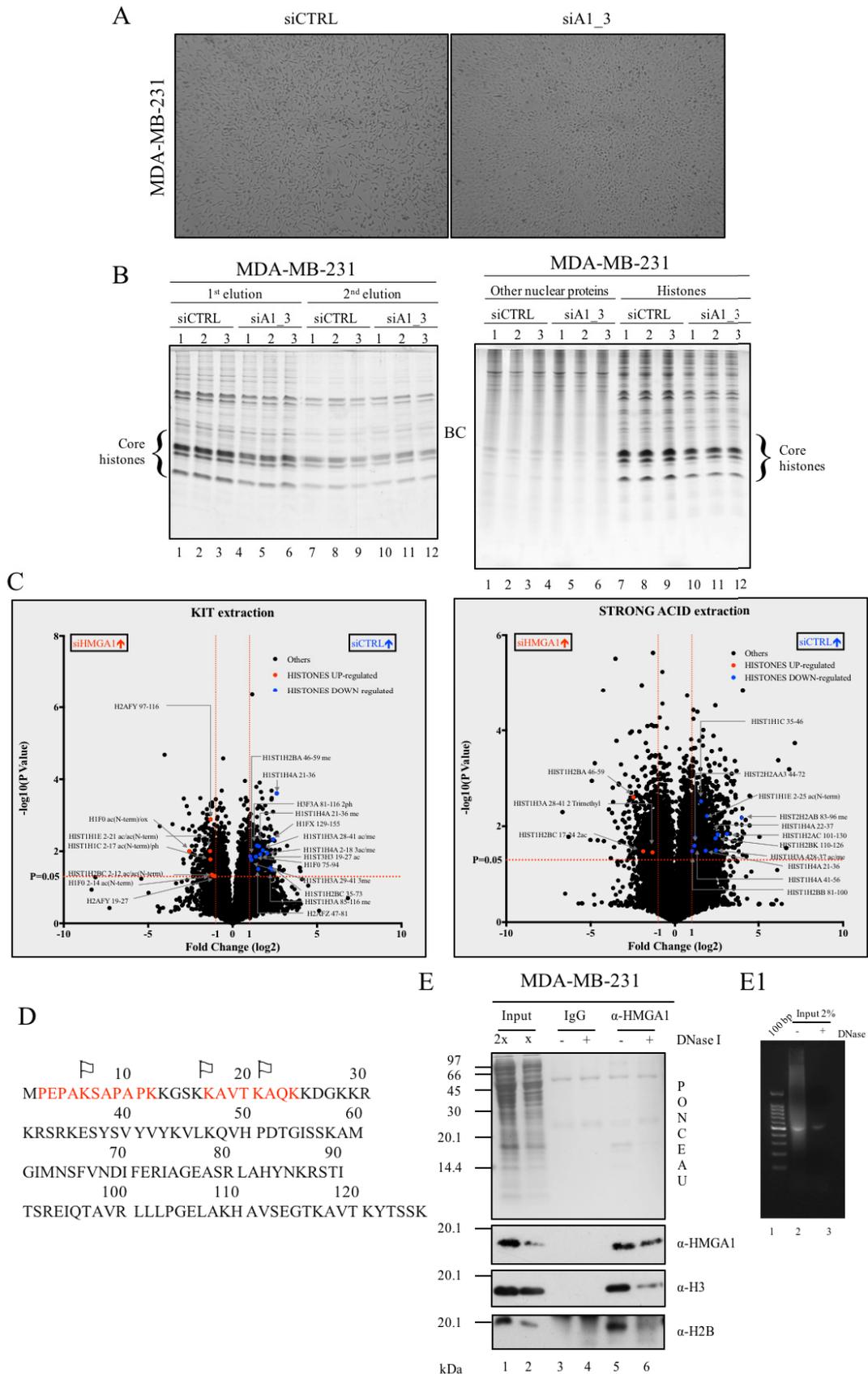


Figure 3.1 HMGA1 interacts with core histones and influences their PTMs in MDA-MB-231. *A*) Optical images of representative cells silenced either with control siRNA (siCTRL) or HMGA1-targeting siRNA (siA1_3) after 72 hours of treatment. *B*) Blue Coomassie stained SDS-PAGE loaded with samples silenced for HMGA1 expression and controls both pre-fractionated by histone extraction performed with *Active Motif* kit (left panel) or by strong acid extraction (right panel). In the left gel are represented the first two column elutions of

core histones with kit extraction; these elutions have been pooled together for MS analyses. In the right gel, first 6 lanes are protein nuclear extracts depleted of histones proteins, whereas samples from lane 7 to lane 12 are correspondent core histones belonging to same nuclei. *C*) Volcano plot of peptides belonging to Arg-C and trypsin digestion of histones fractionated by kit (left panel) or strong acid method (right panel). Normalized peptides intensities obtained by MaxQuant software version 1.2.6.20 have been plotted accordingly to their fold change (\log_2) and p value ($-\log_{10}$) obtained by student t test. Red dots are indicative of peptides significantly up-regulated (>2 folds) upon HMGA1 silencing whereas blue dots are indicative of peptides down-regulated. Black dots are peptides belonging to other non-histone proteins. Text boxes refer to histone peptides and modifications found statistically different between control cells (mesenchymal phenotype) and siA1_3 cells (epithelial phenotype). *D*) HIST1H2BC protein sequence with red letters underling modified peptides over-expressed upon HMGA1 silencing. White flags are indicative of acetylation residues. *E*) Co-immunoprecipitation analysis performed in MDA-MB-231. Two different lysate amounts (X and 2X) were loaded in lane 1 and lane 2 while lanes 3 and 4 display IgG immunoprecipitates (controls) and lanes 5 and 6 HMGA1 immunoprecipitates. DNA digestion was performed in order to verify if histones interaction with HMGA1 is mediated by nucleic acids (lanes 4 and 6). Reported are ponceau-stained membrane and western blot analyses of HMGA1 (panel α -HMGA1), histone H3 (panel α -H3) and histone H2B (panel α -H2B). *E1*) DNase digestion control; DNA of input (2%) either digested or not was loaded into 1.5% agarose gel (1/25000 gel red).

Along with these residues, also H2BK5ac was found determinant in phenotypic transition, and data collected by kit extraction figured out an up-regulation of this acetylated peptide upon HMGA1 knocking down. **Figure 3.1D** shows H2B (HIST1H2BC variant) protein sequence with peptides found up-regulated highlighted in red and white flags on hypothetical differentially acetylated lysines. Unfortunately, information regarding histone H3 tail were lacking, and this could not give a support to previously found alteration on phosphorylation events affected by HMGA1 and discovered by our laboratory (see preliminary data).

HMGA1 was previously shown to be able to interact with histone tails (Reeves & Nissen 1993). We confirmed these data demonstrating HMGA1 interaction with core histones (Li et al. 2006) by co-immunoprecipitation assay. As shown in **figure 3.1E**, upon HMGA1 immunoprecipitation (lanes 5 and 6 panel α -HMGA1), both histones H3 and H2B (the ones that we are interested on) are detected by western blot (WB) analyses of immunoprecipitates (lane 5 panels α -H3 and α -H2B) compared to controls (IgG lanes 3 and 4). Moreover, histone H3 and H2B interactions seem to persist after DNA digestion in HMGA1 immunoprecipitate (lane 6 compared to control lane 4). **Figure 3.1E1** is indicative of DNase digestion efficiency. Collectively these data suggest that HMGA1 could influence histones post translational modifications and that this effect could be due to the direct positioning of HMGA1 onto chromatin.

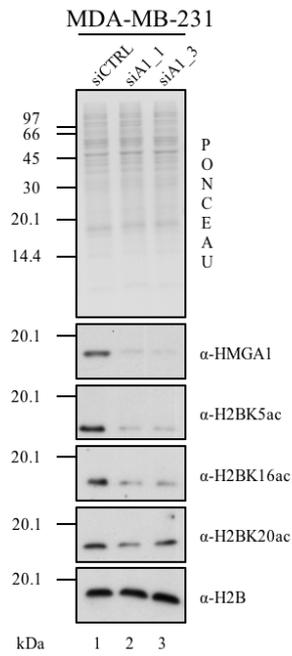
HMGA1 silencing induces downregulation of histone H2B acetylation

In order to confirm HMGA1 role in modulating the acetylation level of histone H2B, western blot (WB) analyses were conducted in MDA-MB-231 lysates transiently silenced for HMGA1 expression. In particular, cells were subjected to treatment with two different siRNA, siA1_3

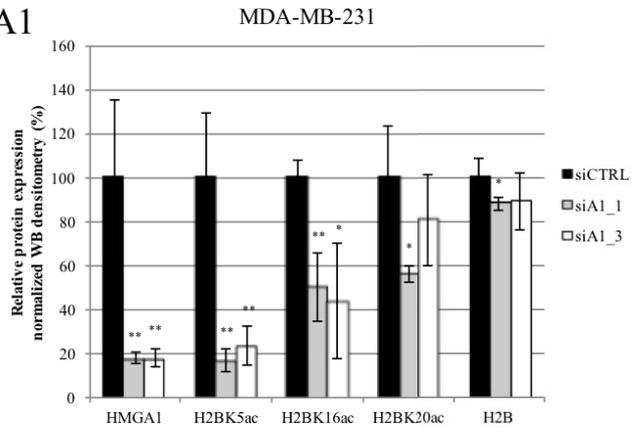
targeting ORF and siA1_1 targeting transcript 3'UTR; this system was designed to avoid undesirable off-target effects. Once having checked HMGA1 protein levels (panel α -HMGA1), levels of H2BK5ac, H2BK16ac, and H2BK20ac were checked along with total histone H2B (**figure 3.2A**). **Figure 3.2A1** reports densitometry values of bands normalized to ponceau densitometry of whole lysate's lanes. All modifications decrease upon HMGA1 knock down, even though a major down-regulation was observed in acetylation of lysine 5 and 16 with both siRNA strategies. Total histone H2B displays a slight reduction with siA1_1 which, however, we suppose could not account for the downfall of acetylation levels alone. Immunofluorescence assays performed on MDA-MB-231 confirmed a significant down-regulation of H2BK5ac that in control cells is homogeneously distributed throughout the cell population (**figure 3.2B** and relative quantification of fluorescence in **figure 3.2B1**). These data clearly disagree with MS data obtained with histones analyses. We suppose that this fact could be due to MS intrinsic limitations: in histone H2B sequence (shown in **figure 3.1C**) the first arginine is located at the 30th position producing an Arg-C digested peptide that could be too long to be analysed and sequenced or this digestion could be inefficient producing a peptide even longer; indeed, the peptides identified (red sequences) are both products of trypsin digestion. Moreover, the lysines that theoretically could be digested by trypsin could be acetylated becoming resistant to digestion. Therefore, we hypothesized that several K residues within the H2B N-terminal tail are differentially modified in control and knocked down breast cancer cells leading to the formation of different sets of peptides that make difficult and possibly misleading the interpretation of MS data.

Acetylation of histone H2B have been linked to epithelial to mesenchymal transition during mammalian development with similar evidences in breast cancer (Abell et al. 2011). We queried whether cell-cell contacts affect levels of H2BK5ac in MDA-MB-231 and whether HMGA1 could preside this mechanism independently to cellular confluence. We silenced MDA-MB-231 seeded at two different confluences, nearby confluent (400.000 cells plated 24 hours before treatment) and the half (200.000 cells plated 24 hours before treatment). Lysates were analysed by western blot and densitometry of bands normalized to total lysate (as previously described) were performed (**figure 3.2C** and **3.2C1**). Although identical quantities of total proteins were loaded as shown by membrane ponceau-staining, levels of acetylated H2BK5 were almost double in cells reaching confluence compared to half-seeded control cells. However, the presence of HMGA1 significantly influences the H2B acetylation level in both conditions. These experimental evidences suggest H2B acetylation could be tuned by multiple independent phenomena and actually HMGA1 expression is one of them.

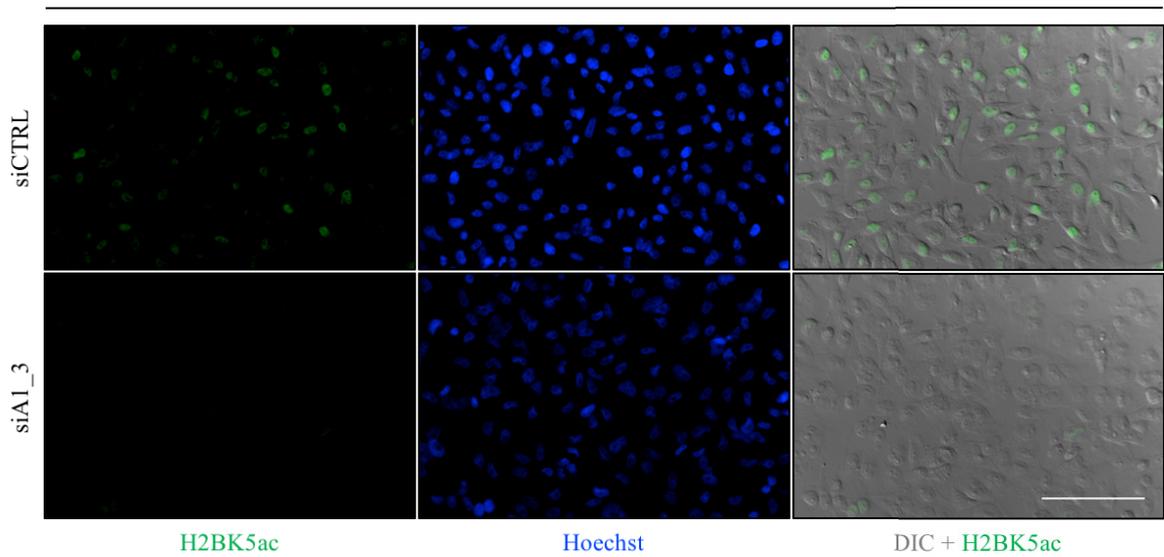
A



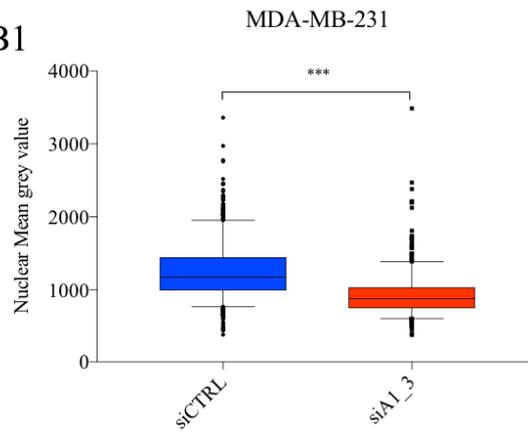
A1



B



B1



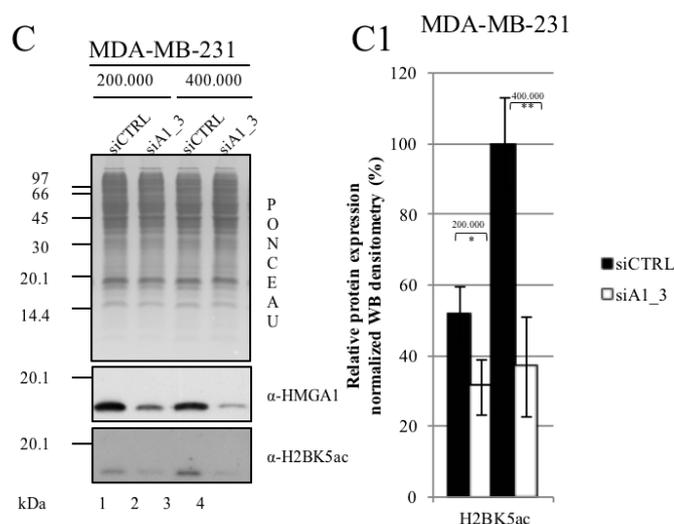


Figure 3.2 HMGA1 levels influence acetylation of histone H2B in MDA-MB-231. *A)* Western blot analyses of MDA MB 231 cell lysate collected at 72 hours upon HMGA1 silencing with siA1_3 (targeting ORF; lane 3), siA1_1 (targeting 3'UTR; lane 2) or control cells (siRNA lacking mammalian targets; lane 1). Panels reported are representative of biological triplicate (at least) and display, in order, the ponceau-stained membrane (illustrative of sample normalization) and western blot analyses of HMGA1, H2BK5ac, H2BK16ac, H2BK20ac and total H2B. Molecular references are reported in kDa on the left of each panel. *A1)* Histograms report band densitometry values normalized to ponceau-stained lysate's lane relative to controls. Mean and standard deviations are illustrated and samples are subjected to student t test; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$. *B)* Immunofluorescences assay of MDA MB 231 cells treated with siCTRL (left) or siA1_3 (right). Images are indicative of two fields of cells seeded on three different cover slips; cells were subjected to immunological detection of H2BK5ac (cyan-green emission of Alexa 488 fluorescent dye; green), nuclei staining with Hoechst (emission in cyan region of visible spectrum; blue) and merged differential interference contrast and H2BK5ac signals; scale bar 100 μm . *B1)* Box plot illustrative of median and quantiles distribution of nuclear mean grey values of H2BK5ac signals; whiskers extend to 5-95 percentiles and outliers are represented as dots. Data were analysed with Shapiro-Wilk normality test and Mann Whitney test; *** $p < 0.001$. *C)* WB analysis of MDA MB 231 cells silenced for HMGA1 expression at two different confluence states (200.000 cells - lane 1 and 2 - versus 400.000 cells - lane 3 and 4). Panels reported are representative of biological triplicate and illustrate ponceau-stained membrane and bands corresponding to HMGA1 and H2BK5ac antibody recognition. Molecular references are reported in kDa on the left of each panel. *C1)* Histogram displays band densitometry values normalized to ponceau-stained lysate's lane relative to controls. Mean and standard deviations are reported and samples are subjected to student t test; * $p < 0.05$; ** $p < 0.01$; *** $p < 0.001$.

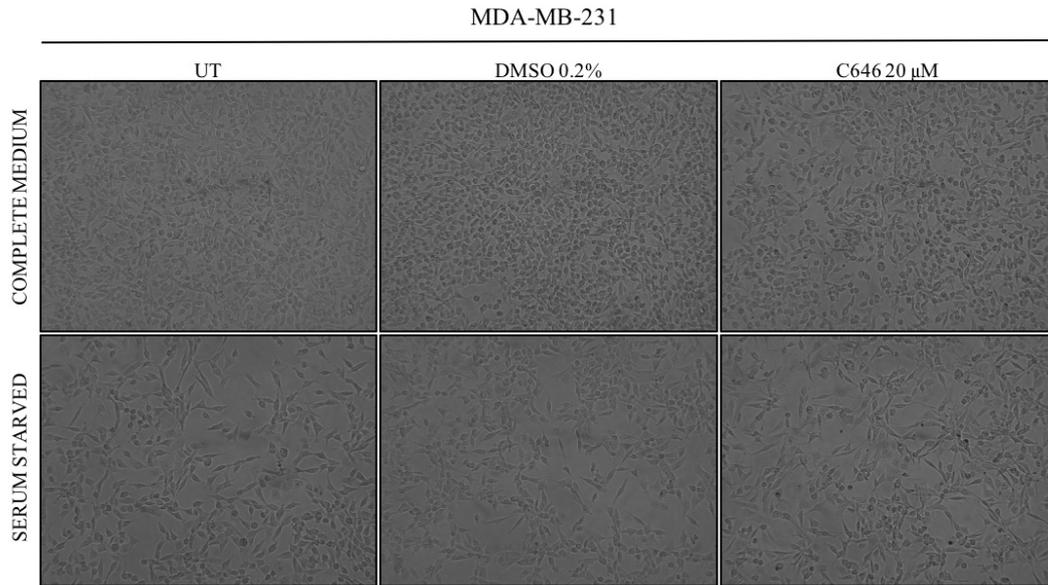
CBP/p300 acetyltransferases are responsible for H2B acetylation in MDA-MB-231

Core histones acetylation has been extensively linked to the activity of two distinct histone acetyltransferases (HATs), the paralog CBP (CREB Binding Protein) and p300 (E1A binding protein p300), which display strong sequence similarity and functional homology and could either play overlapping or unique functions (Chan & La Thangue 2001; Ogryzko et al. 1996). In particular, acetylation of histone H2BK5 has been primary associated with CBP acetyltransferase during cell differentiation (Abell et al. 2011). With this in mind, we investigated if acetylation status of H2B in MDA-MB-231 was affected by CBP/p300 activity and we used a pyrazolone-containing small molecule inhibitor (C646) at sufficient

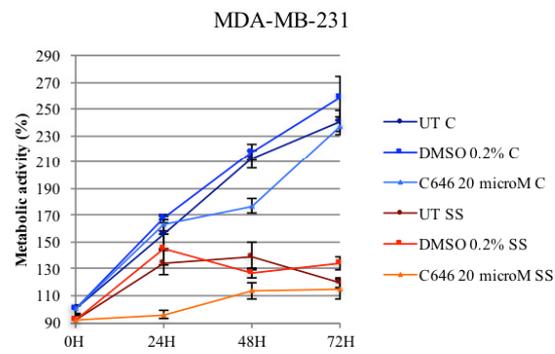
concentrations known to compete with acetyl-CoA for CBP/p300 active site binding (Bowers et al. 2010). Acetyl-CoA concentration has shown to influence HAT activity and selectivity (Henry et al. 2015); therefore, cells have been serum-starved for 24 h prior to C646 treatment so as to increase acetyltransferase inhibition. Cells have been either treated for 24 h with 20 μ M C646 or DMSO (0.2%) as control. As depicted by **figure 3.3A**, cells slow down when subjected to serum starvation; at the same time, treatment with small molecule inhibitor strongly impairs cell metabolism at 24 h only in starved cells whereas no difference could be observed among cells grown with complete medium (**figure 3.3B**). Differences are reduced at 48 h and 72 h between control and C646 treated cells even if a gap persists between serum starved and complete medium grown cells. Western blot analyses performed on cell lysates harvested 24 hours after treatment highlight a notable decrease in acetylation levels of both lysines 5 and 16 of histone H2B leaving total H2B level almost unchanged (**figure 3.3C**). Taken together, these data suggest that CBP/p300 are the acetyltransferases that target histone H2B in MDA-MB-231.

Thereafter, we hypothesised that if these modifications are involved in transcriptional activation in breast cancer cells, by affecting histone H2B acetylation we should obtain overlapping results in terms of gene expression alteration to the signature produced by HMGA1 knock down. After having checked by RT-qPCR that C646 mediated inhibition does not down-regulate CBP and p300 transcripts levels, we looked at some genes whose expression we recently demonstrated to be influenced by HMGA (Pegoraro et al. 2013): AURKB (Aurora kinase B), KIF23 (kinesin-like protein 23), KIF4A (kinesin family member 4A) and CENPF (centromere protein F) all correlating with clinical outcome data. As reported in **figure 3.3D**, inhibition of CBP/p300 resulted in a significant HMGA1 signature down-regulation suggesting a cooperation between these coactivators.

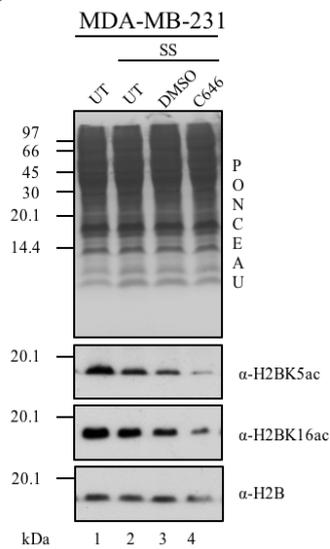
A



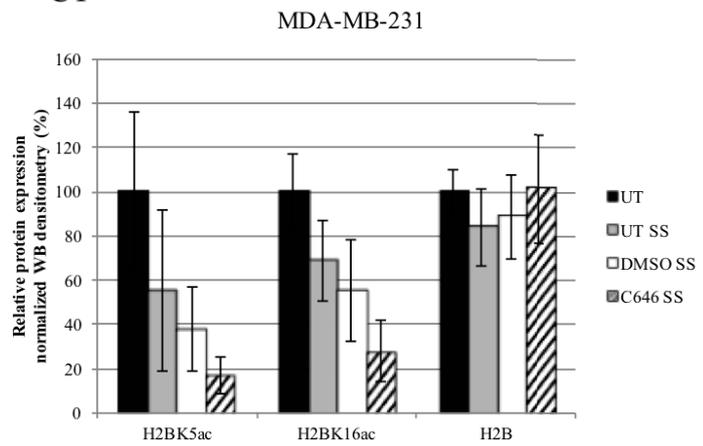
B



C



C1



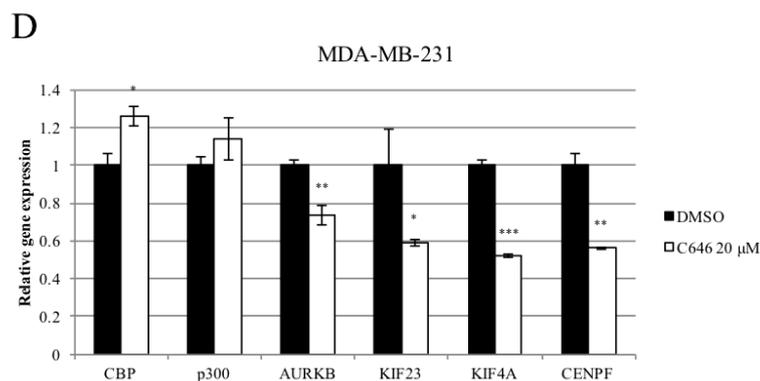


Figure 3.3 CBP/p300 acetyltransferases acetylate histone H2B in MDA-MB-231 cells producing transcriptional alterations overlapping to HMGA1-signature. *A)* Optical microscopy images of MDA-MB-231 grown with complete medium or serum starved. Cells were untreated (UT), treated with 20 μM C646, or incubated with 0.2% DMSO as control. *B)* MTS assay performed on six replicates (minimum and maximum values discarded) after blank subtraction; mean and standard deviation values are reported relative to UT values at time 0. “C” is indicative of complete medium (blue lines) while “SS” indicates serum starvation (red lines). *C)* WB analysis on MDA-MB-231 lysate grown with complete medium (lane 1) or serum starved (SS - lanes 2, 3 and 4) for 24 hours before treatment with 20 μM C646 or equivalent 0.2% DMSO as control (UT are untreated cells). Panels are representative of a biological triplicate. Red ponceau-stained membrane (illustrative of sample normalization) and western blot analyses of H2BK5ac, H2BK16ac, and total H2B are reported. *CI)* Histogram displays band densitometry values normalized to red ponceau-stained lysate’s lane relative to controls. Mean and standard deviations are reported. *D)* Relative gene expression values were obtained by RT-qPCR of MDA-MB-231 cells treated for 24 hours with 20 μM C646 versus control. GAPDH was used as internal reference gene. Mean and standard deviations (±SD) are reported and values are shown with respect to control sample. Experiment was done in biological triplicate and significance assigned by student t test; *p<0.05, **p<0.01, ***p<0.001.

CBP is the acetyltransferase responsible for transcriptional regulation of genes belonging to HMGA1-signature in MDA-MB-231

To dissect the roles played by acetyltransferases in conferring metastatic potential in breast cancer, we silenced the expression of CBP and p300 independently and we examined effects produced in both histone H2B acetylation and gene expression. The knock down of CBP caused high mortality in cell population (**figure 3.4A**) and evaluation of mesenchymal to epithelial transition (MET) was tricky. Anyway, from WB analyses performed in MDA-MB-231 lysates harvested 72 hours after siRNA administration, cells silenced for CBP expression (lane 2) displayed a significant downregulation of H2BK5ac levels compared to control cells (lane 1; **figure 3.4B** and normalized densitometry **figure 3.4B1**). A slight decrease was observed also in total H2B levels even if the entity of acetylation drop appears to be far greater. Exploring gene expression alterations, AURKB seems to be regulated at both mRNA and protein levels (**figure 3.4B**, **figure 3.4B1** and **figure 3.4C**). Moreover, all genes belonging to HMGA1 signature were found significantly down-regulated upon CBP silencing whereas, in a mechanism that suggests feedback looping, p300 mRNA expression increased.

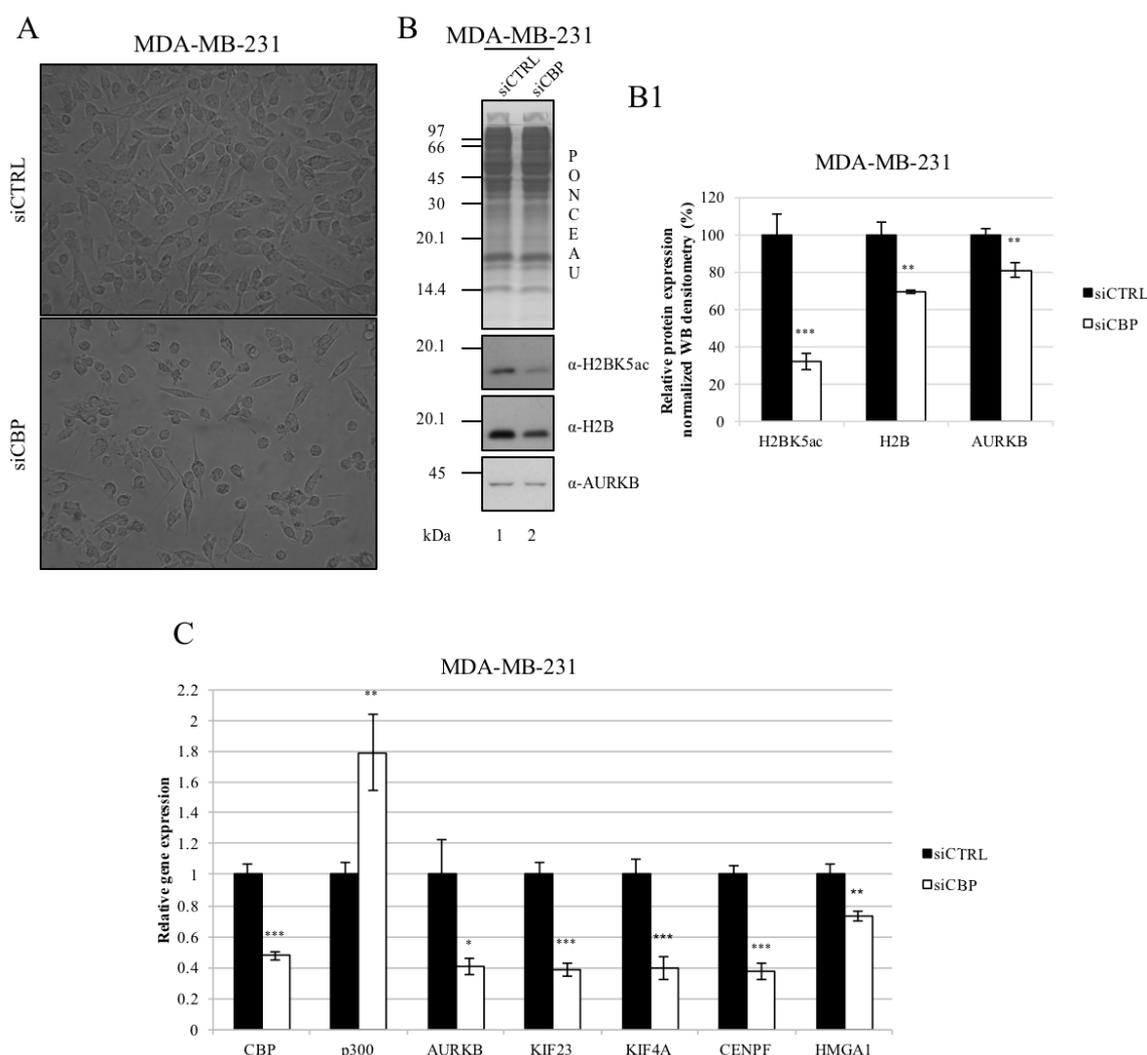


Figure 3.4 Silencing of CBP mirrors HMGA1 silencing effects on histone H2B acetylation and gene expression alteration in MDA-MB-231 cells. *A*) Optical microscopy images of MDA-MB-231 silenced with siCBP and controls. *B*) WB analysis on MDA-MB-231 lysates treated with control siRNA (lane 1) or CBP targeting siRNA (lane 2) for 72 hours. Panels are representative of a biological triplicate. Red ponceau-stained membrane (illustrative of sample normalization) and western blot analyses of H2BK5ac, total H2B, and AURKB are reported. *B1*) Histogram displays band densitometry values normalized to ponceau-stained lysate's lane relative to controls. Mean and standard deviations are reported. Significance was assigned by student t test; * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. *C*) Gene expression results obtained by RT-qPCR of MDA-MB-231 samples treated with siCTRL or siCBP and harvested after 72 hours. Mean and standard deviation (\pm SD) are reported relative to control sample. GAPDH was used as internal reference gene. Experiment was done in biological triplicate and significance assigned by student t test; * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Same experiments were performed on MDA-MB-231 silenced for p300 expression that, analysed by optical microscopy, seemed to acquire organized arrangement and flat morphology, characteristic of epithelial reversion (**figure 3.5A**). WB analyses performed in cell lysates upon silencing demonstrated p300 involvement on acetylation of histone H2B (**figure 3.5B** and **3.5B1**) and analogously to cells silenced for CBP, a parallel mild decrement of total histone H2B. However, both protein and mRNA levels of AURKB appeared

unchanged upon p300 silencing as well as KIF23 mRNA levels (**figure 3.5C**). KIF4A was the only gene belonging to HMGA1-signature significantly down-regulated in MDA-MB-231.

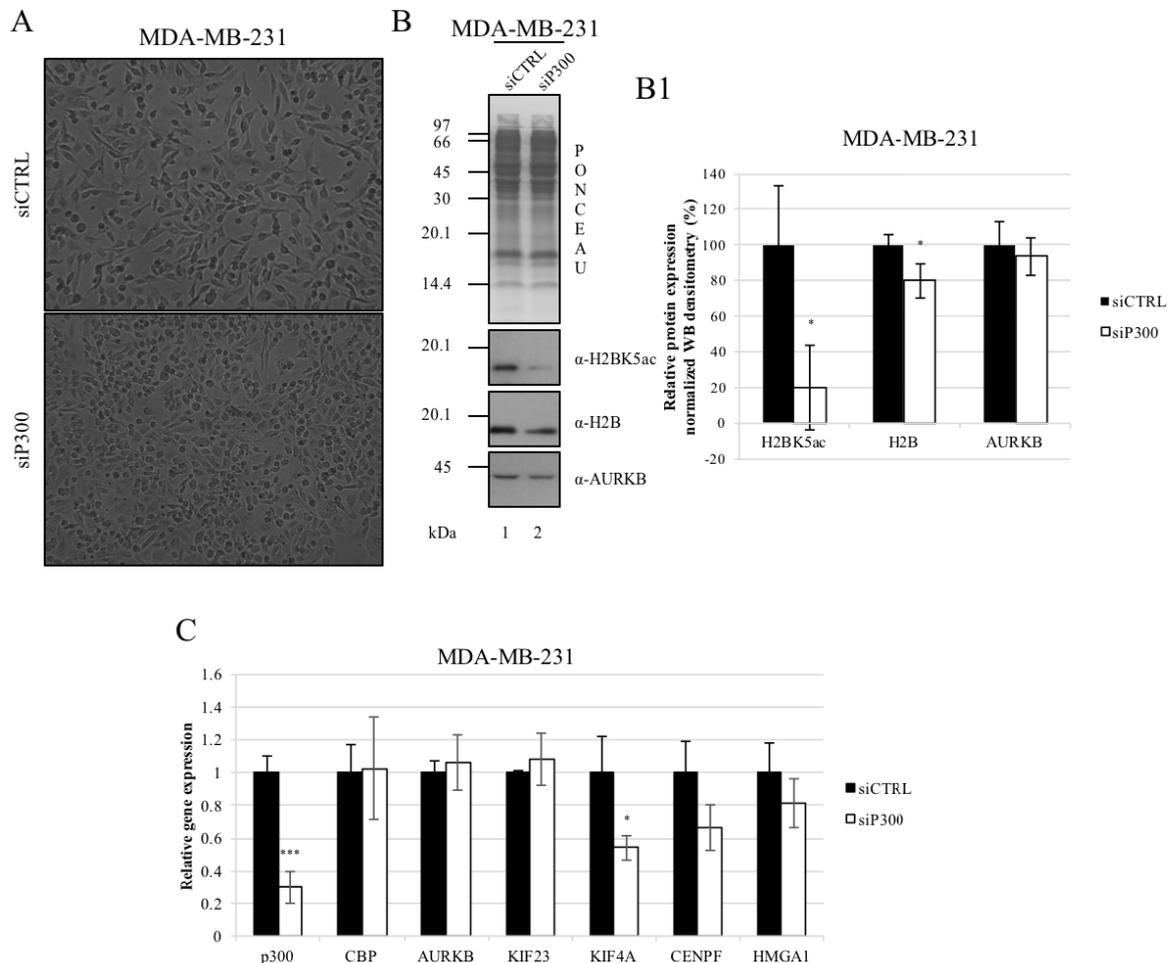
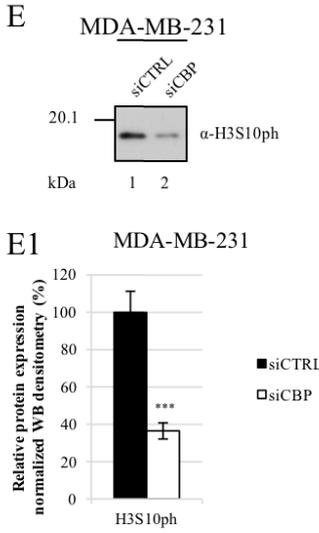
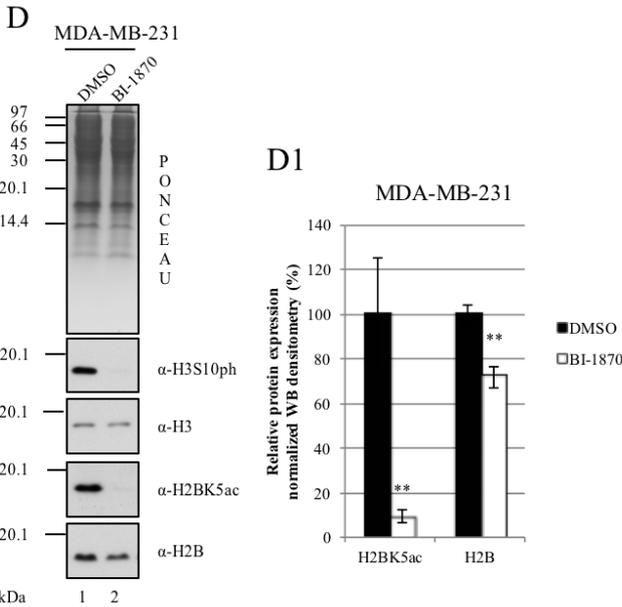
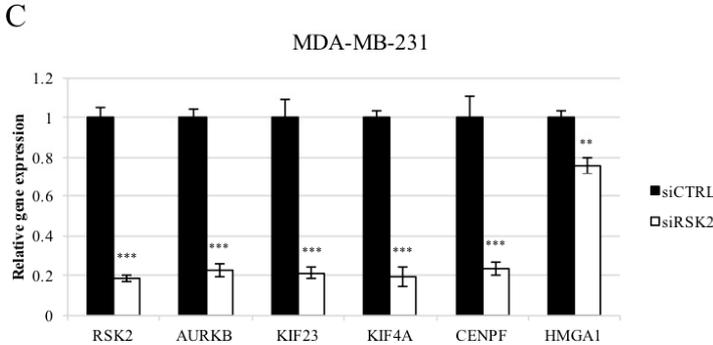
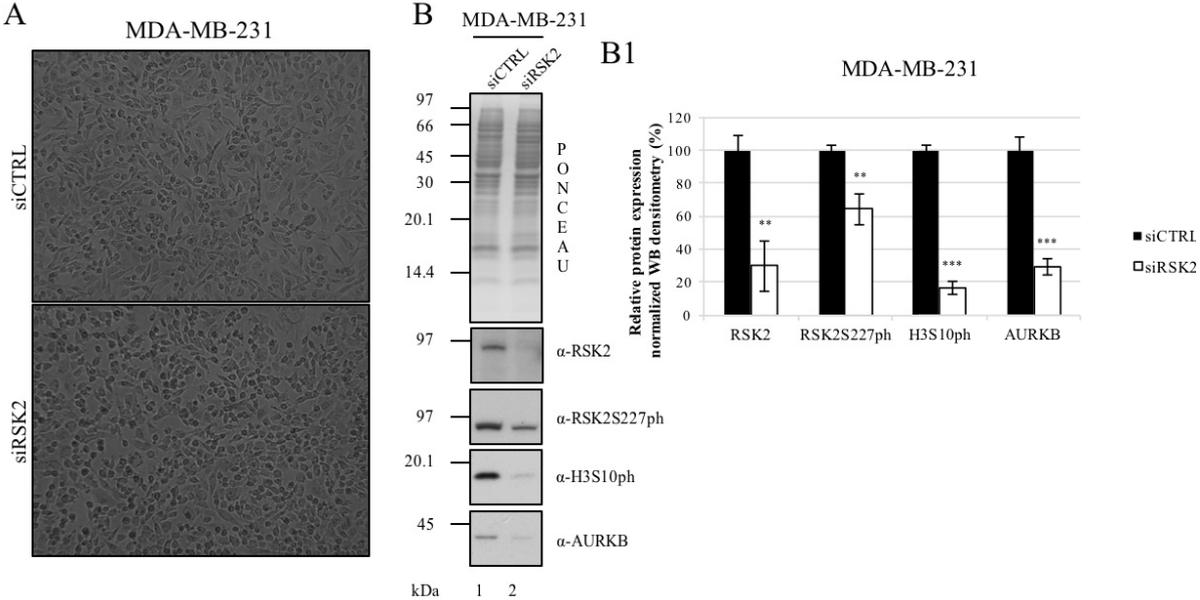


Figure 3.5 Silencing of p300 affect histone H2B acetylation in MDA-MB-231 cells but leaves unchanged the expression of the majority of HMGA1-signature genes. *A*) Optical microscopy images of MDA-MB-231 silenced with siP300 and controls. *B*) WB analysis on MDA-MB-231 lysates treated with control siRNA (lane 1) or siP300 (lane 2) for 72 hours. Panels are representative of a biological triplicate, and reported are ponceau-stained membrane (illustrative of sample normalization) and western blot analyses of H2BK5ac, total H2B and AURKB. *B1*) Histogram displays band densitometry values normalized to ponceau-stained lysate's lane relative to controls. Mean and standard deviations are reported. Significance was assigned by student t test; * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. *C*) Relative gene expression results obtained by RT-qPCR of MDA-MB-231 samples treated with siCTRL or siP300 and harvested after 72 hours. Mean, standard deviation (\pm SD) are reported relative to control sample. CYP33 was used as internal reference gene. Experiment was done in biological triplicate and significance assigned by student t test; * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Collectively these results suggest that CBP is the acetyltransferase responsible for transcriptional regulation of the majority of HMGA1-signature genes and confirmed that besides redundant functions hold by CBP/p300 they could acetylate same residues with different specificity and selectivity driving unique functions (Kalkhoven 2004; Henry et al. 2013).

CBP histone acetylation is functional coupled with RSK2 phosphorylation of H3S10 in MDA-MB-231

Histone acetylation and phosphorylation have been coupled to transcriptional activation also at signalling level: CBP has been shown to interact with un-phosphorylated RSK2 forming a complex where both CBP and RSK2 are kept inactivated. Upon mitogenic stimuli and phosphorylation of RSK2S227, they dissociate and act respectively as histone kinase or HAT driving gene expression activation (Merienne et al. 2001). Data collected in our laboratory provide a solid and robust evidence of HMGA1 involvement in regulating gene expression by modulation of RSK2 (ribosomal protein S6 kinase, 90kDa, polypeptide 3) and phosphorylation of serine 10 of histone H3 (see preliminary data section at the end of this thesis). Indeed, by inhibition of RSK2 with BI-D1870 we obtained striking conformational changes attributable to MET, reduced cell migration and down regulation of HMGA1-signature gene expression in MDA-MB-231. Here we confirmed that H3S10ph decrease upon RSK2 silencing which impairs both kinase mRNA and its activity (RSK2S227ph; **Figure 3.6B** and relative normalized densitometry **figure 3.6B1**). Moreover, the knock down of RSK2 expression produces a gene signature overlapping the one produced by HMGA1 (**figure 3.6C**). Therefore, we investigated if modulating RSK2 activity by use of selective inhibitor BI-D1870, we could regulate CBP acetyltransferase activity. We analysed by WB lysates of MDA-MB-231 cells treated with either BI-1870 10 μ M or DMSO alone as a control. As shown in panel D of **figure 3.6**, H2BK5ac levels were affected by RSK2 inhibition (lane 2 with respect to lane 1 as control). Slight down-regulation was observed also in total H2B protein. **Figure 3.6D1** reported normalized densitometry of bands and student t test results. On the other hand, we decided to check if phosphorylation level of H3S10 was influenced by histone acetyltransferase activity. We analysed by WB lysates of MDA-MB-231 cells silenced for CBP expression and we observed a strong and significant downregulation of H3S10ph (**figure 3.6E** and relative normalized densitometry **figure 3.6E1**). Overall these results suggest that cell lacking HMGA1 expression maintain inactivated both RSK2 and CBP while oncogenic overexpression of HMGA1 could lead to gene expression of factors critical for cell migration and malignancy by aberrant reactivation of kinase and HAT (**figure 3.6F**). Moreover, these data suggest a functional interplay between RSK2 and CBP.



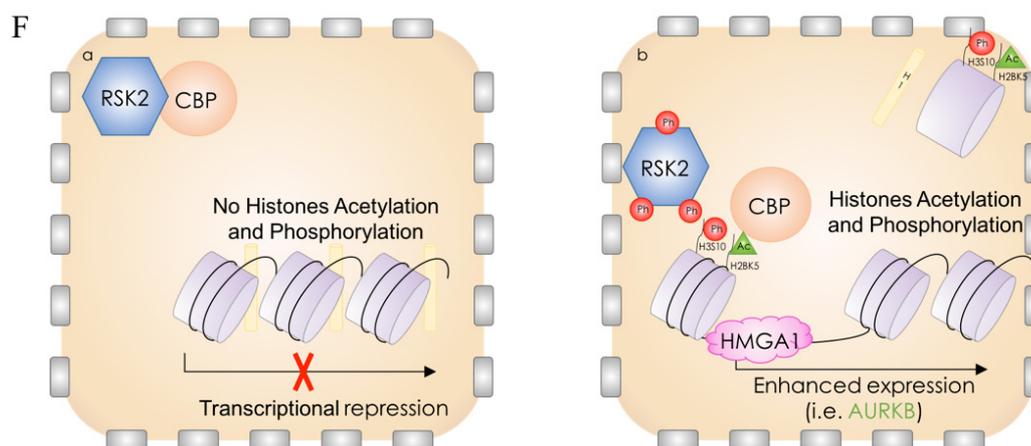


Figure 3.6 Kinase and HAT activities of CBP and RSK2 are functional coupled in modulating gene expression of HMGA1-signature in MDA-MB-231. *A)* Optical images of MDA-MB-231 treated either with siCTRL or siRSK2 for 72 hours. *B)* WB analysis of MDA-MB-231 lysates upon RSK2 silencing (lane 2) or controls (lane 1). Panels are representative of biological triplicate, and reported are ponceau-stained membrane (illustrative of sample normalization) and western blot analyses of RSK2, RSK2S227ph, H3S10ph and AURKB. *B1)* Histogram displays band densitometry values normalized to ponceau-stained lysate's lane relative to controls. Mean and standard deviations are reported. Significance was assigned by student t test; * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. *C)* Relative gene expression results obtained by RT-qPCR of MDA-MB-231 samples treated with siCTRL or siRSK2 and harvested after 72 hours. Mean and standard deviation (\pm SD) are reported relative to control sample. CYP33 was used as internal reference gene. Experiment was done in biological triplicate and significance assigned by student t test; * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. *D)* WB analysis of MDA-MB-231 lysates upon RSK2 inhibition with 10 μ M BI-1870 (lane 2) or DMSO (lane 1). Panels are representative of a biological triplicate, and reported are ponceau-stained membrane (illustrative of sample normalization) and western blot analyses of H3S10ph, total H3, H2BK5ac and total H2B. *D1)* Histogram displays band densitometry values normalized to ponceau-stained lysate's lane relative to controls. *E)* WB analysis of MDA-MB-231 lysates upon CBP silencing (lane 2) or control cells (lane 1). Panels are representative of a biological triplicate, and reported is western blot analysis of H3S10ph. For sample normalization refer to figure 3.3B ponceau-stained membrane. *E1)* Histogram displays band densitometry values normalized to ponceau-stained lysate's lane relative to controls. *F)* Speculative model HMGA1-linked alteration of breast cancer cell epigenetic marks related to RSK2 and CBP.

HMGA1 modulate RSK2 and CBP activities without inferring with CREB phosphorylation

The transcription factor CREB (cAMP response element binding protein) is a downstream effector of different signalling pathways and regulates transcriptional inducibility of several genes. Among other kinases, phosphorylation of CREB serine 133 is mediated by RSK2 and is associated to an increase of CBP acetyltransferase activity (De Cesare et al. 1998; Merienne et al. 2001). Hence, we hypothesised that HMGA1 overexpression could lead to aberrant phosphorylation of CREB which in turn could result in triggered HAT activity of CBP (**Figure 3.7A**). However, by WB analyses performed on MDA-MB-231 cells silenced for HMGA1 expression (for lysates normalization and levels of HMGA1 expression refer to lane 1 - siCTRL - and lane 3 - siA1_3 - of **figure 3.2A**), we observed that levels of total CREB and CREBS133ph are independent from HMGA1 levels (**figure 3.7B** and relative normalized

bands densitometry, **figure 3.7B1**). Therefore, in our model, RSK2 and CBP activities are not linked to CREB.

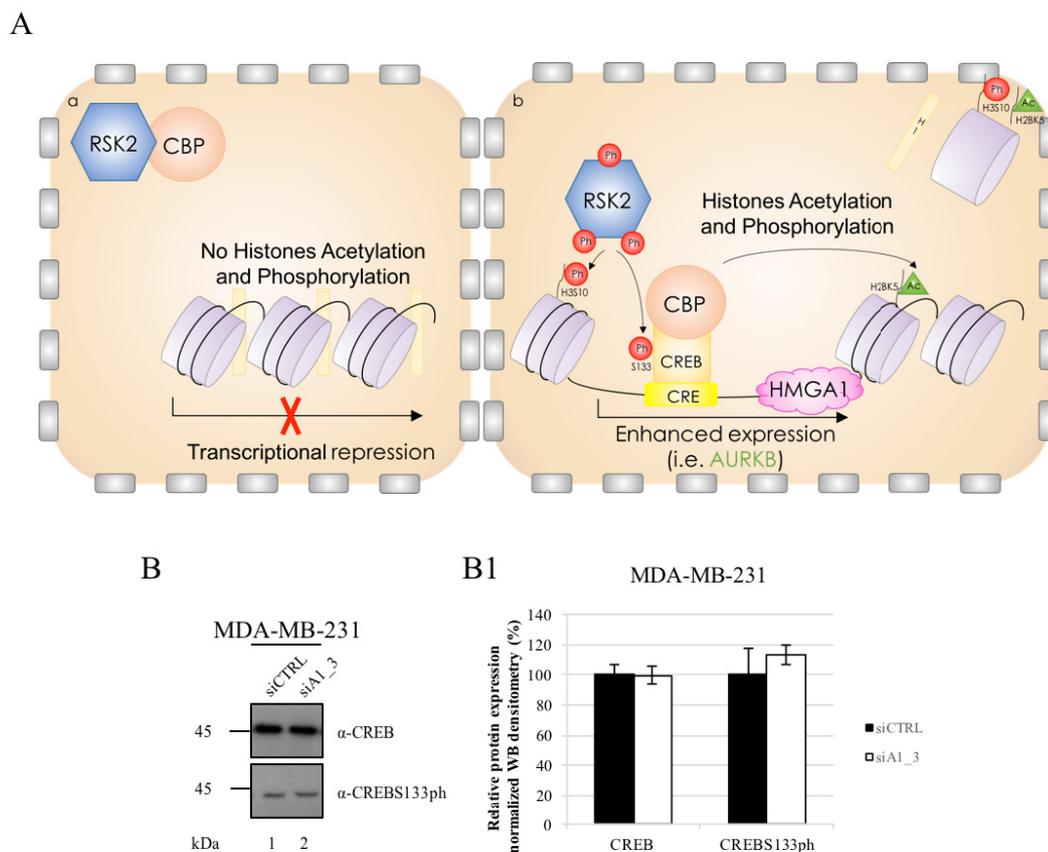


Figure 3.7 HMG1 interplay with CBP and RSK2 does not involve CREB phosphorylation. *A*) Hypothesised model of effects of HMG1 overexpression in RSK2 phosphorylation of CREB and recruitment of CBP for aberrant gene expression in breast cancer cells (panel a). Panel a is indicative of knock down cells (or normal breast cells) where low levels of HMG1 keep CBP and RSK2 inactive in nuclei. *B*) WB analysis performed on MDA-MB-231 harvested 72 hours after HMG1 silencing (lane 2) compared to control cells (lane 1). For normalization and HMG1 intensity levels refer to figure 3.1A. Panels are representative of a biological triplicate and western blot analyses of total CREB and CREBS133ph levels. *B1*) Histogram representative of mean and standard deviations (\pm SD) of bands densitometry normalized to intensity of ponceau-stained lysate membrane.

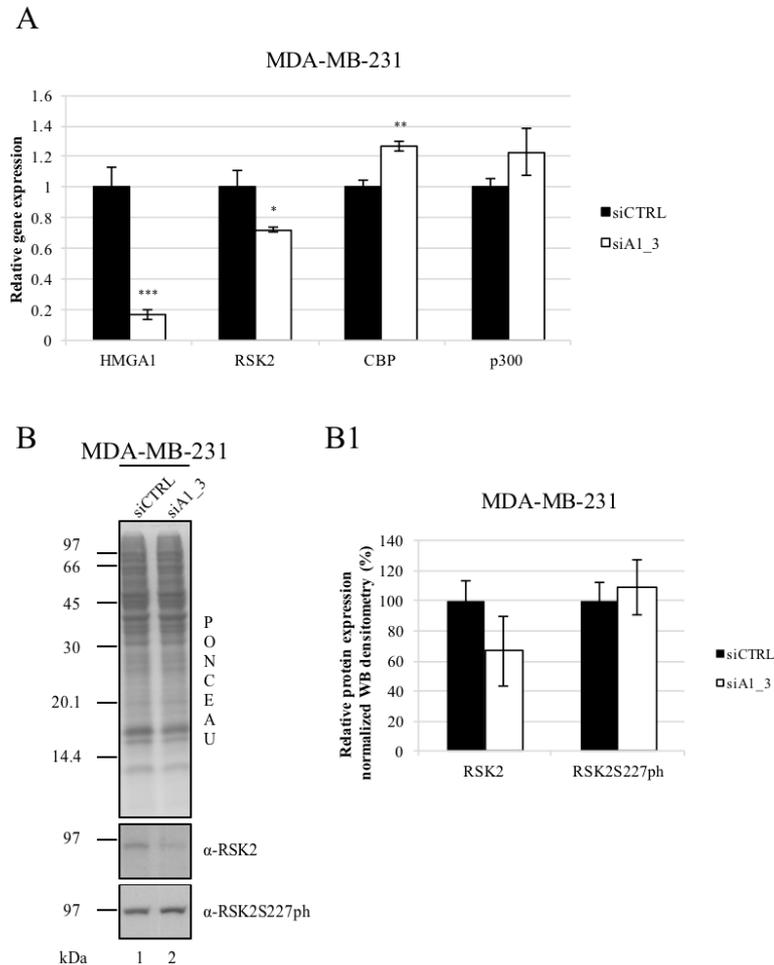
HMG1 regulates epigenetic modifications acting downstream RSK2 activation

We considered that HMG1 could influence histone post-translational modifications by different mechanisms: it could be involved in transcriptional regulation of modifiers expression, in the mechanism of their activation (i.e. interfering with modulators expression or activities) or at chromatin level modulating enzymes accessibility to mediators. Therefore, we silenced HMG1 expression and we glanced at kinase and HATs expression responsible for the modification observed by RT-qPCR assay. As shown in **figure 3.8A**, MDA-MB-231 cells

treated with siA1_3 displayed a slight downregulation of RSK2 expression compared to controls, whereas CBP and p300 appeared even up-regulated upon HMGA1 silencing.

The downregulation of total RSK2 was confirmed by WB analyses performed on cell lysates treated either with siA1_3 or siCTRL for 72 hours even if means of normalized bands densitometries of the biological triplicates were not significantly different (**figure 3.8B**). Hence, we tried to demonstrate that RSK2 expression could be not indicative of differentially activated kinase levels in the two conditions analysed. RSK2 is composed by two kinases domains (N-terminal and C-terminal, respectively NTK and CTK) connected by a linker region of 100 aa. Mechanism of RSK2 activation is complex and includes many phosphorylation events: extracellular signal-regulated kinase (ERK) is responsible for S369 phosphorylation in the linker region and T577 phosphorylation in the activating loop of CTK; last phosphorylation activates the C-terminal domain for the *cis* phosphorylation of S386 which, once phosphorylated, functions as a docking site for 3-phosphoinositide-dependent protein kinase-1 (PDK1). This kinase is responsible for S227 phosphorylation in NTK activation loop which is the one finally responsible for RSK2 substrates phosphorylation (Jensen et al. 1999). Therefore, as a measure of RSK2 activation we decided to query levels of RSK2S227ph in samples silenced for HMGA1 expression. WB analyses performed on MDA-MB-231 highlight that S227 phosphorylation levels are not significantly changed upon HMGA1 depletion despite total protein downregulation (**figure 3.8B** and relative normalized bands densitometry in **figure 3.8B1**). Once activated, RSK2 migrates in the nucleus where it phosphorylates its substrates in a poor described mechanism (Arul & Cho 2013). Hence, we performed immunofluorescence analysis on MDA-MB-231 cell treated with siCTRL, siA1_3 or untreated cells (UT) after 72 hours. As shown by **figure 3.8C** and quantification reported in **figure 3.8C1**, RSK2 seems to be more nuclear after HMGA1 knocking-down, supporting previous results which state that HMGA1 is not responsible for its upstream activation. In particular, obtaining nuclear masks by Hoechst nuclear delimitation and cytoplasmic region by subtracting nuclei area to a region of interest created by a secondary antibody emission (recognizing α -RSK2 primary antibody), we analysed the trend of the curves made by plotting number of pixels (normalized to the sum) with their respective grey values and results suggest a mild increase of RSK2 nuclear translocation upon HMGA1 silencing. These results agree with WB data suggesting a slight boost on RSK2S227ph levels in HMGA1-depleted cells and suggest that HMGA1 could exert its function downstream RSK2 activation, whose action is in turn related to histone acetylation levels. **Figure 3.8C2** is a WB indicative of HMGA1 silencing and is performed on a population of cell lysate used for immunofluorescence

analysis. Finally, we investigated if HMGA1 and RSK2 interact in MDA-MB-231 and a co-immunoprecipitation experiment was done in presence or absence of DNase I in order to discriminate whether their interaction was mediated by DNA. As reported in **Figure 3.8D**, WB analysis was performed after immunoprecipitation with α -HMGA1 antibody (lanes 4 and 5) or IgG as controls (lanes 2 and 3). DNase digestion was performed on lysates 3 and 5 and the input was loaded in lane 1. As shown in the panel relative to RSK2 recognition, this kinase was immunodetected in HMGA1 immunoprecipitate without DNase (lane 4) whereas it is absent when DNA was digested in cell lysate (lane 5). **Figure 3.8D1** shows the control of DNase digestion. As it is possible to observe, the protocol for immunoprecipitation leads to the formation of DNA fragments of about 600 bp. Taken together these results suggest that HMGA1 and RSK2 are found in complex in chromatin, within regions of about 600 bp but their interaction could be mediated by other proteins and/or DNA.



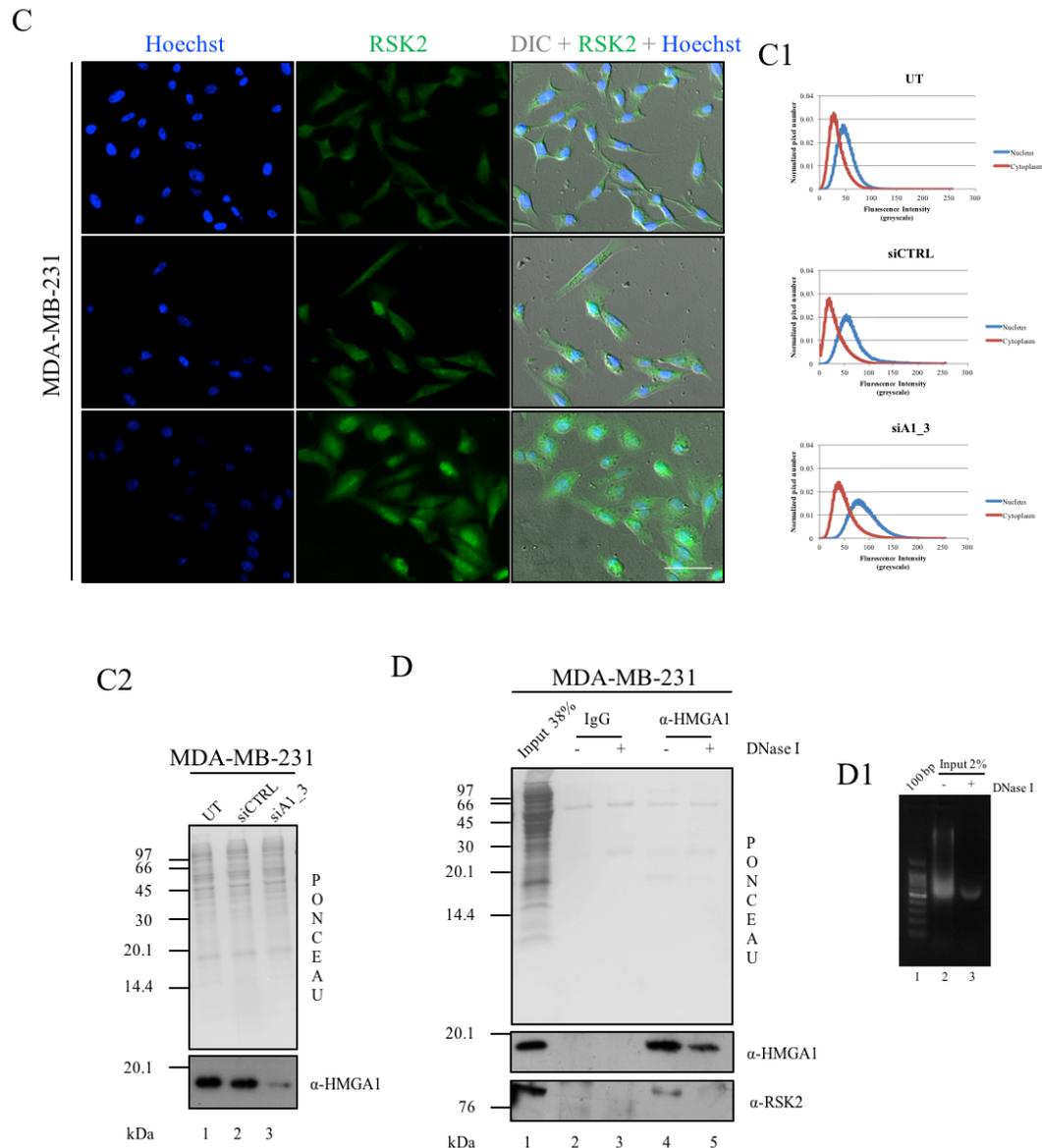


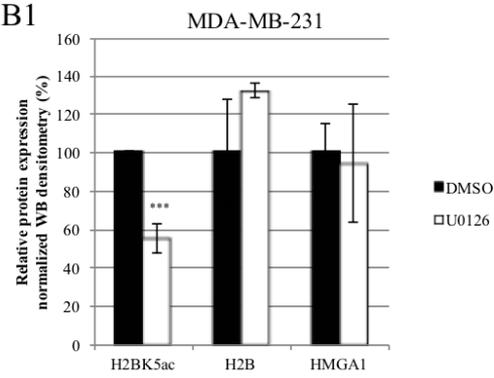
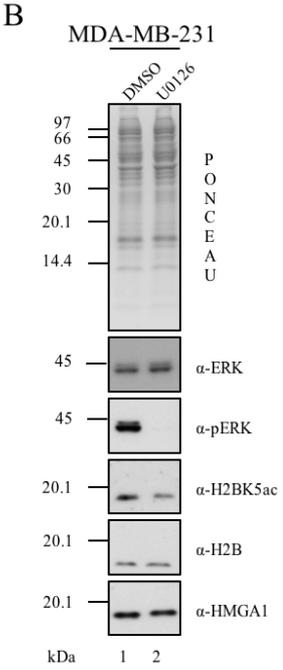
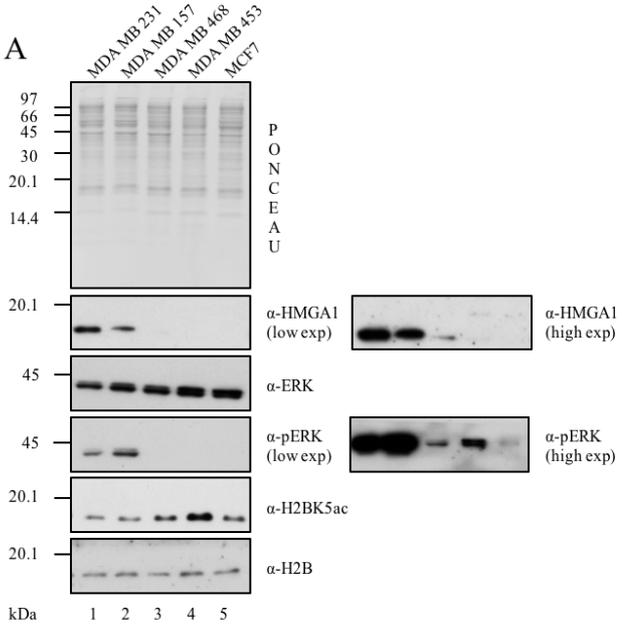
Figure 3.8 HMGA1 modulate histones PTMs acting downstream RSK2 phosphorylation. *A*) Relative gene expression results obtained by RT-qPCR of MDA-MB-231 samples treated with siCTRL or siA1_3 and harvested after 72 hours. Mean, standard deviation (\pm SD) are reported relative to control sample. GAPDH was used as internal reference gene. Experiment was done in biological triplicate and significance assigned by student t test; * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. *B*) WB analysis of MDA-MB-231 lysates upon HMGA1 silencing (lane 2) or controls (lane 1). Panels are representative of biological triplicate, and reported are ponceau-stained membrane (illustrative of sample normalization) and western blot analyses of RSK2 and RSK2S227ph. For HMGA1 levels please refer to image 3.1A lane 1 (CTRL) versus lane 3 (A1_3) panel α -HMGA1. *BI*) Histogram displays band densitometry values normalized to ponceau-stained lysate's lane relative to controls. Mean and standard deviations are reported. *C*) Immunofluorescence analysis performed in untreated MDA-MB-231, same cells treated with siCTRL or siA1_3 placed in a cover slip. Images are indicative of four fields of cells; cells were subjected to immunological detection of RSK2 (cyan-green emission of Alexa 488 fluorescent dye; green), nuclei staining with Hoechst (emission in cyan region of visible spectrum; blue) and merged differential interference contrast and RSK2 and Hoechst signals; scale bar 50 μ m. *C1*) Scatterplot of number of pixels versus their fluorescence intensity (in term of grayscale); blue curve is representative of nuclear fluorescence whereas red curve is representative of cytoplasmic fluorescence. Subcellular regions are defined by Hoechst emission (nuclei) and cyan-green emission of Alexa 488 fluorescent dye (cytoplasmic). Number of pixels are normalized to the sum of pixels measured in the region of interest of each field. *C2*) WB analysis of a population of cells belonging to immunofluorescence analysis to verify HMGA1 knock down. Untreated (UT - lane 1), controls (siCTRL - lane 2) and HMGA1-silenced (siA1_3 - lane 3) are compared and ponceau-stained membrane is

reported indicative of normalization. Second panel shown bands detected in autoradiography films after antibody recognition of HMGA1. *D*) Co-immunoprecipitation analysis performed in MDA-MB-231. The 38% of input lysate was loaded in lane 1 while lanes 2 and 3 display IgG immunoprecipitates (controls) and lanes 4 and 5 HMGA1 immunoprecipitates. DNA digestion was performed in order to verify if RSK2 interaction with HMGA1 is mediated by nucleic acids (lanes 3 and 5). Reported are ponceau-stained membrane and western blot analyses of HMGA1 (panel α -HMGA1) and RSK2 (panel α -RSK2). *D1*) DNase digestion control; DNA of input (2%) either digested or not was loaded into 1.5% agarose gel (1/25000 gel red).

Different HMGA1 abundances reflect different signalling regulation of histone H2B acetylation pattern

Once demonstrate HMGA1 interference with RSK2 and CBP activities in MDA-MB-231 cell line, we decided to investigate if this phenomenon was shared with other breast cancer cell lines and if modification patterns are equivalent in cells with different levels of HMGA1. To this purpose we focused onto HMGA1 connection to histone H2B acetylation and therefore we investigated levels of HMGA1, acetylated H2BK5 and total histone H2B in different breast cancer models: MDA-MB-231 and MDA-MB-157 triple negative breast cancer cell lines classified as mesenchymal-like/claudin-low; MDA-MB-468 triple negative breast cancer cell lines classified as basal-like and displaying amplification of EGFR; MDA-MB-453 negative for ER and PR but HER2 positive; MCF7 positive for both ER and PR negative for HER2 and classified as luminal A. We harvested sub-confluent cells and we performed western blot on their lysates. As shown in **figure 3.9A**, comparing differentially classified breast cancer cells, unexpectedly levels of histone H2BK5ac where higher in breast cancer cells with lower levels of HMGA1 (lanes 3 - 4 - 5, respectively MDA-MB-468, MDA-MB-453 and MCF7 bands in panel α -H2BK5ac) while total H2B levels were comparable in all cell lines used (panel α -H2B). Thus, we questioned if signalling pathway modulators which influence histone modifiers activities where equally abundant in different cell lines used and were associated to HMGA1 levels. First of all, we confirmed that RSK2 activation in MDA-MB-231 was primarily modulated by Ras-RAF-MEK-ERK pathway (Sassone-Corsi et al. 1999) which is responsible for phosphorylation of at least two different residues in RSK2 protein, as told before. We performed WB analysis on cells treated for 24 hours with 10 μ M U0126, a non-competitive inhibitor of MEK (Favata 1998) and results suggest that inhibition of ERK activation (verified in panel α -pERK) significantly affect acetylation of H2BK5 (panel α -H2BK5ac) in U0126-treated cells compared to controls (DMSO) in MDA-MB-231 (**figure 3.9B** and relative normalized densitometry in **figure 3.9B1**). Therefore, we questioned if different cell lines exhibit differential activation of the Ras pathway. In fact, both TNBC mesenchymal-like cell models used displayed greater levels of pERK compared to other cell

lines while they own equal ERK expression levels (lanes 1 and 2 in figure 3.9A, panels α -ERK and α -pERK). We observed that HMGA1 levels were higher in cells with a greater ERK activation, observation confirming literature data that link HMGA1 expression to the activation of Ras pathway (Treff NR et al., 2004). We demonstrated that HMGA1 silencing in MDA-MB-231 does not affect neither phosphorylated nor total ERK levels (see preliminary data section at the end of this thesis). In MDA-MB-231 higher level of pERK is associated with oncogenic KRAS mutation; the other cell lines used have different activating/inactivating mutations in members of other pathways. In particular, MDA-MB-468 and MDA-MB-453 cell lines possess inactivating mutations in tumor suppressor PTEN (phosphatidylinositol-3,4,5-trisphosphate 3-phosphatase), while MCF7 manifest PIK3CA activating mutation, both regulating PI3K (Phosphoinositide 3-kinase) signalling cascade (Hollestelle et al. 2007). Then, we hypothesised that cells lacking HMGA1 expression could have different signalling mechanisms driving regulation of histone acetylation (and phosphorylation) and that these modulators belong to the PI3K pathway. Moreover, PDK1 (3-phosphoinositide-dependent kinase-1), the kinase responsible for RSK2S227ph, is connected to PI3K-AKT pathway activation (Vanhaesebroeck & Alessi 2000). Therefore, we decided to use both MEK inhibitor (U0126) and PI3K inhibitor (LY294002; Vlahos et al. 1994) to address which transduction signal pathway was the one involved in modifying acetylation of histone H2B. We used 10 μ M of each inhibitor for 24 hours in MDA-MB-157 and MDA-MB-453. As shown in **figure 3.9C**, by comparing control cells (lane 1) with cells treated with U0126 (lane 2) and LY294002 (lane 3), levels of H2BK5 acetylation (panel α -H2BK5ac) looked down-regulated upon inhibition of ERK phosphorylation (panel α -pERK) in MDA-MB-157. Both MDA-MB-231 and MDA-MB-157 belong to molecular mesenchymal-like/claudin low TNBC subtype and their pERK levels are comparable (lanes 1 and 2 **figure 3.9A** panel α -pERK). Moreover, inhibition of PI3K does not affect levels of H2BK5ac in both cellular models (MDA-MB-157 lane 3 **figure 3.9C** panel α -H2BK5ac; MDA-MB-231 data not shown). MDA-MB-453 cells belong to a different breast cancer molecular subtype, the HER2 overexpressing one, and display PTEN inactivating mutation.



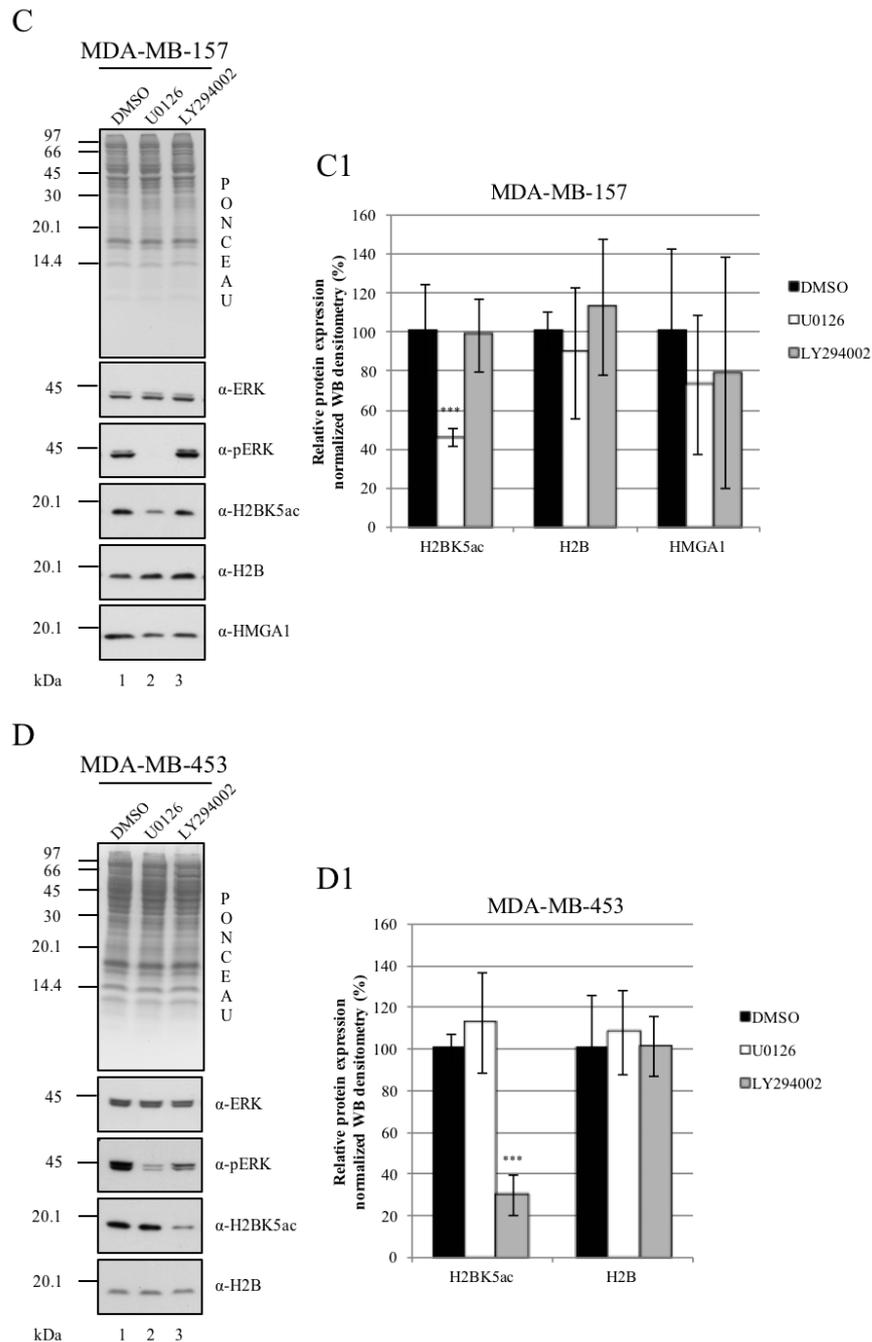


Figure 3.9 Different molecular subtypes of breast cancer display distinct signalling mechanism of histone H2B acetylation which mirror HMGA1 abundances. A) WB analysis of MDA-MB-231, MDA-MB-157, MDA-MB-468, MDA-MB-453 and MCF7 lysates. Panels are representative of biological quadruplicate, and reported are ponceau-stained membrane (illustrative of sample normalization) and western blot analyses of HMGA1 (low and high exposures, respectively left and right panels), ERK, pERK (low and high exposures, respectively left and right panels), H2BK5ac and total H2B levels. B) WB analysis of MDA-MB-231 lysates upon 24 hours treatment with 10 μ M of U0126 inhibitor (lane 2) compared to controls (lane 1; DMSO). Panels are representative of biological triplicate, and reported are ponceau-stained membrane (illustrative of sample normalization) and western blot analyses of ERK, pERK, H2BK5ac, total H2B and HMGA1 levels. B1) Histogram displays band densitometry values normalized to ponceau-stained lysate's lane relative to controls. Mean and standard deviations (\pm SD) are reported. Significance assigned by student t test; * p <0.05, ** p <0.01, *** p <0.001. C) WB analysis of MDA-MB-157 lysates upon 24 hours treatment with 10 μ M of U0126 inhibitor (lane 2), 10 μ M of LY294002 inhibitor (lane 3) and controls (lane 1; DMSO). Panels are representative of biological triplicate, and reported are ponceau-stained membrane (illustrative of sample normalization) and

western blot analyses of ERK, pERK, H2BK5ac, total H2B and HMGA1 levels. *C1*) Histogram displays band densitometry values normalized to ponceau-stained lysate's lane relative to controls. Mean and standard deviations (\pm SD) are reported. significance assigned by student t test; * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. *D*) WB analysis of MDA-MB-453 lysates upon 24 hours treatment with 10 μ M of U0126 inhibitor (lane 2), 10 μ M of LY294002 inhibitor (lane 3) and controls (lane 1; DMSO). Panels are representative of biological triplicate, and reported are ponceau-stained membrane (illustrative of sample normalization) and western blot analyses of ERK, pERK, H2BK5ac and total H2B levels. *D1*) Histogram displays band densitometry values normalized to ponceau-stained lysate's lane relative to controls. Mean and standard deviations (\pm SD) are reported. Significance assigned by student t test; * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$.

Querying the acetylation status of histone H2B we observed that in this cell line the PI3K-AKT pathway inhibition results in decreasing H2B acetylation levels (lane 3 compared to line 1, **figure 3.9D** panel α -H2BK5ac) whereas MEK inhibition does not lead to any effects on this modification (lane 2). Taken together, these data suggest that distinct molecular subtypes show different signalling modulations of histone modifiers and in TNBC cells overexpressing HMGA1 and having the Ras pathway active, HMGA1 could be a specific key modulator of their epigenetic signature.

DISCUSSION

The analysis of histone marks distribution among different breast cancer molecular subtypes provides a specific chromatin signatures defined by the binding pattern of activating or repressing modifications. Genomic distribution of these modifications are indicative of distinct biological pathway activations and give insights into clinical parameters of relapse free survival and outcomes when applied to breast cancer patients (Chen et al. 2016).

Since HMGA proteins bind nucleosome core particles (Reeves & Nissen 1993, Li et al. 2006) and HMGA1 expression inversely correlates with the overall survival in breast cancer (Huang et al. 2015), our study aims to evaluate possible epigenetic mechanisms by which this oncoprotein could promote tumor evolution in triple negative breast cancer, a particularly aggressive and heterogeneous intrinsic molecular subtype of breast cancer with poor prognosis.

HMGA proteins have already been linked to epigenetic alterations, i.e. anomalous histone acetylation pattern in cancer cells. In detail, HMGA2 has been linked to histone acetyltransferase expression regulation in pancreatic ductal adenocarcinoma and is responsible for increased acetylation levels of H3K9 and H3K27 in fibrotic region. Downregulation of HMGA2 expression results in reduction of acetylated histone H3 and sensitizes cells to gemcitabine (Dangi-Garimella et al. 2013). Here, our main finding is that HMGA1-driven transcriptional activation of genes involved in epithelial to mesenchymal transition could happen, at least in part, by the modulation of RSK2/CBP histone modifiers activities.

Among HMGA1 signature taken into consideration are several factors implied in cell division dynamics: the mitotic kinase Aurora Kinase B (AURKB), which is a chromosomal passenger protein; the Kinesin family member 23 (KIF23) and 4A (KIF4A), which are microtubule-dependent motor proteins; and the centromere protein-F (CENPF), which is a cell-cycle regulated protein that is associated with kinetochores. All these proteins are involved in large nuclear rearrangements occurring during mitosis, such as chromosome condensation and segregation, cytokinesis, and spindle dynamics and all have been associated with cancer onset and progression (Portella et al. 2011; Yu & Feng 2010; O'Brien et al. 2007). Moreover, the nucleation of radial microtubules belonging to centrosome has been connected to leading edge protrusion, microenvironment crosstalk and organelle polarization, which are events involved in directional migration and invasion (Ogden et al. 2013). We found these genes considerably down-regulated in HMGA1-depleted MDA-MB-231 cells and their own silencing

significantly impairs cell motility, as detected by wound healing assays (Pegoraro et al. 2013). Therefore, we considered epigenetic mechanisms by which HMGA1 could drive transcriptional regulation affecting epithelial to mesenchymal transition.

In particular, we provided experimental evidences suggesting that HMGA1, perhaps by binding to core histones, promotes histone H3S10 phosphorylation and H2BK5 acetylation. This phenomenon happens leaving unchanged the expression and the activation of the Ras-signalling pathway that is an upstream modulator of both RSK2 and CBP activities.

Contrary to literature data concerning HMGA1 overexpression in luminal MCF7, HMGA1 knocking-down is ineffective towards ERK expression and activation (preliminary data) in MDA-MB-231 TNBC cell model. Furthermore, upon HMGA1 silencing RSK2 levels are slightly downregulated but overall active RSK2 (S227ph) and CBP levels remain the same. If HMGA1 action is carried out at downstream chromatin global level by (i) competing with histone H1, (ii) influencing entry-exit angle of nucleosomal DNA, (iii) by steric changes of core particles or (iv) N-terminal histone tails, or by (v) simply binding to chromatin and recruiting RSK2 and/or CBP, this need to be further elucidated. However, our co-IP data suggest that, at least HMGA1 and RSK2 co-localize within 600 bp DNA fragments.

Moreover, we propose a molecular subtype-dependent mechanism of HMGA1 regulation of RSK2 and CBP activities: in detail, we speculate on HMGA1 regulation of RAS/RAF/MEK/ERK dependent histone writers' activities in triple negative breast cancers with mesenchymal-like phenotype, while we hypothesise a different mechanism in other triple negative subtypes and hormone receptor positive cell lines, whose transformed phenotype relies on the activation of other specific pathways, i.e. PI3K/AKT pathway activation. In fact, very high percentage of TNBC show biologically important somatic alterations of RAS/RAF/MEK/ERK and PI3K/AKT pathways which support targeted therapeutical intervention in clinical trials (Craig et al. 2013). We point out that these molecular alterations could affect modulators activities which in turn influence modifiers activities.

In our analysis, levels of histone H2B acetylation seem to inversely correlate with HMGA1 expression among cell lines used (**figure 3.9 A**). This supports literature data in which authors outlined reduced H2BK5ac on specific genes in claudin-low breast cancer lines compared to normal human mammary epithelial cells (HMEC; Abell et al. 2011). However, our results suggest that HMGA1 expression in MDA-MB-231 could drive HAT specificity towards regulatory regions of genes involved in conferring cells metastatic capabilities. In effect, we erroneously tend to interpret triple negative breast cancers as a phenotype acquired by a

progressive loss of hormone receptor expression from an ancestral benign transformation through all hormone receptor positive intermediates but this is not actually the case.

In this perspective, functions of HATs in cancer onset and progression are still under debate: many studies suggest their role as tumor suppressors, whereas others underlie their tumor promoter nature. In particular, heterozygous germline mutations in CBP gene lead to Rubinstein-Taybi syndrome (RTS) which can result in childhood malignancies of neural crest origin. Moreover, in haematological tumors chromosome translocations frequently produce fused proteins involving N-terminus domain of CBP while mutations on its gene are less commonly found in solid cancers (Iyer et al. 2004). Although these evidences suggest a tumor suppressive features of CBP, melanocytes re-activation of senescence checkpoint upon CBP/p300 inactivation underline their oncogenic driver potential. More deeply, p300 depletion results in cyclin E transcriptional repression by downregulation of histone acetylation on its promoter (Bandyopadhyay et al. 2002). Evidently, CBP/p300 are versatile acetyltransferases with context dependent functions; large multiprotein complexes which they belong to, address for their specificity. In fact, CBP/p300 mediate tumor suppressor transactivation function of their partner p53, FOXO3a, BRCA1 and, on the other hand, as c-Myc, c-Myb and AR coactivators they can promote cancer progression (Wang et al. 2013). In this light, we hypothesise that HMGA1 overexpression can lead to aberrant CBP histone acetyltransferase activity turning out in transcriptional activation of a specific set of onco-promoting factors.

Although both CBP/p300 HATs are able to modulate H2BKac, we determined CBP as the one responsible for HMGA1 signature regulation by small interfering RNA assay and these results underline different specificity gained by different HATs. HMGA1 is indeed a well-known molecular partner of CBP during IFN β gene expression upon viral infection. In particular, HMGA1-driven enhanceosome assembly gives rise to activating surface which in turn recruits CBP/p300 acetyltransferases through multiple protein-protein interactions and this result in synergic activation of transcription (Merika et al. 1998). Moreover, CBP acetylates HMGA1 causing a destabilization and disassembly of IFN β gene enhanceosome (Munshi et al. 1998). HAT inhibition has been pursued as therapeutical tool in cancer fight: bi-substrate inhibitors (mimicking the two HAT substrates, the acetyl CoA and the lysine containing peptide connected by a linker), natural compounds (anacardic acid, curcumin, garcinol), small molecules inhibitors (C646) and protein-protein interaction inhibitors are currently investigated in different tumor types despite their limitation of instability, low cell permeability, selectivity, and low potency (Wapenaar & Dekker 2016). *In vivo*, CBP/p300

inhibitor (L002) suppress tumor growth and recurrence after treatment in MDA-MB-468 xenograft mice models (Yang et al. 2013).

In addition, our results underline a cooperation between RSK2 and CBP in histone modifications regulation. In particular, levels of H3S10ph and H2BK5ac are reciprocally modulated by HAT and kinase activities and both dropped down upon HMGA1 silencing. Moreover, inhibition of CBP or RSK2 mirrors the effects obtained by HMGA1 silencing, i.e. the regulation of a common set of genes and gross morphological cellular changes.

Similarly, loss of HMGN1, a close relative of HMGA proteins, decrease steady state levels of phosphoacetylated histone H3 (H3S10phK14ac) by approximately 35% (Lim JH et al., 2005). Mirroring H2B acetylation trends upon HMGA1 silencing, acetylations of H3K14 and H4K8 have been found both down-regulated in MDA-MB-231 cells treated with siA1_3 (preliminary data and data not shown, respectively). These residues could even be modified by CBP/p300 but also by other HATs and the entity of acetylation reduction appears moderate compared to the one observed in histone H2B.

As for CBP, also RSK2 has emerged as therapeutical target in particular for TNBC metastasis and the use of RSK2 inhibitor has shown reduction in proliferation, survival in a non-adherent environment, and migration (Ludwik et al. 2016). In particular, inhibition of RSK2 suppresses tumor initiating cells growth and promotes cell death (Stratford et al. 2012). Same authors displayed that 85% of TNBC patients exhibit activation of RSK2 (S227ph) in a panel of high grade breast cancer samples. Moreover, RSK2 inhibitor treatment could overcome MEK/ERK inhibitors side effects in clinical trials: in fact, efficiency of inhibition is often lowered by a feedback loop activation of PI3K/AKT pathway.

Translational implications of our work consist on shedding light on downstream molecular events to kinase and HATs activation which could highlight novel therapeutical strategies overriding adverse unspecific actions of upstream proteins inhibitors. In particular, we suggest that in TNBC mesenchymal-like subtypes which hold high levels of HMGA1 expression, iperactivation of Ras pathway result in increased activities of RSK2 and CBP whose chromatin target specificity is driven by HMGA1. In a model created in our laboratory (Pegoraro et al. 2013), the transient silencing of HMGA1 reverts the mesenchymal phenotype to a well-organized, flattened morphology through gene expression alterations of proteins critical for metastatic migratory abilities. Unravelling the underpinning epigenetic molecular events, we suggest that by knocking down HMGA1 expression, these particular genomic regions are closed off to activating epigenetic modifiers. Our data support a direct chromatin-linked role of HMGA1 in the modulation of the epigenetic status of cancer cells: both the

binding of HMGA1 to core histones and the fact that the Ras pathway and its downstream effectors are still kept activated upon HMGA1 silencing, confirm the downstream role played by HMGA1 in ruling accessibility to writer's targets at chromatin level.

Comprehensively these data underscore how different transduction signal pathways could impinge on histone PTMs among triple negative breast cancer subtypes and other hormone receptor positive breast cancer subtypes. Targeted therapies should account for these differences. Architectural factors such as HMGA1 could play essential roles in specific signal transduction pathways and could therefore constitute a good molecular co-target in combined therapeutical approaches.

CONCLUSIONS AND FUTURE PERSPECTIVES

In conclusion, we propose an unconventional method by which HMGA1 could orchestrate aberrant gene expression in TNBC. In particular, we suggest that HMGA1 could play its action downstream RSK2 and CBP activation driving their specificity towards malignancy-promoting oncogenes in which they modulate both H3S10ph and H2BK5ac levels.

Once uncovered functional meaning of these PTMs profile alterations, we focus on alteration of these modifications upon HMGA silencing at specific genes loci. Preliminary data obtained with RSK2 inhibitor (BI-1870) in MDA-MB-231, suggested a reduction of H3S10ph in intragenic region (last exon) of AURKB gene. A slight modulation was observed also in a 5' distal region approximately at -800 bp, a region known to be bound by Forkhead Box M1 transcription factor (FoxM1), which regulate AURKB expression (Wang et al. 2005). However, the positive regulatory region of AURKB promoter phosphorylation levels appeared unchanged upon RSK2 inhibition. Thus, we are currently performing chromatin immunoprecipitation (ChIP) analyses on MDA-MB-231 cell line upon HMGA1 silencing to confirm its involvements in RSK2 (and H3S10ph) and CBP (and H2BK5ac) dynamics regulation; we aim to analyse these genomic areas even through re-ChIP analysis (H3S10ph/H2BK5ac) in order to assess HMGA1 role in both kinase and HAT activities regulation.

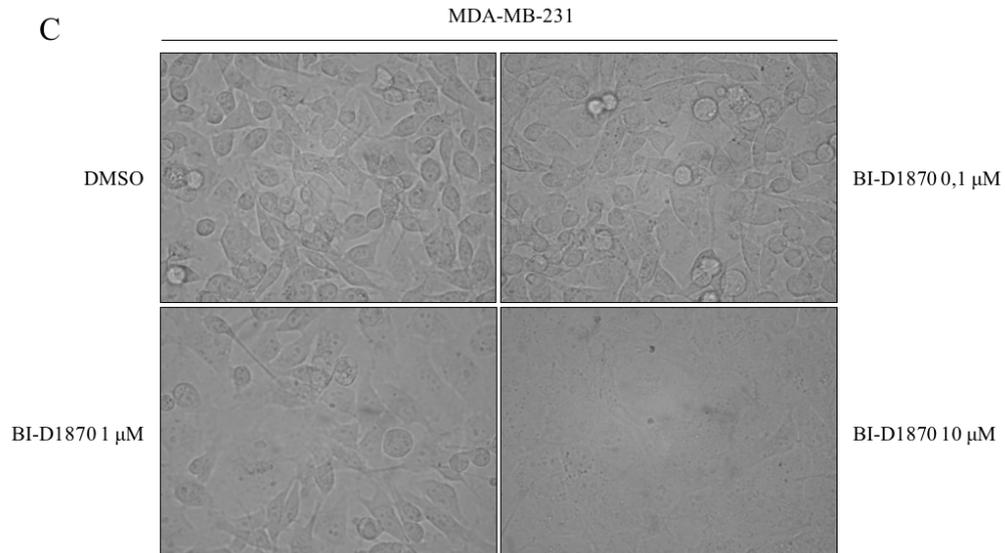
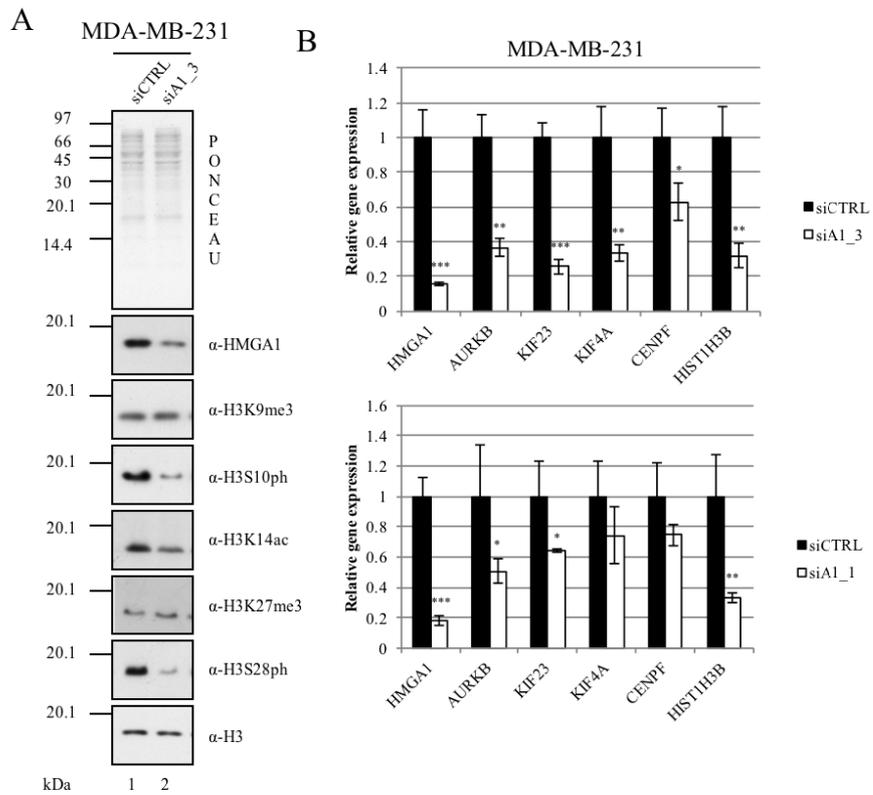
At the same time, we would like to confirm epigenetic alterations promoted by HMGA1 knocking down by mass spectrometry tools; we aim to analyse histone H2B and others core histones by propionic anhydride derivatization of primary amines of lysines and digest with trypsin that produce an Arg-C like digestion profile with higher efficiency and specificity. The propionylation contributes to a parallel increase of the hydrophobicity of peptides produced upon digestion and thereby improve their retention and resolution in RP-HPLC systems (Molden et al. 2015). Moreover, we are aware that our results rely on the use of single antibodies. We will validate these data by other antibodies.

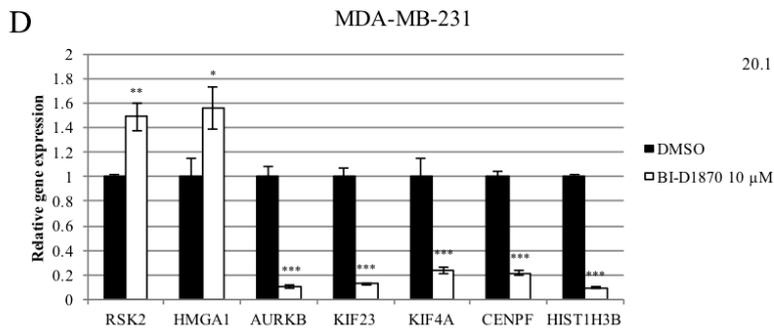
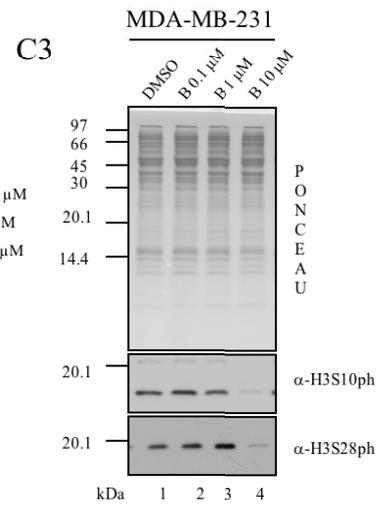
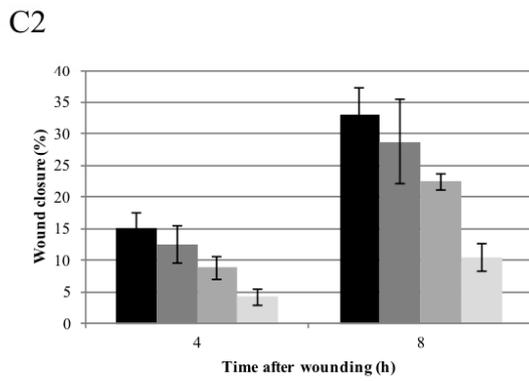
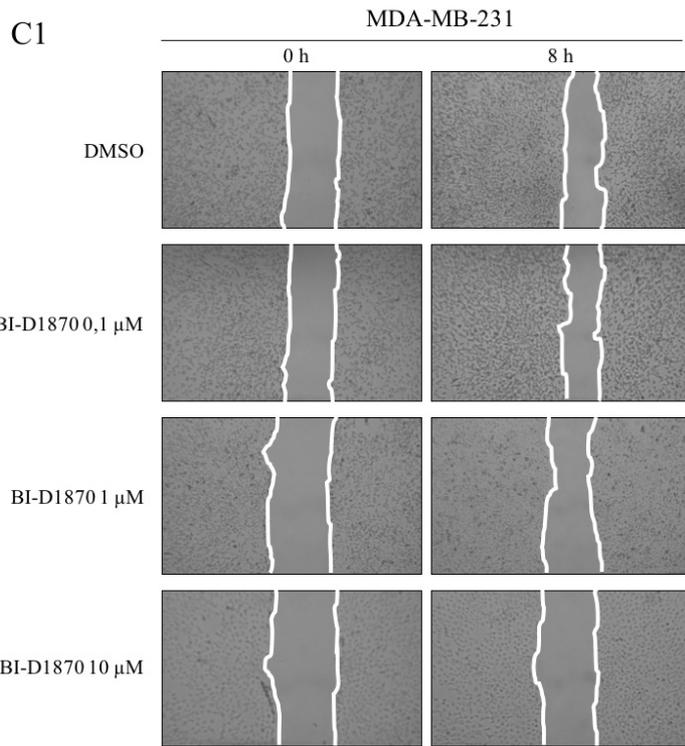
Finally, several questions arose from our work:

- a) HMGA1 is expressed in all TNBC sub-types?
- b) Is the epigenetic role of HMGA1 exclusively linked to the RAS pathway dysregulation?
- c) Have these epigenetic markers (H3S10ph and H2BK5ac) some prognostic and/or predictive value in TNBC?

- d) Could HMGA1 be used as robust biomarker predictor of epigenetic-modifiers targeted therapies responsiveness in TNBC?

Preliminary data





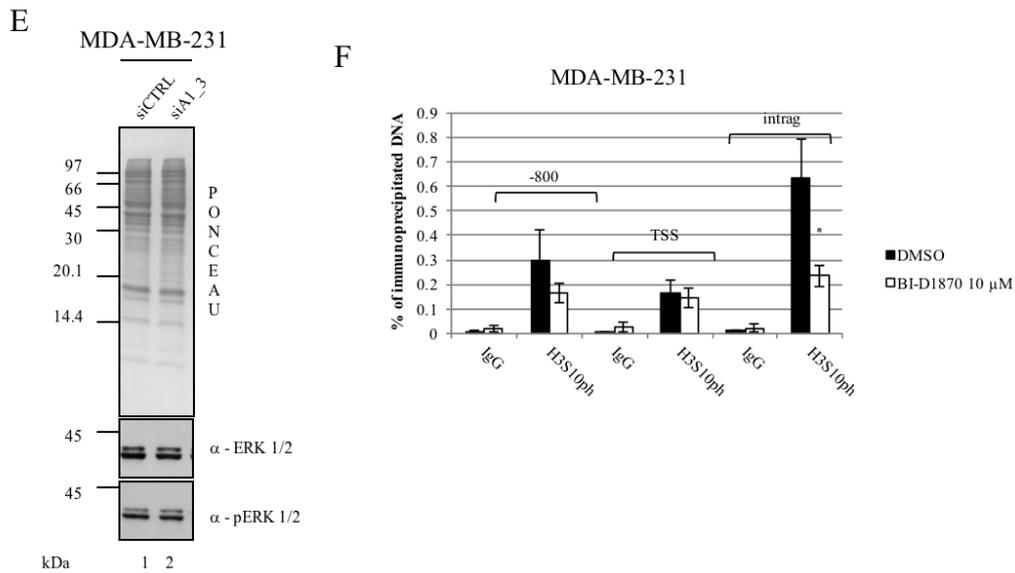


Figure P1 Preliminary data. *A)* Western blot analyses of MDA MB 231 cell lysate collected at 72 hours upon HMGA1 silencing with siA1_3 or controls. Panels reported are representative of biological triplicate and display, in order, the ponceau-stained membrane (illustrative of sample normalization) and western blot analyses of HMGA1, H3S10ph, H3K14ac, H3K27me3, H3S28ph and total H3. Molecular references are reported in kDa on the left of each panel. *B)* Relative gene expression values were obtained by RT-qPCR of MDA-MB-231 cells treated for 72 hours with siA1_3 (or siA1_1 below) and controls. GAPDH was used as internal reference gene. Mean and standard deviations (\pm SD) are reported and values are shown with respect to control sample. Experiment was done in biological triplicate and significance assigned by student t test; * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. *C)* Optical images of representative cells treated with 0.1 - 1 - 10 μ M of BI-D1870 (RSK2 inhibitor) or controls (equal amount of DMSO) for 24 hours. *C1)* Representative optimal images of wound healing assay of MDA-MB-231 treated with 0.1 - 1 - 10 μ M of BI-D1870 or controls (equal amount of DMSO) for 24 hours and results are analysed at 4 and 8 hours upon scratch. Experiment was performed in technical duplicate and biological triplicate. *C2)* Histograms of wound healing data representing mean and standard deviations (\pm SD) of the percentage of wound closure relative to time point 0. *C3)* Western blot analyses of MDA MB 231 cell lysate collected at 24 hours upon treatment with 0.1 - 1 - 10 μ M of BI-D1870 or controls. Panels reported are representative of biological triplicate and display, in order, the ponceau-stained membrane (illustrative of sample normalization) and western blot analyses of H3S10ph and H3S28ph. Molecular references are reported in kDa on the left of each panel. *D)* Relative gene expression values were obtained by RT-qPCR of MDA-MB-231 cells treated for 24 hours with 10 μ M of BI-D1870 (RSK2 inhibitor) or controls (equal amount of DMSO). CYP33 was used as internal reference gene. Mean and standard deviations (\pm SD) are reported and values are shown with respect to control sample. Experiment was done in biological triplicate and significance assigned by student t test; * $p < 0.05$, ** $p < 0.01$, *** $p < 0.001$. *E)* WB analysis of MDA MB 231 upon HMGA1 silencing Panels reported are representative of biological triplicate and illustrate ponceau-stained membrane and bands corresponding to ERK and pERK antibody recognition (refer to *A* for HMGA1 level). Molecular references are reported in kDa on the left of each panel. *F)* ChIP-qPCR experiment performed in cells treated for 24 hours with 10 μ M of BI-D1870 or controls (equal amount of DMSO). Antibody α -H3S10ph was used for the immunoprecipitation and non-specific IgG as negative control. Genomic DNA was amplified with primers targeting 5' distal region (-800 bp), transcription start site (TSS) and 3' distal region (intragenic) of AURKB gene and results are reported as percent of input DNA. Mean and standard deviations (\pm SD) are reported. Experiment was done in biological triplicate and significance assigned by student t test; * $p < 0.05$.

BIBLIOGRAPHY

- Abell, A.N. et al., 2011. MAP3K4/CBP-regulated H2B acetylation controls epithelial-mesenchymal transition in trophoblast stem cells. *Cell Stem Cell*, 8(5), pp.525–537.
- De Abreu, F.B. et al., 2014. Personalized therapy for breast cancer. *Clinical Genetics*, 86(1), pp.62–67.
- Ahn, S.G. et al., 2016. Molecular classification of triple-negative breast cancer. *Journal of Breast Cancer*, 19(3), pp.223–230.
- Akaboshi, S. et al., 2009. HMGA1 Is Induced by Wnt/ β -Catenin Pathway and Maintains Cell Proliferation in Gastric Cancer. *The American Journal of Pathology*, 175(4), pp.1675–1685.
- Ananthula, S. et al., 2016. Geminin overexpression-dependent recruitment and crosstalk with mesenchymal stem cells enhance aggressiveness in triplenegative breast cancers. *Oncotarget*.
- Arnone, P. et al., 2010. The TNM classification of breast cancer: Need for change. *Updates in Surgery*, 62(2), pp.75–81.
- Arul, N. & Cho, Y.-Y., 2013. A Rising Cancer Prevention Target of RSK2 in Human Skin Cancer. *Frontiers in oncology*, 3(August), p.201.
- Baldassarre, G. et al., 2005. HMGA1 protein expression sensitizes cells to cisplatin-induced cell death. *Oncogene*, 24(45), pp.6809–19.
- Baldassarre, G. et al., 2003. Negative regulation of BRCA1 gene expression by HMGA1 proteins accounts for the reduced BRCA1 protein levels in sporadic breast carcinoma. *Molecular and cellular biology*, 23(7), pp.2225–38.
- Bandyopadhyay, D. et al., 2002. Down-regulation of p300/CBP histone acetyltransferase activates a senescence checkpoint in human melanocytes. *Cancer Research*, 62(21), pp.6231–6239.
- Bannister, A.J. & Kouzarides, T., 2011. Regulation of chromatin by histone modifications. *Cell research*, 21(3), pp.381–395.
- Belton, A. et al., 2012. HMGA1 induces intestinal polyposis in transgenic mice and drives tumor progression and stem cell properties in colon cancer cells. *PLoS ONE*, 7(1).
- Benecke, A.G. & Eilebrecht, S., 2015. RNA-mediated regulation of HMGA1 function. *Biomolecules*, 5(2), pp.943–957.
- Berlingieri, M.T. et al., 2002. Thyroid cell transformation requires the expression of the HMGA1 proteins. *Oncogene*, 21(19), pp.2971–2980.
- Bowers, E.M. et al., 2010. Virtual ligand screening of the p300/CBP histone acetyltransferase: Identification of a selective small molecule inhibitor. *Chemistry and Biology*, 17(5), pp.471–482.
- Bussemakers, M.J.G. et al., 1991. Identification of High Mobility Group Protein I (Y) as Potential Progression Marker for Prostate Cancer by Differential Hybridization Analysis . , (8), pp.606–611.
- Cai, F.-F. et al., 2011. Epigenetic therapy for breast cancer. *International journal of molecular sciences*, 12, pp.4465–4487.
- Carels, N. et al., 2016. Toward precision medicine of breast cancer. *Theoretical Biology and Medical Modelling*, 13, p.7.
- Catez, F. et al., 2002. Competition between histone H1 and HMGN proteins for chromatin binding sites. *EMBO*

- Reports*, 3(8), pp.760–766.
- De Cesare, D. et al., 1998. Rsk-2 activity is necessary for epidermal growth factor-induced phosphorylation of CREB protein and transcription of c-fos gene. *Proceedings of the National Academy of Sciences of the United States of America*, 95(21), pp.12202–7.
- Chan, H.M. & La Thangue, N.B., 2001. p300/CBP proteins: HATs for transcriptional bridges and scaffolds. *Journal of cell science*, 114(Pt 13), pp.2363–2373.
- Chau, K. yin et al., 1995. The gene for the human architectural transcription factor HMGI-C consists of five exons each coding for a distinct functional element. *Nucleic Acids Research*, 23(21), pp.4262–4266.
- Chen, K.J. et al., 2014. Reexpression of let-7g microRNA inhibits the proliferation and migration via K-Ras/HMGA2/snail axis in hepatocellular carcinoma. *BioMed Research International*, 2014.
- Chen, X. et al., 2016. A novel subtype classification and risk of breast cancer by histone modification profiling. *Breast Cancer Research and Treatment*, 157(2), pp.267–279.
- Cheng, X. & Blumenthal, R.M., 2008. Mammalian DNA methyltransferases: a structural perspective. *Structure*, 16(3), pp.341–350.
- Chiappetta, G. et al., 1996. High level expression of the HMGI (Y) gene during embryonic development. *Oncogene*, 13(11), pp.2439–46.
- Chiappetta, G. et al., 2004. HMGA1 protein overexpression in human breast carcinomas: Correlation with ErbB2 expression. *Clinical Cancer Research*, 10(22), pp.7637–7644.
- Choe, M.K. et al., 2012. Functional elements demarcated by histone modifications in breast cancer cells. *Biochemical and biophysical research communications*, 418(3), pp.475–82.
- Cleynen, I. & Van de Ven, W.J.M., 2008. The HMGA proteins: A myriad of functions (Review). *International Journal of Oncology*, 32(2), pp.289–305.
- Connolly, R. & Stearns, V., 2012. Epigenetics as a therapeutic target in breast cancer. *Journal of Mammary Gland Biology and Neoplasia*, 17(3–4), pp.191–204.
- Craig, D.W. et al., 2013. Genome and Transcriptome Sequencing in Prospective Metastatic Triple-Negative Breast Cancer Uncovers Therapeutic Vulnerabilities. *Molecular Cancer Therapeutics*, 12(1), pp.104–116.
- D’Angelo, D. et al., 2012. Altered microRNA expression profile in human pituitary GH adenomas: Down-regulation of miRNA targeting HMGA1, HMGA2, and E2F1. *Journal of Clinical Endocrinology and Metabolism*, 97(7).
- Dai, X. et al., 2016. Cancer hallmarks, biomarkers and breast cancer molecular subtypes. *Journal of Cancer*, 7(10), pp.1281–1294.
- DAMASKOS, C. et al., 2017. Histone Deacetylase Inhibitors: An Attractive Therapeutic Strategy Against Breast Cancer. *Anticancer Research*, 37(1), pp.35–46.
- Dangi-Garimella, S. et al., 2013. Three-Dimensional Collagen I Promotes Gemcitabine Resistance In Vitro in Pancreatic Cancer Cells through HMGA2-Dependent Histone Acetyltransferase Expression. *PLoS ONE*, 8(5).
- Davis, N.M. et al., 2014. Deregulation of the EGFR/PI3K/PTEN/Akt/mTORC1 pathway in breast cancer: possibilities for therapeutic intervention. *Oncotarget*, 5(13), pp.4603–4650.
- Desantis, C.E. et al., 2014. Cancer Treatment and Survivorship Statistics , 2014.
- DeSantis, C.E. et al., 2015. International Variation in Female Breast Cancer Incidence and Mortality Rates. *Cancer Epidemiology Biomarkers & Prevention*, 24(10), pp.1495–1506.

- Dworkin, A.M., Huang, T.H.M. & Toland, A.E., 2009. Epigenetic alterations in the breast: Implications for breast cancer detection, prognosis and treatment. *Seminars in Cancer Biology*, 19(3), pp.165–171.
- Eidelman, S. et al., 1989. Expression of the cell-cell adhesion glycoprotein cell-CAM 120/80 in normal human tissues and tumors. *American Journal of Pathology*, 135(1), pp.101–110.
- Ellsworth, R.E. et al., 2016. Molecular heterogeneity in breast cancer: State of the science and implications for patient care. *Seminars in Cell & Developmental Biology*.
- Elsheikh, S.E. et al., 2009. Global histone modifications in breast cancer correlate with tumor phenotypes, prognostic factors, and patient outcome. *Cancer Research*, 69(9), pp.3802–3809.
- Elston, C. w. & Ellis, I. o., 1991. Pathological prognostic factors in breast cancer. I. The value of histological grade in breast cancer: experience from a large study with long-term follow-up. *Histopathology*, 19(5), pp.403–410.
- Favata, M.F., 1998. Identification of a Novel Inhibitor of Mitogen-activated Protein Kinase Kinase. *Journal of Biological Chemistry*, 273(29), pp.18623–18632.
- Fedele, M. et al., 2005. Transgenic mice overexpressing the wild-type form of the HMGA1 gene develop mixed growth hormone/prolactin cell pituitary adenomas and natural killer cell lymphomas. *Oncogene*, 24(21), pp.3427–35.
- Fedele, M. & Fusco, A., 2010. HMGA and Cancer. *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms*, 1799(1–2), pp.48–54.
- Federico, A. et al., 2014. Hmgal / Hmga2 double knock-out mice display a ““ superpygmy ”” phenotype. , pp.372–378.
- Feinberg, A.P., Koldobskiy, M.A. & Göndör, A., 2016. Epigenetic modulators, modifiers and mediators in cancer aetiology and progression. *Nature reviews. Genetics*, 17(5), pp.284–99.
- Ferguson, A.T. et al., 1995. Demethylation of the estrogen receptor gene in estrogen receptor-negative breast cancer cells can reactivate estrogen receptor gene expression. *Cancer research*, 55(11), pp.2279–83.
- Ferlay, J. et al., 2015. Cancer incidence and mortality worldwide: sources, methods and major patterns in GLOBOCAN 2012. *Int J Cancer*, 136(5), pp.E359-386.
- Foulkes, W.D., Smith, I.E. & Reis-Filho, J.S., 2010. Triple-negative breast cancer. *The New England journal of medicine*, 363(20), pp.1938–1948.
- Fusco, A. & Fedele, M., 2007. Roles of HMGA proteins in cancer. *Nature reviews. Cancer*, 7(12), pp.899–910.
- Giancotti, V. et al., 1985. Changes in Nuclear Proteins on Transformation of Rat Epithelial Thyroid Cells by a Murine Sarcoma Retrovirus. , 45(December), pp.6051–6057.
- GOODWIN, G.H. et al., 1985. Fractionation by high-performance liquid chromatography of the low-molecular-mass high-mobility- group (HMG) chromosomal proteins present in proliferating rat cells and an investigation of the HMG proteins present in virus transformed cells. *European Journal of Biochemistry*, 149(1), pp.47–51.
- Gröger, C.J. et al., 2012. Meta-Analysis of Gene Expression Signatures Defining the Epithelial to Mesenchymal Transition during Cancer Progression. *PLoS ONE*, 7(12).
- Guo, L. et al., 2013. Stat3-coordinated Lin-28-let-7-HMGA2 and miR-200-ZEB1 circuits initiate and maintain oncostatin M-driven epithelial-mesenchymal transition. *Oncogene*, 32(October 2012), pp.5272–5282.
- Hennessy, B.T. et al., 2009. Characterization of a naturally occurring breast cancer subset enriched in epithelial-to-mesenchymal transition and stem cell characteristics. *Cancer Research*, 69(10), pp.4116–4124.

- Henry, R.A. et al., 2015. Changing the selectivity of p300 by acetyl-coa modulation of histone acetylation. *ACS Chemical Biology*, 10(1), pp.146–156.
- Henry, R.A., Kuo, Y.M. & Andrews, A.J., 2013. Differences in specificity and selectivity between CBP and p300 acetylation of histone H3 and H3/H4. *Biochemistry*, 52(34), pp.5746–5759.
- Herschkowitz, J. et al., 2007. Identification of conserved gene expression features between murine mammary carcinoma models and human breast tumors. *Genome Biol*, 8(5), p.R76.
- Hervouet, E. et al., 2013. Epigenetic regulation of estrogen signaling in breast cancer. *Epigenetics*, 8(3), pp.237–45.
- Hill, D.A. & Reeves, R., 1997. Competition between HMG-I(Y), HMG-1 and histone H1 on four-way junction DNA. *Nucleic Acids Research*, 25(17), pp.3523–3531.
- Himes, S.R. et al., 2000. The role of high-mobility group I(Y) proteins in expression of IL-2 and T cell proliferation. *Journal of immunology (Baltimore, Md. : 1950)*, 164(6), pp.3157–3168.
- Hock, R. et al., 2007. HMG chromosomal proteins in development and disease. *Trends in Cell Biology*, 17(2), pp.72–79.
- Hollestelle, A. et al., 2007. Phosphatidylinositol-3-OH kinase or RAS pathway mutations in human breast cancer cell lines. *Molecular cancer research : MCR*, 5(2), pp.195–201.
- Holliday, D.L. & Speirs, V., 2011. Choosing the right cell line for breast cancer research. *Breast cancer research : BCR*, 13, p.215.
- Holm, K. et al., 2010. Molecular subtypes of breast cancer are associated with characteristic DNA methylation patterns. *Breast cancer research : BCR*, 12(3), p.R36.
- Huang, R. et al., 2015. Overexpression of HMGA1 correlates with the malignant status and prognosis of breast cancer. *Molecular and Cellular Biochemistry*, 404(1–2), pp.251–257.
- Huang, Y. et al., 2011. Epigenetics in breast cancer: what's new? *Breast Cancer Research*, p.11.
- Huleihel, L. et al., 2014. Let-7d microRNA affects mesenchymal phenotypic properties of lung fibroblasts. *American journal of physiology. Lung cellular and molecular physiology*, 306(6), pp.L534-42.
- Huso, T.H. & Resar, L.M.S., 2014. The high mobility group A1 molecular switch: turning on cancer - can we turn it off? *Expert opinion on therapeutic targets*, 18(5), pp.541–53.
- Huth, J.R. et al., 1997. The solution structure of an HMG-I(Y)-DNA complex defines a new architectural minor groove binding motif. *Nat Struct Biol*, 4(8), pp.657–665.
- Iyer, N.G. et al., 2004. p300/CBP and cancer. *Oncogene*, 23(24), pp.4225–4231.
- Jensen, C.J. et al., 1999. 90-kDa ribosomal S6 kinase is phosphorylated and activated by 3- phosphoinositide-dependent protein kinase-1. *Journal of Biological Chemistry*, 274(38), pp.27168–27176.
- John, S. et al., 1995. Regulation of cell-type-specific interleukin-2 receptor alpha-chain gene expression: potential role of physical interactions between Elf-1, HMG-I(Y), and NF-kappa B family proteins. *Molecular and Cellular Biology*, 15(3), pp.1786–1796.
- Johnson, K.R., Lehn, D. a & Reeves, R., 1989. Alternative processing of mRNAs encoding mammalian chromosomal high-mobility-group proteins HMG-I and HMG-Y. *Molecular and cellular biology*, 9(5), pp.2114–2123.
- Jones, P.A., Issa, J.-P.J. & Baylin, S., 2016. Targeting the cancer epigenome for therapy. *Nature reviews. Genetics*, 17(10), pp.630–41.
- Jovanovic, J. et al., 2010. The epigenetics of breast cancer. *Molecular Oncology*, 4(3), pp.242–254.

- Kaddar, T. et al., 2009. Two new miR-16 targets: caprin-1 and HMGA1, proteins implicated in cell proliferation. *Biology of the cell / under the auspices of the European Cell Biology Organization*, 101(9), pp.511–524.
- Kalluri, R. & Weinberg, R.A., 2009. The basics of epithelial-mesenchymal transition. *Journal of Clinical Investigation*, 119(6), pp.1420–1428.
- Kamiya, T. et al., 2016. Cross Talk Mechanism among EMT, ROS, and Histone Acetylation in Phorbol Ester-Treated Human Breast Cancer MCF-7 Cells. *Oxidative Medicine and Cellular Longevity*, 2016.
- Karlić, R. et al., 2010. Histone modification levels are predictive for gene expression. *Proceedings of the National Academy of Sciences of the United States of America*, 107(7), pp.2926–2931.
- Kreike, B. et al., 2007. Gene expression profiling and histopathological characterization of triple-negative/basal-like breast carcinomas. *Breast Cancer Research*, 9(5), p.R65.
- Kumar, P. & Aggarwal, R., 2016. An overview of triple-negative breast cancer. *Archives of Gynecology and Obstetrics*, 293(2), pp.247–269.
- Lamouille, S., Xu, J. & Derynck, R., 2014. Molecular mechanisms of epithelial-mesenchymal transition. *Nature reviews. Molecular cell biology*, 15(3), pp.178–96.
- Lehmann, B.D.B. et al., 2011. Identification of human triple-negative breast cancer subtypes and preclinical models for selection of targeted therapies. *Journal of Clinical Investigation*, 121(7), pp.2750–2767.
- Li, J.P. et al., 2015. The pan-inhibitor of aurora kinases danusertib induces apoptosis and autophagy and suppresses epithelial-to-mesenchymal transition in human breast cancer cells. *Drug Design, Development and Therapy*, 9, pp.1027–1062.
- Li, O. et al., 2006. High-level expression of DNA architectural factor HMGA2 and its association with nucleosomes in human embryonic stem cells. *Genesis*, 44(11), pp.523–529.
- Liau, S.-S. et al., 2007. Overexpression of HMGA1 promotes anoikis resistance and constitutive Akt activation in pancreatic adenocarcinoma cells. *British journal of cancer*, 96(6), pp.993–1000.
- Liu, K. et al., 2015. Let-7a inhibits growth and migration of breast cancer cells by targeting HMGA1. *Int J Oncol*, 46(6), pp.2526–2534.
- Lombaerts, M. et al., 2006. E-cadherin transcriptional downregulation by promoter methylation but not mutation is related to epithelial-to-mesenchymal transition in breast cancer cell lines. *British journal of cancer*, 94(5), pp.661–71.
- Ludwik, K. et al., 2016. Development of a RSK Inhibitor as a Novel Therapy for Triple Negative Breast Cancer. *Molecular Cancer Therapeutics*, 15(November), pp.2598–2608.
- Mack, G.S. & Marshall, A., 2010. Lost in migration. *Nature Publishing Group*, 28(3), pp.214–229.
- Makki, J., 2015. Diversity of breast carcinoma: Histological subtypes and clinical relevance. *Clinical Medicine Insights: Pathology*, 8(1), pp.23–31.
- Manfioletti, G. et al., 1991. cDNA cloning of the HMGI-C phosphoprotein, a nuclear protein associated with neoplastic and undifferentiated phenotypes. *Nucleic acids research*, 19(24), pp.6793–7.
- Mani, S.A. et al., 2008. The Epithelial-Mesenchymal Transition Generates Cells with Properties of Stem Cells. *Cell*, 133(4), pp.704–715.
- De Martino, M. et al., 2016. HMGA1-pseudogenes and cancer. *Oncotarget*, 7(19). Maurizio, E. et al., 2016. Translating Proteomic Into Functional Data: An High Mobility Group A1 (HMGA1) Proteomic Signature Has Prognostic Value in Breast Cancer. *Molecular & cellular proteomics : MCP*, 15(1), pp.109–23.
- Merienne, K. et al., 2001. Mitogen-Regulated RSK2-CBP Interaction Controls Their Kinase and Acetylase

- Activities. *Molecular and Cellular Biology*, 21(20), pp.7089–7096.
- Merika, M. et al., 1998. Recruitment of CBP/p300 by the IFN β enhanceosome is required for synergistic activation of transcription. *Molecular Cell*, 1(2), pp.277–287.
- Messier, T.L. et al., 2016. Histone H3 lysine 4 acetylation and methylation dynamics define breast cancer subtypes. *Oncotarget*, 7(5).
- Miller, E. et al., 2014. Current treatment of early breast cancer: adjuvant and neoadjuvant therapy. *F1000Research*, 3(0), p.198.
- Molden, R.C. et al., 2015. Multi-faceted quantitative proteomics analysis of histone H2B isoforms and their modifications. *Epigenetics & Chromatin*, 8(1), p.15.
- Morishita, A. et al., 2013. HMGA2 is a driver of tumor metastasis. *Cancer Research*, 73(14), pp.4289–4299.
- Muggerud, A.A. et al., 2010. Frequent aberrant DNA methylation of ABCB1, FOXC1, PPP2R2B and PTEN in ductal carcinoma in situ and early invasive breast cancer. *Breast cancer research : BCR*, 12(1), p.R3.
- Müller, S. et al., 2001. The double life of HMGB1 chromatin protein: Architectural factor and extracellular signal. *EMBO Journal*, 20(16), pp.4337–4340.
- Munshi, N. et al., 1998. Acetylation of HMG I(Y) by CBP Turns off IFN β Expression by Disrupting the Enhanceosome. *Molecular Cell*, 2(4), pp.457–467.
- Narita, M. et al., 2006. A Novel Role for High-Mobility Group A Proteins in Cellular Senescence and Heterochromatin Formation. *Cell*, 126(3), pp.503–514.
- Nitta, H. et al., 2016. The assessment of HER2 status in breast cancer: The past, the present, and the future. *Pathology International*, 66(6), pp.313–324.
- O'Brien, S.L. et al., 2007. CENP-F expression is associated with poor prognosis and chromosomal instability in patients with primary breast cancer. *International journal of cancer. Journal international du cancer*, 120, pp.1434–1443.
- Ogden, A., Rida, P.C.G. & Aneja, R., 2013. Heading off with the herd: How cancer cells might maneuver supernumerary centrosomes for directional migration. *Cancer and Metastasis Reviews*, 32(1–2), pp.269–287.
- Ogryzko, V. V. et al., 1996. The transcriptional coactivators p300 and CBP are histone acetyltransferases. *Cell*, 87(5), pp.953–959.
- Pallante, P. et al., 2015. High mobility group a proteins as tumor markers. *Front Med (Lausanne)*, 2, p.15.
- Palma, G. et al., 2015. Triple negative breast cancer: looking for the missing link between biology and treatments. *Oncotarget*, 6(29), pp.26560–26574.
- Pasquier, J. et al., 2015. Epithelial to Mesenchymal Transition in a Clinical Perspective. *Journal of Oncology*, 2015.
- PDQ Adult Treatment Editorial Board, 2016. Breast Cancer Treatment (PDQ®)—Patient Version. *PDQ Cancer Information Summaries*.
- Pegoraro, S. et al., 2015. A novel HMGA1-CCNE2-YAP axis regulates breast cancer aggressiveness. *Oncotarget*, 6(22), pp.19087–101.
- Pegoraro, S. et al., 2013. HMGA1 promotes metastatic processes in basal-like breast cancer regulating EMT and stemness. *Oncotarget*, 4(8), pp.1293–308.
- Pellarin, I. et al., 2016. The architectural chromatin factor high mobility group A1 enhances dna Ligase IV activity influencing dna repair. *PLoS ONE*, 11(10), pp.1–20.

- Perou, C.M. et al., 2000. Molecular portraits of human breast tumours. *Nature*, 406(6797), pp.747–752.
- Pierantoni, G.M. et al., 2003. HMGA2 locus rearrangement in a case of acute lymphoblastic leukemia. *International journal of oncology*, 23(2), pp.363–367.
- Portella, G., Passaro, C. & Chieffi, P., 2011. Aurora B: a new prognostic marker and therapeutic target in cancer. *Current medicinal chemistry*, 18(4), pp.482–96.
- Rafique, S. et al., 2015. Estrogen-induced chromatin decondensation and nuclear re-organization linked to regional epigenetic regulation in breast cancer. *Genome biology*, 16, p.145.
- Reeves, R., 2001. Molecular biology of HMGA proteins: Hubs of nuclear function. *Gene*, 277(1–2), pp.63–81.
- Reeves, R., 2010. Nuclear functions of the HMG proteins. *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms*, 1799(1–2), pp.3–14.
- Reeves, R., Edberg, D.D. & Li, Y., 2001. Architectural transcription factor HMGI(Y) promotes tumor progression and mesenchymal transition of human epithelial cells. *Molecular and cellular biology*, 21(2), pp.575–94.
- Reeves, R. & Nissen, M.S., 1993. Interaction of high mobility group-I (Y) nonhistone proteins with nucleosome core particles. *J. Biol. Chem.*, 268(28), pp.21137–21146.
- Reis-Filho, J.S. & Tutt, A.N.J., 2008. Triple negative tumours: A critical review. *Histopathology*, 52(1), pp.108–118.
- Rhie, S.K. et al., 2014. Nucleosome positioning and histone modifications define relationships between regulatory elements and nearby gene expression in breast epithelial cells. *BMC genomics*, 15, p.331.
- Saito, A. et al., 2013. An Integrated Expression Profiling Reveals Target Genes of TGF- β and TNF- α Possibly Mediated by MicroRNAs in Lung Cancer Cells. *PLoS ONE*, 8(2).
- Salmon, R., 2016. Breast Cancer Management : Should We Treat Our Patients according to the TNM or the Molecular Classification ? , 3(1), pp.1–2.
- Sassone-Corsi, P. et al., 1999. Requirement of Rsk-2 for epidermal growth factor-activated phosphorylation of histone H3. *Science (New York, N.Y.)*, 285(5429), pp.886–91.
- Scala, S. et al., 2000. Adenovirus-mediated suppression of HMGI(Y) protein synthesis as potential therapy of human malignant neoplasias. *Proc Natl Acad Sci U S A*, 97(8), pp.4256–4261.
- Sepe, R. et al., 2016. HMGA1 overexpression is associated with a particular subset of human breast carcinomas. *Journal of clinical pathology*, 69(2), pp.117–21.
- Servick, K., 2014. Breast cancer: a world of differences. *Science*, 343(6178), pp.1452–3.
- Sgarra, R. et al., 2010. HMGA molecular network: From transcriptional regulation to chromatin remodeling. *Biochimica et Biophysica Acta - Gene Regulatory Mechanisms*, 1799(1–2), pp.37–47.
- Sgarra, R. et al., 2004. Nuclear phosphoproteins HMGA and their relationship with chromatin structure and cancer. *FEBS Letters*, 574(1–3), pp.1–8.
- Shah, S.N. et al., 2012. HMGA1 Reprograms Somatic Cells into Pluripotent Stem Cells by Inducing Stem Cell Transcriptional Networks. *PLoS ONE*, 7(11).
- Shechter, D. et al., 2007. Extraction, purification and analysis of histones. *Nat Protoc*, 2(6), pp.1445–1457.
- Shi, Z. et al., 2016. Silencing of HMGA2 promotes apoptosis and inhibits migration and invasion of prostate cancer cells. *J Biosci*, 41(2), pp.229–36.
- Siitonen, S.M. et al., 1996. Reduced E-cadherin expression is associated with invasiveness and unfavorable prognosis in breast cancer. *American Journal of Clinical Pathology*, 105(4), pp.394–402.

- Soldi, M. et al., 2013. Mass spectrometry-based proteomics for the analysis of chromatin structure and dynamics. *International Journal of Molecular Sciences*, 14(3), pp.5402–5431.
- Sørli, T. et al., 2001. Gene expression patterns of breast carcinomas distinguish tumor subclasses with clinical implications. *Proceedings of the National Academy of Sciences of the United States of America*, 98(19), pp.10869–74.
- Stefansson, O. a & Esteller, M., 2013. Epigenetic modifications in breast cancer and their role in personalized medicine. *The American journal of pathology*, 183(4), pp.1052–63.
- Stratford, A.L. et al., 2012. Targeting p90 ribosomal S6 kinase eliminates tumor-initiating cells by inactivating Y-box binding protein-1 in triple-negative breast cancers. *Stem Cells*, 30(7), pp.1338–1348.
- Tan, E.J. et al., 2012. Regulation of transcription factor twist expression by the DNA architectural protein high mobility group A2 during epithelial-to-mesenchymal transition. *Journal of Biological Chemistry*, 287(10), pp.7134–7145.
- Tan, E.J. et al., 2015. The high mobility group A2 protein epigenetically silences the Cdh1 gene during epithelial-to-mesenchymal transition. *Nucleic Acids Research*, 43(1), pp.162–178.
- Tang, D. et al., 2010. High-mobility Group Box 1 [HMGB1] and Cancer. *Biochemical Biophysical Acta*, 1799, pp.1–22.
- Thiery, J.P. et al., 2009. Epithelial-Mesenchymal Transitions in Development and Disease. *Cell*, 139(5), pp.871–890.
- Thuault, S. et al., 2008. HMGA2 and Smads co-regulate SNAIL1 expression during induction of epithelial-to-mesenchymal transition. *Journal of Biological Chemistry*, 283(48), pp.33437–33446.
- Thuault, S. et al., 2006. Transforming growth factor- β employs HMGA2 to elicit epithelial-mesenchymal transition. *Journal of Cell Biology*, 174(2), pp.175–183.
- Torre, L.A. et al., 2015. Global cancer statistics, 2012. *CA: a cancer journal for clinicians*, 65(2), pp.87–108.
- Treff, N.R. et al., 2004. High-mobility group A1a protein regulates Ras/ERK signaling in MCF-7 human breast cancer cells. *Oncogene*, 23(3), pp.777–85.
- Tufegdzcic Vidakovic, A. et al., 2015. Context-Specific Effects of TGF- β /SMAD3 in Cancer Are Modulated by the Epigenome. *Cell Reports*, 13(11), pp.2480–2490.
- Vad-Nielsen, J. & Nielsen, A.L., 2015. Beyond the histone tale: HP1 alpha deregulation in breast cancer epigenetics. *Cancer Biology and Therapy*, 16(2), pp.189–200.
- Vanhaesebroeck, B. & Alessi, D.R., 2000. The PI3K-PDK1 connection: more than just a road to PKB. *The Biochemical journal*, 346 Pt 3, pp.561–76.
- Veronesi, U. et al., 2006. Rethinking TNM: Breast cancer TNM classification for treatment decision-making and research. *Breast*, 15(1), pp.3–8.
- Visone, R. et al., 2008. Hmga1 null mice are less susceptible to chemically induced skin carcinogenesis. *European journal of cancer (Oxford, England : 1990)*, 44(2), pp.318–25.
- Vlahos, C.J. et al., 1994. A specific inhibitor of phosphatidylinositol 3-kinase, 2-(4-morpholinyl)-8-phenyl-4H-1-benzopyran-4-one (LY294002). *Journal of Biological Chemistry*, 269(7), pp.5241–8.
- Vo, N. & Goodman, R.H., 2001. CREB-binding Protein and p300 in Transcriptional Regulation. *Journal of Biological Chemistry*, 276(17), pp.13505–13508.
- Wang, F., Marshall, C.B. & Ikura, M., 2013. Transcriptional/epigenetic regulator CBP/p300 in tumorigenesis: Structural and functional versatility in target recognition. *Cellular and Molecular Life Sciences*, 70(21),

- pp.3989–4008.
- Wang, I.-C. et al., 2005. Forkhead box M1 regulates the transcriptional network of genes essential for mitotic progression and genes encoding the SCF (Skp2-Cks1) ubiquitin ligase. *Molecular and cellular biology*, 25(24), pp.10875–94.
- Wapenaar, H. & Dekker, F.J., 2016. Histone acetyltransferases: challenges in targeting bi-substrate enzymes. *Clinical epigenetics*, 8(1), p.59.
- Watanabe, S. et al., 2009. HMGA2 maintains oncogenic RAS-induced epithelial-mesenchymal transition in human pancreatic cancer cells. *The American journal of pathology*, 174(3), pp.854–868.
- Wei, J.J. et al., 2011. Regulation of HMGA1 expression by MicroRNA-296 affects prostate cancer growth and invasion. *Clinical Cancer Research*, 17(6), pp.1297–1305.
- Wood, L.J. et al., 2000. HMG-I/Y, a new c-Myc target gene and potential oncogene. *Molecular and cellular biology*, 20(15), pp.5490–502.
- Xu, G. et al., 2014. MiR-142-3p functions as a potential tumor suppressor in human osteosarcoma by targeting HMGA1. *Cellular Physiology and Biochemistry*, 33(5), pp.1329–1339.
- Yang, H. et al., 2013. Small-molecule inhibitors of acetyltransferase p300 identified by high-throughput screening are potent anticancer agents. *Molecular cancer therapeutics*, 12(5), pp.610–20.
- Yang, J. & Weinberg, R.A., 2008. Epithelial-Mesenchymal Transition: At the Crossroads of Development and Tumor Metastasis. *Developmental Cell*, 14(6), pp.818–829.
- Yao, K. et al., 2015. MiR-186 suppresses the growth and metastasis of bladder cancer by targeting NSBP1. *Diagnostic Pathology*, 10(1), p.146.
- Yie, J., Senger, K. & Thanos, D., 1999. Mechanism by which the IFN-beta enhanceosome activates transcription. *Proceedings of the National Academy of Sciences of the United States of America*, 96(23), pp.13108–13113.
- Yokoyama, Y. et al., 2014. Loss of histone H4K20 trimethylation predicts poor prognosis in breast cancer and is associated with invasive activity. *Breast cancer research : BCR*, 16(3), p.R66.
- Yoo, K.H. & Hennighausen, L., 2011. EZH2 methyltransferase and H3K27 methylation in breast cancer. *International Journal of Biological Sciences*, 8(1), pp.59–65.
- Yu, Y. & Feng, Y.M., 2010. The role of kinesin family proteins in tumorigenesis and progression. *Cancer*, 116(22), pp.5150–5160.
- Zeisberg, M. & Neilson, E.G., 2009. Biomarkers for epithelial-mesenchymal transitions. *Journal of Clinical Investigation*, 119(6), pp.1429–1437.
- Zhang, Q. & Wang, Y., 2008. High mobility group proteins and their post-translational modifications. *Biochimica et biophysica acta*, 1784(9), pp.1159–66.
- Zhao, K. et al., 1993. SAR-dependent mobilization of histone H1 by HMG-I/Y in vitro: HMG-I/Y is enriched in H1-depleted chromatin. *The EMBO journal*, 12(8), pp.3237–47.
- Zhao, Q.-Y. et al., 2016. Global histone modification profiling reveals the epigenomic dynamics during malignant transformation in a four-stage breast cancer model. *Clinical epigenetics*, 8, p.34.
- Zhao, X.X. et al., 2015. MicroRNA-26a inhibits proliferation by targeting high mobility group AT-hook 1 in breast cancer. *International Journal of Clinical and Experimental Pathology*, 8(1), pp.368–373.
- Van Zijl, F., Krupitza, G. & Mikulits, W., 2011. Initial steps of metastasis: Cell invasion and endothelial transmigration. *Mutation Research - Reviews in Mutation Research*, 728(1–2), pp.23–34.

Zu, X. et al., 2015. TGF- β 1 induces HMGA1 expression in human breast cancer cells: Implications of the involvement of HMGA1 in TGF- β signaling. *International Journal of Molecular Medicine*, 35(3), pp.693–701.

ACKNOWLEDGEMENTS

We would like to thank Associazione Italiana per la Ricerca sul Cancro (AIRC) and Fondo Sociale Europeo (FSE- regione Friuli Venezia Giulia) for research fundings. Moreover, we would like to thank Prof. Jacek Wiśniewski and Prof. Matthias Mann (Max Planck Institute of Biochemistry, Munich, Germany) for fruitful collaboration and mass spectrometry analyses. I would like to extend my gratitude to all the people who supported my research project by professional advices and contributions, and personal support.